DISSERTATION

METHODOLOGIES TO DETECT LEAKAGES FROM GEOLOGICAL CARBON STORAGE SITES

Submitted by

Ana González-Nicolás Álvarez

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2014

Doctoral Committee:

Advisor: Domenico Baù

Darrell Fontane
Michael Ronayne
Thomas Sale

ABSTRACT

METHODOLOGIES TO DETECT LEAKAGES FROM

GEOLOGICAL CARBON STORAGE SITES

Geological carbon storage (GCS) has been proposed as a favorable technology to reduce carbon dioxide (CO_2) emissions to the atmosphere. Candidate storage formations include abandoned oil and natural gas reservoirs, un-mineable coal seams, and deep saline aquifers. The large global storage capacity and widespread occurrence of deep saline formations make them ideal repositories of large volumes of CO_2, however they generally lack of data for geological characterization in comparison to oil and gas reservoirs. Thus, properties of the injected formation or the sealing formation are unknown, which implies that the evolution and movement of the CO_2 plume are uncertain in these geological formations.

The first part of this research aims to provide an understanding of the main sources of uncertainty during the injection of CO_2 that cause leakage variability and fluid pressure change near the injection well, which could be responsible for fracturing the sealing formation. With this purpose the effect of uncertain parameters such as permeability and porosity of injected aquifer, permeability of CO_2 leakage pathways through the sealing layers, system compressibility, and brine residual saturation are investigated using stochastic and global sensitivity analyses. These analyses are applied to a potential candidate site for GCS located at the Michigan Basin. Results show aquifer permeability and system compressibility are the most influential parameters on fluid overpressure and CO_2 mass leakage. Other parameters, such as rock porosity, permeability of passive wells, and brine residual saturation do not influence fluid overpressure nearby the

injection well. CO_2 mass leakage is found to be sensitive to passive well permeability as well as the type of statistical distribution applied to describe well permeability.

Scarce data of the Michigan Basin exist that can be used directly to describe the spatial distribution at the basin scale of the caprock overlying the candidate site. The continuity of this formation is uncertain. The second part of this investigation explores the application of binary permeability fields for the study of CO_2 leakage from GCS at the candidate site. A sequential indicator simulation algorithm is used to populate binary permeability fields representing a caprock formation with potential leaky areas (or inclusions). Results of the caprock continuity uncertainty conclude that increasing the probability of inclusions occurrence increases the CO_2 leakage. In addition, the correlation length used by the sequential indicator simulator affects the occurrence of inclusions.

The third part investigates the detection and location of the presence of possible brine or carbon leakage pathways at the caprock during the injection operations of a GCS system. A computational framework for the assimilation of changes in head pressure data into a subsurface flow model is created to study the evolution of the CO_2 plume and brine movement. The capabilities of two data assimilation algorithms, the ensemble smoother (ES) and the ensemble Kalman smoother (EnKS), to identify and locate the leaky pathways are examined. Results suggest that the EnKS is more effective than the ES in characterizing caprock discontinuities.

ACKNOWLEDGEMENTS

I would like to give a very special thank you to my Advisor, Dr. Domenico Baù, for his continuous support during these four years and his guidance that made this dissertation possible.

I would like also to thank to my committee members Dr. Darrell Fontane, Dr. Michael Ronayne, and Dr. Thomas Sale for their insights and comments.

This research was funded by the U.S. Department of Energy, National Energy Technology Laboratory. The research conducted was also made possible with the support of Schlumberger Ltd., who kindly donated to Colorado State University the reservoir simulation software suites PETREL and ECLIPSE, together with training opportunities to their use.

Last, I would like to greatly thank my husband Luca for his love, patience, and emotional support.

DEDICATION

This work is dedicated to my family in Spain, my husband Luca, and our daughter Elisa.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Chapter: Introduction

1 Problem Statement

The Earth is experiencing global warming as a consequence of increased greenhouse gas concentrations (IPCC, 2007). Global mean annual surface temperatures have increased between 0.3 and 0.6° C since the late 19th century (Nicholls et al., 1996; Ledley et al., 1999). Carbon dioxide (CO_2) is the most important greenhouse gas produced by human activities (IPCC, 2007). Main sources of this CO_2 are power generation from combustion of fossil fuels (coal, oil, and gas), transportation, and industrial processes. Combustion of fossil fuels accounts about 50% of CO_2 emissions, transport about 25%, and industrial processes together with others contribute for the rest (International Energy Agency, 2002). These emissions exceed the estimated uptake of CO_2 from natural sinks, such as oceans and terrestrial biosphere (Ledley et al., 1999). To mitigate climate change, one suggested goal is to limit atmospheric CO_2 concentration around 500 ± 50 ppm (O'Neill and Oppenheimer, 2002; Béal et al., 2010). This would imply to maintain emissions of CO_2 at the existing level of 7 Giga tons of carbon per year (GtC/year) (Pacala and Socolow, 2004). Different options exist to mitigate CO_2 emissions and/or reduce CO_2 atmospheric concentrations (IPCC, 2005): 1) improve energy efficiency; 2) switch to less carbon-intensive fuels; 3) increased use of low and near-zero-carbon energy sources; 4) sequester CO_2 through the enhancement of natural, biological sinks; and 5) CO_2 capture and storage.

In the last decade, Carbon Capture and Storage (CCS) has been proposed as a technology for reducing CO_2 emissions in the atmosphere. CCS involves the capture of CO_2 at CO_2 sources (e.g. power plant), transporting the CO_2, and finally injecting the CO_2 into a deep formation. Figure 1-1 illustrates an example of a CCS process including several CCS phases. A power plant

equipped with CCS could reduce CO_2 emissions to the atmosphere by approximately 80-90% in comparison with a plant without CCS (IPCC, 2005).

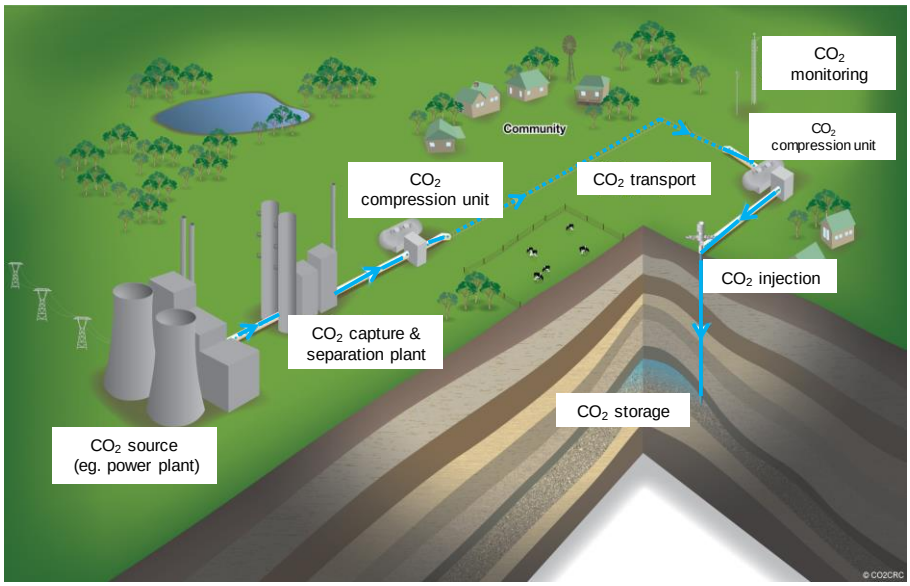

Figure 1-1. Carbon Capture and Storage process. Source: Cooperative Research Center for Greenhouse Gas Technologies (CO2CRC), http://www.co2crc.com.au/imagelibrary3/general.php.

Candidate storage formations include abandoned oil and natural gas reservoirs, un-mineable coal seams, and deep saline aquifers (Bergman and Winter, 1995; Ruether, 1998; Bachu, 2003). The Intergovernmental Panel on Climate Change (IPCC) recognizes the storage of CO_2 in deep oil and gas fields, and saline formations as an economically feasible option (IPCC, 2005). IPCC estimated storage capacity at a minimum of 1,678 $GtCO_2$ and potentially much higher, of which 60% are of deep saline formations (International Energy Agency, 2008). Therefore, deep saline formations are potential alternatives to the lack of petroleum fields. Michael et al. (2010) provide a good review of the experience gained from existing saline aquifer storage operations (pilots and commercials projects) and point out CCS as technologically feasible.

Geological carbon storage (GCS) in a saline aquifer involves the injection of CO_2 in supercritical state into a deep formation with a minimum depth of 800 meters (Van der Meer, 1992; Bachu et al., 1994). CO_2 density increases with depth, so at an adequate depth, CO_2 has higher density and consequently a reduced volume. In supercritical state, CO_2 has a density of a liquid (150 - >800 kg/m^3) (Bachu, 2003), and the viscosity of a gas, and behaves as a gas filling all the pore volume available. Figure 1-2 shows that CO_2 is in supercritical state for temperatures higher than 31.1^{o}C and pressures higher than 73.9 bar (IPCC, 2005).

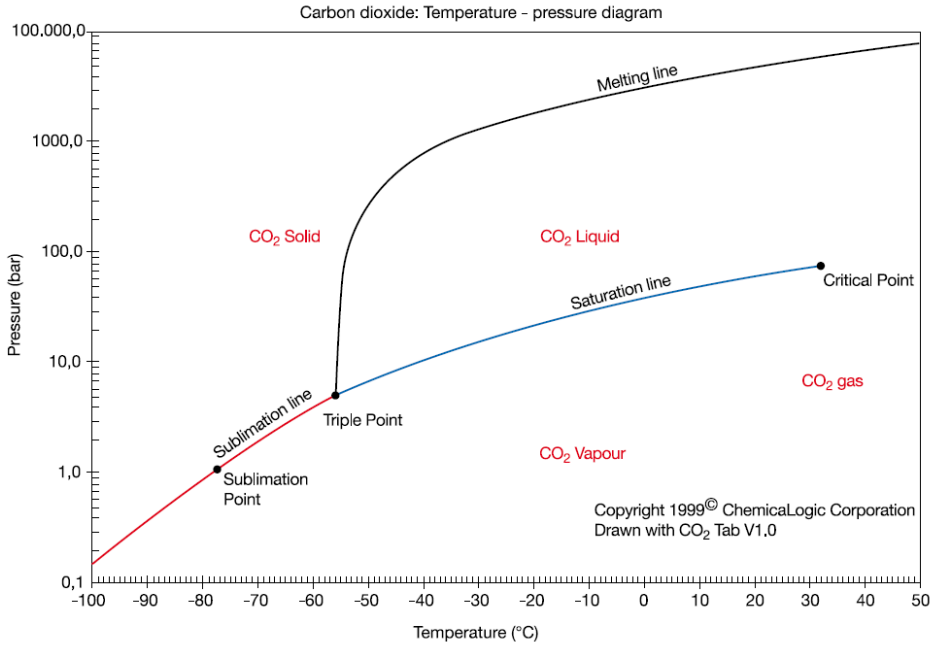

Figure 1-2. CO_2 phase as function of temperature and pressure. Source: IPCC (2005).

Several mechanisms contribute to the CO_2 storage in reservoirs and saline aquifers during the stages of injection and post-injection. These mechanisms, illustrated in Figure 1-3 have been described in the literature (Bachu and Adams, 2003; IPCC, 2005) and can be classified into two categories:

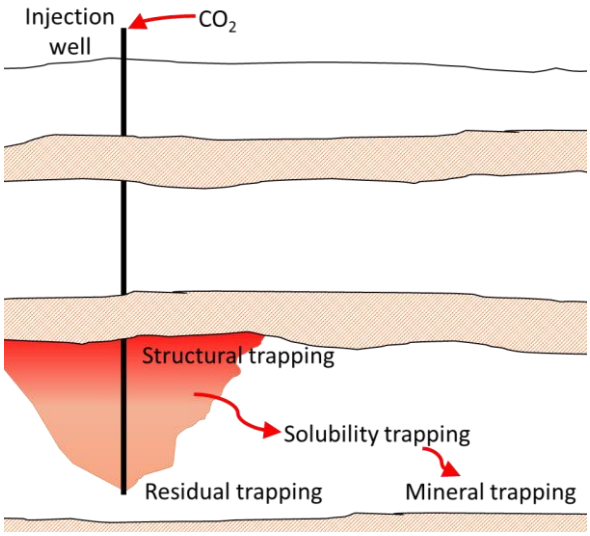

Figure 1-3. Trapping mechanisms during injection and storage of CO_2.

1) Physical trapping:
   - Stratigraphic and structural: a portion of the injected CO_2 is trapped as mobile phase, which migrates laterally and vertically until a sealing formation, a facies change, or an anticline detains the plume movement.
   - Hydrodynamic: part of the injected CO_2 is trapped in the pore space as residual saturation of CO_2 (immobile phase) due to capillary forces when brine reimbibes the CO_2.
2) Geochemical trapping:
   - Dissolution of CO_2 into formation fluids: part of the injected CO_2 is dissolved into the brine giving carbonic acid. When CO_2 is dissolved into the brine, it increases brine density resulting in a convention process of carbonate brine and brine.
   - Chemical reactions with minerals: carbonic acid will react with minerals from the rock matrix and will dissolve silicate minerals at large times. During the reactions, stable precipitates will form.

The time scales associated with each trapping mechanism are very different. Time scales span from the decades of the injection stage to the centuries and millennia of the dissolution and mineral trapping. Figure 1-4 shows time scales corresponding at each trapping mechanism. Storage security is correlated to the time scale of the trapping mechanism, being mineral trapping the slowest and the safest at the same time (Figure 1-4).

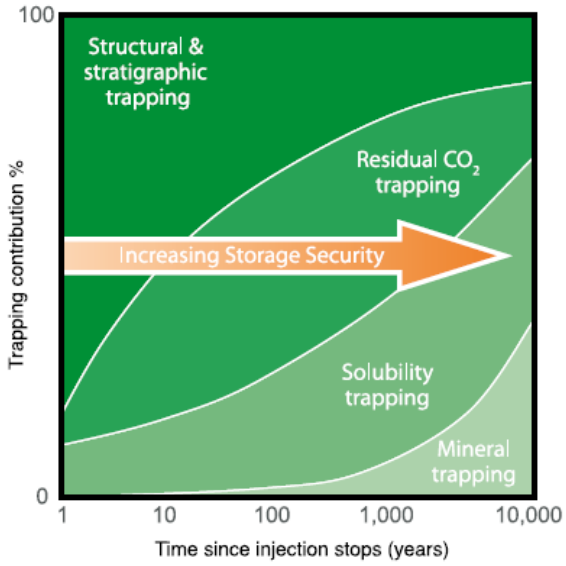

Figure 1-4. Storage security depending on time of post-injection stage and trapping mechanism. Source: IPCC (2005).

One of the criteria of a geological formation candidate for GCS is that the formation must be permeable enough to accept large amounts of injected CO_2, and must be overlain by a low permeability caprock to prevent CO_2 from migrating upward toward drinking water aquifers or back into the atmosphere. If CO_2 finds a potential pathway that leads upwards, CO_2 can produce pH changes of groundwater resources by increasing the concentration of carbonates. This can influence dissolution and sorption of minerals and hazardous trace metals (such as arsenic), and consequently affecting groundwater quality (Kharaka et al., 2006; Apps et al., 2010). Potential pathways are (IPCC, 2005): “weak” areas of the caprock (permeable areas) where CO_2 enter the

caprock if the capillary entry pressure is exceeded, faults and fractures, and poorly completed and abandoned wells.

Deep saline aquifers are barely geologically characterized in comparison to oil and gas reservoirs. Thus, properties of the injected formation or the sealing formation are unknown, which involves that the evolution and movement of the CO_2 plume are uncertain in these geological formations.

This investigation is intended to (i) have developed a scientific basis for understanding of the main sources of uncertainty that cause leakage variability in GCS at the proposed site at the Michigan Basin; (ii) test the capacity of applying a binary permeability fields to study CO_2 leakages from GCS systems; and (iii) improve the performance of GCS by presenting a methodology that detects leakage pathways from fluid pressure head measurements. These goals are met by developing three main parts.

The first part, Chapter 2, aims to provide an understanding of the main sources of uncertainty during the injection of CO_2 that cause leakage variability and fluid pressure change near the injection well, which could be responsible for fracturing the sealing formation. Thereby identifying where uncertainty reduction efforts should be directed during the characterization of a candidate site for GCS. This will allow CCS site managers to predict the feasibility of GCS in barely characterized aquifers. With this purpose the effect of uncertain parameters such as porosity and permeability of injected aquifer, permeability of CO_2 leakage pathways through the sealing layers, system compressibility, and brine residual saturation are investigated using stochastic and global sensitivity analyses. We apply these analyses to a potential site candidate for GCS located at the Michigan Basin.

Scarce data exist in the Michigan Basin exist that can be used directly to describe the spatial distribution at the basin scale of the sealing formation overlying the candidate site at the Michigan Basin. The continuity of this sealing formation is uncertain outside a certain area. The second part of this investigation, Chapter 3, explores the use of a sequential indicator Kriging simulation algorithm to study the sealing properties of the caprock at the Michigan Basin for GCS. Seismic surveys are used to define lithological facies of this real site and used to generate stochastically different distributions of facies over the caprock of the reservoir where no information is available.

The third part, Chapter 4, investigates the detection of the presence of possible brine or carbon leakage pathways during the injection operations of a GCS system, as well as monitor and predict the movement of the CO_2 plume. A computational framework for the assimilation of changes in head pressure data into a subsurface flow model is created to study the evolution of the CO_2 plume and brine movement in deep geologic formations during the CO_2 injection. This will provide a methodology to identify the spatial distribution of the injected CO_2 and detect escapes from the injected formation, allowing GCS site managers to control injection rates or schemes of injections to prevent undesirable leakages.

2 Major Findings

One of the main concerns about GCS is the risk of CO_2 escape from the storage formation through leakage pathways in the sealing formation. The second chapter of this study aims at understanding the main sources of uncertainty affecting the upward migration of CO_2 through pre-existing “passive” wells and the risk of fissuring of target formation during GCS operations, which may create pathways for CO_2 escape. The analysis focuses on a potential GCS

site located within the Michigan Basin, a geologic basin situated on the Lower Peninsula of the state of Michigan. For this purpose, a stochastic approach and a global sensitivity analysis is applied to investigate the influence of uncertain parameters such as permeability and porosity of the injected formation, passive well permeability, system compressibility, and brine residual saturation. A semi-analytical model that reduces the time of simulation and makes the application of stochastic analysis and global sensitivity analysis computationally feasible is used.

From the investigation of potential injection scenarios, it is observed that lower injection rates with longer injection times reduce the probability of producing excessive fluid overpressures in the injected aquifer. As far as CO_2 mass leakage is concerned, there is a small difference among these scenarios. Therefore, injection of CO_2 at low rates and protracted for a longer period of time appears to be the most convenient policy for the safety of the GCS system. Results of the stochastic and global sensitivity analyses show that, among these parameters, the most influential on both fluid overpressure and CO_2 mass leakage are the aquifer permeability and the system compressibility. Uncertainty on passive well permeability is important only for CO_2 mass leakage. Also, passive wells closer to the injection well are found to have a greater impact on CO_2 mass leakage.

The sealing formation outside the boundaries of the test site is barely characterized. Information about the continuity of this layer or its possible discontinuities is highly uncertain. To model this uncertainty, in Chapter 3, a sequential indicator simulator is developed and implemented to represent the sealing properties, namely permeability, of the geological unit overlying the GCS candidate formation. The sequential indicator simulation algorithm is used to create binary fields of caprock with low permeability and inclusions with high permeability where brine and CO_2 could leak out. Due to the lack of geological data, different geostatistical

configurations for the sealing formation are studied to assess the impact of the uncertainty of caprock discontinuity. To simulate injection and potential leakage of CO_2, a semi-analytical multiphase flow model is used. Inclusions located at close distance from one another are grouped and considered as single clusters to reduce the number of leaky points used by the semi-analytical multiphase flow model, thus reducing significantly the computational effort. The results obtained with the semi-analytical model are compared with those obtained using a numerical model to understand the limitations of using the semi-analytical model with large areas of leakage. Results of this comparison validate the semi-analytical multiphase flow.

Results of the caprock continuity uncertainty conclude that increasing the probability of inclusions occurrence increases the CO_2 leakage. In addition, the correlation length used by the sequential indicator simulators affects the occurrence of inclusions. CO_2 leakage is affected by passive wells permeability uncertainty for scenarios with low probability of inclusions.

One of the main concerns of GCS systems is the risk of leakage through "weak" permeable areas of the sealing formation or caprock. Since the fluid pressure pulse travels faster than the carbon dioxide (CO_2) plume across the storage reservoir, a fluid pressure change is inevitably transmitted into overlying permeable formations through "weak" areas of the caprock, and can be potentially detected sooner than actual CO_2 leakage occurs. An inverse modeling method based on fluid pressure measurements collected in strata above the target CO_2 storage formation is proposed, which aims at detecting the presence, the location, and the extent of possible carbon leakage pathways. A three-dimensional subsurface multiphase flow model with ensemble-based data assimilation algorithms are combined to identify potential caprock discontinuities that can undermine the long-term safety of GCS. In Chapter 4, this study examines and compares the capabilities of data assimilation algorithms such as the ensemble

smoother (ES) and the ensemble Kalman smoother (EnKS) to detect the presence of brine and/or CO_2 leakage pathways, potentially in real-time during GCS operations. For the purpose of this study, changes in fluid pressure in the brine aquifer overlying to CO_2 storage formation aquifer are assumed to be observed in monitoring wells, or provided by 4D time-lapse seismic surveys. Caprock discontinuities are typically characterized by higher values of permeability, so that the permeability distribution tends to fit to a non-Gaussian bimodal process, which does not comply with the requirements of the ES and EnKS algorithms. Here, issues related to the non-Gaussianity of the caprock permeability field are investigated by developing and applying a normal score transform procedure.

The performance of ES and EnKS methods is evaluated for two different scenarios: 1) a scenario in which permeability of both storage and overlying aquifer is considered known and homogeneous, and 2) a scenario in which aquifers' permeability are considered uncertain and heterogeneous. Results suggest that the EnKS is more effective than the ES in characterizing caprock discontinuities. Also, since results from ES-NST rely on only one update of the system, it is found that EnKS-NST method allows for better detection of leaky pathways. In the case of the heterogeneous aquifers scenario, more assimilation times are required in order to better identify caprock discontinuities from spurious leakages.

3 Future Research

Brine leakage through a passive well or a 'weak' area from the candidate storage formation may affect the quality of shallow groundwater resources. Although this study focuses the methodologies presented in Chapter 2 and Chapter 3 on CO_2 leakage, it could also be applied to study their impact on brine leakage.

In Chapter 4, it would be of particular interest to extend several aspects of the research presented.

First, apply a sensitivity analysis of the reference caprock log-k variance to verify that EnKS-NST performs better than ES when the non-Gaussianity and nonlinearity of the experiments are increased or decreased.

Second, study the effect on identifying and detecting leakage areas if the number of measurements locations is decreased or increased.

Third, study the possibility of combining measurements from the overlying and storage aquifers if these are provided by 4D time-lapse seismic surveys. This would allow better definition of pressure pulse movement through the formations. Consequently, better identification and location of the discontinuities of the caprock would be expected.

Fourth, apply the same methodology using the EnKF algorithm, which would present a reduced central processor unit time in comparison to the EnKS. In order to avoid the inconsistency of updated state values that EnKF may provide, only system parameters would be updated.

Last, apply the methodology presented in Chapter 4 to a real scenario.

One of the main risks associated with carbon geological storage in deep brine aquifers is the escape of CO_2 through potential pathways of the sealing layers such as poorly completed and abandoned wells. The vast number of abandoned wells present in certain areas and the lack of information about their cementation integrity convert them in a main source of leakage uncertainty. Based on a similar data assimilation-framework presented in Chapter 4, it would be possible to identify which wells are leaking in a certain field. This would provide a method to

detect escapes through poor cemented wells from the injected formation and a better control of injection rates or schemes of injections to prevent undesirable leakages.

4 Organization

This dissertation is organized in the following three sections:

- Chapter 2 is entitled ‘Stochastic and global sensitivity analyses of uncertain parameters affecting the safety of geological carbon storage in saline aquifers of the Michigan Basin’ and includes an article by González-Nicolás, Baù, Cody, and Alzraiee, being submitted to “International Greenhouse Gases Journal”.
- Chapter 3 is entitled ‘Application of binary permeability fields for the study of CO_2 leakage from geological carbon storage in saline aquifers of the Michigan Basin’ and includes an article by González-Nicolás, Cody, Baù. In preparation, to submit to “Hidrogeology Journal”.
- Chapter 4 is entitled ‘Detection of potential leakage pathways from geological carbon storage by fluid pressure data assimilation’ and includes an article by González-Nicolás, Baù, and Alzraiee. In preparation, to submit to Special Issue in “Advances in Water Resources”.

5 References

Apps, J., Zheng, L., Zhang, Y., Xu, T., Birkholzer, J., 2010. Evaluation of potential changes in groundwater quality in response to CO_2 leakage from deep geologic storage. Transport in Porous Media 82, 215-246.

Bachu, S., 2003. Screening and ranking of sedimentary basins for sequestration of CO_2 in geological media in response to climate change. Environmental Geology 44, 277-289.

Bachu, S., Adams, J.J., 2003. Sequestration of CO2 in geological media in response to climate change: capacity of deep saline aquifers to sequester CO2 in solution. Energy Conversion and Management 44, 3151-3175.

Bachu, S., Gunter, W., Perkins, E., 1994. Aquifer disposal of CO< sub> 2</sub>: Hydrodynamic and mineral trapping. Energy Conversion and Management 35, 269-279.

Béal, D., Brasseur, P., Brankart, J.-M., Ourmières, Y., Verron, J., 2010. Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis. Ocean Science 6, 247-262.

Bergman, P.D., Winter, E.M., 1995. Disposal of carbon-dioxide in aquifers in the US. Energy Conversion and Management 36, 523-526.

International Energy Agency, 2002. World energy outlook, 2002 ed. OECD/IEA, Paris.

International Energy Agency, 2008. Carbon Capture and Storage: Meeting the Challenge of climate change, Paris.

Kharaka, Y., Cole, D., Hovorka, S., Gunter, W., Knauss, K., Freifeld, B., 2006. Gas-water-rock interactions in Frio Formation following CO2 injection: Implications for the storage of greenhouse gases in sedimentary basins. Geology 34, 577-580.

Ledley, T.S., Sundquist, E.T., Schwartz, S.E., Hall, D.K., Fellows, J.D., Killeen, T.L., 1999. Climate change and greenhouse gases. Eos, Transactions American Geophysical Union 80, 453.

Metz, B., Intergovernmental Panel on Climate Change. Working Group III., 2005. IPCC Special Report on Carbon Dioxide Capture and Storage. Cambridge University Press, for the Intergovernmental Panel on Climate Change, Cambridge.

Michael, K., Golab, A., Shulakova, V., Ennis-King, J., Allinson, G., Sharma, S., Aiken, T., 2010. Geological storage of CO_2 in saline aquifers-A review of the experience from existing storage operations. International Journal of Greenhouse Gas Control 4, 659-667.

Nicholls, N., Gruza, G., Jouzel, J., Karl, T., Ogallo, L., Parker, D., 1996. Observed climate variability and change. Cambridge University Press.

O'Neill, B.C., Oppenheimer, M., 2002. Dangerous climate impacts and the Kyoto Protocol. Science 296, 1971-1972.

Pacala, S., Socolow, R., 2004. Stabilization wedges: Solving the climate problem for the next 50 years with current technologies. Science 305, 968-972.

Ruether, J.A., 1998. FETC Programs for Reducing Greenhouse Gas Emissions, Other Information: PBD: Feb 1998, p. Medium: ED; Size: 20 p.; Other: FDE: PDF; PL:.

Solomon, S., Intergovernmental Panel on Climate Change., Intergovernmental Panel on Climate Change. Working Group I., 2007. Climate change 2007 : the physical science basis : contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge ; New York.

Van der Meer, L., 1992. Investigations regarding the storage of carbon dioxide in aquifers in the Netherlands. Energy Conversion and Management 33, 611-618.

2 Chapter: Stochastic and Global Sensitivity Analyses of Uncertain Parameters Affecting the Safety of Geological Carbon Storage in Saline Aquifers of the Michigan Basin

Summary

Geological carbon storage (GCS) has been proposed as a favorable technology to reduce carbon dioxide (CO_2) emissions to the atmosphere. One of the main concerns about GCS is the risk of CO_2 escape from the storage formation through leakage pathways in the sealing formation. This study aims at understanding the main sources of uncertainty affecting the upward migration of CO_2 through pre-existing "passive" wells and the risk of fissuring of target formation during GCS operations, which may create pathways for CO_2 escape. The analysis focuses on a potential GCS site located within the Michigan Basin, a geologic basin situated on the Lower Peninsula of the state of Michigan. For this purpose, we apply a stochastic approach and a global sensitivity analysis to investigate the influence of uncertain parameters such as: permeability and porosity of the injected formation, passive well permeability, system compressibility, and brine residual saturation. We use a semi-analytical model that reduces the time of simulation and makes the application of stochastic analysis and global sensitivity analysis computationally feasible. Results show that, among these parameters, the most influential on both fluid overpressure and CO_2 mass leakage are the aquifer permeability and the system compressibility. Uncertainty on passive well permeability is important only for CO_2 mass leakage. Passive wells closer to the injection well are found to have a greater impact on CO_2 mass leakage.

1 Introduction

The Earth's atmosphere is experiencing global climate change caused by increasing greenhouse gas concentrations. Carbon dioxide (CO_2) is the most important greenhouse gas produced by human activities (Solomon et al., 2007). In the last decade, geological carbon storage (GCS) has been identified as a promising technology for reducing CO_2 emissions to the atmosphere. Candidate storage formations include depleted oil and natural gas reservoirs, unmineable coal seams, and deep saline aquifers (Bergman and Winter, 1995; Ruether, 1998; Bachu, 2003). The latter represent potential alternatives to the lack of petroleum fields and constitute 60% of the estimated storage capacity worldwide (International Energy Agency, 2008). GCS in saline aquifers involves the injection of supercritical CO_2 into deep brine-saturated formations. Supercritical CO_2 is less dense and less viscous than the brine residing in saline formations, which causes gravity override as well as viscous fingering. Thus, supercritical CO_2 tends to migrate upwards driven by buoyancy until low-permeable layers, or caprock, detains its ascent. However, if the injected CO_2 finds a potential leakage pathway through the caprock, it may adversely affect shallow fresh groundwater resources or even migrate to the land surface.

Characteristics of the caprock overlying the injected formations are critical elements for the effectiveness and safety of GCS operations. Nevertheless, unlike petroleum reservoirs, saline aquifers have never contained oil or gas. Consequently, there are less data associated with saline aquifers than petroleum reservoirs. In addition, information about the sealing properties of the caprock might be scarce or inexistent. Typically, physical properties of potential candidate GCS sites are highly uncertain. Host rock permeability, spatial distribution of potential leakage pathways, and increase of fluid pressure in the injected formations may directly influence CO_2

leakage. Leakage pathways may also be created during the CO_2 injection process due to caprock fracturing associated with increased fluid pressure and the ensuing reduction in effective stress. Therefore, assessing the risk of CO_2 leakage given the uncertainty on these parameters is vital prior to the implementation of GCS systems.

Carbon injection into deep saline aquifers involves complex processes of two-phase flow in confined geological formations, which make its modeling a demanding endeavor. Complexities associated with multiphase flow and transport processes, such as non-linearity, induced fingering, convective mixings, etc., create the need for computationally efficient assessment approaches. Several analytical and semi-analytical solutions have appeared in the literature, including Saripalli and McGrail (2002), Nordbotten et al. (2005a), Gasda et al. (2008) and Dentz and Tartakovsky (2009) among others, which rely on a number of simplifying assumptions. Modifications of some of these solutions accounting for CO_2 compressibility are presented in Vilarrasa et al. (2010). The main advantage of analytical and semi-analytical models is that they allow simulations to be performed in a very short central processor unit (CPU) time (of the order of seconds), which makes stochastic and global sensitivity analyses, requiring on the order of thousands of model runs, computationally viable.

Risk assessment is an important tool for decision making during the initial stages of GCS projects. Some algorithms have been developed to predict performance and risk of GCS systems (e.g., LeNeveu, 2008; Stauffer et al., 2008; Oldenburg et al., 2009; Dobossy et al., 2011), where potential candidate sites are selected for evaluation of their safety and effectiveness. Several studies have been published that statistically analyze the uncertainty of leakage associated with parameters of the injected aquifer in a GCS system. For example, Celia et al. (2009) investigated

the influence of the injection depth on leakage risk and showed that this risk decreases when injection depth increases.

CO_2 injection performance and sequestration efficiency have also been investigated in published literature. For example, Celia et al. (2011) found that CO_2 injection rates are reduced by higher brine residual saturations and are influenced by the relative permeability of CO_2. Gupta and Bryant (2011) found that more CO_2 trapping is achieved when the gravity number (*i.e.* the ratio between buoyancy and viscous forces) is low, leading to enhanced lateral displacement of the CO_2 plume. On the other hand, high gravity numbers lead to stronger gravity override, resulting in both less trapping of CO_2 and less contact between the CO_2 plume and ambient brine. Middleton et al. (2012) showed that uncertainties from permeability, porosity, and formation thickness significantly affected capacity and cost calculations.

Studies that analyze the uncertainty of leakage associated with abandoned wells can also be found. Kopp et al. (2010) concluded that the increased risk of leakage is produced by a long injection time, small distance between injection wells and leaky wells, high permeability anisotropy, high geothermal gradient, and shallow depth. In order to show that potential leakage depends on formation properties, as well as location and number of leaky wells, Nogues et al. (2012) conducted a Monte Carlo simulation where the main uncertainty was the effective well permeability.

A sensitivity analysis on the long-term behavior of CO_2 in a multilayered aquifer was conducted by Kano and Ishido (2011). Parameters that had more influence in the long-term were geothermal gradient and thicknesses of layers, as well as the capillary pressure, relative permeability and permeability. Aoyagi et al. (2011) presented an example of a sensitivity analysis of productivity index and fault permeability affecting the leakage of CO_2 through wells

or faults. They found that the fault permeability value is more relevant when leakage starts. Zhao et al. (2010) determined that CO_2 dissolution increased when the vertical-to-horizontal permeability ratio, critical gas saturation, or brine salinity are decreased, and when brine saturation is increased.

Alternative methods for quantifying uncertainty by stochastic simulation can be found, for example, in the works of Oladyshkin et al. (2011) and Walter et al. (2011). Oladyshkin et al. (2011) used an integrative probabilistic collocation method to reduce the computational cost associated with stochastic approaches. Walter et al. (2011) used this method to study the pressure increase in a channel system during injection of CO_2.

All these studies investigate uncertainties of multiple factors to aid the decision making of best injection strategies. The aim of this study is to provide an understanding of the main sources of uncertainty that affect leakage through potential escape pathways and fluid overpressure variability, thereby identifying where data collection efforts should be directed to improve the characterization of a candidate site for GCS. With this purpose, we perform stochastic simulations and conduct global sensitivity analyses to investigate the effect of several parameters − such as permeability and porosity of injected aquifers, passive well permeability, system compressibility, and brine residual saturation − on the potential fracturing of the storage formation during CO_2 injection and the upward migration of CO_2 through passive wells. This study is applied to a potential GCS site embedded in the Michigan Basin.

To simulate the CO_2 injection, we use ELSA-IGPS, a semi-analytical model implemented by Cody et al. (2014), which builds upon the semi-analytical model developed by Celia and Nordbotten (2009) and Nordbotten et al. (2009). We apply these analyses to study the influence of these uncertain parameters on: (i) the maximum fluid overpressure produced by carbon

injection; and (ii) the mass of CO_2 that migrates into overlying formations in relation to the total mass of injected CO_2. Results from the stochastic approach are used to estimate the probability of fracturing the caprock, and the probability of leaked mass to exceed predefined threshold values. The global sensitivity analysis aims at identifying the most influential and the least relevant uncertain parameters.

This paper is organized as follows. First, the multiphase flow model is presented, followed by a description of the methodologies used for stochastic analysis and global sensitivity analysis. Results of the application of these methodologies to the Michigan Basin test site are thus presented and discussed. Last, a summary and conclusions of this work are given.

2 Multiphase Flow Semi-Analytical Model

The algorithm used in this study is called ELSA-IGPS (Estimating Leakage Semi-Analytically- Iterative Global Pressure Solution) (Cody et al., 2014), and constitutes a modified version of the semi-analytical model ELSA devised by Celia and Nordbotten (2009) and Nordbotten et al. (2009).

By solving the partial differential equations for two-phase immiscible flow, Nordbotten et al. (2005b) developed a semi-analytical model to estimate the leakage of brine and CO_2 flux through permeable caprock locations resulting from GCS. Permeable caprock locations are conceptualized as segments of pre-existing, abandoned wells and represent cylindrical portions of the caprock layers having low, yet non-negligible, permeability. These are referred to as "passive" wells and are assumed to be the only pathways for fluid flux exchange between permeable layers.

In ELSA-IGPS, the domain is structured into a stack of L aquifers separated by $L+1$ caprock layers, perforated by M carbon injection wells and N passive wells. Aquifers are assumed to be horizontal, homogenous, and isotropic. Injection wells are theoretically able to inject into any aquifer. Initially, fluid is not flowing through any of the passive wells because the entire domain is assumed saturated with brine under hydrostatic pressure conditions. Additional assumptions include perfectly horizontal flow, and negligible capillary pressure, which results in a sharp CO_2-brine interface. The CO_2 plume thickness at any given location is assumed to be the thickness from all sources and sinks in the aquifer. Pressure response from sources and sinks can be superimposed in each aquifer, and the capacity of the formation to store the injected CO_2 remains constant during GCS operations.

At the start of injection, aquifer fluid pressures begin to change throughout the domain resulting in pressure differentials across caprock layers and fluid flux through passive wells. It is therefore very important to understand aquifer fluid pressure response resulting from changes in the mass storage of CO_2 and brine. Celia et al. (2011) express this radial pressure response p [$ML^{-1}T^{-2}$] at the bottom of an aquifer in which a single well injects CO_2 as:

$$p = p_0 + (\rho_b - \rho_c) g H \Delta p' \tag{2-1}$$

where p_0 is the initial fluid pressure at the bottom of the aquifer, ρ_α is the fluid density [ML^{-3}] (α denotes the phase type, b for brine and c for CO_2), g is the gravitational acceleration [LT^{-2}] and H is the aquifer thickness [L]. In Equation (2-1), $\Delta p'$ [/] is defined as:

$$\Delta p'(\chi) = \begin{cases} 0 & \text{for } \chi \geq \psi \\ -\frac{1}{2\Gamma}\ln\left(\frac{\chi}{\psi}\right) + \Delta p'(\psi) & \text{for } \psi > \chi \geq 2\lambda \\ \frac{1}{\Gamma} - \frac{\sqrt{\chi}}{\Gamma\sqrt{2\lambda}} + \Delta p'(2\lambda) + F(h') & \text{for } 2\lambda > \chi \geq \frac{2}{\lambda} \\ -\frac{1}{2\lambda\Gamma}\ln\left(\frac{\chi\lambda}{2}\right) + \Delta p'\left(\frac{2}{\lambda}\right) & \text{for } \frac{2}{\lambda} > \chi \end{cases} \quad (2\text{-}2)$$

where,

$$\chi = \frac{2\pi H\varphi(1 - s_b^{res})r^2}{Qt} \quad (2\text{-}3)$$

$$\Gamma = \frac{2\pi(\rho_b - \rho_c)gkH^2}{\mu_b Q} \quad (2\text{-}4)$$

$$\psi = \frac{4.5\pi H\varphi k(1 - s_b^{res})}{\mu_b c_{eff} Q} \quad (2\text{-}5)$$

$$h' = \frac{h(\chi)}{H} = \frac{1}{\lambda - 1}\left(\frac{\sqrt{2\lambda}}{\sqrt{\chi}} - 1\right) \quad (2\text{-}6)$$

$$F(h') = \frac{-\lambda}{\lambda - 1}\left[\mathrm{h}' - \frac{ln[(\lambda - 1)h' + 1]}{\lambda - 1}\right] \quad (2\text{-}7)$$

In Equations (2-2)-(2-7): h is the CO_2 plume thickness [L], h' [/] is the CO_2 plume thickness relative to the aquifer thickness H; s_b^{res} is the residual saturation of the brine [/]; t represents time [T]; k is the aquifer permeability [L^2]; μ_b is the dynamic viscosity of the brine [$ML^{-1}T^{-1}$]; φ is the aquifer porosity [/]; Q is the total CO_2 volumetric well flux [L^3T^{-1}]; c_{eff} is the effective compressibility of the fluid and the solid matrix [$M^{-1}LT^2$]; and r is the radial distance [L]. Also, $F(h')$ is an offset term related to the vertical pressure distribution (Celia et al., 2011), and the mobility ratio is defined as $\lambda = \lambda_c/\lambda_b$, where $\lambda_\alpha = k_{r,\alpha}/\mu_\alpha$ and $k_{r,\alpha}$ is the relative permeability of phase α.

To determine the fluid overpressure at any given time throughout the aquifer system Nordbotten et al. (2005b) apply superposition of effects derived from the application of Equation (2-1) for all the volumetric sources and sinks corresponding to CO_2 injection wells and passive wells. Consequently, the fluid pressure at any given time t, at the bottom of the generic aquifer l (l=1,2,..,L) and for each passive well j (j=1,2,..,N) can be expressed as:

$$p_{j,l} = p_{0_l} + (\rho_b - \rho_c) g H_l \left[\sum_{iw=1}^{M} \Delta p'(\chi_{iw,j,l}) + \sum_{i=1}^{N} \Delta p'(\chi_{i,j,l}) \right] \tag{2-8}$$

where $\chi_{iw,j,l} = 2\pi H_l \varphi_l (1 - s_{b,l}^{res}) r_{iw,j}^2 / (Q_{iw,l} t)$ and $\chi_{i,j,l} = 2\pi H_l \varphi_l (1 - s_{b,l}^{res}) r_{i,j}^2 / \int_o^t (Q_{j,l} - Q_{j,l+1}) d\tau$ (see Equation (2-3)). r_{iw} and Q_{iw} are the injection well radius [L] and the total CO_2 volumetric injection well flux [L^3T^{-1}] respectively. With this approach, the fluid pressures at the bottom of each aquifer and at each passive well can be grouped into the following $(N \cdot L) \times 1$ vector:

$$\boldsymbol{p}(t) = \boldsymbol{p}[\boldsymbol{P}_1, \boldsymbol{M}(t)] \tag{2-9}$$

which is a function of the array:

$$\boldsymbol{P}_1 \equiv \left[\boldsymbol{H}, \boldsymbol{\varphi}, \boldsymbol{k}, \boldsymbol{s}_b^{res}, \boldsymbol{r}, \boldsymbol{Q}_{iw}, \rho_c, \rho_b, g, \lambda, \pi, \mu_b, c_{eff} \right] \tag{2-10}$$

where the $L \times 1$ vectors $\boldsymbol{H}$, $\boldsymbol{\varphi}$, $\boldsymbol{s}_b^{res}$ and $\boldsymbol{k}$ include the thicknesses, porosities, brine residual saturations and permeabilities of all aquifers; $\boldsymbol{Q}_{iw}$ is a $(M \cdot L) \times 1$ vector including the CO_2 inflow rates for each aquifer l (l=1,2,..,L) and for each injection well iw (iw=1,2,..,M); and the $(M + N) \times (M + N)$ matrix $\boldsymbol{r}$ includes the relative distances between all injection and passive wells.

In addition, $\boldsymbol{p}(t)$ (Equation (2-9)) is a function of the $N \times L$ array $\boldsymbol{M}(t)$, whose generic component $M_{j,l}(t)$ represents the net cumulative fluid mass transferred into aquifer l through passive well j until time t, calculated as:

$$M_{j,l}(t) = \int_0^t \rho_{eff,j,l}(\tau)\left[Q_{j,l}(\tau) - Q_{j,l+1}(\tau)\right]d\tau \tag{2-11}$$

where $\rho_{eff,j,l}(\tau)$ is the effective fluid density in aquifer l at passive well j. This density is time-dependent since the composition of the leaking fluid is a function of the phase saturations of CO_2 and brine, which vary based on the CO_2 plume location. The effective fluid density is estimated as $\rho_{eff,j,l} = \rho_c s_{c\,j,l} + \rho_b\left(1 - s_{c\,j,l}\right)$, where $s_{c\,j,l}$ denotes CO_2 saturation.

Since the application of Equation (2-9) requires the temporal evolution of leakage rates through passive wells (Equation (2-11)) to be known, Nordbotten et al. (2005b) propose to use the sum of the flow rates $Q_{\alpha\,j,l}$ for each phase α calculated using the multiphase version of Darcy's law across each confining layer l for each passive well j:

$$Q_{j,l} = \sum_{\alpha=b,c}\left[\pi r_{pw\,j,l}^2 \frac{k_{r,\alpha\,j,l} k_{pw\,j,l}}{\mu_\alpha B_l}\left(p_{j,l-1} - \rho_\alpha g B_l - g\rho_\alpha H_{l-1} - p_{j,l}\right)\right] \tag{2-12}$$

In Equation (2-12), $r_{pw\,j,l}$ is the passive well radius and $k_{pw\,j,l}$ is the single phase passive well permeability for passive well j (j=1,2,..,N) and aquitard layer l. Note that in order to apply Equation (2-12), fluid pressures $p_{j,l}$ as well as CO_2 relative thicknesses in passive well pathways must be known to estimate pressure gradients, fluid saturations and relative permeability values. Given Equation (2-12), the flow rates across each aquitard l (l=1,2,..,L) for each passive well j (j=1,2,..,N) can be grouped into the $(N \cdot L) \times 1$ vector $\boldsymbol{Q}$:

$$\boldsymbol{Q}(t) = \boldsymbol{Q}[\boldsymbol{P}_2, \boldsymbol{p}(t)] \tag{2-13}$$

where the array $\boldsymbol{P}_2$ is given by:

$$\boldsymbol{P}_2 \equiv \left[\boldsymbol{B}, \boldsymbol{H}, \boldsymbol{r}_{pw}, \boldsymbol{k}_{pw}, \boldsymbol{k}_{r,c}, \boldsymbol{k}_{r,b}, \rho_c, \rho_b, \mu_c, \mu_b, g\right] \tag{2-14}$$

In Equation (2-14), the $(L+1)\times 1$ vector $\boldsymbol{B}$ includes the aquitard thicknesses [L], the $N \times (L+1)$ matrices $\boldsymbol{r}_{pw}$ and $\boldsymbol{k}_{pw}$ contain the passive well radii and the permeabilities, and the

$N \times (L+1)$ matrices $\boldsymbol{k}_{r,c}$ and $\boldsymbol{k}_{r,b}$ include the relative permeabilities of CO_2 and brine at passive wells, respectively.

By combining Equations (2-9) and (2-13), a set of $2 \cdot N \cdot L$ non-linear equations in $2 \cdot N \cdot L$ unknowns is obtained. These unknowns are the fluid pressures at the bottom of each aquifer and at each passive well (Equation (2-8)), and the flow rates (Equation (2-12)) across each aquitard for each passive well.

To solve this system, a computationally efficient fixed-point iterative scheme (Takahashi, 2000) is implemented. Note that since Equation (2-9) requires the calculation of the cumulative fluid mass transferred through passive wells, time-stepping is applied by discretizing the time interval [0; t_{end}], over which CO_2 injection occurs, into a number of time steps of length Δt. This approach allows for approximating the evolution of the pressure distribution, passive well fluxes, as wells CO_2 plume locations and thicknesses over the interval [0; t_{end}].

A fixed-point iteration scheme is applied to solve the system of non-linear equations introduced above at a generic time *t* given the solution at time $t - \Delta t$ obtained at the previous time step. By denoting with (f) the iteration index, the scheme uses the initial (for f=1) assumption that passive well flow rates for the current time-step remain the same as in the previous time step:

$$\boldsymbol{Q}(t)^{(f)} = \boldsymbol{Q}(t - \Delta t) \tag{2-15}$$

Next, the vector of the cumulative fluid mass transferred through passive wells is approximated as:

$$\boldsymbol{M}(t)^{(f)} = \boldsymbol{M}(t - \Delta t) + \left[(1-\omega)\boldsymbol{\rho}_{eff}(t - \Delta t)\boldsymbol{Q}(t - \Delta t) + \omega \boldsymbol{\rho}_{eff}(t)^{(f)}\boldsymbol{Q}(t)^{(f)}\right]\Delta t \tag{2-16}$$

where ω is a relaxation coefficient such that $0 < \omega \leq 1$. This coefficient is implemented to smooth out oscillations and speed up convergence. In Equation (2-16), $\boldsymbol{\rho}_{eff}(\mathrm{t})^{(f)}$ includes the fluid densities at passive wells, which are assumed to be equal to $\boldsymbol{\rho}_{eff}(t-\Delta t)$ for the first iteration (f= 1). The current estimate of the vector $\boldsymbol{M}(t)^{(f)}$ is substituted in Equation (2-9) to provide:

$$\boldsymbol{p}(t)^{(f)} = \boldsymbol{p}\left[\boldsymbol{P}_1, \boldsymbol{M}(t)^{(f)}\right] \tag{2-17}$$

This pressure distribution at passive wells is used in Equation (2-13) to obtain the updated passive flow rate vector:

$$\boldsymbol{Q}(t)^{(f+1)} = \boldsymbol{Q}\left[{\boldsymbol{P}_2}^{(f)}, \boldsymbol{p}(t)^{(f)}\right] \tag{2-18}$$

Note that before applying Equation (2-18), the array ${\boldsymbol{P}_2}^{(f)}$ must be re-calculated since it depends on the relative permeabilities of CO_2 and brine at passive wells (Equation (2-14)).

At this point, the iteration index is increased ($f \leftarrow f$+1) and the sequence of Equations (2-16)-(2-18) is repeated. The iteration proceeds until the maximum norm of the vector including the relative error between the preceding and current iteration's flow rate becomes smaller than a prescribed tolerance coefficient, ε_{max}:

$$\varepsilon = max\left\{\left|\frac{{Q_{1,1}}^{(f+1)} - {Q_{1,1}}^{(f)}}{{Q_{1,1}}^{(f+1)}}\right|, \cdots, \left|\frac{{Q_{j,l}}^{(f+1)} - {Q_{j,l}}^{(f)}}{{Q_{j,l}}^{(f+1)}}\right|, \cdots, \left|\frac{{Q_{N,L}}^{(f+1)} - {Q_{N,L}}^{(f)}}{{Q_{N,L}}^{(f+1)}}\right|\right\} \leq \varepsilon_{max} \tag{2-19}$$

In order to ensure time-step convergence stability, a maximum passive well flow rate, $Q_{pw,max}$, is specified to dampen artificially high-magnitude pressure differentials calculated when using large time- step intervals or closely-spaced passive well positions. $Q_{pw,max}$ is typically of the order of one tenth the volumetric CO_2 injection rate.

This ELSA-IGPS algorithm (Cody et al., 2014) described above allows for drastically reducing the computational effort (a complete simulation takes CPU times on the order of seconds or minutes) making possible the application of this solution within a stochastic simulation (or Monte Carlo) approach or a global sensitivity analysis such as those described in the following sections.

In this study, ELSA-IGPS is used to explore the uncertainty and sensitivity of the input parameters on the uncertainty and variability of two states variables of interest: i) the fluid overpressure nearby the injection well, and ii) the percentage of CO_2 mass leakage into overlying formations.

In these analyses, fluid overpressure Δp_{iw} is defined as the difference between the final (at final time t_{end}) and initial fluid pressures in proximity of the injection well. In our analyses, the number of injection wells M is set equal to 1, and injection occurs into the deepest aquifer (l=1). Therefore, based on Equation (2-8), the fluid overpressure nearby the injection well at final time is calculated as:

$$\Delta p_{iw} = \Delta p_{1,1}(r'_{iw}, t_{end}) = (\rho_b - \rho_c) g H_1 \left[\Delta p'(\chi_{1,1,1}) + \sum_{i=1}^{N} \Delta p'(\chi_{i,1,1}) \right] \tag{2-20}$$

where r'_{iw} is a radial distance nearby the injection well, equal to 5 m in this study.

The total CO_2 mass leakage is given by (Equation (2-11)):

$$M_{leak}(t_{end}) = \int_0^{t_{end}} \left[\sum_{i=1}^{N} \rho_c s_{c,i,2}(\tau) Q_{i,2}(\tau) \right] d\tau \tag{2-21}$$

Thus, the percent of CO_2 mass leakage $\%M_{leak}$ is defined as the ratio between the mass of CO_2 that escapes from the injected aquifer into overlying formations and the total mass of injected CO_2 at time t_{end}:

$$\%M_{leak} = \frac{M_{leak}(t_{end})}{\rho_c Q_{1,1} t_{end}} 100 \tag{2-22}$$

3 Stochastic Analysis

Stochastic, or Monte Carlo, simulation is a mathematical method that allows for the analysis of complex systems while accounting for uncertainty in quantitative terms. The main advantage of stochastic analysis (SA) is that it produces statistical distributions of possible outcomes resulting from uncertain input parameters. The number of total simulations required by SA depends in general on the number of uncertain parameters and their degree of uncertainty or range of variability. Values of the uncertain parameters are sampled randomly from their respective probability distribution functions (PDF), which are meant to reproduce the uncertainty of the parameter. Since the sampling is random, the outputs of the model are random as well.

In the SA presented here, the uncertain input parameters that may affect the state variables of interest, that is, Δp_{iw} (Equation (2-20)) and $\%M_{leak}$ (Equation (2-22)) are: permeability and porosity of injected aquifers, passive well permeability, system compressibility, and brine residual saturation. These uncertainties are modeled conceptually using a series of independents PDFs representing their typical range of uncertainty. In the case of passive well permeability, up to three different PDFs are considered.

In this study, ensembles of uncertain parameters are used within the mathematical model (see Section 2) to simulate how parameter uncertainty affects the uncertainty in the state variables of interest. Output ensembles of the state variables are used to produce cumulative distribution function (CDF) plots. The CDF of the generic state variable Y, either Δp_{iw} or $\%M_{leak}$, is obtained from the output of N_{MC} model simulations, where N_{MC} is the size of the ensemble. After ordering the Y values in ascending order, $Y_1 < Y_2 < \cdots < Y_{N_{MC}}$, the

corresponding CDF values are calculated as $CDF(Y) = (i - 0.5)/N_{MC}$ (i=1,2,…,N_{MC}) (Hahn, 1967).

By analyzing the statistics of the output ensembles ($Y_1 < Y_2 < \cdots < Y_{N_{MC}}$) the information that can be drawn is, for example, PDF type and its parameters, ensemble spread, quantiles, confidence bounds, and percentile values. In the case of state variables such as Δp_{iw} (Equation (2-20)) and $\%M_{leak}$ (Equation (2-22)), percentile values can be used to estimate the probability of fracturing the caprock formations and the probability of leaked mass to exceed predefined threshold values.

4 Global Sensitivity Analysis

Modeling CO_2 injection into a deep saline aquifer requires a large number of parameters that often are difficult to obtain and consequently present large uncertainties. Therefore, separating the most significant input parameters from the non-relevant, and assessing their relative contributions to the overall output uncertainty can be extremely helpful to focus research means effectively. To achieve this goal, global sensitivity analysis (GSA) may be used.

GSA (Saltelli, 2008) differs from typical sensitivity analysis methods in that it computes incremental ratios of a generic output state variable Y over an input parameter Z_i (i=1,2,…, n; where n is the total number of uncertain input parameters) such as $[Y(Z_i + \Delta Z_i) - Y(Z_i)]/\Delta Z_i$, and is able to explore all the space of the input parameters, also called input factors. Contrarily, sensitivity analysis computes these ratios as derivatives of Y with respect to Z_i (*i.e.* $\partial Y/\partial Z_i$) centered on a single data point of the input space.

GSA methods can be classified into two groups (Sudret, 2008): (1) regression-based methods, and (2) variance-based methods, better known as ANOVA techniques. The first group

includes, among others, standardized regression coefficient and partial correlation coefficient methods (Helton, 1993). The second group includes methods such as Sobol's indices (Sobol, 1993) and the Fourier amplitude sensitivity test (FAST) indices (Cukier et al., 1978; Saltelli et al., 1999; Saltelli, 2008).

In this study, we apply the extended Fourier Amplitude Sensitivity Test (FAST) introduced by Saltelli (1999). Extended FAST is a GSA variance-based method, whose main characteristic is it can capture not only the uncertain parameters having more influence on the variance of the model output, but also the interaction effect among the input parameters. For each uncertain parameter Z_i, the extended FAST method provides two sensitivity measures: the first-order index and the total effect index. The first-order index S_i represents the main effect contribution of each model input parameter Z_i to the variance of the generic model output Y (Δp_{iw} or $\%M_{leak}$). In practice, S_i quantifies how much the variance of Y would be reduced if the uncertain input parameter Z_i was fixed. This index is calculated as (Saltelli, 2008):

$$S_i = \frac{V[E(Y|Z_i)]}{V(Y)} \tag{2-23}$$

where $V(\cdot)$ indicates the variance operator and $E(Y|Z_i)$ indicates the expected value of Y conditioned to Z_i, and $V[E(Y|Z_i)]$ is the first-order effect.

Two or more input parameters present interaction when the sum of their first-order indices cannot explain their effect on Y. $V(Y)$ can thus be decomposed into first-order and higher-order effect terms:

$$V(Y) = \sum_{i=1}^{n} V_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} V_{ij} + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{l=j+1}^{n} V_{ijl} + \cdots \tag{2-24}$$

where $V_i = V[E(Y|Z_i)]$ is the first-order effect of Z_i, and $V_{ij} = V\left[E\left(Y\middle|Z_i, Z_j\right)\right] - V[E(Y|Z_i)] - V[E\left(Y\middle|Z_j\right)]$ is the second-order effect between parameters Z_i and Z_j, etc. The reader is referred

to Saltelli (Saltelli, 2008) to find equations for terms of order greater than second. If both sides of Equation (2-24) are divided by $V(Y)$, the following identity is obtained:

$$\sum_{i=1}^{n} S_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} S_{ij} + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{l=j+1}^{n} S_{ijl} + \cdots = 1 \tag{2-25}$$

where S_i is the first-order index of Z_i (Equation (2-23)), S_{ij} is the second-order index of parameters of Z_i and of Z_j, etc. It is possible to show that the total number of terms at the left-hand side of Equation (2-25) is $2^n - 1$, which increases exponentially with n. This makes the calculation of higher-order indices computationally intensive.

As an alternative to compute higher-order indices, GSA computes the total effect index S_{T_i} (i=1,2,…, n), which detects the interaction of the parameter Z_i with all other parameters and represents its total contribution to the model output. In other words, S_{T_i} is equal to the first-order index S_i plus the interaction of Z_i with other uncertain parameters, and is calculated as (Saltelli, 2008):

$$S_{T_i} = 1 - \frac{V[E(Y|\boldsymbol{Z}_{\sim i})]}{V(Y)} \tag{2-26}$$

where $\boldsymbol{Z}_{\sim i}$ is the vector including all input parameters but Z_i. For example, for a model composed by three uncertain parameters, Z_1, Z_2, and Z_3, the total effect index of parameter Z_1 is defined as:

$$S_{T_1} = S_1 + S_{12} + S_{13} + S_{123} \tag{2-27}$$

Consequently, the sum of higher-order effects S_{I_i} (i=1,2,…, n) can be defined as:

$$S_{I_i} = S_{T_i} - S_i \tag{2-28}$$

The index S_{I_i} quantifies the importance of the interaction of parameter Z_i with the all other input parameters. Consequently, if S_{I_i} is negligible, then the interaction of the uncertain

parameter Z_i with other parameters is non-relevant ($S_{I_i} \approx 0$ and $S_{T_i} \approx S_i$). On the other hand, the value of S_{T_i} provides information concerning the relevance of the input parameter Z_i. If S_{T_i} is zero or close to zero, then the parameter Z_i can be set to any value of its range of variability without having any impact on the output variance of the model.

Extended FAST is a Monte-Carlo based numerical procedure, where the variances and the conditional variances of the model response with respect to uncertain input parameters are estimated from the output of an ensemble of model runs. To apply the extended FAST and compute first-order and total effect indices of each parameter the SIMLAB package (SIMLAB, 2007) is employed. The total number of executions that the extended FAST method requires is equal to $D \cdot (2 + n)$, where D is the size of the ensemble used for each input parameter, which can range from a few hundreds to a few thousands. For example, for a number of 5 input parameters and an ensemble size of 1,000, extended FAST would require 7,000 runs. Inevitably, for a model with a large number n of input parameters, this method requires large ensembles of executions, and is viable only for simulation models that are not computationally intensive, as is the case of the multiphase flow simulator presented in Section 2.

5 Application to the Michigan Basin Test Site

5.1 Site description

The SA and GSA introduced in Sections 3 and 4 are applied to a geological test site located near the town of Thompsonville, MI. The storage formation proposed for GCS is embedded in the northern reef trend of the Michigan Basin. These reefs are evaporate-encased and, up until a few decades ago, significantly contributed to the production of hydrocarbons in Michigan. Most of these formations are associated with the reef buildups of Middle Silurian age.

Figure 2-1 shows a cross-section of the Michigan Basin in the area of interest with available log-wells.

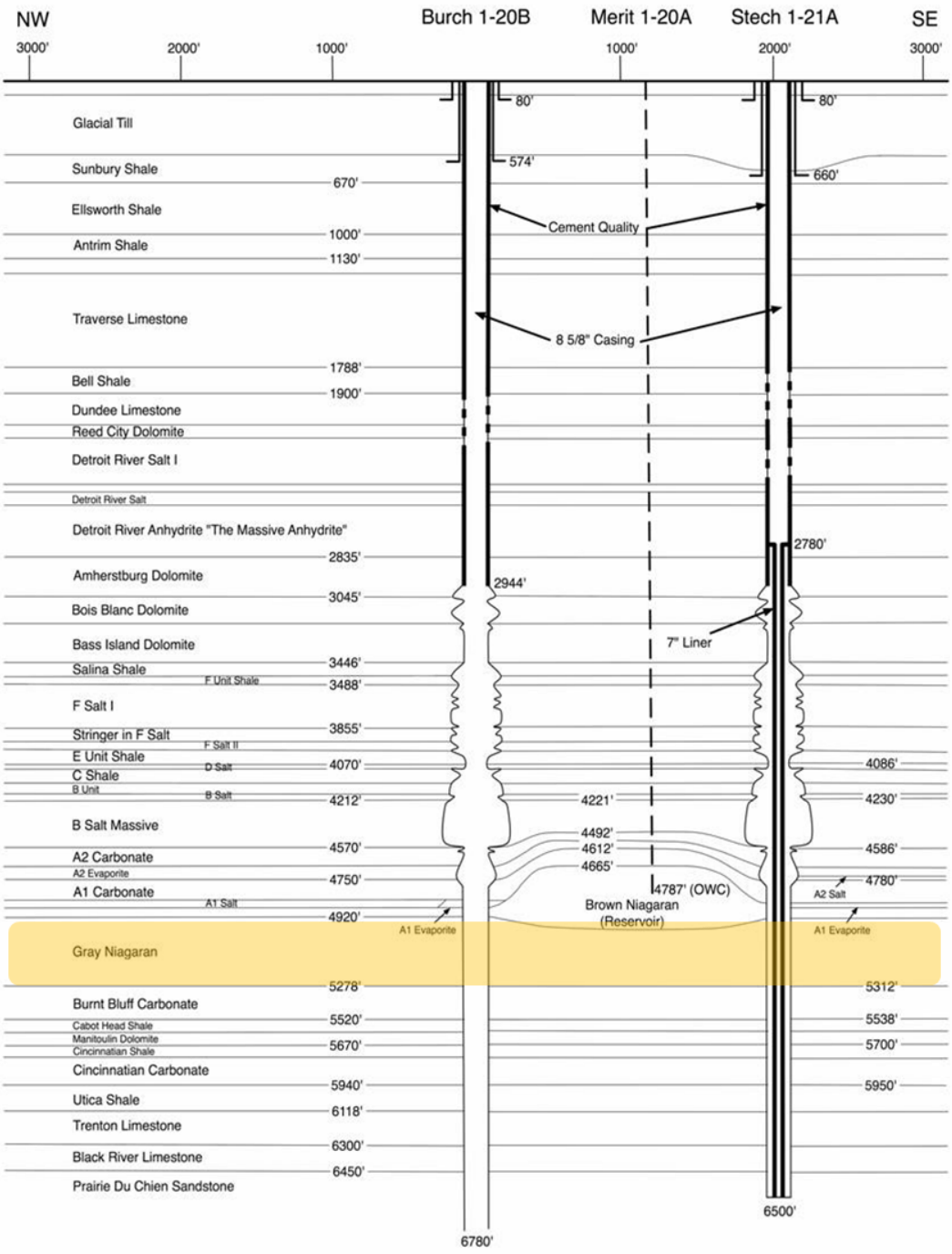

Figure 2-1. Cross-section of the Michigan Basin test site (adapted from (Turpening et al., 1992). The Gray Niagaran formation highlighted in yellow, is selected as potential candidate for GCS.

The Gray Niagaran formation, highlighted in yellow, lies below the Brown Niagaran pinnacle, a depleted oil reservoir currently used by Michigan Technological University for geophysical research (Figure 2-1). This formation is chosen as a candidate to store supercritical CO_2 because it is underneath the Brown Niagaran pinnacle, and thus the sealing capacity in that region is somehow warranted. In addition, the Gray Niagaran formation is already perforated by two exploration wells (Burch 1-20B and Stech 1-21A), which could possibly serve as CO_2 injection wells. The top and the bottom of this formation lie at a depth of 1,500 m and 1,619 m below ground, respectively. These characteristics make this formation a good candidate for storage of CO_2 in supercritical state.

To simplify the simulation of CO_2 injection, the system is modeled as an aquifer (the Gray Niagaran formation) confined by one sealing caprock (Evaporites), and another aquifer with lower permeability (Carbonate formation) located above the sealing caprock. Supercritical CO_2 is injected within the lower aquifer from a single well. The thicknesses of the Gray Niagaran formation and the overlying aquifer are 119 m and 35 m, respectively. The caprock has a thickness of 17 m and is assumed impermeable except where there are passive wells. The area of interest covers a horizontal extent of about 9,000 m × 9,000 m around the Brown Niagaran pinnacle and comprehends a total of 80 passive, and potentially leaky wells drilled across the Gray Niagaran formation. The locations of these wells have been obtained from DEQ (2013). If these wells are deteriorated or not well cemented, they may represent a pathway for upward leakage of both brine and CO_2 from the Gray Niagaran formation. In this study, all formations are assumed initially saturated with brine under hydrostatic pressure conditions.

Wells logs are available for the two boreholes shown in Figure 2-1. Log-porosity values are extracted from neutron porosity hydrogen index from the available logs (SCH, 1983;

Halliburton, 1990; SCH, 1991). Given the lack of data, permeabilities k, in millidarcy (mD; 1mD $\equiv 1\times10^{-15}$ m^2), for the candidate formation and the overlying aquifers are estimated from porosity φ as (Trebin, 1945):

$$k = 2e^{31.6\varphi} \quad \text{if } 100\varphi < 12\%$$
$$k = 4.94(100\varphi)^2 - 763 \quad \text{if } 100\varphi > 12\% \tag{2-29}$$

The residual saturation of brine, s_b^{res}, is assumed to be equal to 0.3 (Zhou et al., 2009).

In the analyses presented here, a reference case is considered with the hydro-geomechanical parameters provided in Table 2-1.

Table 2-1. Hydro-geomechanical parameters of the reference case. Parameters of this table remain unchanged (deterministic) unless the parameter of interest is considered uncertain.

Parameter	**Symbol**	**Value**	**Units**
Brine density	ρ_b	1,045	kg m^{-3}
CO_2 density	ρ_c	575	kg m^{-3}
Brine viscosity	μ_b	4.5×10^{-4}	Pa s
CO_2 viscosity	μ_c	4.6×10^{-5}	Pa s
System compressibility	c_{eff}	4.6×10^{-10}	Pa^{-1}
Injected aquifer porosity	φ_1	0.084	/
Overlying aquifer porosity	φ_2	0.05	/
Brine residual saturation	s_b^{res}	0.3	/
Injected aquifer permeability	k_1	2.8×10^{-14}	m^2
Overlying aquifer permeability	k_2	9.6×10^{-15}	m^2
Passive wells permeability	k_{pw}	1.0×10^{-14}	m^2

The uncertain parameters of interest are: permeability k_1 and porosity φ_1 of the injected aquifer, passive wells permeability k_{pw}, system compressibility c_{eff}, and brine residual saturation s_b^{res}. A PDF is prescribed for each of these parameters to represent their uncertainty for the candidate formation. These PDFs are given in Table 2-2.

Table 2-2. Probability distribution functions (PDFs) for uncertain parameters.

Parameter (unit)		Distribution	Median value	Log standard deviation	Minimum value	Maximum value	Realizations
Aquifer permeability (m^2)		lognormal	2.8×10^{-14}	0.5	-	-	1,000
Aquifer porosity (/)		uniform	-	-	0.05	0.35	1,000
Passive well permeability (m^2)	Case 1	lognormal	1.0×10^{-14}	1	-	-	1,000
	Case 2	bimodal	1.0×10^{-14}	-	$1.0\times10^{-17*}$	$1.0\times10^{-11*}$	1,000
	Case 3	bimodal	1.0×10^{-14}	-	$1.0\times10^{-16*}$	$1.0\times10^{-12*}$	1,000
System compressibility (Pa^{-1})		lognormal	1.0×10^{-9}	1	-	-	1,000
Brine residual saturation (/)		uniform	-	-	0.00	0.40	1,000

*50% chance for bimodal distributions

To characterize the permeability of passive wells, three different PDFs sharing the same median value are considered (see Table 2-2). In Case 1, a lognormal PDF with a median permeability of $1.0\text{x}10^{-14}$ m^2 and a log-standard deviation of 1 log-m^2 is adopted (Nordbotten et al., 2009). For Cases 2 and 3, bimodal distributions, in which each passive well permeability may assure two values, each with a 50% probability of occurrence. In Case 2, the minimum value of permeability assigned to well-cemented passive well is 1×10^{-17} m^2, and the maximum value of permeability assigned to a passive well is 1×10^{-11} m^2. In Case 3, the permeabilities for a well-cemented well and a leaky passive well are 1×10^{-16} m^2 and 1×10^{-12} m^2, respectively. In both cases, the permeability corresponding to a cemented well never exceeds the maximum

permeability recommended for a well-sealed passive well, which is 2×10^{-16} m^2 according to Kutchko et al. (2009).

In the SA, the 80 passive wells are considered individually, whereas in the GSA, these wells are clustered into 20 equivalent leakage pathways to reduce the number of input parameters and, consequently, the computational cost of the procedure. Figure 2-2 shows the position of the passive wells (DEQ, 2013) located in the area under consideration and the position of each cluster of wells.

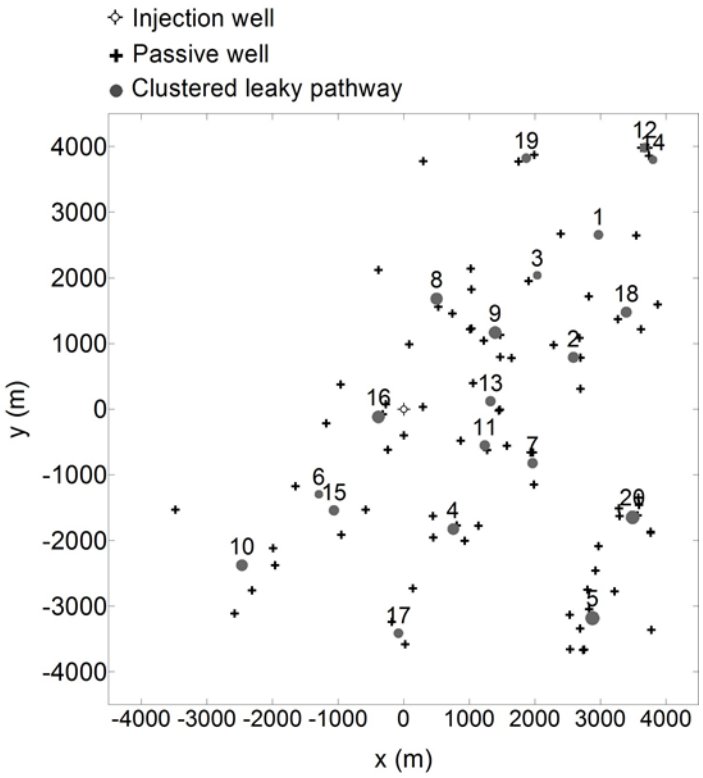

Figure 2-2. Location of the 80 passive wells (indicated as crosses) that reach the Niagaran formation and location of the 20 equivalent leakage pathways (indicated as circles) used in the GSA and obtained after clustering the 80 passive wells. The injection well is located at the center of the domain.

Well clusters are identified with an optimization procedure that minimizes the sum of the Euclidean distances of the passive wells forming a cluster and the cluster centroid. The equivalent leaky area considered for each cluster of wells is equal to the sum of the cross section

areas of the wells included in that group. Since the focus is on investigating the contribution of passive well permeability to the variance of the outputs, compared to the single contribution of individual passive well permeability to the output, it is expected that this clustering process will not affect results significantly.

Hence, in the GSA, n=24 input parameters are considered, characterized by the PDFs given in Table 2-2. These parameters are: the permeability k_1 and the porosity φ_1 of the aquifer, the permeability k_{pw} of the 20 clusters of passive wells (Figure 2-2), the system compressibility c_{eff}, and the brine residual saturation s_b^{res}. The GSA is applied to study the impact of these 24 parameters on the maximum fluid overpressure reached around the injection well Δp_{iw} (Equation (2-20)) and on the percent of CO_2 mass leakage $\%M_{leak}$ (Equation (2-22)).

Preliminary tests are run to figure out the minimum ensemble size beyond which CDFs remain substantially stationary. Based on the results of these tests (*i.e.* the SA and the GSA) sample sizes of N_{MC}=1,000 and D=1,000 are selected respectively. Thus, in the GSA we use a total of 26,000 simulations ($D \cdot (2 + n)$, see Section 4).

5.2 Results and discussion

This section includes first a preliminary analysis carried out to select the CO_2 injection rate and the duration of CO_2 injection. Next, we present and discuss the results of the stochastic and global sensitivity analyses applied to the Michigan Basin deep saline aquifer introduced in Section 5.1.

5.2.1 Selection of CO_2 injection rates

To choose an appropriate injection rate, a set of multiphase flow simulations is performed for a hypothetical system representative of the Michigan Basin. This hypothetical system

consists of the same permeable formations and caprock considered for this study plus two overlying aquifers of thickness equal to 37 m and 75 m, underlain by two 18-m and 109-m thick aquitards, corresponding to the formations illustrated in Figure 2-1. Deterministic values of porosity and permeability are assigned to each layer based on values from available log-wells (SCH, 1983; Halliburton, 1990; SCH, 1991). The aquifers are named L1, L2, L3 and L4 from the deepest to the shallowest. Aquitards are assumed impermeable except where passive wells are present. 24 hypothetical leaky passive wells are included in a 5×5 square regular grid, with the injection well positioned at its center. The N-S and E-W distances between these passive wells are equal to 1 km. Three main scenarios with different CO_2 mass injection rates Q_m and durations are simulated. Table 2-3 summarizes these scenarios. Note that the final injected mass of CO_2 is the same in all scenarios and equal to about 63 Mt.

Table 2-3. CO_2 injection rates and duration of injection of the multiphase flow simulations of the hypothetical system representative of the Michigan Basin.

Scenario	Q_m **(kg/s)**	t_{end} **(years)**
S1	100	20
S2	50	40
S3	33.33	60

Fluid overpressure values Δp_{iw} nearby the injection well at final time t_{end} (Equation (2-20)) resulting from multiphase flow simulations for the three scenarios of Table 2-3 are reported in Figure 2-3. Scenario S1 produces the greatest overpressure around the injection well with a value of 78 bar (1 bar ≡ 1×10^5 Pa) at t_{end}= 20 years. On the other hand, scenario S3 produces the lowest overpressure nearby the injection well with a value of 33 bar at t_{end}= 60

years. Scenario S2 produces intermediate results between S1 and S3, with a Δp_{iw}= 46 bar at t_{end}= 40 years.

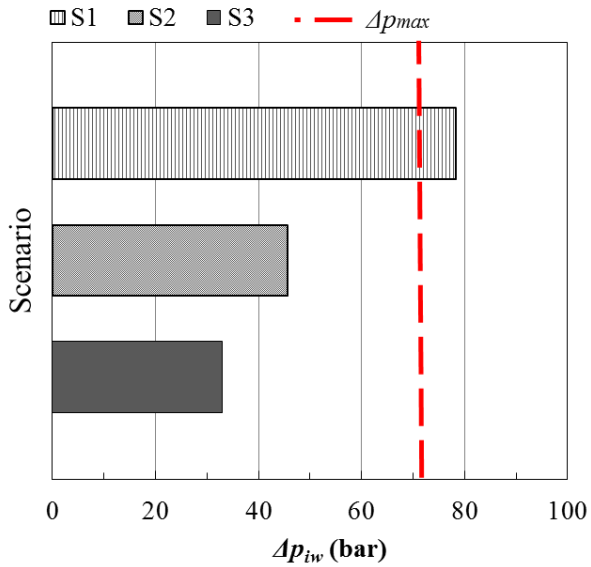

Figure 2-3. Fluid overpressure results nearby the injection well from the multiphase flow simulations of the hypothetical system based on the Michigan Basin for scenarios S1, S2, and S3 (Table 2-3). The vertical dashed line represents the maximum overpressure allowed at the injected formation.

In Figure 2-3, the vertical dashed line represents the maximum admissible overpressure, Δp_{max}, that is, the overpressure threshold beyond which the caprock is likely to fissure. This threshold value is calculated as (Teatini et al., 2010):

$$\Delta p_{max} = \sigma_z' \frac{\upsilon}{1-\upsilon} \tag{2-30}$$

where: υ is the Poisson ratio and σ_z' is the estimated effective vertical stress at the caprock depth under pressostatic undisturbed conditions. Assuming a Poisson ratio of υ=0.25, Δp_{max} is estimated to be equal to 72 bar. The results in Figure 2-3 show that the maximum overpressure allowed is exceeded only in Scenario S1. In practice, these results indicate that S2 and S3 would be the safest for the GCS system not to fissure the sealing formation.

Figure 2-4 shows results concerning the amount of CO_2 mass leaked into the overlying formations (L2, L3, and L4) estimated by multiphase flow simulation of the considered hypothetical system. This figure presents the $\%M_{leak}$ that escapes from the injected formation L1 and is stored into the overlying formations, L2, L3, and L4. Scenario S1 produces the lowest total percentage of CO_2 leaked from the injected aquifer with $\%M_{leak}$=0.014% at t_{end}= 20 years, while S3 produces the highest leakage with $\%M_{leak}$=0.020% at t_{end}= 60 years. Although scenario S1 produces the highest fluid overpressure nearby the injection well (Figure 2-3), it has the lowest CO_2 leakage. It is, however, important to observe that the resulting values of $\%M_{leak}$ for the three scenarios are very similar to one another. From Figure 2-4, it can also be noticed that in all scenarios S1, S2, and S3, most of the leaked CO_2 tends to be stored in the lower aquifer L2.

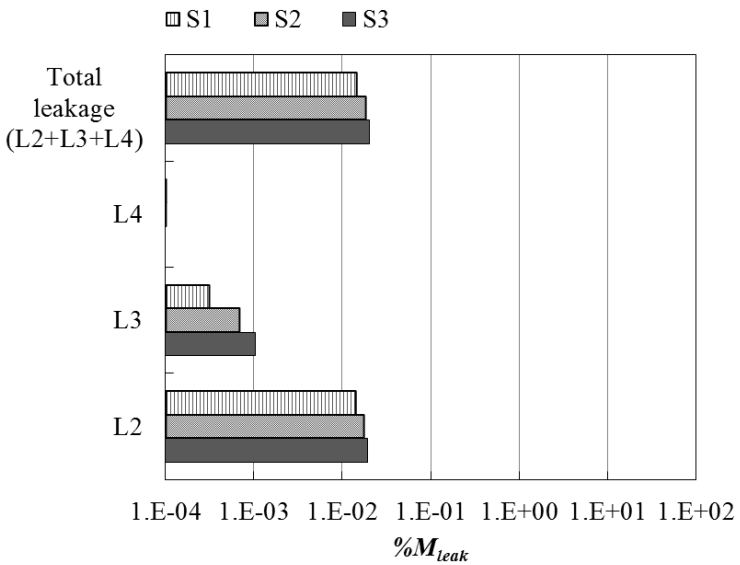

Figure 2-4. Percent of CO_2 mass leaked to overlying formations (L2, L3, and L4) from the multiphase flow simulations of the hypothetical system based on the Michigan Basin for scenarios S1, S2, and S3 (Table 2-3).

The fact that scenario S3 produces greater leakage than scenario S1 is explained by the CO_2 plume spread. In both scenarios, the final shape and distribution of the CO_2 plume are very

similar since they only depend on the mass of CO_2 injected into the formation and the mass that escapes into overlying formations (Equation (2-6)). In scenario S1, the plume advances faster since the injection rate is greater than in scenario S3, therefore the plume reaches passive wells sooner than in scenario S3. However, the period of time in which a passive well is exposed to the CO_2 plume in scenario S1 is shorter than the period of time in which the same passive well is exposed to the CO_2 plume in scenario S3.

Figure 2-5 shows the temporal evolution of CO_2 mass leakage through one passive well located at a distance of 1,000 m from the injection well for scenarios S1 and S3. In scenario S1, the CO_2 plume reaches the passive well at t= 2 years, whereas in scenario S3 at t= 6 years. Due to different injection rates and simulation times, in scenario S1 the leakage occurs over 18 years, while in scenario S3 it occurs over 54 years, with a greater CO_2 mass leakage occurring in the latter.

In summary, these tests show that scenario S3 produces significantly lower fluid overpressure nearby the injection well than the other two scenarios. At the same time, the differences in $\%M_{leak}$ at the end of the injection time may be considered negligible. Therefore, scenario S3 will be the only one used in the following analyses.

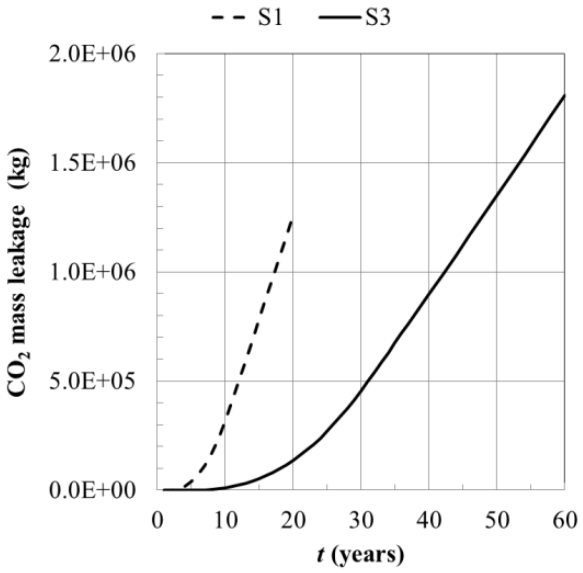

Figure 2-5. Comparison of CO_2 mass leakage through one passive well located at 1,000 m from injection well for scenarios S1 (dashed line) and S3 (solid line). See Table 2-3 for descriptions of these two scenarios.

5.2.2 Stochastic Analysis Results

In this section, we present results obtained from the SA. We study the effects of uncertainty on aquifer permeability k_1, aquifer porosity φ_1, passive well permeability k_{pw}, system compressibility c_{eff}, and brine residual saturation s_b^{res} on the two state variables of interest: Δp_{iw} (Equation (2-20)) and $\%M_{leak}$ (Equation (2-22)).

Results from stochastic flow simulations are used to derive CDFs (see Section 3) for these state variables. These CDFs may be used to estimate the probability of fracturing the caprock, and the probability of CO_2 mass leakage not to exceed given threshold values. To analyze the risk of fracturing the caprock formations we consider "safe" conditions when the 95th percentile of Δp_{iw} is below Δp_{max} as estimated by Equation (2-30). To investigate the risk of CO_2 mass leakage, we consider "safe" conditions when the 95th percentile of $\%M_{leak}$ does not exceed limits derived from maximum CO_2 leakage rates of 1% per one year as suggested by

Pacala (2003). To have a consistent metric with this threshold and $\%M_{leak}$ results, the thresold must be converted to $\%M_{leak}$, resulting in 1%. It is important to emphasize that this estimate is rather conservative since the limit proposed by Pacala (2003) represents CO_2 leakage rates back to the atmosphere, whereas we consider the mass of CO_2 that escapes the target storage formation as leaked and do not account for the processes of storage and attenuation that CO_2 may undergo within the overburden formations.

Effect of aquifer permeability. The permeability of the aquifer is expected to have a significant influence on the fluid overpressure, with low permeability values producing large overpressure. SA results for aquifer permeability as the uncertain input parameter are shown in Figure 2-6.

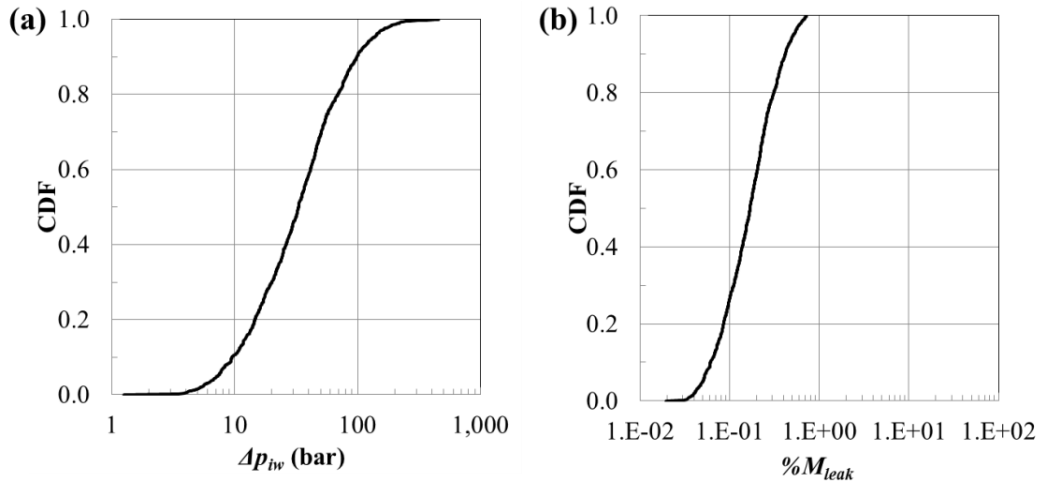

Figure 2-6. CDF of the (a) fluid overpressure nearby the injection well, and (b) $\%CO_2$ mass leakage associated with the uncertainty on aquifer permeability.

Figure 2-6a shows the CDF of Δp_{iw} obtained by sampling the aquifer permeability from the PDF described in Table 2-2. Aquifer permeability uncertainty affects significantly the spread of the CDF. Its range varies between 1 bar and 450 bar (more than two orders of magnitude). Figure 2-6b shows the CDF of $\%M_{leak}$ obtained by assuming uncertain aquifer permeability (Table 2-2). This figure shows that uncertainty on k_1 has also a significant influence on CO_2

leakage, with a CDF whose spread spans over more than two orders of magnitude, from a minimum value of 0.02% to a maximum value of 0.72%.

Figure 2-7 shows the percent of CO_2 mass leaked as a function of aquifer permeability. In general, lower permeability values correspond to increased mass leakage since higher fluid overpressures are obtained, which ease the escape of CO_2 through passive wells. On the contrary, the CO_2 plume advances more easily through the injected formation with higher permeabilities, easing its storage in the injected aquifer.

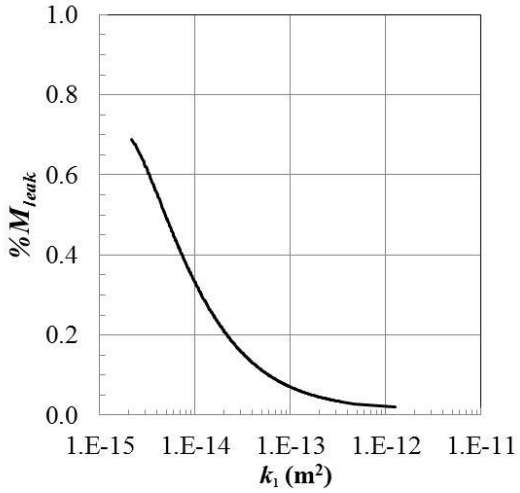

Figure 2-7. %CO_2 mass leakage as a function of aquifer permeability.

Effect of aquifer porosity. SA results for uncertain aquifer porosity (Table 2-3) are presented in Figure 2-8a for Δp_{iw} and Figure 2-8b for $\%M_{leak}$, respectively. The CDF in Figure 2-8a shows that uncertainty on formation porosity has a weak impact on the statistical variability of the maximum fluid overpressure at the injection well. Figure 2-8a reveals that Δp_{iw} varies between approximately 30 and 41 bars, resulting in a very small spread of the CDF. In general, larger porosities produce larger values of overpressure. Propagation of the overpressure pulse depends on porosity (Equation (2-2)), in such a way that the same amount of CO_2 occupies a smaller region of the aquifer, hence retarding the attenuation of the overpressure pulse.

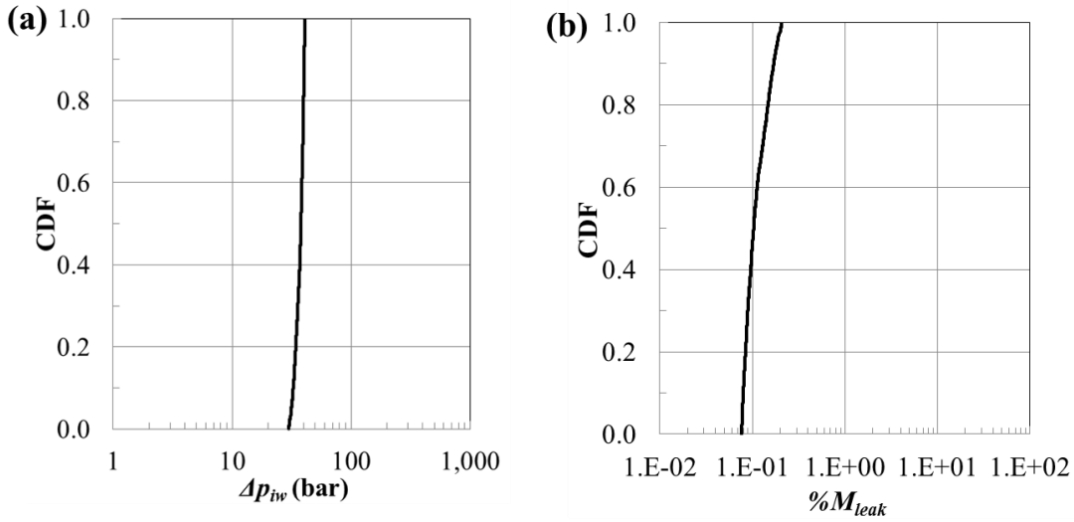

Figure 2-8. CDF of the (a) fluid overpressure nearby the injection well, and (b) %CO_2 mass leakage associated with the uncertainty on aquifer porosity.

Uncertainty on formation porosity has also a minor effect on the variability of CO_2 mass leakage. Smaller porosities are generally expected to result in larger leakage rates. Indeed, the shape of the plume depends on porosity (Equation (2-6)) and lower porosities result in faster plume propagation and a higher likelihood of encountering leakage pathways. However, Figure 2-8b shows that the variability of CO_2 mass leakage is relatively contained since the CDF spread is less than one order of magnitude. Comparison of Figure 2-6 and Figure 2-8 indicates that uncertainty on porosity φ_1 has a much lower influence on fluid overpressure and CO_2 mass leakage than the uncertainty on injected formation permeability k_1.

Effect of passive well permeability. Stochastic simulation results indicate that the fluid overpressure nearby the injection well is rather insensitive to passive well permeabilities k_{pw}. The CDF of Δp_{iw} is consequently not presented here. This result can be easily explained by noting that, for the geological setting investigated here (Section 5.1), fluid overpressure depends upon "local" conditions around the injection well, such as injection rate and formation permeability, rather than on conditions in regions of the domain "away" from the well.

On the other hand, the SA shows that uncertainty on leakage passive well permeability has a strong impact on CO_2 mass leakage.Figure 2-9 displays the CDFs of $\%M_{leak}$ corresponding to the three PDFs for the passive well permeability k_{pw} given in Table 2-2.

Although the three PDFs have the same median value of k_{pw}, the CDFs for CO_2 mass leakage are substantially different. In Figure 2-9 one may observe that in Case 1 (Table 2-2, lognormal PDF) the probability of CO_2 leakage is typically the smallest except for values of k_{pw} sampled from the upper tail of its distribution. Case 1 also presents the largest CDF spread (more than one order of magnitude), whereas in Cases 2 and 3 the spread of the CDF is hardly noticeable. In Cases 2 and 3, k_{pw} is sampled from bimodal distributions (Table 2-2) characterized by two equally likely values, 1×10^{-17} m^2 and 1×10^{-11} m^2 in Case 2, and 1×10^{-16} m^2 and 1×10^{-12} m^2 in Case 3. Figure 2-9 shows that CO_2 mass leakage is probabilistically larger in Case 2, which indicates that the intensity of leakage is largely affected by the presence of highly permeable discontinuities.

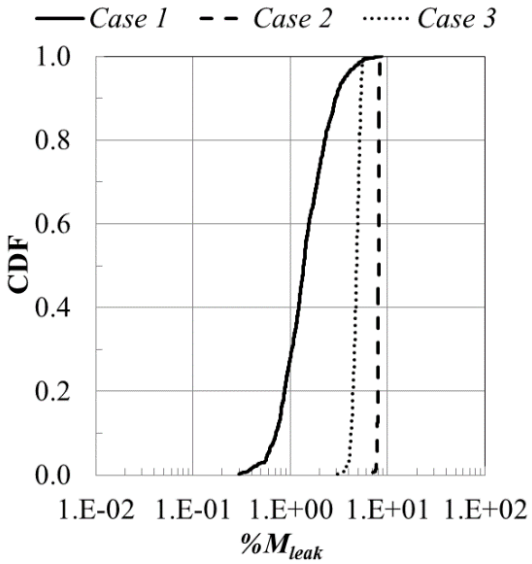

Figure 2-9. CDF of $\%CO_2$ mass leakage associated with the uncertainty on passive well permeability. See Table 2-2 for descriptions of Case 1 – Case 3.

Effect of system compressibility. The system compressibility is expected to have an impact on the fluid overpressure and mass leakage, with low values of c_{eff} producing greater values of fluid overpressure Δp_{iw}, and consequently higher $\%M_{leak}$. SA results under uncertain c_{eff} (Table 2-2) are presented in Figure 2-10. Figure 2-10a shows the CDF of Δp_{iw}, which varies between 1 and 68 bar.

The CDF of $\%M_{leak}$ shown in Figure 2-10b, indicates that system the compressibility has a significant impact on the variability of CO_2 mass leakage (about two orders of magnitude). However, the comparison of Figure 2-6b and Figure 2-10b reveals that the spread of the CDF is smaller than that obtained with uncertain aquifer permeability.

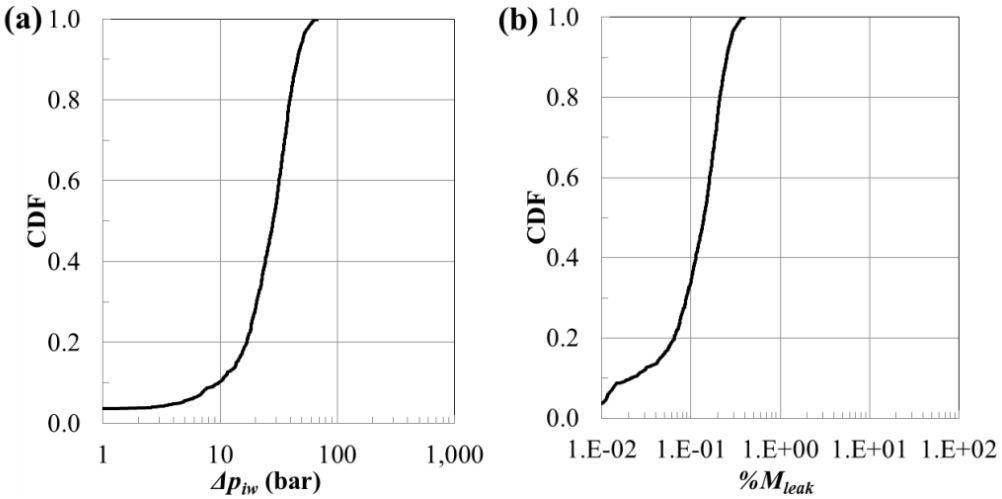

Figure 2-10. CDF of the (a) fluid overpressure nearby the injection well, and (b) $\%CO_2$ mass leakage associated with the uncertainty on system compressibility.

Figure 2-11 shows fluid overpressure and percent of CO_2 mass leakage as functions of system compressibility, suggesting that lower values of system compressibility lead to larger fluid overpressure and larger leakage. In general, larger values of c_{eff} results in lower values of Δp_{iw} since the propagation of the pressure pulse depends on system compressibility (Equations (2-2) and (2-5)), and the outer boundary of the pressure pulse will be smaller (Equation (2-5)).

Hence, a smaller region of the aquifer accepts the same amount of CO_2 because of the larger storage capacity deriving from the deformability of the porous medium.

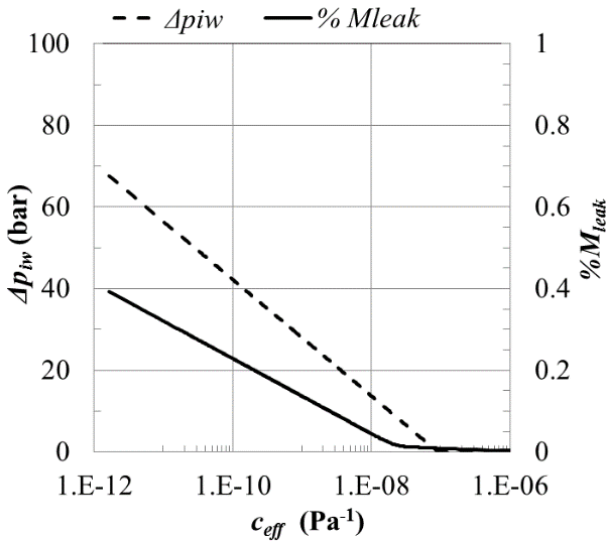

Figure 2-11. Fluid overpressure (left vertical axis) and %CO_2 mass leakage (right vertical axis) as functions of system compressibility.

Effect of brine residual saturation. SA results show that uncertainty from brine residual saturation s_b^{res} has a negligible effect on the maximum fluid overpressure Δp_{iw}. Similar results are obtained for $\%M_{leak}$, which is not significantly affected by the uncertainty on s_b^{res}. For these reasons, the CDFs of Δp_{iw} and $\%M_{leak}$ are not shown here. In general, greater s_b^{res} values result in slightly larger leakage rates. Certainly, the extension of the CO_2 plume depends on brine residual saturation (Equation (2-6)) and greater values of s_b^{res} result in a more pronounced plume propagation and a higher likelihood of encountering leakage pathways.

General considerations from the SA applied to the Michigan Basin test site. In order to make general considerations on the feasibility of GCS for the Michigan Basin test site, a SA under Scenario S3 is carried out considering all parameters of Table 2-2 uncertain at the same time. For passive well permeability, the PDF of Case 1 (Table 2-2) is considered since is the

situation that produces the greatest CDF spread as well as the largest values of $\%M_{leak}$. The CDFs of Δp_{iw} and $\%M_{leak}$ calculated from this SA are given in Figure 2-12.

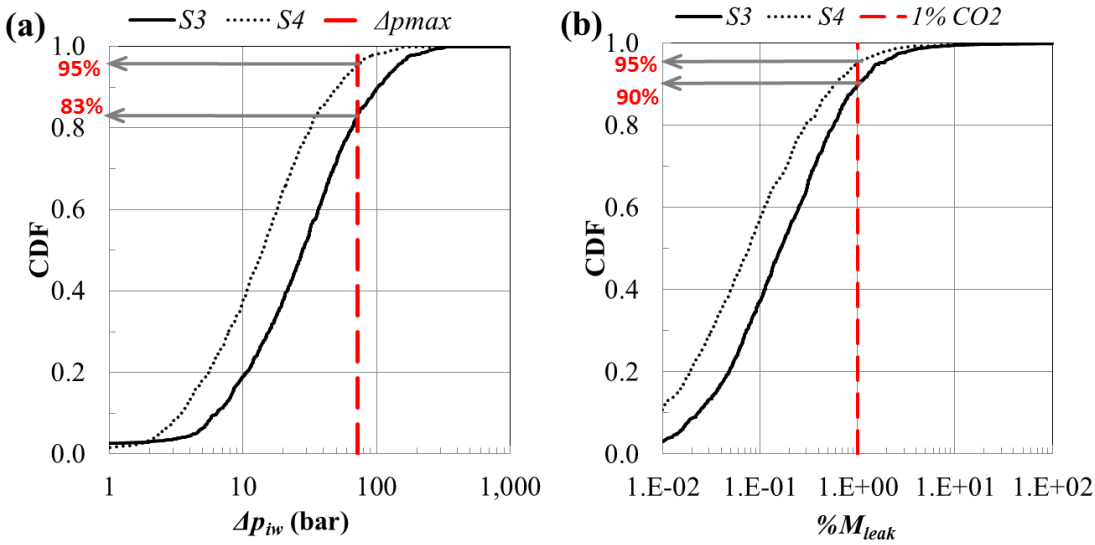

Figure 2-12. CDF of the (a) fluid overpressure nearby the injection well, and (b) $\%CO_2$ mass leakage associated with the uncertainty on all uncertain parameters for scenarios S3 and S4. The vertical dashed line in (a) represents the maximum overpressure allowed at the injected formation and in (b) the $\%CO_2$ mass leakage threshold.

The solid black line in Figure 2-12a represents the CDF of Δp_{iw} under Scenario S3. In the same graph, the vertical dashed line represents the maximum fluid overpressure Δp_{max}=72 bar allowed in the formation in order not to fracture the caprock (Equation (2-30)). The intersection of this vertical line with the CDF of Δp_{iw} shows that in scenario S3 there is an 83% probability of not exceeding Δp_{max}.

Likewise, the solid black line in Figure 2-12b represents the CDF of $\%M_{leak}$ under Scenario S3. The 1% CO_2 mass leakage threshold defined by Pacala (2003) is represented by the vertical dashed line. Based on the CDF of $\%M_{leak}$, there appears to be a 90% probability of not exceeding such threshold.

In order to increase both the 83% probability of not fissuring the caprock and the 90% probability of not exceeding the 1% CO_2 mass leakage threshold to 95%, a new injection

scenario S4 is investigated. In this scenario, the total amount of injected CO_2 is reduced by 70%, with an injection rate Q_m= 14 kg/s and an injection period t_{end}= 40 years. The SA for Scenario S4 leads to the CDFs of Δp_{iw} and $\%M_{leak}$ represented by the dotted profiles in Figure 2-12a and Figure 2-12b, respectively. Under this new scenario, the probabilities of not exceeding both Δp_{max}= 72 bar and $\%M_{leak}$= 1% are increased to 95%.

The need to reduce the total mass injected from Scenario S3 to S4 in order to meet the prescribed safety constraints on Δp_{iw} and $\%M_{leak}$ is due to a "conflict" existing between these constraints when injecting a given mass of CO_2 ($\rho_c Q_m t_{end}$). Indeed, increasing the CO_2 injection rate Q_m and decreasing the injection time t_{end} is beneficial towards reducing $\%M_{leak}$, but also increases the probability that Δp_{iw} exceeds Δp_{max}. Vice versa, decreasing Q_m and increasing t_{end} reduces the probability of fracturing the caprock, but increases the probability of violating the 1% threshold for $\%M_{leak}$. Therefore, in order to comply with the requirement of both safety constraints, Δp_{max}= 72 bar and $\%M_{leak}$= 1%, the total mass of injected CO_2 must be necessarily reduced by adequately decreasing both the injection rate Q_m and the injection time t_{end}.

5.2.3 Results of Global Sensitivity Analysis

In this section, we present results of the application of the extended FAST methodology to the Michigan Basin deep saline aquifer. The sensitivities of the 24 uncertain parameters (aquifer permeability, aquifer porosity, permeability of 20 potential passive well pathways, system compressibility, and brine residual saturation) on the variability of the outputs Δp_{iw} (Equation (2-20)) and $\%M_{leak}$ (Equation (2-22)) are studied. For the permeability of passive wells, the PDF of Case 1 (Table 2-2) is chosen since, in the SA, this has been shown to produce the largest spread of the $\%M_{leak}$ CDF (see Section 5.2.2).

The GSA results are presented in Figure 2-13 and in Table 2-4. Figure 2-13 shows pie charts for Δp_{iw} and $\%M_{leak}$, where each total effect index S_{T_i} (Equation (2-26)) is represented by the "normalized" percentage:

$$\%S_{T_i} = \frac{S_{T_i}}{\sum_{i=1}^{n} S_{T_i}} 100 \tag{2-31}$$

where n is the total number of uncertain input parameters, in this case equal to 24. In this figure, the combined effect of the 20 leakage pathways is grouped and denoted as k_{pw1-20}. Figure 2-13 displays the first-order sensitivity indices S_i as a percentage of the total effect indices S_{T_i} for both Δp_{iw} and $\%M_{leak}$.

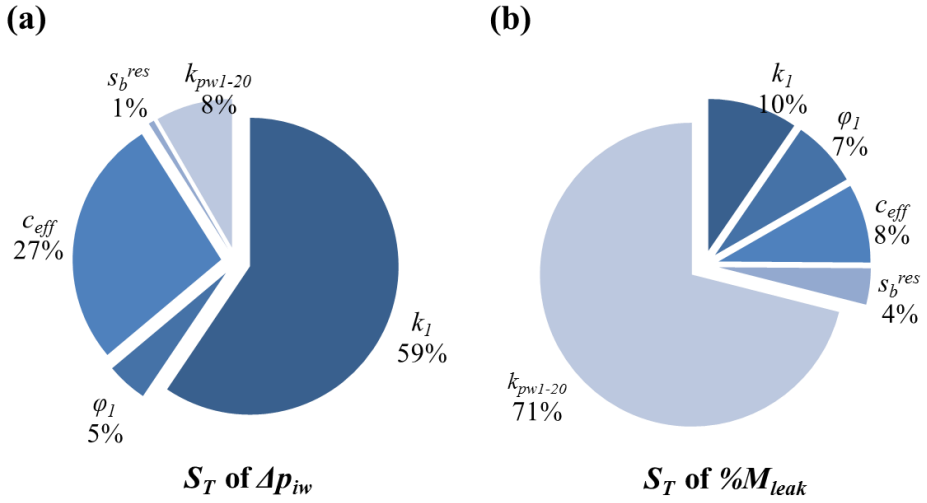

Figure 2-13. Total effect indices normalized after applying extended FAST method on: (a) maximum fluid overpressure at the vicinity of the injection well, and (b) $\%CO_2$ mass leakage.

Fluid overpressure nearby the injection well. Figure 2-13a illustrates the normalized total effect indices $\%S_{T_i}$ for the fluid overpressure nearby the injection well. This figure shows that the variability of Δp_{iw} is mainly influenced by only two parameters, that is, the aquifer permeability k_1 and the system compressibility c_{eff}, which altogether account for about 87% of the overall Δp_{iw} variance. Of this 87%, 59% is due to k_1 and 28% is due to c_{eff}.

Table 2-4. First-order effect as a percentage of the total effect obtained applying the Extended FAST method for fluid overpressure at the vicinity of the injection well, and %CO_2 mass leakage.

Uncertain parameter	%S_i of S_{Ti} Δp_{iw}	%S_i of S_{Ti} %M_{leak}	Uncertain parameter	%S_i of S_{Ti} Δp_{iw}	%S_i of S_{Ti} %M_{leak}
k_1	68.6	19.4	k_{pw9}	0.7	4.8
φ_1	49.8	12.5	k_{pw10}	0.0	0.6
c_{eff}	37.1	20.8	k_{pw11}	0.6	8.3
s_b^{res}	27.7	3.1	k_{pw12}	0.7	1.9
k_{pw1}	1.0	1.0	k_{pw13}	0.4	2.4
k_{pw2}	0.4	1.0	k_{pw14}	0.7	2.3
k_{pw3}	0.1	1.6	k_{pw15}	0.3	3.4
k_{pw4}	0.6	1.2	k_{pw16}	1.9	16.4
k_{pw5}	0.4	0.2	k_{pw17}	1.0	2.1
k_{pw6}	0.3	12.3	k_{pw18}	0.6	1.6
k_{pw7}	0.1	0.7	k_{pw19}	0.3	2.1
k_{pw8}	0.4	2.9	k_{pw20}	0.8	0.7

The prominent influence of these two parameters is somehow expected since the propagation of the pressure pulse is mainly governed by the aquifer permeability and the system compressibility (Equations (2-2) and (2-5)). When k_1 has a large value, the overpressure pulse can propagate easily through the injected formation moving away from the injection well and producing lower Δp_{iw} values and vice versa. When c_{eff} is larger the aquifer exhibits a larger capacity and consequently a lower increase in pressure throughout the domain.

Figure 2-13a indicates that the porosity has a small impact on the variability of Δp_{iw}, with a $\%S_{T_{\varphi 1}}$ equal to 4%. Residual saturation accounts for only 1% of the total variance

(%$S_{T_{s_b^{res}}}$= 1%). The compound effect of leakage permeabilities, %$S_{T_{k_{pw1-20}}}$, accounts for 8%, with each individual leakage pathway permeability accounting for less than 1%. These results indicate that the total effect indices of porosity, brine residual saturation, and pathway permeability are negligible, so that their uncertainty has a limited impact on the variability of the fluid overpressure nearby the injection well.

In Table 2-4 one may observe that only aquifer permeability k_1 affects the variability of Δp_{iw}, mostly through the first-order index, with S_{k_1}= 68.6% of the total effect index. The contribution of system compressibility c_{eff} to the variability of Δp_{iw} derives from the interaction with other parameters ($S_{c_{eff}}$= 37.1% < $S_{I_{c_{eff}}}$= 62.9%). The contribution of porosity φ_1 is about 50% from the first-order effect and 50% from higher-order effects. The contribution of brine residual saturation s_b^{res} and leakage pathway permeabilities k_{pw_1-20} to the total effect derives from the interaction with other parameters. It must be pointed that the respective total effect indices of φ_1, s_b^{res}, and k_{pw_1-20} are insignificant.

CO_2 mass leakage. Figure 2-13b illustrates results of total effect index normalized on %M_{leak} variance. This figure shows that the spread of the CO_2 mass leakage output is mainly influenced by the passive well permeability as a group with %$S_{T_{k_{pw1-20}}}$= 71%. However, the maximum total effect index coming from an individual pathway permeability cluster is 8%. The clusters of passive wells located closer to the injection well (clusters 6, 8, 11, 13, and 16 in Figure 2-2) have the largest values of the total effect index. The total effect index normalized of the aquifer permeability is %$S_{T_{k1}}$= 10%, followed by the system compressibility with %$S_{T_{c_{eff}}}$= 8%, and the aquifer porosity with %$S_{T_{\varphi 1}}$= 7%. Therefore, the aquifer permeability has a larger total effect index than the pathway permeability of any of the 20 clusters has. The brine residual

saturation contributes to the $\%M_{leak}$ variability for about 4%. Although the contribution of s_b^{res} to the $\%M_{leak}$ variability is greater than the contribution to the Δp_{iw} variability, it is less important than the effect due to other uncertain parameters. Therefore, the total effect index for s_b^{res} can be considered negligible. In other words, any value of s_b^{res} selected from the PDF presented in Table 2-2 would produce the same variability of CO_2 mass leakage.

Table 2-4 lists the first-order sensitivity index for each of the 24 uncertain parameters as a percentage of the total effect on $\%CO_2$ mass leakage. One can observe that the main contribution to the $\%M_{leak}$ variability comes from higher-order effects. Aquifer permeability and system compressibility have the largest contribution from the first-order index with a value of about 20%. Higher-order effects from porosity, residual saturation, and pathways permeability have more impact on the variability of $\%M_{leak}$ than their respective first-order effects. The leakage pathway permeability that presents the greatest contribution from the first-order sensitivity index is k_{pw_16} with a value of 16%. Indeed this is the passive well cluster closest to the injection well, which shows that the location of leakage pathways is an important component on the contribution to $\%M_{leak}$ variability.

General considerations from the GSA applied to the Michigan Basin test site. GSA results for fluid overpressure confirm observations already made in the SA. A large portion of the Δp_{iw} variability is attributed to only two parameters: aquifer permeability and system compressibility. By far, aquifer permeability is the most influential parameter as it ranks in first position with a normalized S_T equal to 59%. In order to significantly reduce the prediction of Δp_{iw}, acquiring accurate data of aquifer permeability is necessary.

From the GSA results for $\%M_{leak}$, one can conclude that aquifer permeability, system compressibility, and pathway permeability have the most significant impact on the variability of

the output. Aquifer porosity and brine residual saturation also show an impact, although this is minor in comparison to the other parameters. Location of leakage pathways closer to the injection well shows a significant effect on $\%M_{leak}$ with significantly higher first-order indices respect to passive wells located farther away. Therefore, when interested in studying uncertainty and risk of CO_2 leakage, an effort to acquire data concerning aquifer permeability, system compressibility, and location and permeability of potential leakage pathways is essential to reduce the uncertainty in the simulation of $\%M_{leak}$.

6 Summary and Conclusions

In this work, we analyzed the variability of fluid overpressure in proximity of injection wells and CO_2 mass leakage of a candidate site for GCS located within the Michigan Basin. This study relied on a stochastic analysis and a global sensitivity analysis accounting for the uncertainty on the following: permeability and porosity of injected aquifers, permeability of passive wells, system compressibility, and brine residual saturation. From the investigation of potential injection scenarios, it was observed that lower injection rates with longer injection times reduced the probability of producing excessive fluid overpressures in the injected aquifer. As far as CO_2 mass leakage is concerned, there was a small difference among these scenarios. Therefore, injection of CO_2 at low rates and protracted for a longer period of time appears to be the most convenient policy for the safety of the GCS system.

The stochastic analysis showed that the most influential parameter on both fluid overpressure and CO_2 mass leakage is the aquifer permeability, followed by the system compressibility, although with less intensity. Fluid overpressure in proximity of injection well seems unaffected by uncertainty on porosity, permeability of passive wells, or brine residual

saturation. On the other hand, CO_2 mass leakage is shown to be particularly sensitive to passive well permeability and the type of statistical distribution used to characterize uncertainty in it. The stochastic analysis also revealed that constraints on maximum overpressure and maximum leakage are competing against one another when injecting the same mass of CO_2.

Results from the extended FAST global sensitivity analysis confirmed the influences already observed in the stochastic analysis. Aquifer permeability and system compressibility had an impact on both the variability of the fluid overpressure and CO_2 mass leakage. Porosity had a greater impact on CO_2 mass leakage; however its impact is relatively minor. When studying the variability on the fluid overpressure, permeability of the leakage pathways and brine residual saturation resulted to be non-relevant parameters. Therefore, any value included in the variability studied could be assigned to k_{pw} and s_b^{res}, and results on fluid overpressure nearby the injection well would not be affected. On the other hand, the effect of leakage pathways with respect to the variability on CO_2 mass leakage is significant and cannot be neglected.

From this analysis, we can conclude that efforts to obtain further information about influent parameters, such aquifer permeability and system compressibility are necessary when studying their impact on fluid overpressure and CO_2 mass leakage for the test site under consideration. In addition, recollection of leakage pathway information especially from pathways located closer to the injection well is needed for quantifying potential CO_2 mass leakages with higher degree of confidence.

7 References

Aoyagi, R., Kitamura, O., Itaoka, K., Igawa, S., Suzuki, S., 2011. Study on role of simulation of possible leakage from geological CO_2 storage in sub-seabed for environmental impact assessment. Energy Procedia 4, 3881-3888.

Bachu, S., 2003. Screening and ranking of sedimentary basins for sequestration of CO_2 in geological media in response to climate change. Environmental Geology 44, 277-289.

Bergman, P.D., Winter, E.M., 1995. Disposal of carbon-dioxide in aquifers in the US. Energy Conversion and Management 36, 523-526.

Celia, M.A., Nordbotten, J.M., 2009. Practical Modeling Approaches for Geological Storage of Carbon Dioxide. Ground Water 47, 627-638.

Celia, M.A., Nordbotten, J.M., Bachu, S., Dobossy, M., 2009. Risk of leakage versus depth of injection in geological storage. Energy Procedia 1, 2573-2580.

Celia, M.A., Nordbotten, J.M., Court, B., Dobossy, M., Bachu, S., 2011. Field-scale application of a semi-analytical model for estimation of CO_2 and brine leakage along old wells. International Journal of Greenhouse Gas Control 5, 257-269.

Cody, B., Baù, D., González-Nicolás, A., 2014. Improved Semi-Analytical Simulation of Geological Carbon Sequestration (in review). Computational Geosciences.

Cukier, R., Levine, H., Shuler, K., 1978. Nonlinear sensitivity analysis of multiparameter model systems. Journal of computational physics 26, 1-42.

Dentz, M., Tartakovsky, D.M., 2009. Abrupt-Interface Solution for Carbon Dioxide Injection into Porous Media. Transport in Porous Media 79, 15-27.

DEQ, 2013. Oil and Gas Database. Department of Environmental Quality, Michigan.Gov.

Dobossy, M.E., Celia, M.A., Nordbotten, J.M., 2011. An efficient software framework for performing industrial risk assessment of leakage for geological storage of CO_2. Energy Procedia 4, 4207-4214.

Gasda, S.E., Celia, M.A., Nordbotten, J.M., 2008. Upslope plume migration and implications for geological CO_2 sequestration in deep, saline aquifers. The IES journal. Part A, Civil & structural engineering 1, 15.

Gupta, A.K., Bryant, S.L., 2011. Analytical Correlations for Risk Parameters Involved in CO_2 Storage. Energy Procedia 4, 3849-3856.

Hahn, G.J., 1967. Statistical models in engineering. Wiley, New York.

Halliburton, 1990. Log 21101375660000 Stech Upper Half (Date: 09/04/1990).

Helton, J.C., 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliab Eng Syst Safe 42, 327-367.

International Energy Agency, 2008. Carbon Capture and Storage: Meeting the Challenge of climate change, Paris.

Kano, Y., Ishido, T., 2011. Numerical simulation on the long-term behavior of CO_2 injected into a deep saline aquifer composed of alternating layers. Enrgy Proced 4, 4339-4346.

Kopp, A., Binning, P.J., Johannsen, K., Helmig, R., Class, H., 2010. A contribution to risk analysis for leakage through abandoned wells in geological CO_2 storage. Adv Water Resour 33, 867-879.

Kutchko, B.G., Strazisar, B.R., Huerta, N., Lowry, G.V., Dzombak, D.A., Thaulow, N., 2009. CO2 reaction with hydrated class H well cement under geologic sequestration conditions: effects of flyash admixtures. Environ Sci Technol 43, 3947-3952.

LeNeveu, D.M., 2008. CQUESTRA, a risk and performance assessment code for geological sequestration of carbon dioxide. Energy Conversion and Management 49, 32-46.

Middleton, R.S., Keating, G.N., Viswanathan, H.S., Stauffer, P.H., Pawar, R.J., 2012. Effects of geologic reservoir uncertainty on CO_2 transport and storage infrastructure. International Journal of Greenhouse Gas Control 8, 132-142.

Nogues, J.P., Court, B., Dobossy, M., Nordbotten, J.M., Celia, M.A., 2012. A methodology to estimate maximum probable leakage along old wells in a geological sequestration operation. International Journal of Greenhouse Gas Control 7, 39-47.

Nordbotten, J.M., Celia, M.A., Bachu, S., 2005a. Injection and storage of CO_2 in deep saline aquifers: Analytical solution for CO2 plume evolution during injection. Transport in Porous Media 58, 339-360.

Nordbotten, J.M., Celia, M.A., Bachu, S., Dahle, H.K., 2005b. Semianalytical solution for CO_2 leakage through an abandoned well. Environ Sci Technol 39, 602-611.

Nordbotten, J.M., Kavetski, D., Celia, M.A., Bachu, S., 2009. Model for CO_2 Leakage Including Multiple Geological Layers and Multiple Leaky Wells. Environ Sci Technol 43, 743-749.

Oladyshkin, S., Class, H., Helmig, R., Nowak, W., 2011. An integrative approach to robust design and probabilistic risk assessment for CO_2 storage in geological formations. Computational Geosciences 15, 565-577.

Oldenburg, C.M., Bryant, S.L., Nicot, J.P., 2009. Certification framework based on effective trapping for geologic carbon sequestration. International Journal of Greenhouse Gas Control 3, 444-457.

Pacala, S.W., 2003. Global Constraints on Reservoir Leakage, in: Gale, J., Kaya, Y. (Eds.), Greenhouse Gas Control Technologies - 6th International Conference. Pergamon, Oxford, pp. 267-272.

Ruether, J.A., 1998. FETC Programs for Reducing Greenhouse Gas Emissions, Other Information: PBD: Feb 1998, p. Medium: ED; Size: 20 p.; Other: FDE: PDF; PL:.

Saltelli, A., 2008. Global sensitivity analysis the primer. John Wiley, Chichester, England.

Saltelli, A., Tarantola, S., Chan, K.-S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41, 39-56.

Saripalli, P., McGrail, P., 2002. Semi-analytical approaches to modeling deep well injection of CO_2 for geological sequestration. Energy Conversion and Management 43, 185-198.

SCH, 1983. Log 21101365880000 Burch Main Suite (Date 06/19/1983).

SCH, 1991. Log 21101375660000 Stech Lower Half (Date: 06/29/1991).

SIMLAB, 2007. Ver 2.2, http://ipsc.jrc.ec.europa.eu/?id=756, European Commission, Joint Research Centre, Ver 2.2 ed.

Sobol, I.M., 1993. Sensitivity estimates for nonlinear mathematical models. Mathematical Modelling and Computational Experiments 1, 407-414.

Solomon, S., Intergovernmental Panel on Climate Change., Intergovernmental Panel on Climate Change. Working Group I., 2007. Climate change 2007 : the physical science basis : contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge ; New York.

Stauffer, P.H., Viswanathan, H.S., Pawar, R.J., Guthrie, G.D., 2008. A system model for geologic sequestration of carbon dioxide. Environ Sci Technol 43, 565-570.

Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. Reliab Eng Syst Safe 93, 964-979.

Takahashi, W., 2000. Nonlinear Functional Analysis, Fixed point theory and its applications, p. iv+ 276. Yokohama, Yokohama, Japan.

Teatini, P., Ferronato, M., Gambolati, G., Baù, D., Putti, M., 2010. Anthropogenic Venice uplift by seawater pumping into a heterogeneous aquifer system. Water Resour Res 46, W11547.

Trebin, F.A., 1945. Permeability to oil of sandstone reservoir.

Turpening, R., Toksöz, M., Born, A., al., e., 1992. Reservoir Delineation Consortium Annual Report, Massachusetts Institute of Technology, Cambridge.

Vilarrasa, V., Bolster, D., Dentz, M., Olivella, S., Carrera, J., 2010. Effects of CO_2 Compressibility on CO_2 Storage in Deep Saline Aquifers. Transport in Porous Media 85, 619-639.

Walter, L., Oladyshkin, S., Class, H., Darcis, M., Helmig, R., 2011. A study on pressure evolution in a channel system during CO_2 injection. Energy Procedia 4, 3722-3729.

Zhao, H.J., Liao, X.W., Chen, Y.F., Zhao, X.L., 2010. Sensitivity analysis of CO_2 sequestration in saline aquifers. Petroleum Science 7, 372-378.

Zhou, Q., Birkholzer, J.T., Mehnert, E., Lin, Y.-F., Zhang, K., 2009. Modeling Basin- and Plume-Scale Processes of CO2 Storage for Full-Scale Deployment. Ground Water 48, 494-514.

3 Chapter: Application of Binary Permeability Fields for the Study of CO_2 Leakage from Geological Carbon Storage in Saline Aquifers of the Michigan Basin

Summary

Global mean annual surface temperature has increased about 0.3-0.6^{o}C since the late 19th century due to the proliferation of greenhouse gas concentrations from anthropogenic emissions; particularly from carbon dioxide (CO_2). Several strategies have been advanced to stabilize CO_2 emissions into the atmosphere, among them geological carbon storage (GCS). One of the requirements for GCS is that there must be a sealing formation to prevent CO_2 from escaping from the storing formation. A test site located in the Michigan Basin is proposed as a potential candidate for GCS. Since the sealing formation outside the boundaries of the test site is barely characterized, information about the continuity of this layer or its possible discontinuities is highly uncertain. To model this uncertainty, a sequential indicator simulator is here developed and implemented to represent the sealing properties, namely permeability, of the geological unit overlying the GCS candidate formation. The sequential indicator simulation algorithm is used to create binary fields of caprock with low permeability and inclusions with high permeability where brine and CO_2 could leak out. To simulate injection and potential leakage of CO_2, a semi-analytical multiphase flow model is used. Inclusions located at close distance from one another are grouped and considered as single clusters to reduce the number of leaky points used by the semi-analytical multiphase flow model, thus reducing significantly the computational effort. The results obtained with the semi-analytical model are compared with those obtained using a

numerical model to understand the limitations of using the semi-analytical model with large areas of leakage. Results of this comparison validate the semi-analytical multiphase flow. Results of the caprock continuity uncertainty study conclude increasing the probability of inclusions occurrence increases the CO_2 leakage. The occurrence of inclusions is affected by the correlation length used by the sequential indicator simulator. CO_2 leakage is affected by passive wells permeability uncertainty for scenarios with low probability of inclusions.

1 Introduction

Increase of global average temperatures in air and ocean are documented around the world (IPCC, 2007), with a global mean annual surface temperature increase of 0.3-0.6^{o}C since the late 19th century (Nicholls et al., 1996). This phenomenon is due to the proliferation of greenhouse gas concentrations from anthropogenic emissions, and particularly from carbon dioxide (CO_2), which is the most important greenhouse gas produced by human activities (IPCC, 2007). To stabilize CO_2 emissions into the atmosphere several strategies have been suggested, among them geological carbon storage (GCS). GCS is advanced as a promising approach to reduce CO_2 emissions from power plants without the necessity of fuel switching (IPCC, 2005). Suitable reservoirs for GCS are deep saline formations, depleted oil and gas reservoirs, and unmineable coal seams (Bergman and Winter, 1995; Ruether, 1998; Bachu, 2003). Deep saline formations are widespread and offer 60% of the estimated storage capacity (IEA, 2008). However compared to oil and gas reservoirs, they are barely characterized and available information about their geological properties is not expected to have.

One of the requirements for GCS is that there must be a sealing formation that prevents the stored CO_2 from escaping from the injected formation (IPCC, 2005) and guarantees a long

term sequestration. Thus, deep saline aquifers, seemingly an appropriate option to store CO_2 due to their widespread availability around the world and large storage capacities, have the inconvenience of being typically unexplored. So that little is known about the properties of the sealing formations, which are potentially compromised by the presence of leakage pathways, such as faults or fractures, permeable areas of the caprock, and poorly completed existing wells (IPCC, 2005).

Several studies that investigate the importance of CO_2 leakage associated with faults and existing wells have been documented. For instance, Chang et al. (2008) studied the CO_2 leakage through faults when flow properties of faults are uncertain. They found that lateral CO_2 migration through overlying permeable formations attenuates CO_2 leakage through faults. The effect of faults, fault permeability, and flow velocity of groundwater on the migration of CO_2 plume was studied by Sakamoto et al. (2011). Zhang et al. (2010) proposed a method to calculate the probability of CO_2 leakage through fractures and faults in a two-dimensional system.

For high well density areas, abandoned wells may represent a significant escape pathway for the injected CO_2. Gasda et al. (2004) observed that a CO_2 plume could impact 20 to several hundred abandoned wells depending on the well density of the studied area. Kopp et al. (2010) concluded that high risk of leakage through abandoned wells was produced by long injection times, small distances between an injection well and leaky well, high permeability anisotropy, high geothermal gradient, and low depth. In Celia et al. (2011), permeability of abandoned wells was identified as the most influential parameter resulting in CO_2 leakage from GCS. Nogues et al. (2012) implemented a Monte Carlo simulation where the main uncertainty was the effective well permeability. They showed that results on leakage depended on formation properties, location, and number of leaky wells. In González-Nicolás et al. (2014), results showed that the

most influential parameters on CO_2 mass leakage are the storage formation permeability, the system compressibility, and uncertainty on passive well permeability. Also, passive wells closer to the injection well are found to have a greater impact on CO_2 mass leakage.

In this work, CO_2 leakage through weak caprock areas generated with a sequential indicator simulation algorithm is studied. Here, the weak areas are referred to as 'discontinuities' or 'inclusions'. Also the influence of CO_2 leakage through existing abandoned wells located in the area of interest is studied.

A discontinuity of the sealing formation, as defined in this paper, is a localized deposition of higher permeability materials. These analyses are applied to a potential candidate site located at the Michigan Basin, whose sealing properties of the caprock are unknown. Therefore the location, size and frequency of the discontinuities are practically unknown. A sequential categorical indicator Kriging simulation algorithm is developed and applied to generate an ensemble of realizations of the caprock permeability field with two types of facies: 1) sealing formation (areas with low permeability), and 2) inclusions (areas with high permeability). The caprock permeability ensemble is thus used in a Monte Carlo simulation to simulate stochastically injection of CO_2 and study probabilistically its leakage through the weak caprock areas. Due to the unavailability of geological data, different geostatistical configurations for the sealing formation are studied to assess the impact of the uncertainty of caprock discontinuity. Areas of high permeability having relatively similar spatial locations are grouped together and considered a cluster to reduce the number of leaky points used by the semi-analytical multiphase flow model, thus reducing the computational effort. To understand the limitations of the clustering approach the semi-analytical multiphase flow model results are compared with those obtained using a numerical model.

The organization of this paper is as follows. First, the methodology used is presented. This includes the multiphase flow semi-analytical algorithm, the sequential categorical indicator Kriging, the clustering approach, and the statistical analysis. Then the application of the methodology to the Michigan Basin test site and results are presented. Lastly, a summary and conclusion of this study are given.

2 Methodology

2.1 Multiphase Flow Semi-Analytical Model

In this section the multiphase flow semi-analytical model used in this study is introduced. ELSA-IGPS (Estimating Leakage Semi-analytically - Iterative Global Pressure Solution) (Cody et al., 2014) is a revised version of the semi-analytical model ELSA developed by Celia and Nordbotten (2009) and Nordbotten et al. (2009). ELSA-IGPS is able to simulate the injection of supercritical CO_2 into a deep saline formation and compute the leakage of brine and CO_2 through permeable segments located on the caprocks. These segments represent poorly-sealed wells and are called 'passive wells'.

The domain is structured as a stack of L aquifers separated by $L + 1$ caprock layers, perforated by M carbon injection wells and N passive wells. The model assumes that aquifers are horizontal, homogeneous, and isotropic; caprocks are impermeable except where a passive well is located; initially the domain is saturated with brine at hydrostatic pressure; flow is horizontal through the aquifers; there is no transition zone between brine and supercritical CO_2 since capillary pressure is neglected; there is vertical equilibrium in pressure distributions; and the pressure response from sources and sinks can be superimposed in each aquifer. More details about the model assumptions can be found in Celia and Nordbotten (2009).

ELSA-IGPS, as the original model, applies superposition of effects for the fluid flux across sources iw (injection wells, iw=1,2,…,M) and sinks j (passive wells, j=1,2,…,N) to solve the fluid pressure $p_{j,l}$ [$ML^{-1}T^{-2}$] at any given time t [T] at the bottom of the generic aquifer l (l=1,2,..,L) and for each passive well j. Therefore, fluid pressure can be expressed as:

$$p_{j,l} = p_{0_l} + (\rho_b - \rho_c) g H_l \left[\sum_{iw=1}^{M} \Delta p'(\chi_{iw,j,l}) + \sum_{i=1}^{N} \Delta p'(\chi_{i,j,l}) \right] \tag{3-1}$$

where: p_{0_l} is the initial fluid pressure [$ML^{-1}T^{-2}$] at the bottom of the aquifer l, ρ_α is the fluid density [ML^{-3}] (α denotes the phase type, b for brine and c for CO_2), g is the gravitational acceleration [LT^{-2}], H_l is the aquifer thickness [L] of aquifer l, and:

$$\Delta p'(\chi) = \begin{cases} 0 & \text{for } \chi \geq \psi \\ -\frac{1}{2\Gamma} \ln\left(\frac{\chi}{\psi}\right) + \Delta p'(\psi) & \text{for } \psi > \chi \geq 2\lambda \\ \frac{1}{\Gamma} - \frac{\sqrt{\chi}}{\Gamma\sqrt{2\lambda}} + \Delta p'(2\lambda) + F(h') & \text{for } 2\lambda > \chi \geq \frac{2}{\lambda} \\ -\frac{1}{2\lambda\Gamma} \ln\left(\frac{\chi\lambda}{2}\right) + \Delta p'\left(\frac{2}{\lambda}\right) & \text{for } \frac{2}{\lambda} > \chi \end{cases} \tag{3-2}$$

where:

$$\chi = \frac{2\pi H \varphi (1 - s_b^{res}) r^2}{Qt} \tag{3-3}$$

$$\Gamma = \frac{2\pi(\rho_b - \rho_c) g k H^2}{\mu_b Q} \tag{3-4}$$

$$\psi = \frac{4.5\pi H \varphi k (1 - s_b^{res})}{\mu_b c_{eff} Q} \tag{3-5}$$

$$h' = \frac{h(\chi)}{H} = \frac{1}{\lambda - 1}\left(\frac{\sqrt{2\lambda}}{\sqrt{\chi}} - 1\right) \tag{3-6}$$

$$F(h') = \frac{-\lambda}{\lambda - 1}\left[\mathrm{h}' - \frac{ln[(\lambda - 1)h' + 1]}{\lambda - 1}\right] \tag{3-7}$$

where: h is the CO_2 plume thickness [L]; h' [/] is the CO_2 plume thickness relative to the aquifer thickness H; s_b^{res} is the residual saturation of the brine [/]; k is the aquifer permeability [L^2]; μ_b is the dynamic viscosity of the brine [$ML^{-1}T^{-1}$]; φ is the aquifer porosity [/]; Q is the total volumetric well flux [L^3T^{-1}]; c_{eff} is the effective compressibility of the fluid and solid matrix [$M^{-1}LT^2$]; and r is the radial distance [L]. *F(h')* is an offset term related to the vertical pressure distribution (Celia et al., 2011) and the mobility ratio λ [/] .

This derives to a system of equations where the unknowns are the fluid pressures $p_{j,l}$ at the bottom of each aquifer l and at each passive well j, and the flow rates $Q_{j,l}$ across each caprock for each passive well. $Q_{j,l}$ is calculated using the multiphase version of Darcy's law:

$$Q_{j,l} = \sum_{\alpha=b,c}\left[\pi r_{pw_{j,l}}^2 \frac{k_{r,\alpha_{j,l}} k_{pw_{j,l}}}{\mu_\alpha B_l}\left(p_{j,l-1} - \rho_\alpha g B_l - g\rho_\alpha H_{l-1} - p_{j,l}\right)\right] \tag{3-8}$$

where: $r_{pw_{j,l}}$ is the passive well radius [L], $k_{r,\alpha_{j,l}}$ is the relative permeability of phase α [/], and $k_{pw_{j,l}}$ is the single phase passive well permeability [L^2] for passive well j and aquitard layer l, and B is the caprock thickness [L].

The fluid pressure (Equation (3-1)) at the bottom of each aquifer and at each passive well can be grouped into a $(N \cdot L) \times 1$ vector. Similarly, the flow rates (Equation (3-8)) across each aquitard for each passive well can be grouped into another $(N \cdot L) \times 1$ vector. By combining these two vectors a set of $2 \cdot N \cdot L$ non-linear equations with $2 \cdot N \cdot L$ unknowns is obtained.

Domains having large numbers of passive wells (N) and/or layers (L) produce very large sets of equations; resulting in significantly higher simulation run times.

To solve this system of non-linear equations at a time t, a computational efficient fixed-point iterative scheme (Takahashi, 2000) is developed. For more details about ELSA-IGPS, the reader is referred to Cody et al. (2014) and González-Nicolás et al. (2014).

In this work, to investigate CO_2 mass leakage from caprock continuity uncertainty and wells permeability uncertainty, we focus on the following state variable: percent of CO_2 mass leakage into overlying formations. In some scenarios, to better understand results of CO_2 mass leakage, fluid overpressure near the injection well is also considered.

In these analyses, CO_2 injection takes place in the deepest formation (l=1) through a single injection well (M=1), with only one overlying aquifer (l=2) above the injected aquifer considered (see Section 3.1 for more details on the conceptualized model).

The percent of CO_2 mass leakage ($\%M_{leak}$) is defined as the ratio between the mass of CO_2 that escapes from the injected formation into layer l=2 and the total amount of injected CO_2 at final time t_{end}:

$$\%M_{leak} = \frac{M_{leak}(t_{end})}{\rho_c Q_{1,1} t_{end}} 100 \tag{3-9}$$

where $M_{leak}(t_{end})$ is given by the net cumulative CO_2 mass transferred into aquifer l=2 through all passive wells i (i=1,2,…,N):

$$M_{leak}(t_{end}) = \int_0^{t_{end}} \left[\sum_{i=1}^{N} \rho_c s_{c,i,2}(\tau) Q_{i,2}(\tau) \right] d\tau \tag{3-10}$$

where $s_{c,i,2}$ is the CO_2 saturation at passive well i and layer l=2.

Fluid overpressure Δp_{iw} is defined as the difference between the final (at t_{end}) and initial (at t=0) fluid pressures in proximity of the injection well based on Equation (3-1):

$$\Delta p_{iw} = \Delta p_{1,1}(r'_{iw}, t_{end}) = (\rho_b - \rho_c) g H_1 \left[\Delta p'(\chi_{1,1,1}) + \sum_{i=1}^{N} \Delta p'(\chi_{i,1,1}) \right] \quad (3\text{-}11)$$

where r'_{iw} is a radial distance near the injection well (r'_{iw}= 5 m in this study).

Originally, ELSA-IGPS was developed to simulate multi-phase flow and estimate the leakage of both brine and CO_2 flux along existing passive wells. That is to say leakage always occurs through small cross-sectional areas of the caprock (radii between 0.15 m – 1 m). On the other hand, here, ELSA-IGPS is used to simulate escapes through larger weak areas of the caprock (minimum radius of 56.4 m in our example). A comparison with a numerical code is used to understand the limitations the semi-analytical model when is used in this way.

2.2 Multiphase Flow Numerical Model

Results of ELSA-IGPS are compared with results obtained using ECLIPSE (Schlumberger, 2010). ECLIPSE is a commercial numerical multi-phase flow model based on a 3D finite-difference discretization and widely used in the gas and oil industry. The comparison is carried out using the CO2SOL option of ECLIPSE (Schlumberger, 2010). The compositional version (E300) of ECLIPSE is here used to perform two-phase compositional simulation of gas-brine systems, which computes mass balances for each component. Our interest focuses on the CO_2 and H_2O components, since the salinity of water is neglected in order to compare results of ELSA-IGPS with ECLIPSE.

Assuming the presence of only two fluid phases, a CO_2-rich gas phase denoted as g, and H_2O-rich liquid phase denoted as w, C_{jg} represents the mass fraction of component j present in the gas phase and C_{jw} represents the mass fraction of component j in the liquid phase. Based on mass continuity, in a system of N_c components mass fractions must be such that:

$$\sum_{j=1}^{N_c} C_{jg} = 1 \tag{3-12a}$$

$$\sum_{j=1}^{N_c} C_{jw} = 1 \tag{3-12b}$$

The compositional option E300 of ECLIPSE follows the formulation of Trangenstein and Bell (1989), in which the components flux is defined as the sum of the phases of the molar densities $\hat{\rho}_{j\alpha}$ (moles per volume, where a mole is the mass divided by the molecular weight), times its flow rate v_α / s_α. Hence, the PDE for one-dimensional is written as:

$$\frac{\partial}{\partial x}\left(\hat{\rho}_{jg}\frac{v_g}{s_g} + \hat{\rho}_{jw}\frac{v_w}{s_w}\right) = \frac{\partial}{\partial t}\left(\varphi\left(\hat{\rho}_{jg} + \hat{\rho}_{jw}\right)\right) \quad (j = 1, \dots, N_c) \tag{3-13}$$

subject to the conditions:

$$s_g + s_w = 1 \tag{3-14}$$

$$\hat{\rho}_j = \hat{\rho}_{jg} + \hat{\rho}_{jw} \tag{3-15}$$

where: x represents the spatial coordinate, φ is the porosity of the medium, and s_α and v_α are the saturation and the Darcy velocity of phase α (α denotes the phase type, either water w or gas g), respectively. Darcy's velocity of phase α is expressed as:

$$v_\alpha = -\frac{k k_{r\alpha}}{\mu_\alpha}\frac{\partial p_\alpha}{\partial x} \tag{3-16}$$

where: k is the intrinsic permeability of the porous medium, $k_{r\alpha}$, μ_α and p_α are the relative permeability, the dynamic viscosity, and the partial pressure of phase α, respectively. The capillary pressure p_c is defined as:

$$p_c = p_g - p_w \tag{3-17}$$

The relative permeabilities and capillary pressure depend upon saturation values, which can be calculated using different models such as Van Genuchten's model (Van Genuchten, 1980), or they can be obtained from experimental data.

Based on Equations (3-13) and , the one-dimensional flow of N_c fluid components is governed by the following system of PDEs:

$$\frac{\partial}{\partial x}\left(\frac{\hat{\rho}_{jg}}{s_g}\frac{kk_{rg}}{\mu_g}\frac{\partial p_g}{\partial x}+\frac{\hat{\rho}_{jw}}{s_w}\frac{kk_{rw}}{\mu_w}\frac{\partial p_w}{\partial x}\right)=\frac{\partial}{\partial t}\left(\varphi\left(\hat{\rho}_{jg}+\hat{\rho}_{jw}\right)\right)\quad (j=1,\dots,N_c) \tag{3-18}$$

To obtain comparable results between ELSA-IGPS and ECLIPSE, capillary pressure is neglected in ECLIPSE. CO2SOL option is used to model immiscible fluids. This option of ECLIPSE does not allow water to dissolve in the gas phase. In contrast, CO_2 can be present in both gas and liquid phase. Therefore, solubility of CO_2 into water must be specified to a negligible value.

2.3 Sequential Indicator Simulation Algorithm

In this section we introduce the indicator Kriging (Krige, 1951) simulation algorithm developed and implemented to generate the ensemble of realization fields of permeability at the sealing formation. This algorithm creates two types of facies: 1) sealing formation (areas with low permeability), and 2) inclusions (areas with high permeability) where CO_2 may leak.

Indicator Kriging (IK) is a geostatistical approach to geospatial modeling. IK is useful when the data are non-normally distributed, highly skewed, or essentially of categorical nature (e.g., success/failure drilling, soil type, vegetation type). IK does not assume a normal distribution at unsampled locations; instead IK builds the cumulative distribution function (CDF) at each unsampled location with the information available from the neighborhood locations. IK needs a series of thresholds (or cutoff) values that comprise the data and uses them to build the

CDF at the unsampled locations. For each threshold, IK transforms the neighborhood data to indicators and calculates their correlation based on an interpolation method such as simple Kriging (Krige, 1951) or ordinary Kriging (Matheron, 1963). With this information, IK estimates the probability of the variable at the unsampled location to be less or equal to the threshold. The final result of IK represents the uncertain distribution at the unsampled locations.

To sample a value at the unsampled locations, first, a conditional technique must be applied. IK uses an indicator transform. For a random continuous variable $X(u)$ at position u, the corresponding indicator transform can be defined as (Journel et al., 1989):

$$\begin{aligned} I(u, x_t) &= 1, \quad if \quad X(u) \leq x_t \quad t = 1, \dots, n_t \\ &= 0, \quad if \quad X(u) > x_t \end{aligned} \tag{3-19}$$

And similarly, for a discrete variable $K(u)$:

$$\begin{aligned} I(u, k_t) &= 1, \ if \ K(u) = k_t \quad t = 1, \dots, n_t \\ &= 0, \ \text{otherwise} \end{aligned} \tag{3-20}$$

where: I is the indicator value, $X(u)$ or $K(u)$ is the value of the variable at position u, x_t and k_t are the threshold or category value, and n_t is the total number of thresholds or categories. This indicator transform has the characteristic that its expected value is equal to the cumulative probability of the variable (Olea, 1999) and $I(u, x_t)$ is a random variable itself. That is to say, the IK method, instead of predicting the $X(u)$ value, predicts $I(u, x_t)$ for the selected thresholds, and provides a least-squares estimate of the conditional CDF at the threshold x_t (Deutsch and Journel, 1997):

$$[I(u, x_t)]^* = E\{I(u, x_t|(n))\}^* = P^*\{X(u) \leq x_t|(n)\} \tag{3-21}$$

where: E indicates the expectancy of $I(u, x_t)$ conditioned to the information (n) available in the neighborhood of location u, and P is the probability. Therefore $I(u, x_t)$ means the probability

that $X(u) \leq x_t$. The indicator value needs to be calculated at each threshold, which discretizes the interval of variability of property $X(u)$. Depending on the value of $X(u)$, the indicator variable will take a value of 0 or 1 at each threshold, as Equation (12) states.

Once the indicator values are assigned at each threshold, the indicator values at the unsampled locations can be estimated. To calculate the indicator values at the unsampled locations u_0, a Kriging algorithm is used. If simple Kriging (SK) is used, for a continuous and discrete variables the following equations are written (Kelkar et al., 2002), respectively:

$$I^*(u_0, x_t) = w_0 + \sum_{j=1}^{n} w_j I(u_j, x_t) \tag{3-22}$$

$$I^*(u_0, k_t) = w_0 + \sum_{j=1}^{n} w_j I(u_j, k_t) \tag{3-23}$$

where: n is the number of sampled locations to calculate the indicator values and w are the SK weights corresponding to the thresholds. Since the indicator values represent the CDF, they must satisfy for a continuous variable:

$$I(u, x_1) \leq I(u, x_2), \qquad if\ x_1 \leq x_2 \tag{3-24}$$

and for a discrete variable:

$$\sum_{t=1}^{n_t} I(u, k_t) = 1 \tag{3-25}$$

2.3.1 Simple Indicator Kriging

If the random continuous variable $X(u)$ is stationary with constant mean $\bar{m}$, and the covariance function $C(\boldsymbol{D}) = C(u, u + \boldsymbol{D})$, for any u, the SK algorithm is reduced to its stationary version (Deutsch and Journel, 1997):

$$X_{SK}^{*}(u) = \sum_{j=1}^{n} w_j(u)X(u_j) + \left[1 - \sum_{j=1}^{n} w_j(u)\right]\overline{m} \tag{3-26}$$

$$\sum_{i=1}^{n} w_i(u)C(u_i - u_j) = C(u - u_j) \qquad j = 1, \dots, n \tag{3-27}$$

where $\boldsymbol{D}$ is the lag-distance between two points.

The expectancy E of the random variable $I(u, x_t)$ is the CDF of the random function $X(u)$:

$$\begin{aligned} E\{I(u,x_t)\} &= 1 \cdot P\{X(u,x_t) \le x_t\} + 0 \cdot P\{X(u\,,x_t\,) > x_t\} \\ &= 1 \cdot P\{I(u,x_t) = 1\} + 0 \cdot P\{I(u\,,x_t\,) = 0\} \\ &= 1 \cdot P\{I(u,x_t) = 1\} = CDF(x_t) \end{aligned} \tag{3-28}$$

where: $CDF(x_t)$ is the CDF value for threshold x_t. According to Equations (3-21), (3-26), and (3-28) the SK estimate of the indicator transform is written as:

$$\begin{aligned} &[I(u,x_t)]_{SK}^{*} = [P\{X(u,x_t) \le x_t|(n)\}]_{SK}^{*} = \textstyle\sum_{j=1}^{n} w_j(u,x_t) \cdot \\ &I(u_j,x_t) + \left[1 - \textstyle\sum_{j=1}^{n} w_j(u,x_t)\right] CDF(x_t) \end{aligned} \tag{3-29}$$

where: the weights w_j are given by a SK system of type Equation (3-27):

$$\textstyle\sum_{i=1}^{n} w_i(u,x_t)C_I(u_i - u_j, x_t) = C_I(u - u_j, x_t) \quad j = 1, \dots, n \tag{3-30}$$

where: $C_I(\boldsymbol{D}, x_t) = Cov\{I(u,x_t), I(u+\boldsymbol{D}, x_t)\}$ is the indicator covariance at the threshold x_t.

Once the variograms are estimated and modeled, a selection of a path to visit the unsampled locations is needed to create a field of the random variable. For this purpose, a random sequential condition simulation method can be used.

2.3.2 Sequential Indicator Kriging Simulation Method

The sequential conditional simulation methods are grid-based and Kriging-based methods, in which unsampled locations are sequentially visited randomly until all locations are visited. The sequential simulation extends the conditioning to include all data available in the neighborhood of u, which includes the original data and values previously simulated. The sequential indicator simulation can be used for both continuous and discrete variables.

To generate the path to follow, a random number generator is used to create a sequence of random numbers corresponding to the total number of gridblocks. The sequence of the path is random to avoid artifacts resulting from a restricted search and a regular path. Based on the order, all unsampled locations are visited following the path. When an unsampled location is visited, a search neighborhood is applied where both the sampled points and previously simulated values are selected. To estimate the value of the unsampled location, a Kriging method is used. For SK the system to be solved is based on Equation (3-29) where the weights have been replaced by Equation (3-30). Therefore, for location $(n+1)$ the probability of the variable $X(u_{n+1}) \leq x_t$ is:

$$[I(u_{n+1}, x_t)]^*_{SK} = CDF(u_{n+1}, x_t) + \boldsymbol{C}^T_{n+1,n}\boldsymbol{C}^{-1}_{n,n}[\boldsymbol{I}(u_n, x_t) - \boldsymbol{CDF}(u_n, x_t)] \quad (3\text{-}31)$$

where: $[I(u_{n+1}, x_t)]^*_{SK}$ is the marginal probability of category x_t, $\boldsymbol{I}(u_n, x_t)$ is a column vector including the indicator value calculated at the locations previously simulated, $\boldsymbol{CDF}(u_n, x_t)$ is a column vector including the CDF value for threshold x_t calculated at the previous n locations, $\boldsymbol{C}_{n,n}$ is the covariance matrix for the n location previously simulated, $\boldsymbol{C}_{n+1,n}$ is the covariance matrix between the current simulation location $n+1$ and the n locations already simulated. Next, the simulated value is added to the data set and next location is simulated.

In this work, we use CIKSIM (Baù, 2012) a sequential indicator simulator based on SK that can generate several categories. Therefore, CIKSIM generates N inclusions of the caprock to introduce in the multiphase flow semi-analytical model explained in Section 2.1. If larger number of N inclusions is generated for each field of the ensemble, the computational cost of the model will increase. To avoid higher simulation run times, a clustering method of the inclusions is applied.

2.4 Clustering approach

To enhance the performance of the semi-analytical model, a method that groups these inclusions is applied. Grouping the inclusions by clusters considerably decreases the number of leaky areas to be introduced into the semi-analytical model of Section 2.1.

It is considered a cluster when two or more inclusion blocks are in contact. Thereby, two inclusions that are in contact only diagonally belong to the same group. That is the distance between the centers of their grid-blocks is less or equal to $\sqrt{2\Delta x^2}$. For our grid (Section 3.2) this distance is 142.4 m.

The size and distribution of these clusters depend on the parameters assigned for their generation. Each cluster is modeled as a single circular area of leakage with the equivalent area of the cluster. The position of the circular area is calculated as the average x and y coordinates of the gridblock centers that constitute this cluster.

One example of grouping the clusters at the caprock is shown in Figure 3-1. This example has 84 grid-blocks that are inclusions (orange grid-blocks), after clustering this number is reduced to only 16 clusters. The equivalent area of the clusters are shown as a black circles in

Figure 3-1. Each of these clusters is used as a single leaky point in the semi-analytical model ELSA-IGPS.

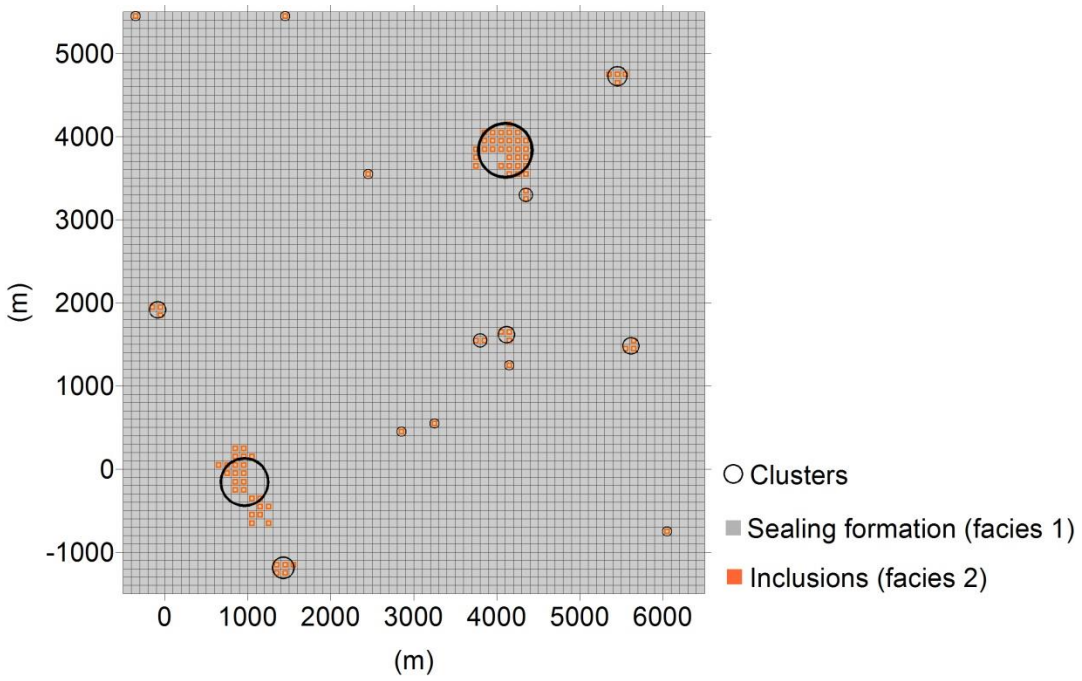

Figure 3-1. Representation of the clusters circular equivalent area (in black) of one realization of the caprock generated with CIKSIM. In this example, the number of 84 inclusions-blocks (in orange) is reduced to 16 clusters.

2.5 Statistical analysis

Output ensembles of the state variables are used to produce CDF plots. CDF of the state variable Y (either $\%M_{leak}$ or Δp_{iw}) is obtained from the output of N_{MC} model simulations. After ordering the Y values in ascending order, $Y_1 < Y_2 < \cdots < Y_{N_{MC}}$, the corresponding CDF values are calculated as $CDF(Y) = (i - 0.5)/N_{MC}$ (i=1,2,…,N_{MC}) (Hahn, 1967).

To optimize the performance of the simulations, preliminary tests are run to find the minimum ensemble size N_{MC} beyond which CDFs remain substantially stationary. A sample size

of N_{MC}= 500 is selected, therefore 500 fields of caprock are generated for each scenario of Table 3-2.

A flow chart of the methodology is summarized in Figure 3-2. First, the sequential indicator simulator CIKSIM is applied to the grid domain conditioned to the available previous information. As a result of this step, an ensemble of caprock binary fields containing the two types of facies is obtained. Then the clustering approach is applied to the caprock binary fields in order to decrease the number of leaky areas to introduce in the multiphase flow semi-analytical model. Upon clustering process, the ELSA-IGPS Monte Carlo simulations are run and the output ensembles of the state variables are obtained. Last, the statistical analysis is applied to obtain the CDF results.

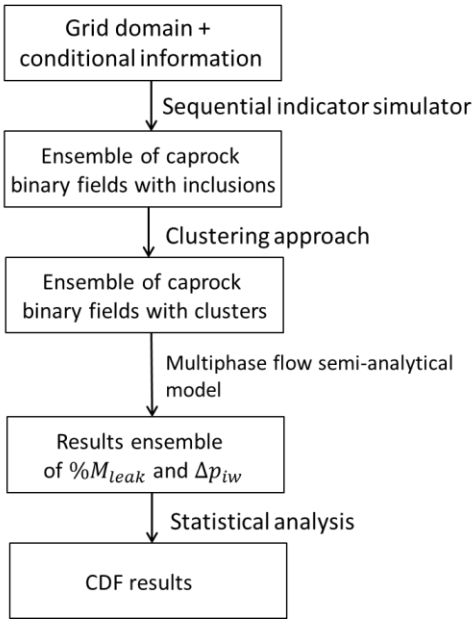

Figure 3-2. Flow chart of the methodology.

3 Application to the Michigan Basin Test Site

3.1 Study Area

The methodology introduced in Section 2 is applied to a geological test site located within the Michigan Basin close to Thompsonville (MI). The candidate formation proposed for GCS is known as the Gray Niagaran formation. Figure 3-3 shows a cross-section of the Michigan basin in the area of interest with the candidate storage formation highlighted in yellow. The Gray Niagaran formation lies below an almost depleted hydrocarbon reservoir (Brown Niagaran pinnacle in Figure 3-3), which is currently used by Michigan Technological University for geophysical research.

The Gray Niagaran formation has a thickness of 119 m (390 ft) with its top at 1,500 m (4,920 ft) below ground surface, making this formation a good repository for GCS. The choice to store supercritical CO_2 in this formation is justified by the sealing capacity of the formation above the Brown Niagaran pinnacle. In addition, the Gray Niagaran formation is already perforated by two exploration wells, Burch 1-20B and Stech 1-21A (Figure 3-3), which could possibly serve as CO_2 injection wells.

A relevant source of uncertainty in choosing the Gray Niagaran formation as a candidate for GCS is the continuity of its caprock highlighted in green in Figure 3-3. Although some data are available from monitoring wells at the test site such as Burch 1-20B and Stech 1-21A (SCH, 1983; Halliburton, 1990; SCH, 1991), the data that can be used directly to describe the spatial distribution of the sealing properties of the caprock formation at the basin scale are scarce.

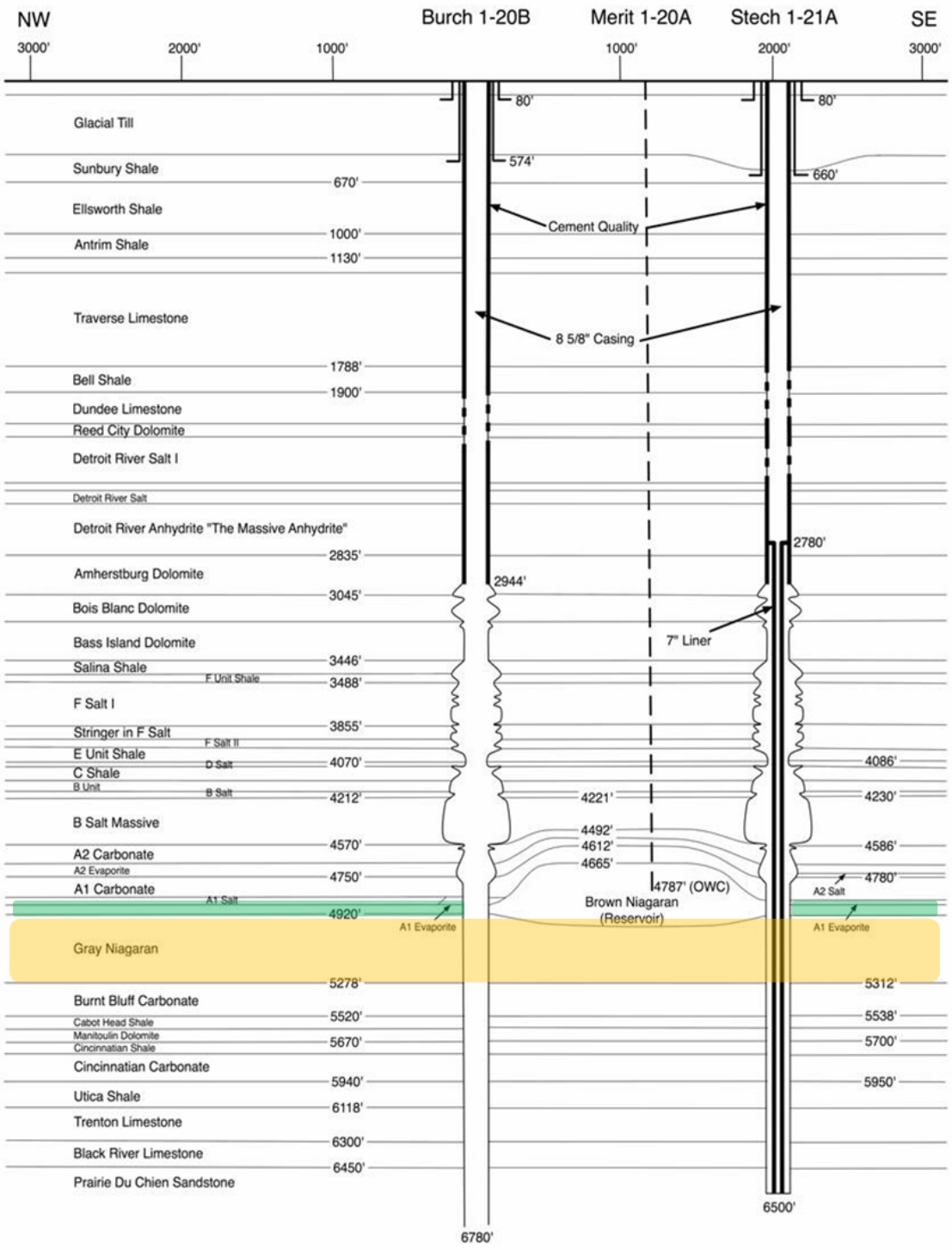

Figure 3-3. Cross-section of the Michigan Basin test site proposed for GCS (Turpening et al., 1992). Candidate formation is highlighted in yellow and caprock in green.

The model system is conceptualized in ELSA-IGPS as a stack of two aquifers (L=2): the Gray Niagaran formation (119 m thick) and the Carbonate formation (35 m thick). The two aquifers are separated by a caprock: the Evaporites formation (17 m thick). The enumeration of the aquifers starts from the deepest one. Supercritical CO_2 is injected into aquifer l=1 from a single well (M=1). The mass injection rate is Q_m= 30 kg s^{-1} and remains constant throughout the simulation time t_{end}= 10 years.

In ECLIPSE, the geological model is conceptualized as two aquifers separated by a caprock. Thicknesses of these formations follow thicknesses applied in ELSA-IGPS. Tthe model domain is divided into 100 m × 100 m grid-blocks horizontally. Vertically, each formation is divided into four layers. A single CO_2 injection vertical well is considered at the center of the model domain and screened within the deepest formation. The surrounding area of the injection well is refined to achieve an appropriate size for a well (≈0.5 m). The CO_2 injection well is assumed to operate at a rate of 30 kg s^{-1} for a period of 10 years. Initially all formations are saturated only with brine and in hydrostatic pressure conditions. To simulate a laterally infinite aquifer system, the pore-volume of the boundary blocks is multiplied by a factor of 1x10^6.

In both models, ELSA-IGPS and ECLIPSE, the caprock is assumed impermeable except for the location of inclusions or passive wells located in the area of interest. Initially all formations are saturated with brine and pressure is hydrostatically distributed. Due to the lack of data availability, Van Genuchten model (Van Genuchten, 1980) is used to calculate relative permeabilities of CO_2 and brine. The parameters used for this model are a brine residual saturation of s_b^{res}=0.3 and a fitting parameter of 0.41 (Zhou et al., 2009). Representations of the relative permeability curves of CO_2 and brine are shown in Figure 3-4.

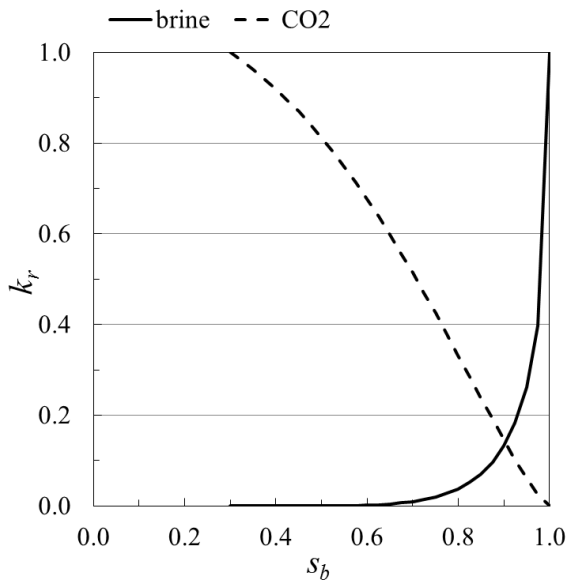

Figure 3-4. Relative permeability curves of CO_2 (dashed line) and brine (solid line) as a function of brine saturation.

Porosity values are extracted from Burch 1-20B and Stech 1-21A log-wells (SCH, 1983; Halliburton, 1990; SCH, 1991). The injected aquifer and the overlying formation are assumed to have a permeability equal to 2.8×10^{-14} m^2 and 9.6×10^{-15} m^2, respectively, calculated according to Trebin (1945) as:

$$\begin{aligned} k &= 2e^{31.6\varphi} && \text{if } 100\varphi < 12\% \\ k &= 4.94(100\varphi)^2 - 763 && \text{if } 100\varphi > 12\% \end{aligned} \tag{3-32}$$

where: k is the permeability in millidarcy (mD, 1mD ≡ 1×10^{-15} m^2), and φ is the porosity (/). For simplification, the weak areas of the caprock are considered to have the same set of properties as the injected aquifer l=1 ($k_2 = k_{l_1}$). The hydro-geomechanical parameters used in this study are provided in Table 3-1. In order to obtain comparable results with ELSA-IGPS, CO2SOL option of ECLIPSE is used to model the flow of immiscible fluids. Also capillary pressure is neglected and the wetting phase is assumed as pure water to exclude the impact of salt on the results.

Table 3-1. Hydro-geomechanical parameters.

Parameter	Symbol	Value	Units
Brine density	ρ_b	1,000	kg m^{-3}
CO_2 density	ρ_c	575	kg m^{-3}
Brine viscosity	μ_b	4.5×10^{-4}	Pa s
CO_2 viscosity	μ_c	4.6×10^{-5}	Pa s
System compressibility	c_{eff}	4.6×10^{-10}	Pa^{-1}
Injected aquifer porosity	φ_{l_1}	0.084	/
Overlying aquifer porosity	φ_{l_2}	0.05	/
Brine residual saturation	s_b^{res}	0.3	/
Injected aquifer permeability	k_{l_1}	2.8×10^{-14}	m^2
Overlying aquifer permeability	k_{l_2}	9.6×10^{-15}	m^2
Weak areas/inclusions	k_2	2.8×10^{-14}	m^2

3.2 Sequential indicator simulation applied to the Michigan Basin

In order to generate permeability fields of the caprock formation according to a bimodal stochastic process, CIKSIM is implemented to generate weak areas of the caprock where data are not available. For this purpose, a grid covering an area of 7 km × 7 km is considered with the hydrocarbon reservoir located at its center (Figure 3-5). Each grid-block is 100 m × 100 m, yielding a total of 4,900 blocks.

The thickness of the caprock above the Gray Niagaran formation is relatively small when compared to the horizontal extension of this formation (30.5 m thickness of caprock vs. 7,000 m of estimated grid extension). For this reason, the permeability is assumed to be homogeneous in

the vertical direction z, whereas in the horizontal directions, x and y, the permeability is assumed heterogeneous.

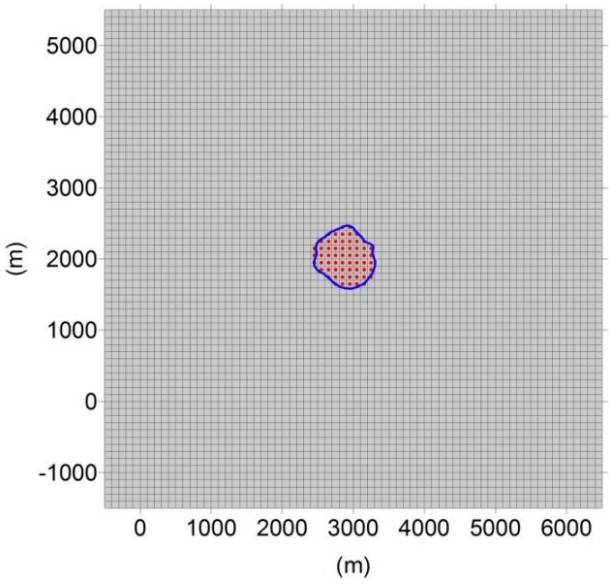

Figure 3-5. Area of 7 km × 7 km used by CIKSIM. Each grid-block is 100 m × 100 m. Limit of the hydrocarbon reservoir is shown by the blue line (Brown Niagaran pinnacle in Figure 3-3). Red spots correspond to the sealing caprock above the reservoir, which are conditional locations of facies 1 with low permeability.

Assuming the sealing formation (facies 1 of low permeability) comprises uncertain inclusions of high permeability where CO_2 could leak (facies 2 with a high permeability), the permeability in the horizontal direction can be considered a binary field. The horizontal permeability of the confining formation is low compared to permeability values of the inclusions, and the inclusions represent randomly distributed zones of high permeability within the caprock.

The nearly depleted reservoir is already known to be covered by a sealing formation, since it contained oil originally. Thus the sealing capacity in that region is somehow warranted. Hence the blocks in the reservoir are assumed to be sealing formation (conditioning information), whereas the formation of the others blocks (unsampled locations) are unknown and

need to be simulated stochastically. Figure 3-5 shows the lateral boundary of the reservoir by the blue line and red dots correspond to the conditional locations of facies 1.

3.2.1 Uncertainty from caprock continuity

To represent the uncertainty from caprock continuity, an ensemble of realizations of the caprock permeability field is generated with the sequential indicator simulation algorithm of Section 2.3. The generation of the caprock ensembles follows the properties provided in Table 3-2.

An exponential covariance model is used to generate random permeability fields of the caprock containing inclusions. Several probabilities of the occurrence of P_2 are applied for facies 2 (inclusions) ranging between 0.0005 and 0.02, as well as correlation lengths l_{xy} extending between 200 m and 1,500 m (where xy denotes that correlation lengths in the x and y direction are equal). Facies 1 (sealing formation) have a probability of $P_1 = 1 - P_2$, and a correlation length constant for all scenarios with a value of l_{xy}= 1,000 m.

To test the application of CIKSIM to generate the caprock binary fields, two relationships as a function of the inclusions correlation length l_{xy} are defined for scenarios of Table 3-2. The two relationships are the following: (i) relationship between correlation length and distances from the centers of the clusters to the injection well D, and (ii) relationship between correlation length and number of inclusion blocks ratio $ratio\_lc$.

The average distance D between centers of the clusters and injection well for one ensemble of realizations of Table 3-2 is defined as:

$$D = \frac{1}{N_{MC}} \sum_{j=1}^{N_{MC}} \frac{\sum_{i=1}^{N_{cl}} d_i}{N_{cl}} \tag{3-33}$$

Table 3-2. Parameters used for the generation of caprock fields with CIKSIM (Baù, 2012). All considered scenarios are assumed to have a correlation length $\boldsymbol{l_{xy}}$=1,000 m for facies 1.

Scenario	Correlation model	N_{MC}	$P_2{}^*$	l_{xy} (m)
1.1	Exponential	500	0.0005	200
1.2				400
1.3				600
1.4				1,500
2.1	Exponential	500	0.001	200
2.2				400
2.3				600
2.4				1,500
3.1	Exponential	500	0.005	200
3.2				400
3.3				600
3.4				1,500
4.1	Exponential	500	0.01	200
4.2				400
4.3				600
4.4				1,500
5.1	Exponential	500	0.02	200
5.2				400
5.3				600
5.4				1,500

*Where facies=2 corresponds to inclusions. Probability of facies 1 (sealing formation) is $P_1 = 1 - P_2$.

where N_{cl} is the total number of clusters present in one realization j (i=1,2,.., N_{cl}), and d_i is the distance between the center of the cluster i and the injection well.

The ratio of inclusion blocks $ratio\_lc$ of one ensemble of realizations is defined as the ratio between the average number of inclusion blocks of one ensemble of realizations and the expected number of inclusion blocks of one realization:

$$ratio\_lc = \frac{\frac{\sum_{i=1}^{N_{MC}} l_{c,i}}{N_{Mc}}}{P_2\, N_{gc}} \tag{3-34}$$

where N_{gc} is the total number of grid-blocks considered for the generation of the caprock (N_{gc}= 4,900, see section 3.2), and $l_{c,i}$ is the number of inclusion blocks or leaky blocks (with facies 2) of realization i. For instance, for a probability P_2= 0.01, the expected number of inclusion blocks is 49 ($P_2 \times N_{gc}$).

Also in this section, the influence of the injected formation permeability and inclusions permeability on CO_2 leakage is studied. Different pairs of permeability values of the injected formation and inclusions for scenarios 1.1, 2.1, 3.1, 4.1, and 5.1 are considered. The range of permeabilities of the injected formation k_{l_1} and inclusions k_2 studied spans from 1×10^{-15} m^2 (1 mD) to 1×10^{-12} m^2 (1,000 mD). Results of this analysis report the CO_2 leakage percentage at the 95th percentile.

3.2.2 Uncertainty from caprock continuity and passive wells permeability

The area considered in Section 3.1 comprises 60 wells that perforate the candidate formation to store CO_2. The locations of these wells are obtained from the Michigan Department of Environmental Quality database (DEQ, 2013). However, the conditions of these wells are not known. A deteriorated or poorly cemented well can create a leaky pathway for brine and/or CO_2.

The significant number of wells at this location results in a source of uncertainty that cannot be disregarded. Consequently, they are included in the uncertainty analysis of caprock continuity to study their influence on CO_2 leakage.

Associated with the use of the semi-analytical model, these 60 passive wells are grouped into 20 equivalent leaky pathways. Figure 3-6 shows the position of the passive wells located in the area under consideration and the position of each cluster of wells. These clusters are identified through an optimization procedure that minimizes the sum of the Euclidean distances of the passive wells that form a cluster and the cluster centroid. The equivalent leaky area considered for each cluster of wells is equal to the sum of the cross sectional areas of the wells included in that group.

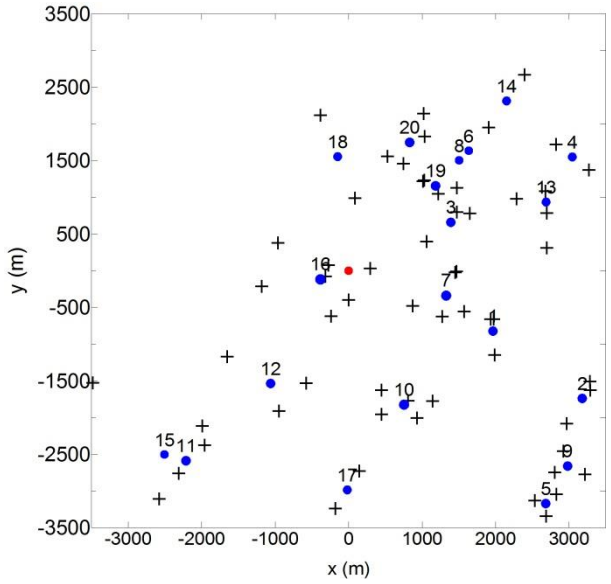

Figure 3-6. Location of the 60 passive wells (indicated as crosses) that reach the candidate storage formation and location of the 20 leaky pathways (indicated as circles) after clustering the existing passive wells. The injection well is located at the center of the domain (indicated as a red point).

The location of these clusters of wells is deterministic in each of the realizations of the caprock binary fields of Table 3-2. On the other hand, the permeability of the clusters of wells is considered stochastic, since permeability of the passive wells is unknown. We assume that all

passive wells follow the same lognormal probability distribution function with a log-mean of $\log(1\times10^{-14}\ m^2)$ and a log-standard deviation of 1 log-m^2 (Nordbotten et al., 2009).

4 Results and Discussion

This section includes results and discussions of the analyses conducted on the area of study introduced in Section 3.1. First, we present results of an ad-hoc analysis conducted to validate ELSA-IGPS when large areas of leakage are introduced in the model. Next, we present and discuss results of the caprock continuity uncertainty study followed by a sensitivity analysis of the permeability of the injected formation and the inclusions affecting CO_2 mass leakage. Finally, we investigate the effect that uncertainty from passive wells permeability has on CO_2 leakage produced by caprock continuity uncertainty.

4.1 Validation of ELSA-IGPS for large areas of CO_2 leakage

We compare the results of the semi-analytical algorithm ELSA-IGPS with those produced with the numerical code ECLIPSE when large areas of leakage are introduced in the semi-analytical model.

To compare the results of the semi-analytical solution (Section 2) with the numerical model, we randomly select one realization from the ensemble of realizations of the permeability fields available for scenarios 1.1, 2.1, 3.1, 4.1, and 5.1 (Table 3-2). Both ELSA-IGPS and ECLIPSE are used to simulate these realizations. Inclusions of these realizations are clustered following the approach explained in Section 2.4. Information of the number of inclusions, the number of clusters after the clustering approach, minimum and maximum radius of the clusters is provided in Table 3-3.

Table 3-3. Information of the number of inclusions, clusters, and their minimum and maximum radius.

Scenario (one realization)	Number of inclusions	Number of clusters	Minimum radius of the clusters	Maximum radius of the clusters
1.1	0	-	-	-
2.1	12	2	112.8	159.6
3.1	24	3	56.4	239.4
4.1	68	7	56.4	418.4
5.1	132	15	56.4	343.2

Comparison of the results obtained with both codes ELSA-IGPS and ECLIPSE are shown in Figure 3-7, where each subpanel presents the cumulative mass leakage of CO_2 over time for each scenario. Note that the y-axis of the cumulative CO_2 mass leakage is presented in logarithmic scale.

Scenario 1.1, with the lowest probability of finding an inclusion P_2 and a correlation length l_{xy}=200 m, does not produce CO_2 mass leakage using either code. This is an obvious result since the realization simulated does not include any inclusion at the caprock (see Table 3-3). Consequently, these results are not shown here.

In Scenario 2.1 and Scenario 3.1, Figure 3-7a and Figure 3-7b respectively, CO_2 mass leakage produced by ECLIPSE starts earlier than the leakage produced by ELSA-IGPS. However, final cumulative CO_2 mass leakage for each code is very similar.

In Scenario 4.1 (Figure 3-7c) and Scenario 5.1 (Figure 3-7d), which have the largest probabilities (P_2) of finding an inclusion, and consequently the largest probability of leakage, the two codes present similar CO_2 mass leakages. Moreover, CO_2 leakage seems to start approximately at the same time for both ELSA-IGPS and ECLIPSE.

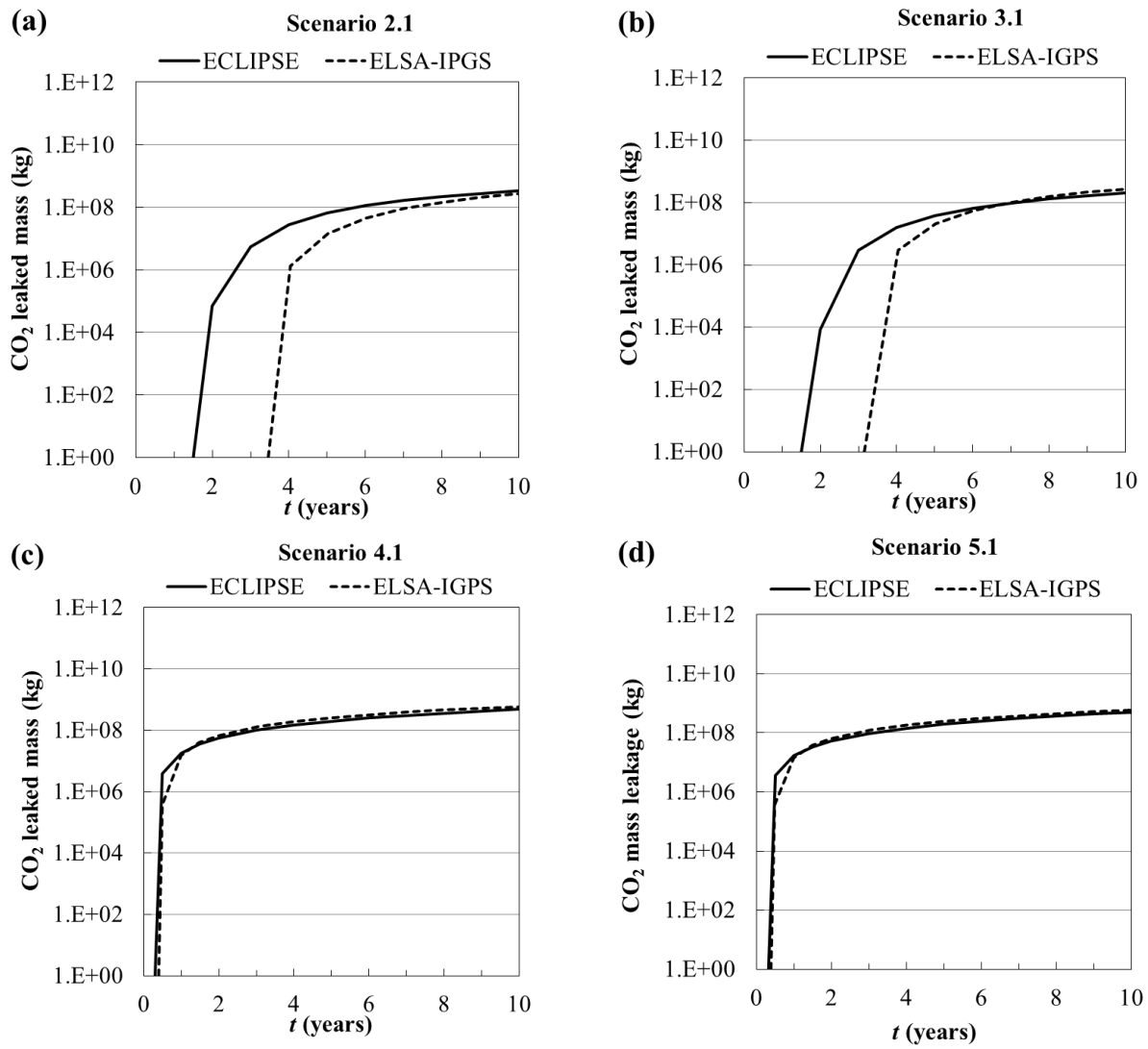

Figure 3-7. ECLIPSE and ELSA-IGPS comparison of CO_2 mass leakage results over time for one realization of the caprock ensemble: (a) Scenario 2.1, (b) Scenario 3.1, (c) Scenario 4.1, and (d) Scenario 5.1. See Table 3-2 for description of the scenario.

The fact that CO_2 leakage with ELSA-IGPS starts later than in ECLIPSE for Scenario 2.1 and 3.1 (Figure 3-7a and Figure 3-7b, respectively) can be explained by the difference in the evolution of fluid pressure. Results of the comparison of fluid overpressure near the injection well for the two codes are reported in Figure 3-8. We notice that initial fluid overpressure pulse in ECLIPSE is greater than in ELSA-IGPS for the first years, which is associated with a faster propagation of the CO_2 plume in the injected aquifer. Therefore, the CO_2 plume reaches the

clustered areas sooner in ECLIPSE. As time increases fluid overpressure produced by ECLIPSE decreases, until it reaches similar overpressure values produced by ELSA-IGPS at t_{end}.

In general, the cumulative CO_2 mass leakage produced with both models is of the same order of magnitude at later times, hence showing a good agreement between the two models. Therefore, the approach of clustering the inclusions-blocks and introducing them into ELSA-IGPS as a large equivalent leaky area can be considered valid.

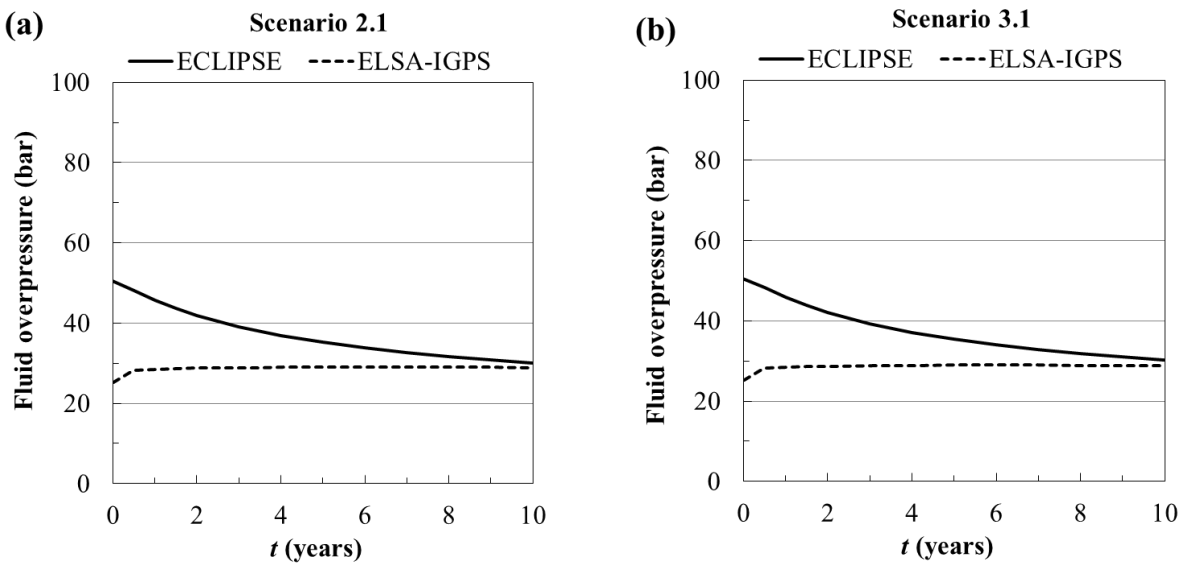

Figure 3-8. ECLIPSE and ELSA-IGPS comparison of fluid overpressure nearby the injection well results over time for one realization of the caprock ensemble: (a) Scenario 2.1, and (b) Scenario 3.1. See Table 3-2 for description of the scenario.

4.2 Results of the study of caprock continuity uncertainty

In this section, we present results of the study of uncertainty from caprock continuity.

4.2.1 CIKSIM tests results

Here, results of the relationships D and $ratio\_lc$ (Section 3.2.1) as function of correlation length are reported.

Figure 3-9 reports the relationship between the correlation length and the average distance between cluster centers and injection well distances for probabilities P_2 equal to 0.005, 0.01, and 0.02.

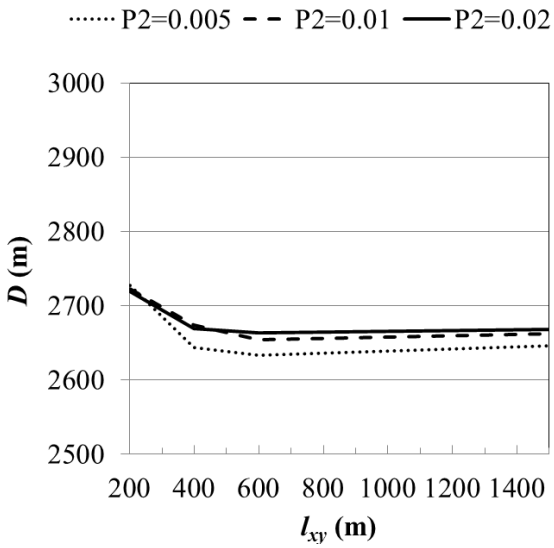

Figure 3-9. Relationship between correlation length and the average distance between cluster centers and injection well.

Figure 3-9 shows that the D distance decreases when l_{xy} is increased. Correlation length l_{xy}= 200 m has the greatest D for the different probabilities P_2. When the correlation length of facies 2 is increased, larger inclusions are expected to be generated, which occupy more space in the domain and consequently their locations are more restricted in the domain. On the other hand, smaller correlation lengths generate smaller inclusions, which can be more spread out through the domain and further from the injection well. As a result, same probability P_2 but different correlation lengths, the lowest l_{xy} produces less CO_2 mass leakage. CO_2 leakage is lower for the lowest l_{xy} because the average distance D that the CO_2 plume has to travel through the storage formation is longer, therefore the time to reach the inclusions of the caprock is larger too.

Figure 3-10 displays the relationship between correlation length l_{xy} and the ratio of inclusion blocks $ratio\_lc$ for probabilities P_2 equal to 0.005, 0.01, and 0.02. This figure shows that the average number of inclusion blocks of one ensemble of generated caprock fields is equal to the expected inclusion blocks; that is to say that $ratio\_lc$ is equal to 1 only when the correlation length is very small (l_{xy}= 0.1 m).

The ratio of inclusion blocks increases with the correlation length, meaning that the correlation length affects the number of inclusion blocks of one ensemble of caprock fields. This dependency does not seem to increase significantly for correlation lengths greater than l_{xy}= 400 m, and it seems to reach a value between $ratio\_lc$= 1.6 and $ratio\_lc$= 1.8 for the different probabilities.

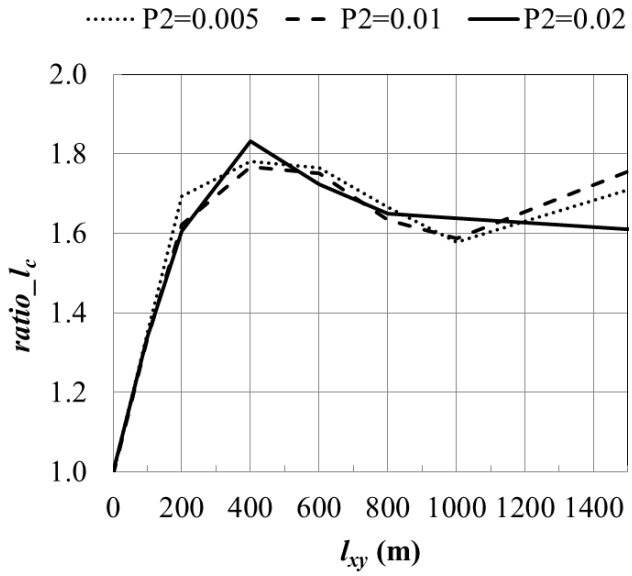

Figure 3-10. Relationship between correlation length and the ratio of inclusion blocks.

4.2.2 CO_2 leakage results

In this section, we present results concerning the uncertainty analysis for caprock continuity. The effects of correlation length and probability of facies 2 on the two state variables

of interest: $\%M_{leak}$ (Equation (3-9)) and Δp_{iw} (Equation (2-20)) are studied. Output ensembles of the state variables are used to produce the CDF plots.

Figure 3-11 presents results of $\%M_{leak}$ and Δp_{iw} obtained for correlation lengths l_{xy}= 200 m and l_{xy}= 400 m, and different values of the probability P_2. In Figure 3-11a one may note that CO_2 mass leakage is higher for larger P_2 values. For instance, 81% of realizations in Scenario 1.1 (P_2= 0.0005 and l_{xy}= 200 m) present $\%M_{leak}$ lower or equal to $1x10^{-3}$%, whereas in Scenario 5.1 (P_2= 0.02 and l_{xy}= 200 m) only 1% of the realizations has $\%M_{leak}$ lower than 1×10^{-3}%. This agrees with the fact that higher probability P_2 is expected to present greater CO_2 leakage since the probability of the CO_2 plume to find an inclusion through the caprock is higher. Similarly, Figure 3-11a shows that CO_2 mass leakage increases with P_2 when correlation length is l_{xy}= 400 m.

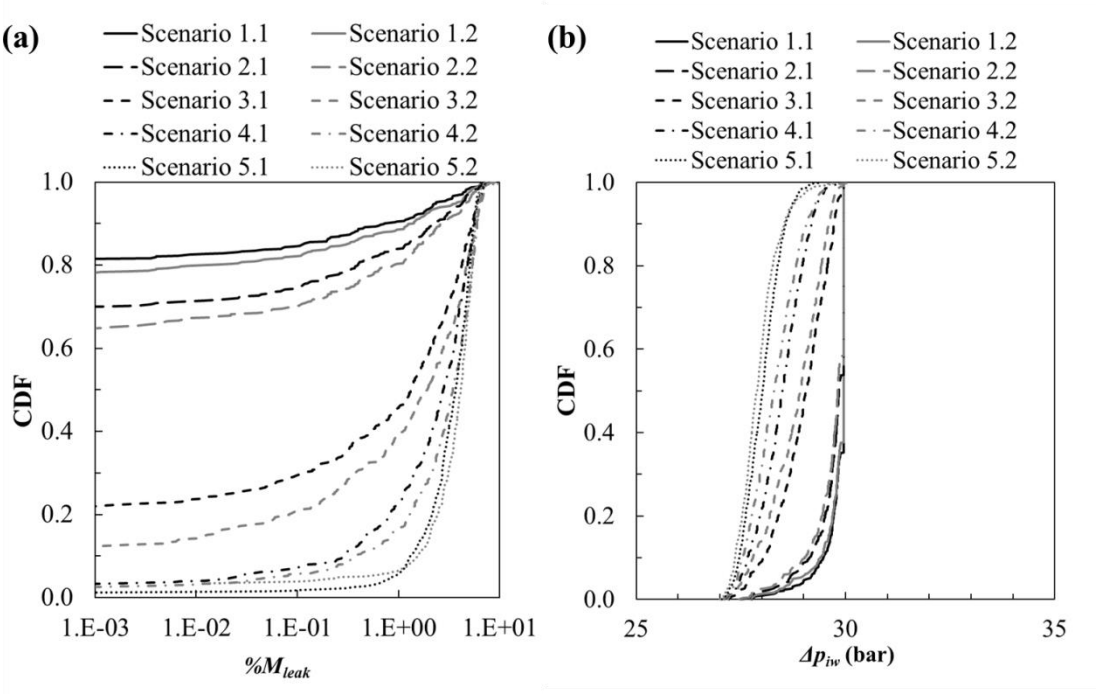

Figure 3-11. ELSA-IGPS results of scenario with correlation length l_{xy}= 200 m (in black) and l_{xy}= 400 m (in grey): (a) percent of CO_2 mass leakage, and (b) fluid overpressure near the injection well. See

Table 3-2 for description of the scenarios.

When comparing the $\%M_{leak}$ of both correlation lengths l_{xy}= 200 and l_{xy}= 400 m in Figure 3-11a, we observe that $\%M_{leak}$ is greater for the latter correlation length (Scenarios 1.2, 2.2, 3.2, 4.2, and 5.2). Greater CO_2 leakages are expected for higher correlation lengths since the inclusions generated by CIKSIM are larger in size and, consequently, the CO_2 mass leakage is larger as the carbon plume reaches the inclusions.

Figure 3-11b shows fluid overpressure in bars (1 bar ≡ 1×10^5 Pa). Δp_{iw} decreases when the probability of finding an inclusion is greater; therefore, CDF curves for a P_2= 0.02 have the lowest Δp_{iw} in Figure 3-11b (dotted lines).

Correlation length also influences the fluid overpressure. When comparing the same probability P_2 for l_{xy}= 200 m and l_{xy}= 400 m, the greatest correlation length yields to a lower fluid overpressure. This agrees with the fact that greater correlation lengths have inclusions with greater leakage areas, and consequently when the CO_2 plume reaches the inclusion there is more leakage from the injected aquifer to the overlying formation. As a result a larger relief in fluid pressure takes place.

Figure 3-12 shows mass leakage CDFs obtained with probabilities P_2= 0.005 and P_2= 0.01 for different correlation lengths of facies 2, ranging from l_{xy}= 200 m to l_{xy}= 1,500 m. Figure 3-12a shows that $\%M_{leak}$ increases with the correlation length l_{xy} and the probability P_2.

Results of fluid overpressure are illustrated in Figure 3-12b. Δp_{iw} decreases when increasing the correlation length and when increasing the probability of finding an inclusion P_2. Thus, the greatest fluid overpressure is produced by Scenario 1.1 (P_2= 0.005 and l_{xy}= 200 m, in black solid line), whereas the lowest fluid overpressure is produced by Scenario 5.4 (P_2= 0.01 and l_{xy}= 1,500 m, in grey dotted line).

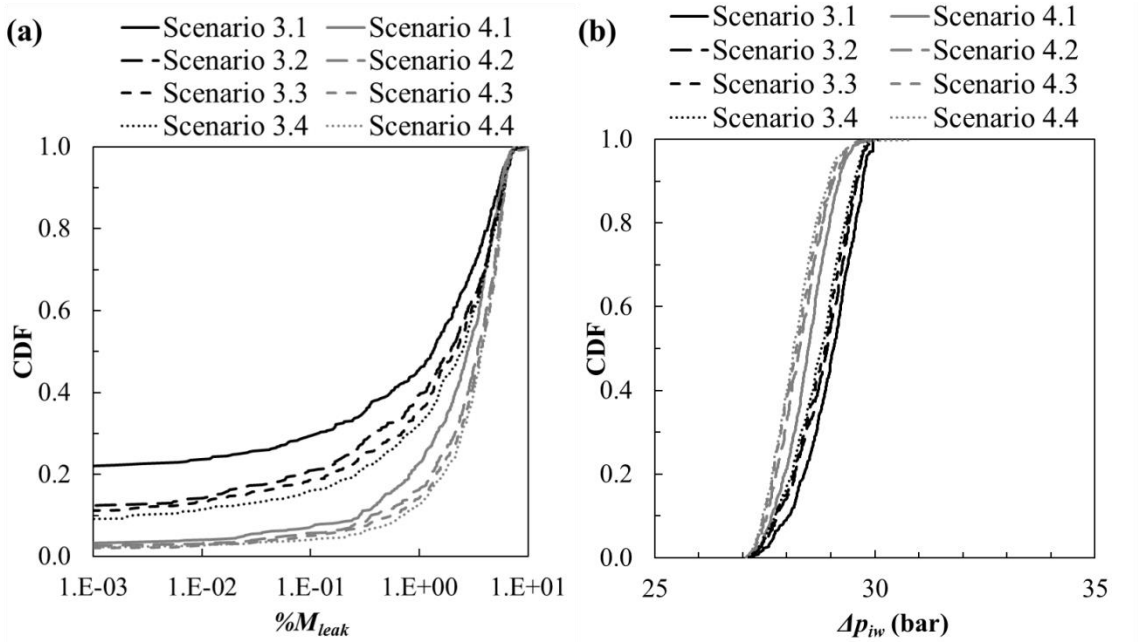

Figure 3-12. ELSA-IGPS results of scenarios with probability P_2= 0.005 (in black) and P_2=0.01 (in grey): a) percent of CO_2 mass leakage, and b) fluid overpressure nearby the injection well. See Table 3-2 for description of the scenarios.

Figure 3-11 indicates that scenarios with the same correlation length and with higher P_2 have greater $\%M_{leak}$ and lower Δp_{iw}. In Figure 3-11, when comparing the amount of leaked CO_2 for the two correlation lengths (l_{xy}= 200 m and l_{xy}= 400 m), we observe that $\%M_{leak}$ is greater for l_{xy}= 400 m. This agrees with two facts: 1) the distance from the center cluster to the injection well D is lower for correlation length l_{xy}= 400 m (see Figure 3-9), and 2) the ratio of inclusion blocks is greater for l_{xy}= 400 m (see Figure 3-10). Therefore, in the scenario of l_{xy}= 400 m, there are more inclusion blocks on average for one stochastic simulation, and the distance that the CO_2 plume must travel until reaches the center of the cluster is shorter, favoring the CO_2 leakage.

Increasing the ratio of inclusion blocks facilitates the CO_2 leakage to aquifer l=2, relieving pressure in the injected aquifer l=1. Therefore, Δp_{iw} results with correlation length

l_{xy}= 400 m present slightly lower fluid overpressures than results with l_{xy}= 200 m, since $ratio\_lc$(200 m) < $ratio\_lc$(400 m).

4.2.3 Influence of permeability values of the injected formation and inclusions on CO_2 mass leakage

To study the influence of the injected formation permeability k_{l_1} and inclusions permeability k_2 on the maximum probable amount of leaked CO_2, different pairs of permeability values of k_{l_1} and k_2 are considered for Scenarios 1.1, 2.1, 3.1, 4.1, and 5.1 (Table 3-2). These results are presented in terms of the $\%M_{leak}$ 95th percentile and shown in Figure 3-13 for different values of k_{l_1} and k_2.

Each subpanel in Figure 3-13 corresponds to one of the scenarios. All scenarios present low CO_2 mass leakages when k_{l_1} is high and k_2 is low. In general, high permeability of the injected formation corresponds to less escape of CO_2 through weak areas. The CO_2 plume advances more easily through the injected formation when k_{l_1} is high, easing its storage instead of CO_2 escape (González-Nicolás et al., 2014).

Scenarios 1.1 and 2.1 produce the lowest CO_2 mass leakages. In Scenarios 4.1 (Figure 3-13d) and 5.1 (Figure 3-13e), considerable amounts of CO_2 leakage are observed when the inclusion's permeability is greater than 3.16×10^{-13} m^2 (log-k_2= -12.5).

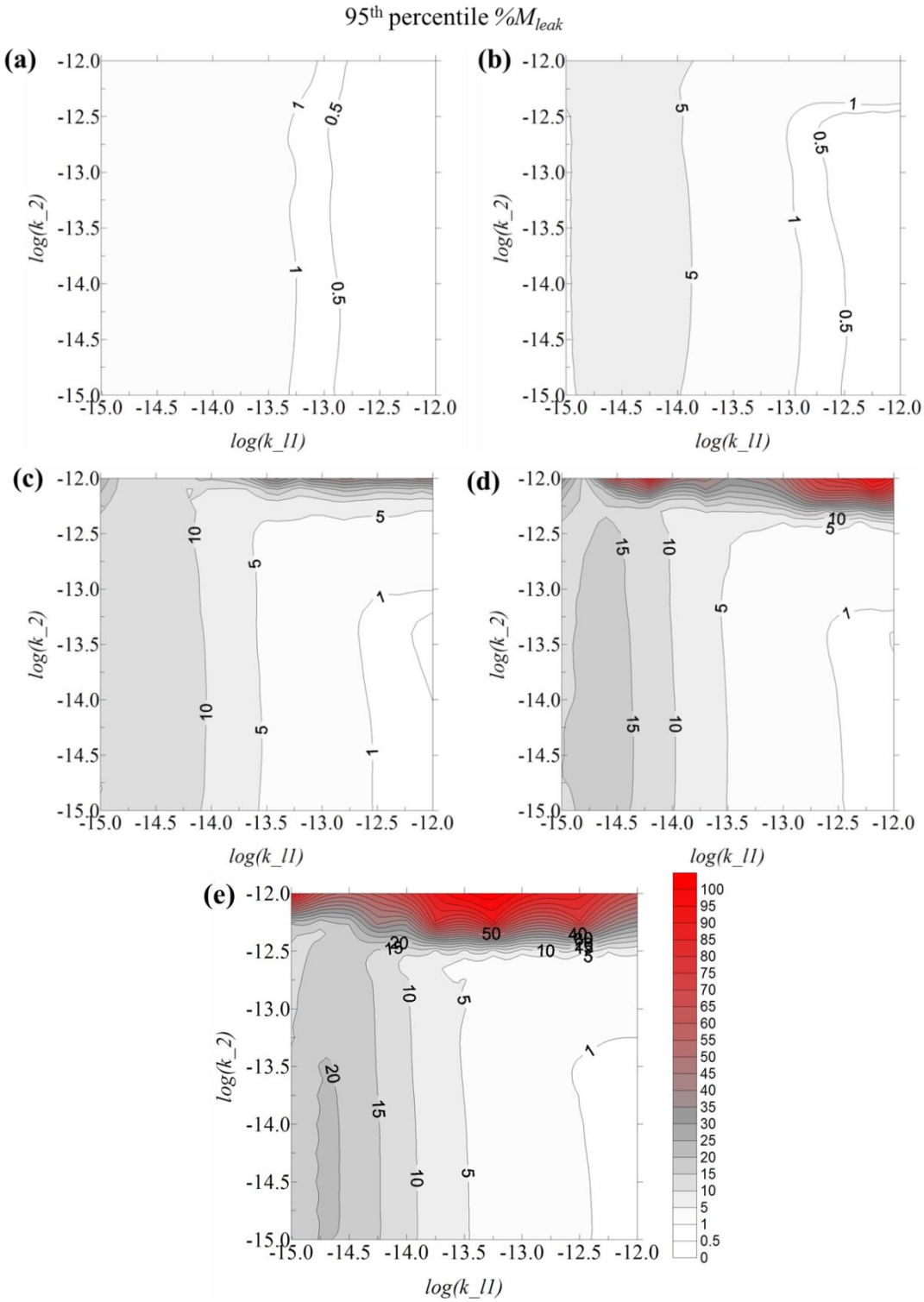

Figure 3-13. Maximum probable CO_2 leakage within a confidence of 95th as a function of the injected formation permeability (k_{l_1}) and the inclusions permeability (k_2): (a) Scenario 1.1, (b) Scenario 2.1, (c) Scenario 3.1, (d) Scenario 4.1, and (e) Scenario 5.1. See Table 3-2 for description of the scenarios.

$\%M_{leak}$ values of Figure 3-13 can be compared to the metric reported by Pacala (2003) that limits the amount of CO_2 leakage returning to the atmosphere to 1% per one year. In Scenario 1.1 (Figure 3-13a), where the probability of finding an inclusion is the lowest, $\%M_{leak}$ would be less than or equal to 1% per one year for values of k_{l_1} greater than 5.01×10^{-14} m^2 (log-$k_{l_1} \geq$ -13.3). On the other hand, if P_2 is increased to 0.01 (Figure 3-13d), in order to maintain the maximum probable CO_2 leakage below the 1% per year threshold, the minimum permeability required for the injected formation and the inclusions should be 3.98×10^{-13} m^2 (log-k_{l_1}= -12.4) and 6.31×10^{-14} m^2 (log-k_2= -13.2), respectively.

This analysis shows geostatistical data such as the probability P_2 and the correlation length l_{xy} play a critical role for the risk assessment prior to the injection of CO_2 into a candidate reservoir. For instance, from Figure 3-13 it is deduced that a P_2 greater than 0.001 with l_{xy}= 200 m (Scenarios 2.1, 3.1, 4.1, and 5.1) is likely to produce CO_2 leakages greater than 1% per year, in which case the injections of CO_2 into the candidate storage formation should not be recommended.

It is also important to emphasize that these estimates are conservative since the limit proposed by Pacala (2003) consists of CO_2 leakage rates back to the atmosphere, whereas in this study the CO_2 mass leakage considered is the CO_2 that escapes the target storage formation l=1. Therefore, additional processes of storage and attenuation that CO_2 may undergo in the overburden formations are not accounted for.

4.3 Results of the study of uncertainty from caprock continuity and passive wells permeability

Uncertainty from passive wells permeability does not impact Δp_{iw} results when this uncertainty is added to caprock continuity uncertainty. These results are coincident to results

when only caprock uncertainty is considered. Therefore, Δp_{iw} results are not shown here. On the other hand, uncertainty from passive wells affects CO_2 mass leakage, especially in scenarios in which CO_2 leakage from the caprock discontinuities is expected to be low.

Figure 3-14 shows a comparison of $\%M_{leak}$ results obtained when considering only the uncertainty of caprock, and both uncertainties (caprock continuity and passive wells permeability) for Scenarios 1.1, 2.1, 3.1, 4.1, and 5.1 (Table 3-2).

Uncertainty from passive wells permeability does not affect CO_2 mass leakage results of caprock continuity uncertainty, independently of P_2 value, when CO_2 mass leakage is greater than 1% as Figure 3-14a–Figure 3-14e show. Both scenarios, with and without passive wells permeability uncertainty, produce similar results of $\%M_{leak}$.

Figure 3-14a reports the results of Scenario 1.1, which has the lowest probability of finding an inclusion on the caprock. We can see that for the scenario considering only caprock continuity uncertainty (dashed line), there is an 82% probability that $\%M_{leak}$ will be equal to or lower than $1x10^{-3}$%. Whereas when taking into account passive well permeability uncertainty (solid line) yields a 0% probability that $\%M_{leak}$ will be equal to or lower than $1x10^{-3}$%.

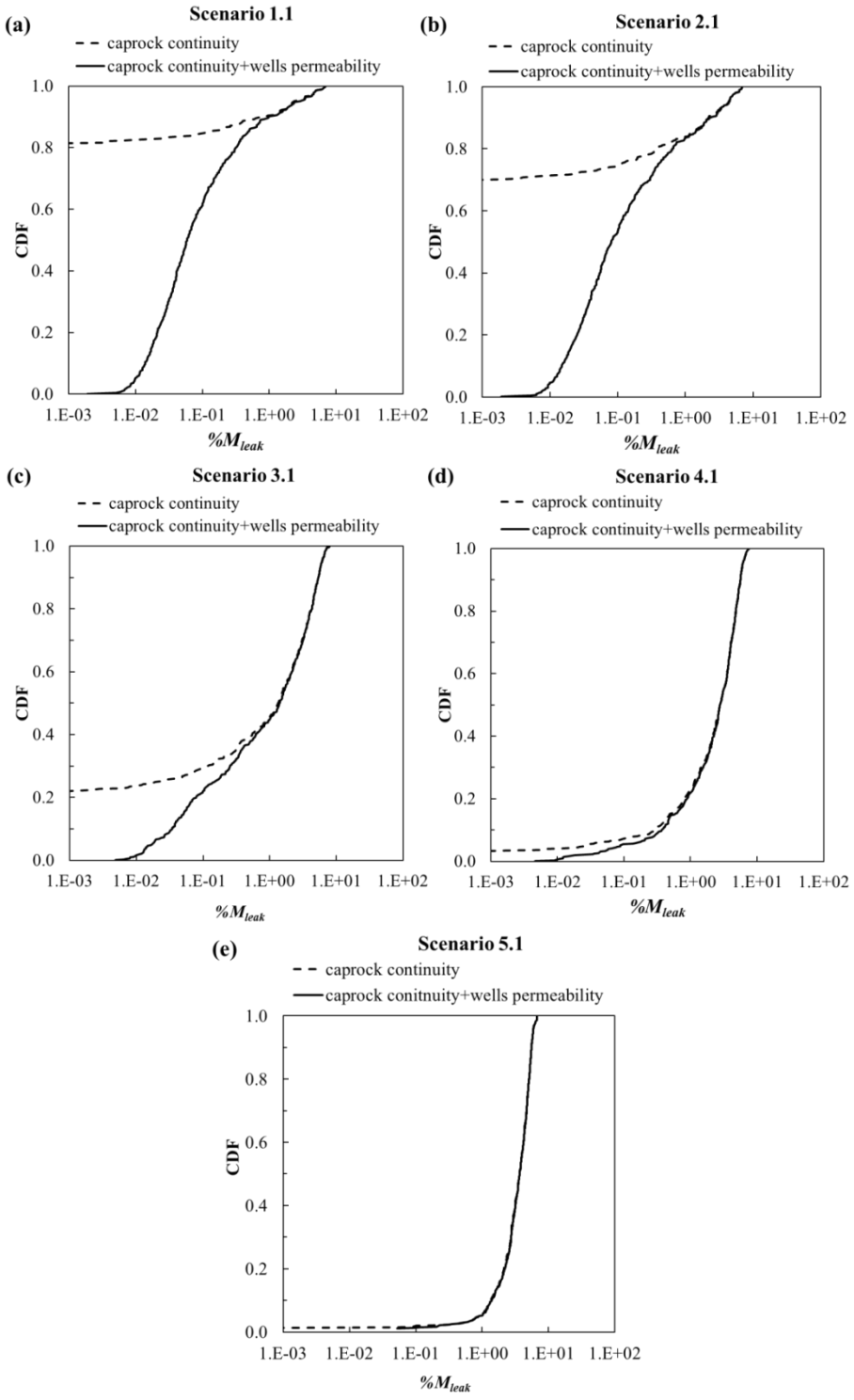

Figure 3-14. CO_2 mass leakage considering uncertainty from the caprock continuity and uncertainty from passive wells permeability. Solid line represents scenario with both uncertainties, and dashed line scenario with only caprock continuity uncertainty: (a) Scenario 1.1, (b) Scenario 2.1, (c) Scenario 3.1, (d) Scenario 4.1, and (e) Scenario 5.1. See Table 3-2 for description of the scenarios.

Scenarios 2.1, 3.1, 4.1 and 5.1 (Figure 3-14b–Figure 3-14e, respectively) present the same tendency seen in Scenario 1.1. Uncertainty from passive wells permeability is noticeable only for low CO_2 mass leakages.

Scenarios with high probability of finding inclusions, such as Scenario 4.1 (Figure 3-14d) and 5.1 (Figure 3-14e), present small differences on their CDFs even for low values of CO_2 mass leakage. The influence on leakage produced by passive wells permeability uncertainty is negligible in comparison to the leakage produced through the weak areas of the caprock for these scenarios.

To study the effect of the correlation length on passive wells permeability uncertainty, Scenarios 2.2, 2.3, and 2.4 are simulated including this uncertainty. Figure 3-14b and Figure 3-15 report the comparison of CO_2 mass leakage results obtained when considering the uncertainty of caprock continuity together with wells uncertainty for the same P_2= 0.001 and different l_{xy} (see Table 3-2 for their description). Results of these figures show that independent of the correlation length, uncertainty of passive wells permeability has an impact on low $\%M_{leak}$. For CO_2 mass leakages greater than 1% this influence is not noticed.

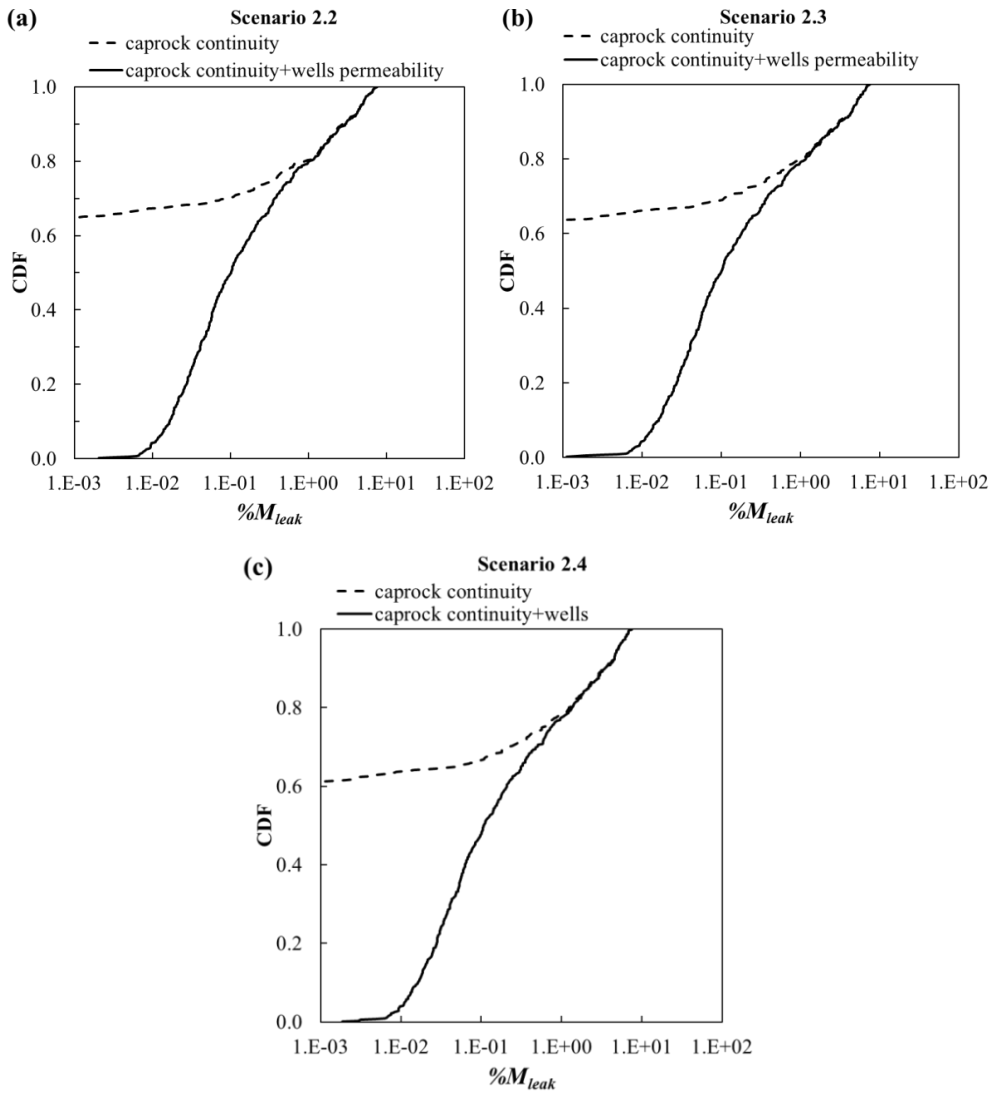

Figure 3-15. CO_2 mass leakage considering uncertainty from the caprock continuity and uncertainty from passive wells permeability. Solid line represents scenario with both uncertainties, and dashed line scenario with only caprock continuity uncertainty. (a) Scenario 2.2, (b) Scenario 2.3, and (c) Scenario 2.4. See Table 3-2 for description of the scenarios.

Figure 3-16 displays the comparison of CO_2 mass leakage obtained for Scenarios 2.1, 2.2, 2.3, and 2.4 when considering both uncertainties (caprock continuity and passive wells permeability). It shows with better detail how the uncertainty of passive wells permeability produces the same $\%M_{leak}$ at low leakages. It also shows that the influence l_{xy} of facies 2 (i.e. the size of the inclusions) is noticed about 0.1%; being more prominent at 1%.

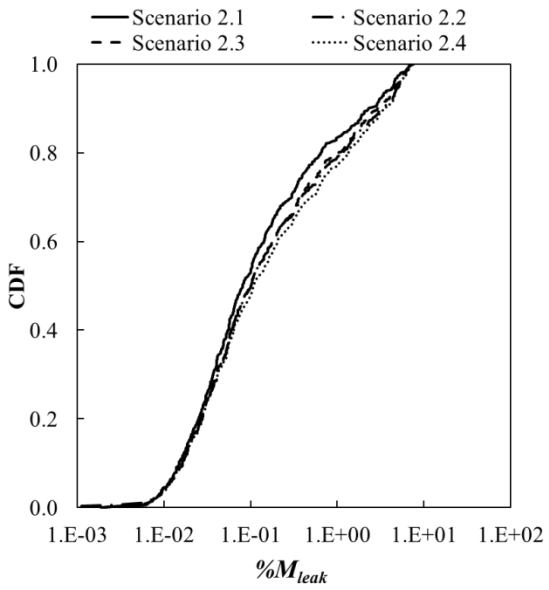

Figure 3-16. Comparison of CO_2 mass leakage considering uncertainty from the caprock continuity and uncertainty from passive wells of Scenarios 2.1, 2.2, 2.3, and 2.4. See Table 3-2 for description of the scenarios.

5 Summary and Conclusions

A sequential indicator simulation algorithm is used to study the uncertainty on the continuity of the geological sealing unit of a candidate formation for GCS located at the Michigan Basin. This algorithm creates binary fields of caprock with low permeability and inclusions with high permeability. The inclusions represent areas where CO_2 could leak through the caprock. A semi-analytical multiphase flow model is used to simulate the CO_2 injection and study the CO_2 leakage. Inclusions having similar spatial locations are grouped together and considered a cluster to reduce the number of leaky points introduced in the multiphase flow semi-analytical algorithm in order to improve its computational efficiency.

Originally, the semi-analytical model is built to simulate escapes through small tubular diameters such as abandoned wells. To understand the limitations of applying the semi-analytical to simulate leakage through large areas of the caprock, a comparison of the semi-analytical

algorithm with a numerical code was carried out. Results show that, in general, the cumulative CO_2 mass leakage produced with both models is of the same order at later times, showing there is a good agreement between their results.

We studied the impact of caprock continuity uncertainty with and without the impact of passive wells permeability uncertainty on CO_2 mass leakage results. From the results of the study of the caprock continuity uncertainty, we can conclude that:

i) Increasing the probability of inclusions occurrence increases the CO_2 leakage to overlying formations, relieving more pressure in the injected aquifer.

ii) Correlation length affects the number of the expected inclusion blocks of one realization that the sequential indicator simulation algorithm generates. Also the distance that the CO_2 plume must travel until it reaches the center of the cluster for high correlation length is shorter, thus favoring the CO_2 leakage. This dependency on correlation length does not seem to increase significantly for correlation scales greater than 400 m.

And iii) influence of the inclusion permeability and injected aquifer permeability on CO_2 leakage is significant, especially in scenarios where a high probability of inclusion occurs, in which CO_2 leakage can be very high.

From the results of the study of the caprock continuity uncertainty and passive wells permeability uncertainty, we conclude that uncertainty from passive wells permeability seems not to have influence on fluid overpressure results when this uncertainty is added to caprock continuity uncertainty. Also, the uncertainty from passive wells permeability does not affect CO_2 mass leakage results of caprock continuity uncertainty when CO_2 mass leakage is high; it only has influence at low values of CO_2 leakages.

6 References

Bachu, S., 2003. Screening and ranking of sedimentary basins for sequestration of CO_2 in geological media in response to climate change. Environmental Geology 44, 277-289.

Baù, D., 2012. Multi-Objective Optimization Approaches for the Design of Carbon Geological Sequestration Systems. DE-FE0001830 Progress Report III, DOE, National Energy Technology Laboratory.

Bergman, P.D., Winter, E.M., 1995. Disposal of carbon-dioxide in aquifers in the US. Energy Conversion and Management 36, 523-526.

Celia, M.A., Nordbotten, J.M., 2009. Practical Modeling Approaches for Geological Storage of Carbon Dioxide. Ground Water 47, 627-638.

Celia, M.A., Nordbotten, J.M., Court, B., Dobossy, M., Bachu, S., 2011. Field-scale application of a semi-analytical model for estimation of CO_2 and brine leakage along old wells. International Journal of Greenhouse Gas Control 5, 257-269.

Chang, K.W., Minkoff, S., Bryant, S., 2008. Modeling Leakage Through Faults of CO_2 Stored in an Aquifer.

Cody, B., Baù, D., González-Nicolás, A., 2014. Improved Semi-Analytical Simulation of Geological Carbon Sequestration (in review). Computational Geosciences.

DEQ, 2013. Oil and Gas Database. Department of Environmental Quality, Michigan.Gov.

Deutsch, C.V., Journel, A.G., 1997. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, USA.

Gasda, S.E., Bachu, S., Celia, M.A., 2004. Spatial characterization of the location of potentially leaky wells penetrating a deep saline aquifer in a mature sedimentary basin. Environmental Geology 46, 707-720.

González-Nicolás, A., Baù, D., Cody, B.M., Alzraiee, A., 2014. Stochastic and Global Sensitivity Analyses of Uncertain Parameters Affecting the Safety of Geological Carbon Storage in Saline Aquifers of the Michigan Basin. International Journal of Greenhouse Gas Control (submitted).

Hahn, G.J., 1967. Statistical models in engineering. Wiley, New York.

Halliburton, 1990. Log 21101375660000 Stech Upper Half (Date: 09/04/1990).

International Energy Agency, 2008. Carbon Capture and Storage: Meeting the Challenge of climate change, Paris.

Journel, G., A., Alabert, F., 1989. Non-Gaussian data expansion in the Earth sciences. Terra Nova 1, 123-134.

Kelkar, M., Perez, G., Chopra, A., 2002. Applied Geostatistics for Reservoir Characterization. Society of Petroleum Engineers.

Kopp, A., Binning, P.J., Johannsen, K., Helmig, R., Class, H., 2010. A contribution to risk analysis for leakage through abandoned wells in geological CO_2 storage. Adv Water Resour 33, 867-879.

Krige, D.G., 1951. A statistical approach to some mine valuations and allied problems on the Witwatersrand, Master's thesis of the University of Witwatersrand. M.Sc. Eng. Thesis of University of Witwatersrand, Johannesburg.

Matheron, G., 1963. Principles of geostatistics. Economic Geology 58, 1246-1266.

Metz, B., Intergovernmental Panel on Climate Change. Working Group III., 2005. IPCC Special Report on Carbon Dioxide Capture and Storage. Cambridge University Press, for the Intergovernmental Panel on Climate Change, Cambridge.

Nicholls, N., Gruza, G., Jouzel, J., Karl, T., Ogallo, L., Parker, D., 1996. Observed climate variability and change. Cambridge University Press.

Nogues, J.P., Court, B., Dobossy, M., Nordbotten, J.M., Celia, M.A., 2012. A methodology to estimate maximum probable leakage along old wells in a geological sequestration operation. International Journal of Greenhouse Gas Control 7, 39-47.

Nordbotten, J.M., Kavetski, D., Celia, M.A., Bachu, S., 2009. Model for CO_2 Leakage Including Multiple Geological Layers and Multiple Leaky Wells. Environ Sci Technol 43, 743-749.

Olea, R.A., 1999. Geostatistics for Engineers and Earth Scientists. Kluwer Academic.

Pacala, S.W., 2003. Global Constraints on Reservoir Leakage, in: Gale, J., Kaya, Y. (Eds.), Greenhouse Gas Control Technologies - 6th International Conference. Pergamon, Oxford, pp. 267-272.

Ruether, J.A., 1998. FETC Programs for Reducing Greenhouse Gas Emissions, Other Information: PBD: Feb 1998, p. Medium: ED; Size: 20 p.; Other: FDE: PDF; PL:.

Sakamoto, Y., Tanaka, A., Tenma, N., Komai, T., 2011. Numerical study on flow behavior of CO_2 in an aquifer for risk assessment of carbon capture and storage. Energy Procedia 4, 4170-4177.

SCH, 1983. Log 21101365880000 Burch Main Suite (Date 06/19/1983).

SCH, 1991. Log 21101375660000 Stech Lower Half (Date: 06/29/1991).

Schlumberger, 2010. Eclipse technical description, v. 2010.1, Report, Abingdon, U.K.

Solomon, S., Intergovernmental Panel on Climate Change., Intergovernmental Panel on Climate Change. Working Group I., 2007. Climate change 2007 : the physical science basis : contribution of Working Group I to the Fourth Assessment Report of the

Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge ; New York.

Takahashi, W., 2000. Nonlinear Functional Analysis, Fixed point theory and its applications, p. iv+ 276. Yokohama, Yokohama, Japan.

Trangenstein, J.A., Bell, J.B., 1989. Mathematical structure of compositional reservoir simulation. SIAM journal on scientific and statistical computing 10, 817-845.

Trebin, F.A., 1945. Permeability to oil of sandstone reservoir.

Turpening, R., Toksöz, M., Born, A., al., e., 1992. Reservoir Delineation Consortium Annual Report, Massachusetts Institute of Technology, Cambridge.

Van Genuchten, M.T., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils1. Soil Science Society of America Journal 44, 892.

Zhang, Y., Oldenburg, C.M., Finsterle, S., 2010. Percolation-theory and fuzzy rule-based probability estimation of fault leakage at geologic carbon sequestration sites. Environ Earth Sci 59, 1447-1459.

Zhou, Q., Birkholzer, J.T., Mehnert, E., Lin, Y.-F., Zhang, K., 2009. Modeling Basin- and Plume-Scale Processes of CO2 Storage for Full-Scale Deployment. Ground Water 48, 494-514.

4 Chapter: Detection of potential leakage pathways from geological carbon storage by fluid pressure data assimilation

Summary

One of the main concerns of geological carbon storage (GCS) systems is the risk of leakage through "weak" permeable areas of the sealing formation or caprock. Since the fluid pressure pulse travels faster than the carbon dioxide (CO_2) plume across the storage reservoir, a fluid pressure change is inevitably transmitted into overlying permeable formations through "weak" areas of the caprock, and can be potentially detected sooner than actual CO_2 leakage occurs. In this work, an inverse modeling method based on fluid pressure measurements collected in strata above the target CO_2 storage formation is proposed, which aims at detecting the presence, the location, and the extent of possible carbon leakage pathways. We combine a three-dimensional subsurface multiphase flow model with ensemble-based data assimilation algorithms to identify potential caprock discontinuities that can undermine the long-term safety of GCS. This work examines and compares the capabilities of data assimilation algorithms such as the ensemble smoother (ES) and the ensemble Kalman smoother (EnKS) to detect the presence of brine and/or CO_2 leakage pathways, potentially in real-time during GCS operations. For the purpose of this study, changes in fluid pressure in the brine aquifer overlying to CO_2 storage formation aquifer are assumed to be observed in monitoring wells, or provided by 4D time-lapse seismic surveys. Caprock discontinuities are typically characterized by higher values of permeability, so that the permeability distribution tends to fit to a non-Gaussian bimodal process, which does not comply with the requirements of the ES and EnKS algorithms. Here,

issues related to the non-Gaussianity of the caprock permeability field are investigated by developing and applying a normal score transform procedure. Results suggest that the EnKS is more effective than the ES in characterizing caprock discontinuities.

1 Introduction

In the last decades, geological carbon storage (GCS) has been identified as a technology of great potential for reducing anthropogenic emissions of carbon dioxide (CO_2) in the atmosphere. However, while technically feasible, GCS must be carefully evaluated with respect to environmentally threatening side effects, such as the leakage of CO_2 through sealing formations or caprock. When CO_2 is injected underground it displaces the resident fluid in the geological formation, which, in the case of deep aquifers, is constituted mainly by high-density saline water or brine. If the brine and/or CO_2 find a pathway through the caprock, they migrate into overlying formations, which may negatively affect the quality of shallow fresh water resources (Birkholzer et al., 2009). In particular, CO_2 can produce pH changes of groundwater resources by increasing the concentration of carbonates, which can influence dissolution and sorption of minerals and hazardous trace metals, and consequently deteriorate groundwater quality (Kharaka et al., 2006; Apps et al., 2010). The 2005 report of the Intergovernmental Panel on Climate Change (IPCC, 2005) provided a list of potential CO_2 leakage pathways: (1) "weak" areas of the caprock (permeable areas) where CO_2 breaks into the caprock if capillary entry pressure is exceeded; (2) faults and fractures, and (3) poorly completed and/or abandoned wells. In order to monitor and/or detect leakage sources different techniques have been proposed and classified into three main groups (DOE/NETL, 2009): *i)* atmospheric monitoring techniques, *ii)* near-surface monitoring techniques, and *iii)* subsurface monitoring techniques. Seto and McRae

(2011) advanced an integrated framework that provides an understanding of the type of monitoring techniques available for GCS.

Measuring fluids pressure is a key component of subsurface monitoring (Group III). Since the fluid pressure pulse travels faster than the CO_2 plume across the storage reservoir, if leakage of brine occurs through "weak" areas of the caprock, the fluid pressure change propagates into overlying permeable formations and can be identified sooner than CO_2 leakage. This fluid pressure change can be detected, for example, by pressure-monitoring wells (DOE/NETL, 2009) and time-lapse seismic data (Cole et al., 2002; Cole et al., 2006; MacBeth et al., 2006).

A number of studies related to monitoring pressure changes in observation wells have been published (Chabora and Benson, 2009; Nogues et al., 2011; Zeidouni et al., 2011; Park et al., 2012; Sun and Nicot, 2012). Chabora and Benson (2009) presented a method to assess the usefulness of pressure monitoring in overlying formations based on the correlation between calculated pressure changes and a proposed detection factor. Similarly, Zeidouni et al. (2011) developed an analytical solution to detect leakage through pressure-monitoring wells screened in the geologic formations overlying the injection point. They showed that their model is capable of locating and quantifying the leakage of CO_2 and found a positive correlation between the accuracy of the estimation and the number of available monitoring wells. Also, Nogues et al. (2011) developed an analytical solution to estimate the CO_2 and brine leakage from pressure variations observed at monitoring wells and investigated optimal location of the monitoring wells to improve leakage detection. Park et al. (2012) proposed a methodology to detect CO_2 leakage by measuring pressure changes at monitoring wells with a constrained distribution. Sun and

Nicot (2012) presented an inversion method based on a global optimization algorithm to identify CO_2 leakage from pressure anomalies observed in the layers overlying the injected aquifer.

While analytical solutions and optimization algorithms constitute appealing tools for leakage detection, improved accuracy of estimation and reduced computational cost may be achieved by resorting to ensemble-based Kalman filter (KF) (Kalman, 1960) data assimilation methods. Some advantages of these methods are, for instance: (1) reduced computational demand in comparison to global optimization methods; (2) deterministic optimization techniques typically produce a unique solution, possibly suboptimal, for an ill-posed inverse problem and thus ignore the existence of infinite number of possible solutions; and (3) typical analytical solutions are based on highly idealistic assumptions and do not account for system heterogeneities.

Another alternative to analytical solutions and optimization algorithms is offered by conventional inverse approaches such as Monte Carlo inverse methods (e.g., Hendricks Franssen (2009)); however, the intensive central processor unit (CPU) effort and their inability to assimilate real-time data may hinder their application for CO_2 leakage detection problems. In contrast, ensemble-based KF data assimilation methods require relatively less CPU time and allow for incorporating available measurements into model results as they are collected. Data assimilation methods are commonly used in various disciplines to update model states (Nævdal et al., 2002; Keppenne and Rienecker, 2003; Chen and Zhang, 2006), and system parameters (Chen and Zhang, 2006; Zhou et al., 2012) based on field observations.

The classical KF (Kalman, 1960) provides an optimal solution in the case of linear Gaussian systems and unbiased measurements. Expanding the applicability of KF to nonlinear problems can be achieved by using an ensemble of realizations to approximate the prior

uncertainty in states and parameters (Evensen, 1994). According to Evensen (2009), ensemble-based KF methods can be subdivided into three main categories depending on the scheme adopted to assimilate measurements: (1) ensemble Kalman smoother (EnKS) algorithms, (2) ensemble Kalman filter (EnKF) algorithms, and (3) ensemble smoother (ES) algorithms. Some examples of successful application of EnKS can be found in the literature (Brusdal et al., 2003; Dunne and Entekhabi, 2006; Ngodock et al., 2006; Durand et al., 2008). Examples of EnKF applications are reported by Chen and Zhang (2006), Hendricks Franssen and Kinzelbach (2008), and Li et al. (2012). Skjervheim and Evensen (2011) effectively applied the ES to solve the history-matching problem in a petroleum reservoir and compared it with the EnKF. Bailey and Bau (2012) used ES iteratively to obtain the hydraulic conductivity through assimilation of water table height and stream flow rate data. Herrera and Simuta-Champo (2012) applied the ES to optimize the three-dimensional location of sampling wells in an aquifer and estimate contaminant concentrations.

To provide an optimal solution, these ensemble-based KF methods require: (a) unbiased and uncorrelated observation errors, (b) parameters and state variables to fit to multiGaussian distributions, and (c) a linear relationship between predicted data and model data. Since in practical applications at least one of these assumptions is often not satisfied, there is a need to devise approaches to circumvent these limitations. For example, to avoid the problem of non-Gaussianity, Gu and Oliver (2006) successfully applied the normal score transform (NST) (Goovaerts, 1997) to saturation data in a one-dimensional multiphase reservoir. Other more recent studies have addressed issues of non-Gaussianity of state variables or parameters (e.g. Simon and Bertino, 2009; Béal et al., 2010; Zhou et al., 2011; Schöniger et al., 2012; Zhou et al., 2012; Crestani et al., 2013).

In this work, we propose an inverse modeling method based on the assimilation of fluid pressure data collected in strata above the target carbon storage formation. The objective is to detect the presence, location and extent of potential CO_2 leakage pathways through the caprock formation. The inverse modeling framework relies on the combination of a subsurface multiphase flow model with two ensemble-based data assimilation algorithms. In particular, we investigate and compare the capabilities of the ES and the EnKS to identify the presence of brine and/or CO_2 leakage pathways during GCS operations. Issues related to the non-Gaussianity of caprock permeability field are also examined. For the purpose of this study, we assumed that changes in pressure in the upper aquifer are either observed at monitoring wells or provided by 4D seismic time-lapse. To pose the CO_2 detection problem as an inverse estimation of permeability from known pressure measurements, we assumed that regions with high brine or CO_2 leakage can be represented as regions with high permeability, resulting in a non-Gaussian bimodal distribution of the caprock permeability. To overcome issues of non-Gaussianity of the caprock permeability spatial distribution, such parameter is transformed using a NST procedure. The multiphase flow model used in this work is ECLIPSE (Schlumberger, 2010), a reservoir simulator widely used in the petroleum industry, which is also able to simulate injection of supercritical CO_2 into the subsurface.

This paper is organized as follows. Section 2 presents first the multiphase flow governing equations, followed by the ES and EnKS approach. Section 3 demonstrates the application of these two methods for identifying potential leakage areas of the caprock during GCS operations. Finally, Section 4 summarizes the major findings of the present work.

2 Methodology

In this section, a basic description of the multiphase flow model ECLIPSE (Schlumberger, 2010) used in the numerical experiments is given. Next the ES and the EnKS algorithms are described, followed by the definition of parameters that can be used to assess their performance.

2.1 Multiphase flow model

The multiphase flow equations are solved using the CO2STORE option of ECLIPSE (Schlumberger, 2010). The model is written in Fortran90, and is based on a finite-difference discretization of the subsurface system. The compositional version (E300) of ECLIPSE is here used to perform two-phase compositional simulation of gas-brine systems, which computes mass balances for each component. Our interest focuses on the CO_2 and H_2O components, whereas salt components, such as NaCl, $CaCl_2$, and $CaCO_2$ are neglected.

Assuming the presence of only two fluid phases, a CO_2-rich gas phase denoted as g, and H_2O-rich liquid phase denoted as w, C_{jg} represents the mass fraction of component j present in the gas phase and C_{jw} represents the mass fraction of component j in the liquid phase. Based on mass continuity, in a system of N_c components, mass fractions must be such that:

$$\sum_{j=1}^{N_c} C_{jg} = 1 \tag{4-1a}$$

$$\sum_{j=1}^{N_c} C_{jw} = 1 \tag{4-1b}$$

The mass balance of component j for one-dimensional systems is expressed by the following partial differential equation (PDE):

$$-\frac{\partial}{\partial x}\left(C_{jg}\rho_g v_g + C_{jw}\rho_w v_w\right) = \frac{\partial}{\partial t}\left[\varphi\left(C_{jg}\rho_g s_g + C_{jw}\rho_w s_w\right)\right] \tag{4-2}$$

subject to the condition:

$$s_g + s_w = 1 \tag{4-3}$$

where: x represents the spatial coordinate, φ is the porosity of the medium, and ρ_α, s_α and v_α are the density, the saturation, and the Darcy velocity of phase α (α denotes the phase type, either water w or gas g), respectively. Darcy's velocity of phase α is expressed as:

$$v_\alpha = -\frac{kk_{r\alpha}}{\mu_\alpha}\frac{\partial p_\alpha}{\partial x} \tag{4-4}$$

where: k is the intrinsic permeability of the porous medium, $k_{r\alpha}$, μ_α and p_α are the relative permeability, the dynamic viscosity, and the partial pressure of phase α, respectively. The capillary pressure p_c is defined as:

$$p_c = p_g - p_w \tag{4-5}$$

The relative permeabilities and capillary pressure depend upon saturation values, which can be calculated using different models such as Van Genuchten's model (Van Genuchten, 1980) and Brooks and Corey model (Brooks and Corey, 1964), or they can be obtained from experimental data.

Based on Equations (4-2) and (4-4), the one-dimensional flow of N_c fluid components is governed by the following system of PDEs:

$$\begin{aligned}&\frac{\partial}{\partial x}\left(C_{jg}\rho_g\frac{kk_{rg}}{\mu_g}\frac{\partial p_g}{\partial x} + C_{jw}\rho_w\frac{kk_{rw}}{\mu_w}\frac{\partial p_w}{\partial x}\right)\\ &\qquad = \frac{\partial}{\partial t}\left[\varphi\left(C_{jg}\rho_g s_g + C_{jw}\rho_w s_w\right)\right] \quad (j = 1, \dots, N_c)\end{aligned} \tag{4-6}$$

The compositional option E300 of ECLIPSE (Schlumberger, 2010) follows the formulation of Trangenstein and Bell (1989), in which the components flux is defined as the sum

of the phases of the molar densities $\hat{\rho}_{j\alpha}$ (moles per volume, where a mole is the mass divided by the molecular weight), times its flow rate v_α/s_α. Therefore the mass conservation law (PDE (4-6)) can be rewritten as:

$$\frac{\partial}{\partial x}\left(\frac{\hat{\rho}_{jg}}{s_g}\frac{kk_{rg}}{\mu_g}\frac{\partial p_g}{\partial x}+\frac{\hat{\rho}_{jw}}{s_w}\frac{kk_{rw}}{\mu_w}\frac{\partial p_w}{\partial x}\right)=\frac{\partial}{\partial t}\left(\varphi\left(\hat{\rho}_{jg}+\hat{\rho}_{jw}\right)\right)\quad (j=1,\dots,N_c) \tag{4-7}$$

where the total amount of moles of each component j per pore volume is constant: $\hat{\rho}_j=\hat{\rho}_{jg}+\hat{\rho}_{jw}$.

In this work, we adopt a finite-difference discretization in time based on a fully implicit Euler scheme (selected by the E300 option FULLIMP) to ensure stability of the numerical solution. Newton's method is used to linearize and solve a set of residual (mass balance error) non-linear equations derived from the three-dimensional form of Equation (3-18). The equation of the residual R for each fluid component in each gridblock cell at each time step may be expressed as:

$$R=\frac{M_{t+\Delta t}-M_t}{dt}+F(p_{t+\Delta t},\hat{\rho}_{t+\Delta t})+Q(p_{t+\Delta t},\hat{\rho}_{t+\Delta t})=0 \tag{4-8}$$

where: M is the mass accumulated during the current time step, F is the net flow rate into neighboring grid blocks, and Q is the net flow rate into wells during the time step.

A set of solution variables is required to solve Equation (4-8). The number of independent variables has to be equal to the number of residual conditions. In E300, the primary solution variables are pressure p and molar densities of each component $\hat{\rho}_j$. In Equation (4-8), pressures and molar densities are solved implicitly at the end of the time step and, as previously mentioned, using Newton's method. The linear equations produced by Newton's method are preconditioned and solved simultaneously by a nested-factorization procedure that uses the Orthomin iterative method (Vinsome, 1976).

The CO2STORE option of ECLIPSE allows the two phases to contain the components CO_2 and H_2O, and can account for the processes of dissolution, evaporation, and condensation. The partitioning of CO_2 and H_2O is computed by following a fugacity equilibrium between the two phases (Spycher and Pruess (2005) and (2010)). The fugacity measures how the available energy of a real fluid differs from the available energy of an ideal gas. To calculate CO_2 fugacity ECLIPSE uses a modified Redlich-Kwong equation of state (Redlich and Kwong, 1949), whereas water fugacity is computed using Henry's law. ECLIPSE also computes densities and viscosities of CO_2 and water phases. CO_2 gas density is calculated by a cubic equation of state according to Spycher and Pruess (2005), whereas CO_2 gas viscosity is computed based on relationships developed by Vesovic et al. (1990) and Fenghour et al. (1998).

ECLIPSE output includes, among others, the spatial distributions of fluid phase pressures and saturations in the model domain. For each cell, centered at the generic location (x, y, z), the equivalent fluid pressure $\bar{p}$ at time t is calculated as the average of the phase pressures, p_w and p_g, weighted with respect to the phase saturations, s_w and s_g (Bishop, 1959; Gray and Miller, 2007):

$$\bar{p} = p_w s_w + p_g s_g \tag{4-9}$$

Therefore the fluid pressure change at location (x, y, z) and time t is defined as:

$$\Delta p = \Delta p(x, y, z; t) = \bar{p}(x, y, z; t) - p_0(x, y, z) \tag{4-10}$$

where $p_0(x, y, z)$ is the initial fluid pressure (at time t = 0).

2.2 Ensemble based data assimilation

The problem of detecting CO_2 leakage is approached herein as a data assimilation problem, in which observations of fluid pressure at a number of given locations and subsequent times are assimilated into the multiphase model results to identify potential carbon leakage

pathways in the caprock. Typically, the amount of available pressure data is less than the unknowns, that is, the model output at the cells of the finite difference grid, leading to an ill-posed problem. In this case, since the inverse problem has no unique solution, a Bayesian framework is adopted. Prior information on the system state, characterized by the vector $\boldsymbol{\phi}$ including the fluid pressure distribution within the model domain at a given time, can be updated based on pressure data collected at that time using Bayes' law:

$$pdf(\boldsymbol{\phi}|\boldsymbol{m},\boldsymbol{I}) = \frac{pdf(\boldsymbol{m}|\boldsymbol{\phi})pdf(\boldsymbol{\phi},\boldsymbol{I})}{pdf(\boldsymbol{m}|\boldsymbol{I})} \tag{4-11}$$

where the vector $\boldsymbol{m}$ includes local pressure observations, $pdf(\boldsymbol{\phi}|\boldsymbol{m},\boldsymbol{I})$ represents the posterior probability distribution function (PDF) of $\boldsymbol{\phi}$ given the observations $\boldsymbol{m}$ and a generic prior information $\boldsymbol{I}$ on the system, $pdf(\boldsymbol{m}|\boldsymbol{\phi})$ is the so-called "likelihood" PDF, $pdf(\boldsymbol{\phi},\boldsymbol{I})$ is the prior PDF of $\boldsymbol{\phi}$, and $pdf(\boldsymbol{m}|\boldsymbol{I})$ constitutes a normalization term. An exact solution to Equation (4-11) can be obtained when the measurements $\boldsymbol{m}$ are related to the state $\boldsymbol{\phi}$ through a linear relationship, and when all PDFs conform to Gaussian joint distributions. This solution is referred to as the Kalman filter (KF) (Kalman, 1960). In the KF, the assimilation of system observations into the system state is carried out with a two-step process, defined by a forecast stage and an update stage. In the forecast stage, a forward in time prediction of the state and its error covariance is made. This prediction is carried out at the time in which system observations are made. In the update stage, the system state is corrected based to resemble field measurements.

In addition to being limited to Gaussian linear systems, the application of the KF to high dimensional system states is computationally expensive. To address these shortcomings, Evensen (1994) devised the ensemble Kalman filter (EnKF) in order to extend the KF to nonlinear systems. Within the EnKF, the prior PDF of $\boldsymbol{\phi}$, $pdf(\boldsymbol{\phi},\boldsymbol{I})$ (Equation (4-11)) is approximated by an ensemble of realizations that characterize the prior uncertainty in the system state. The EnKF

can be used to correct not only the system state, but also the system parameters. This may be achieved by "augmenting" the prior ensemble of the state vectors $\boldsymbol{\phi}$ with a corresponding ensemble of the system parameters $\boldsymbol{P}$, thus resulting in a state-parameter forecast ensemble, $[\boldsymbol{\phi}\ \boldsymbol{P}]$.

Similar to the classic KF, the EnKF (Evensen, (1994)) calculates system updates sequentially as system measurements are made available; the forward model forecast simulation proceeds from the current to the next data assimilation time. In contrast, the ES (Van Leeuwen and Evensen, (1996)) uses a forecast state formed by the system states at all measurement times and by the system parameters. All available data are assimilated simultaneously, as opposed to sequentially in time. With the EnKS (Evensen and Van Leeuwen, 2000), the system state and its parameters are updated as measurements become available, in a fashion similar to the EnKF. However, after each update step, the forward model is restarted using the updated parameters, from the initial time to the time at which new systems observations are made.

Figure 4-1 illustrates the procedural structures of (a) the ES, (b) the EnKS, and (c) the EnKF. The horizontal red arrows indicate forecast simulation stages. The vertical blue arrows represent both collection and assimilation of data into the forecast model results. For the ES and EnKS schemes (Figures 1a and 1b), the horizontal green arrows indicate the "restart" of the forward model from the initial time after each update step.

With the ES (Figure 1a), the forward forecast stochastic simulation is performed only once, whereas the EnKS (Figure 1b) requires re-running the forward model after each update step from time t=0 to the next data assimilation time. Consequently, the ES is computationally more advantageous than the EnKS, particularly in the case of "expensive" simulation models. From this perspective, the EnKF (Figure 1c) involves a computational burden similar to that of

the ES, since it does not require restarting the model, and the forward simulation continues from the current data assimilation time to the next.

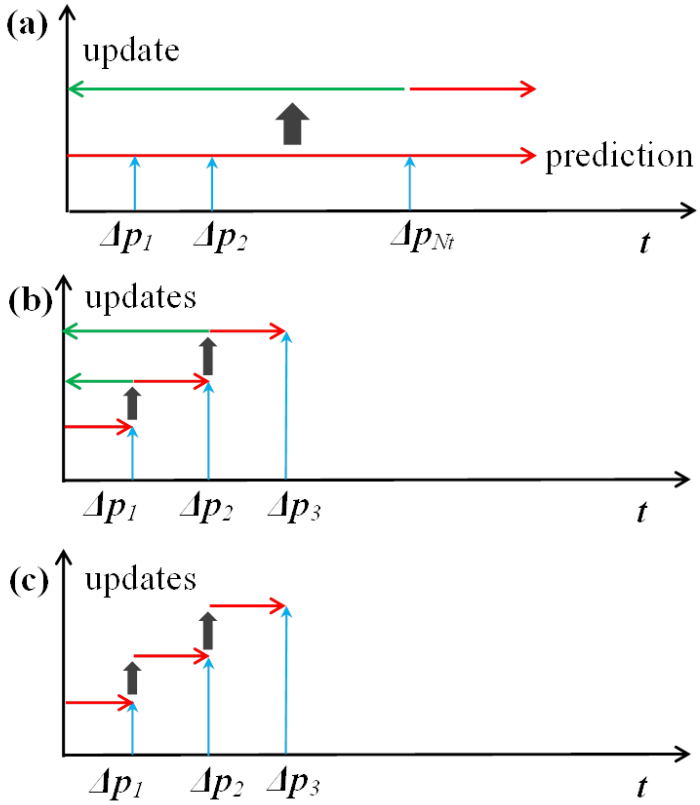

Figure 4-1. Data assimilation processes for (a) the ES, (b) the EnKS, and (c) the EnKF (adapted from Evensen (2009)). Horizontal red arrows indicate forecast simulation stages. Vertical blue arrows represent both collection and assimilation of data. Horizontal green arrows indicate the "restart" of the forward model from the initial time after each update step. $\Delta \boldsymbol{p_i}$ denotes the vector $\boldsymbol{m}$ of fluid overpressure data collected at a number $\boldsymbol{N_t}$ of generic times $\boldsymbol{t_i}$ ($\boldsymbol{i}$=1,2,…,$\boldsymbol{N_t}$).

A well-known inconsistency of the EnKF, when applied to non-linear processes, is constituted by its tendency to provide corrected state values that are not coherent with the physical laws behind the model equations. This inconsistency is further accentuated if the EnKF is used to update system parameters as well as the system states, and may have a significant effect on the accuracy of the forward forecast simulation performed from the current data

assimilation step to the next. For this reason, the EnKF is not considered in this study, and only the ES and the EnKS are investigated for the identification of possible leakage pathways of brine/CO_2 through the caprock.

2.2.1 Forecast of system parameters

In this work, the uncertain parameters of interest are the spatial distribution of permeability in the caprock layer overlying the CO_2 storage formation. As explained above, we use data assimilation algorithms to estimate system parameters by considering "augmented" state-parameter ensemble $[\boldsymbol{\phi}\ \boldsymbol{P}]$. For the purpose of this study, the spatial distribution of the caprock permeability k is simulated geostatistically, using as a two-dimensional, stationary, spatially-distributed random process (Marsily, 1986). According to this process, at any given location the log-transformed permeability, Y=log-k, is assumed to fit to a bimodal PDF, characterized by two average values μ_{Y_1} and μ_{Y_2} with overall probabilities of occurrence P_1 and $P_2 = 1 - P_1$, respectively:

$$\log k = Y = B(\mu_{Y_1}, P_1; \mu_{Y_2}, P_2) \tag{4-12}$$

The first value, μ_{Y_1}, denotes a sealing caprock, whereas the second value, μ_{Y_2}, denotes a permeable, potentially leaky, "weak" area of the caprock ($\mu_{Y_1} < \mu_{Y_2}$). In addition, a local variability of caprock Y is prescribed by adding a "nugget" Gaussian noise, with zero mean and standard deviation equal to σ'. Figure 4-2 shows a demonstrative example of a bimodal cumulative distribution function (CDF) for the caprock permeability, obtained using P_1= 0.8, P_2= 0.2, μ_{Y_1}= -20, μ_{Y_2}= -13, and σ'= 0.5.

The spatial variability of caprock k is characterized by two exponential isotropic covariance functions, denoted as $C_{Y_1Y_1}(d;\sigma_{Y_1}^2,\lambda_{Y_1})$ and $C_{Y_2Y_2}(d;\sigma_{Y_2}^2,\lambda_{Y_2})$ for sealing caprock facies and weak caprock facies, respectively:

$$C_{Y_iY_i}(d;\sigma_{Y_i}^2,\lambda_{Y_i}) = \sigma_{Y_i}^2 exp\left(-\frac{d}{\lambda_{Y_i}}\right) \quad (i = 1,2) \tag{4-13}$$

The scalar d represents the horizontal distance between any two points; $\sigma_{Y_1}^2$, $\sigma_{Y_2}^2$, and λ_{Y_1}, λ_{Y_2} represent the variances and the correlation lengths of the two facies. Note that $\sigma_{Y_1}^2 = P_1(1-P_1)$ and $\sigma_{Y_2}^2 = P_2(1-P_2)$.

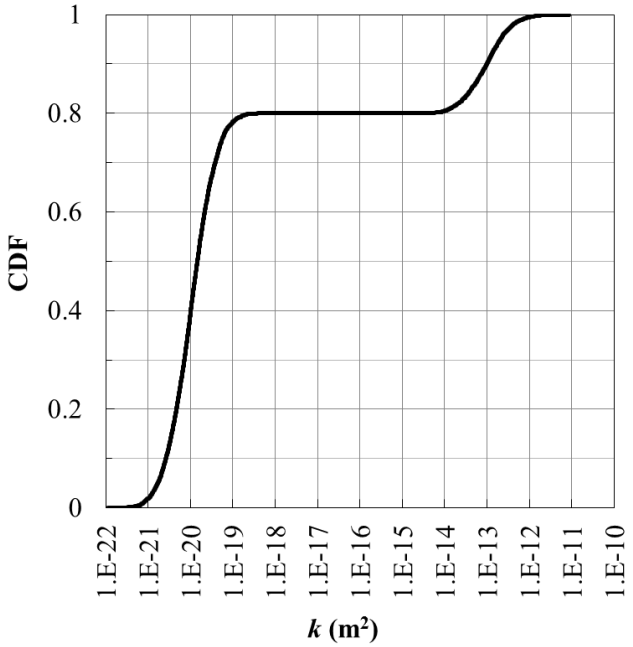

Figure 4-2. Example bimodal CDF of the caprock permeability obtained using parameters of Table 4-1, but probabilities P_1= 0.8 and P_2= 0.2.

Similar to the caprock permeability, the aquifers' permeability is simulated geostatistically, as a three-dimensional, stationary Gaussian process (Marsily, 1986) in a log-transformed space, characterized by an anisotropic exponential covariance model:

$$\log k = Y = N(\mu_Y;\sigma_Y) \tag{4-14}$$

$$C_{YY}(\boldsymbol{d};\sigma_Y^2,\lambda_1,\lambda_2,\lambda_3)=\sigma_Y^2 exp\left(\sqrt{\sum_{i=1}^{3}\frac{d_i^2}{\lambda_i^2}}\right) \quad (4\text{-}15)$$

where: μ_Y and σ_Y^2 are the mean and variance of the aquifer log-transformed permeability Y; d_1, d_2, and d_3 are the components of the distance vector $\boldsymbol{d}$ along the coordinate directions x, y, and z, respectively; and λ_1, λ_2, and λ_3 are the corresponding spatial correlation scales.

Using the geostatistical models described above, it is possible to generate an ensemble of N_{MC} equally likely realizations of the Y field. This is denoted as $\boldsymbol{P}=[\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_{N_{MC}}]$, where $\boldsymbol{Y}_j\ (j=1,\dots,N_{MC})$ is a $n\times 1$ column vector, and n is the number of cells in the finite-difference grid used to discretize the model domain. The resulting ensemble can be seen as a discrete approximation of the prior joint PDF of Y.

In this work, a sequential categorical indicator Kriging simulator (CIKSIM) algorithm (Baù, 2012) is used to generate the ensemble of caprock permeability fields. The ensemble of aquifer permeability fields is generated using a sequential Gaussian simple Kriging (SKSIM) algorithm (Baù and Mayer, 2008). Note that, with this approach, statistical independence between the permeability distributions in the aquifers and in the caprock is intrinsically assumed. It is worth pointing out that the set of geostatistical parameters of the models (Equations (4-12)-(4-15)) $\left(\mu_{Y_1},P_1,\mu_{Y_2},P_2,\sigma',\lambda_{Y_1},\lambda_{Y_2};\mu_Y,\sigma_Y^2,\lambda_1,\lambda_2,\lambda_3\right)$ necessary to generate the forecast parameter ensemble $\boldsymbol{P}$ are assumed to be known, and truly constitute the prior information $\boldsymbol{I}$ of this problem (Equation (4-11)).

2.2.2 Forecast of system states

The forecast ensemble $\boldsymbol{\phi}$ of fluid pressure distributions in the model domain is obtained through a stochastic, or Monte Carlo, multiphase flow simulation performed with ECLIPSE

(Schlumberger, 2010). In this step, the three-dimensional version of PDE (3-18) is solved numerically for each realization $\boldsymbol{Y}_j\ (j = 1, \dots, N_{MC})$ of the forecast parameter ensemble $\boldsymbol{P}$ in order to compute the corresponding forecast state ensemble $\boldsymbol{\phi} = [\boldsymbol{p}_1, \dots, \boldsymbol{p}_{N_{MC}}]$, where $\boldsymbol{p}_j\ (j = 1, \dots, N_{MC})$ is the $n \times 1$ vector including the fluid pressure values (Equation (4-10)) at the cells of the finite-difference grid.

The augmented forecast state-parameter matrix is constructed as $\boldsymbol{X}_f = [\boldsymbol{\phi}\ \boldsymbol{P}]^{\mathrm{T}}$. The size of this matrix is $n_f \times N_{MC}$, where n_f is the number of degrees of freedom used to characterize the state-parameter system. In the case of the ES, the state is formed by the system states at all N_t measurement times and by the system parameters, therefore $n_f = n(N_t + 1)$. In the case of the EnKS, the state is formed by the system state at the current assimilation time and by the system parameters, therefore $n_f = 2n$.

The forecast ensemble $\boldsymbol{X}_f$ is used to calculate the following $n_f \times n_f$ ensemble prior covariance matrix:

$$\boldsymbol{C}_f = \frac{(\boldsymbol{X}_{f,t} - \overline{\boldsymbol{X}}_f)(\boldsymbol{X}_{f,t} - \overline{\boldsymbol{X}}_f)^T}{N_{MC} - 1} \tag{4-16}$$

where $\overline{\boldsymbol{X}}_f$ is the prior ensemble mean obtained as $\overline{\boldsymbol{X}}_f = \boldsymbol{X}_f \overline{\boldsymbol{1}}_{N_{MC}}$ and $\overline{\boldsymbol{1}}_{N_{MC}}$ is a matrix of size $n_f \times N_{MC}$, where each element is equal to $1/N_{MC}$.

2.2.3 Update of system states and parameters

Following a Bayesian least-square estimate, the state measurements are used to "correct" the forecast matrix $\boldsymbol{X}_f$ and thus obtain updated ensembles of system states and parameters. The measurements are assembled into a $m \times 1$ vector $\boldsymbol{m}$. In the investigated problem, it is assumed that fluid pressure measurements are made at a number n_m of locations over N_t measurement times, therefore the size m is equal to $n_m N_t$ in the case of the ES, and n_m in the case of the

EnKS. Based on the KF, the updated state-parameter matrix $\boldsymbol{X}_u$ $[n_f \times N_{MC}]$ and the updated covariance matrix $\boldsymbol{C}_u$ $[n_f \times n_f]$ are calculated as follows:

$$\boldsymbol{X}_u = \boldsymbol{X}_f + \boldsymbol{K_t}(\boldsymbol{D} - \boldsymbol{H}\boldsymbol{X}_f) \tag{4-17}$$

$$\boldsymbol{C}_u = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{C_f}(\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})^T + \boldsymbol{K}\boldsymbol{R}\boldsymbol{K}^T \tag{4-18}$$

where: $\boldsymbol{D}$ is a $m \times N_{MC}$ matrix obtained by perturbing the measurement vector $\boldsymbol{m}$ with a Gaussian noise representing unbiased measurement errors**,** $\boldsymbol{H}$ is a $m \times n_f$ matrix that maps each measurement to its location in the finite difference grid and in time, so that the product $\boldsymbol{H}\boldsymbol{X}_f$ holds model results at measurement locations and times. $\boldsymbol{R}$ is the $m \times m$ measurement error covariance matrix that includes the variance of the measurement values:

$$\boldsymbol{R} = \frac{\boldsymbol{E}\boldsymbol{E}^T}{N_{MC} - 1} \tag{4-19}$$

where $\boldsymbol{E}$ $[m \times N_{MC}]$ is a matrix including the ensemble of measurement errors, with mean equal to zero and prescribed standard deviations. $\boldsymbol{K}$ is the $n_f \times m$ Kalman Gain matrix, and is computed as:

$$\boldsymbol{K_t} = \boldsymbol{C_f}\boldsymbol{H^T}(\boldsymbol{H}\boldsymbol{C_f}\boldsymbol{H^T} + \boldsymbol{R})^{-\mathbf{1}} \tag{4-20}$$

Note that, in the context of parameter estimation addressed here, we are interested in the update of the spatial distribution of permeability, rather than the fluid pressure distribution within the model domain. As a result, the updated ensemble of log-transformed permeability is extracted from the update state as $\boldsymbol{Y}_u = \boldsymbol{X}_u(n_f - n + 1{:}\, n_f, N_{MC})$. The mean of the log-transformed permeability posterior ensemble is computed as $\overline{\boldsymbol{Y}}_u = \boldsymbol{Y}_u\overline{\mathbf{1}}_{N_{MC}}$**,** where $\overline{\mathbf{1}}_{N_{MC}}$ is a $N_{MC} \times 1$ vector with all components equal to $1/N_{MC}$.

2.3 Assessment of ES and EnKS performance

To evaluate the performance of the ES and the EnKS, "synthetic" fluid-pressure pressure data are collected from a "true" reference state $\boldsymbol{\phi}_{ref}$. The reference state is obtained with the multiphase flow model (Section 2.1) using a corresponding "true" parameter set $\boldsymbol{Y}_{ref}$. Doing so allows for computing the deviation from the "true" parameter distribution on a cell-by-cell basis for both the forecasted and updated parameter ensembles, which quantifies the degree to which the forecast is corrected by data assimilation. This correction is quantified by the two associated global parameters AE (absolute error) and AEP (average ensemble precision) (Hendricks Franssen and Kinzelbach, 2008):

$$AE(\boldsymbol{Y}) = \frac{1}{N_{MC}n}\sum_{j=1}^{N_{MC}}\sum_{i=1}^{n}\left|Y_{i,j} - Y_{i,ref}\right| \tag{4-21}$$

$$AEP(\boldsymbol{Y}) = \frac{1}{N_{MC}n}\sum_{j=1}^{N_{MC}}\sum_{i=1}^{n}\left|Y_{i,j} - \bar{Y}_i\right| \tag{4-22}$$

where: $Y_{i,ref}$ is the true value at location i, $Y_{i,j}$ is the value at location i and realization j, and $\bar{Y}_i$ is the average value of all realizations at location i. The absolute error AE measures the deviation between the reference parameter and the updated parameter ensemble. On the other hand, AEP measures the deviation between the updated ensemble and its mean. Lower values of AE correspond to a model state approaching the reference state; whereas lower values of AEP signify reduced uncertainty in the model state.

For a visual assessment, the maps of the updated ensemble median (50th percentile), maps of the 10th and 90th percentile parameter distributions may be plotted and compared to the reference $\boldsymbol{Y}_{ref}$ map. Maps of the 10th percentile of Y should show the permeability value with a 90% probability of being exceeded. Therefore, if that value is high, of the order of Y=-13, then

there is a 90% probability that this spot is leaky. On the other hand, maps of the 90th percentile of Y should show the permeability value with a high probability of not being exceeded. So if the value is small, of the order of Y=-20, then there is a 90% probability that the caprock is sealing. Also, in some situations, a map of the local ensemble spread $AEP_i = \sum_{j=1}^{N_{MC}} \left|Y_{i,j} - \bar{Y}_i\right| / N_{MC}$ may be used to analyze the spatial variability of the updated parameter ensemble.

3 Numerical Experiments

This section presents the results of numerical tests performed to analyze the effectiveness of the propose methodologies. First, we introduce the model setup, and then we present two scenarios in which the ES and the EnKS are applied and, last, we show and discuss the obtained results.

3.1 Model setup

The methodology is applied to a hypothetical geological system consisting of a 100-m thick deep saline aquifer, denoted as "Aquifer 1", which constitutes the target CO_2 storage formation, overlain by another 100-m thick aquifer, denoted as "Aquifer 2". A low permeable 30-m thick layer, representing the caprock, separates the two aquifers. The bottom of Aquifer 1 lies at a depth of 1,600 m, whereas the top of Aquifer 2 lies at a depth of 1,370 m. The extension of the three-dimensional domain is 4,100 m × 4,100 m × 230 m. Figure 4-3 shows a plan view and a cross section of the model domain. Within the model grid, each formation (Aquifer 1, its caprock, and Aquifer 2) is divided vertically into 4 layers for a total of L=12 layers.

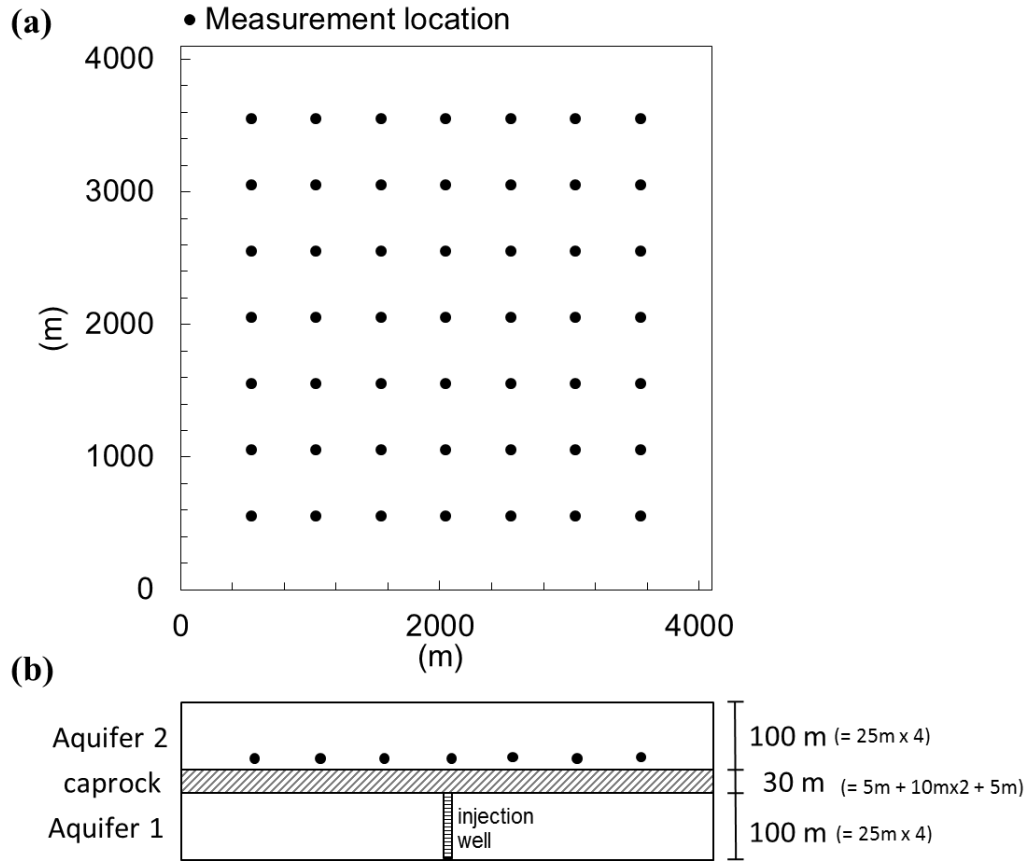

Figure 4-3. (a) Plan view of Aquifer 2, and (b) Vertical cross section of the domain. Measurement locations of Aquifer 2 are indicated as black dots. Fully penetrating injection well is located at the center of the domain in Aquifer 1. Note that vertical cross section is not in scale.

Horizontally, the model domain is uniformly discretized with 100 m × 100 m grid blocks. Overall, 20,172 cells (41×41×12) make up the model grid. A single CO_2 injection vertical well is considered at the center of the model domain and screened within Aquifer 1. The resolution of the finite-difference grid in the area surrounding the injection well is increased in order to achieve an appropriate size for a well (≈0.5 m). The CO_2 injection well is assumed to operate at a constant surface volume injection rate of 1,395,575 m^3/d for an indefinite period of time. For demonstrative purposes, we focus on assimilating fluid pressure data collected during the first two years of operation.

Initially, all formations are saturated only with brine and at hydrostatic pressure conditions. To simulate a laterally infinite aquifer system, the pore-volume of boundary cells is multiplied by a factor of 1×10^6. The brine phase is assumed as fresh water to exclude the impact of salts on results.

Van Genuchten's model (Van Genuchten, 1980) is used to calculate capillary pressure and relative permeability of CO_2 and brine, using data from Zhou et al. (2009). Figure 4-4 illustrates k_{rg}, k_{rw}, and p_c as functions of s_w obtained using Van Genuchten's model.

A porosity of 0.15 is considered for both aquifers and 0.10 for the caprock.

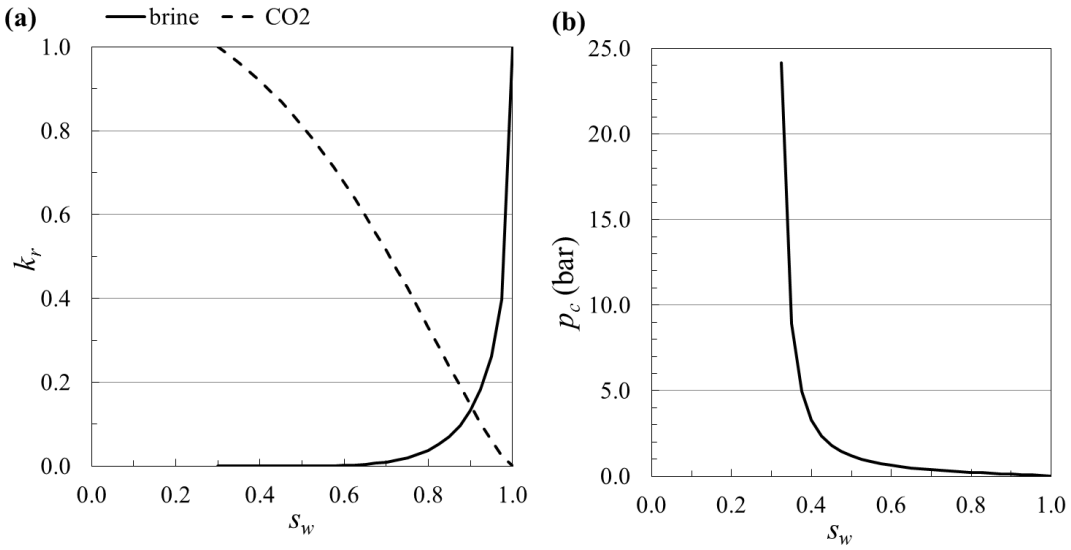

Figure 4-4. (a) Relative permeability curves of CO_2 (dashed line) and brine (solid line) and (b) Capillary pressure as functions of brine saturation.

Fluid-pressure data are assumed to be available at the n_m=49 nodes of a uniform 7×7 square grid, whose locations are indicated in Figure 4-3. These data are collected at the bottom layer of Aquifer 2 every three months, for a total of N_t=8 measurement times over two years. Therefore, in the ES a total of m=392 measurements (49×8) are assimilated into the simulation model response simultaneously; whereas in the EnKS only 49 measurements are assimilated at

each of the N_t=8 observation times. Fluid pressure measurements are characterized by an error that follows a Gaussian distribution with mean equal to zero and a standard deviation of 0.09 bar.

In this work, the uncertain parameters of interests consist of the spatial distribution of permeability in the caprock separating Aquifer 1 and Aquifer 2. The parameters $\left(\mu_{Y_1}, P_1, \mu_{Y_2}, P_2, \sigma', \lambda_{Y_1}, \lambda_{Y_2}\right)$ identifying the bimodal geostatistical model used to simulate the spatial distribution of Y in the caprock (Section 2.2.1) are given in Table 4-1. Note that this model is two-dimensional and no variability of the caprock permeability is hypothesized in the vertical direction z.

Table 4-1. Parameters defining the bimodal distribution to generate the log-transformed permeability at the caprock and weak areas.

Category	**Facies**	**Covariance Type**	$\boldsymbol{\mu_Y}$ **(log-m^2)**	$\boldsymbol{P}$ **(/)**	$\boldsymbol{\lambda_x}$ **(m)**	$\boldsymbol{\lambda_y}$ **(m)**	$\boldsymbol{\sigma'}$ **(log-m^2)**
1	Caprock	Exponential	-20	0.99	1,000	1,000	0.5
2	Weak areas	Exponential	-13	0.01	100	100	0.5

With respect the permeability k in the Aquifers 1 and 2, two scenarios are considered. In the first scenario (Scenario 1), both aquifers are assumed homogeneous and isotropic, with a single value of permeability known deterministically and equal to 1×10^{-13} m^2. In the second scenario (Scenario 2), the aquifers' permeability is uncertain and modeled accordingly with the three-dimensional geostatistical process presented in Section 2.2.1 (Equations (4-14)-(4-15)). The parameters $(\mu_Y, \sigma_Y^2, \lambda_1, \lambda_2, \lambda_3)$ of this process are given in Table 4-2. The aquifer permeability in the vertical direction z is assumed to be 1/10 of the aquifer permeability in the horizontal directions x and y.

Table 4-2. Parameters defining the log-transformed permeability of the storage formation Aquifer 1 and 2.

Scenario	**Covariance Type**	μ_Y **(log-m^2)**	σ_Y^2 **(log-m^2)^2**	$\lambda_x = \lambda_1$ **(m)**	$\lambda_y = \lambda_2$ **(m)**	$\lambda_z = \lambda_3$ **(m)**
1	-	-13	0	-	-	-
2	Exponential	-13	0. 25	1,000	1,000	500

The prior ensemble of k distributions in the caprock layer is generated using CIKSIM (Baù, 2012) for both Scenarios 1 and 2. SKSIM (Baù and Mayer, 2008) is applied to generate the forecast ensemble of permeability fields for Aquifers 1 and 2 in Scenario 2. The size N_{MC} of the ensemble is either 200 or 1,000 depending on the numerical test under consideration, and is discussed in the next section.

Extra realizations of the k spatial distribution in the caprock and in the two aquifers are generated to produce a "true" synthetic reference system from which collection of fluid-pressure data can be simulated, and with respect to which the performances of the ES and the EnKS can be assessed and compared (see Section 2.3). Figure 4-5 shows the spatial distribution of Y in the caprock for the "true" reference system in both Scenarios 1 and 2. This map reveals the presence of two clearly distinct potential leakage pathways, one in proximity of the CO_2 injection well and one in proximity of the lower right corner of the model domain. Figure 4-6 shows the Y fields for Aquifer 1 and Aquifer 2 for the "true" reference system in Scenarios 2.

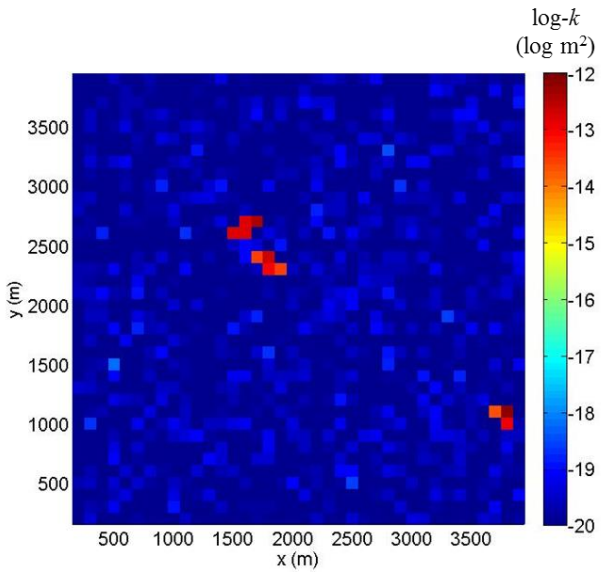

Figure 4-5. "True" reference spatial distribution of log-k in the caprock for the Scenarios 1 and 2.

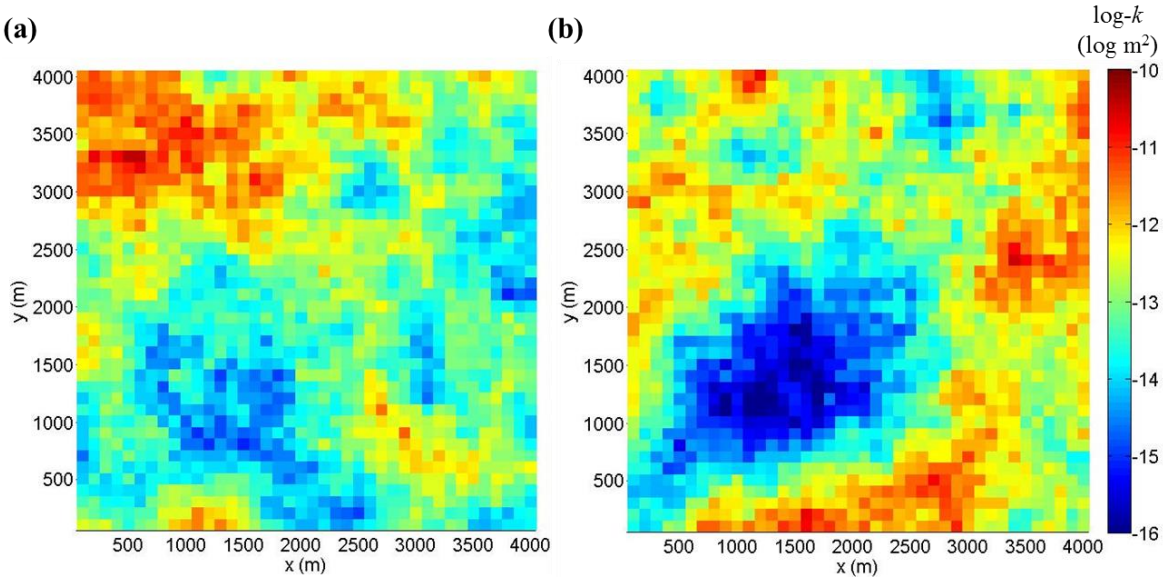

Figure 4-6. "True" reference spatial distributions of log-k in (a) Aquifer 1 and (b) Aquifer 2 for Scenario 2.

The temporal evolutions of the fluid pressure distribution within the model domain for the "true" reference systems in Scenarios 1 and 2 are obtained with two multiphase flow

ECLIPSE simulations making use of the corresponding "true" reference Y fields presented above.

Figure 4-7a and Figure 4-7b show the spatial distributions of fluid overpressure Δp (Equation (4-10)) at t=2 years in Aquifer 2, where measurements are taken, for the reference fields of Scenarios 1 and 2, respectively. The Δp pulse is more spread out and pronounced in Scenario 2 than in Scenario 1 as a consequence of the aquifers' heterogeneity. Values of Δp in excess of 3 bar can be found in Scenario 2, whereas in Scenario 1 the maximum Δp values are around 2.5 bar. In both scenarios larger fluid pressure changes are produced in correspondence the leakage area closer of the CO_2 injection well (Figure 4-5 and Figure 4-7) indicating that brine leakage is taking place. However, low-pressure changes are also noticed in proximity of the leakage pathway located on the lower right corner of the model domain (Figure 4-7).

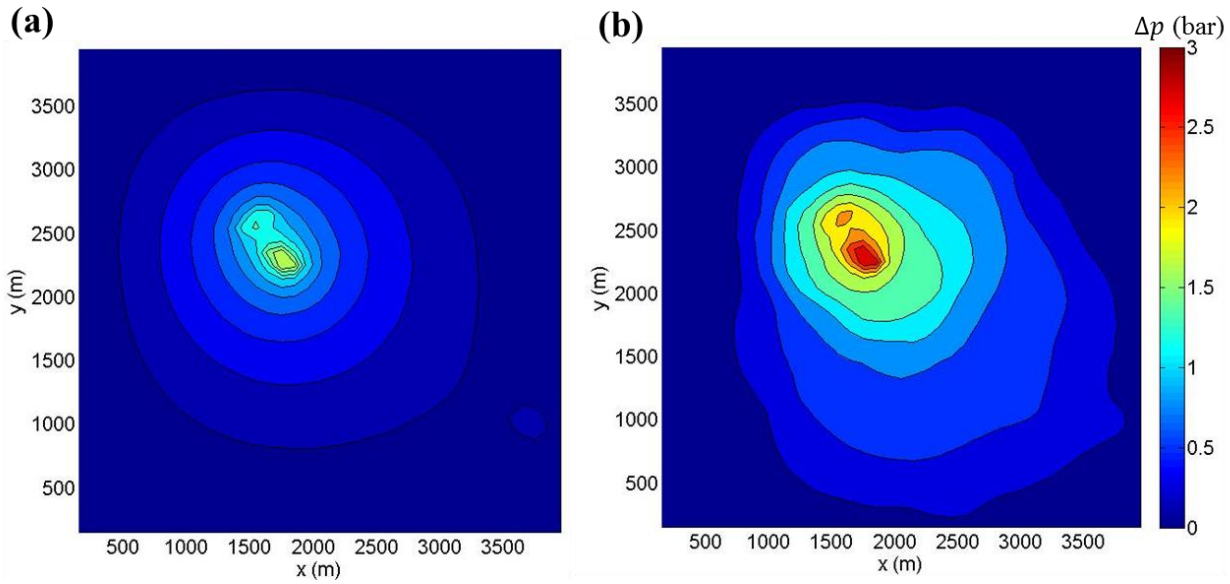

Figure 4-7. Fluid pressure change evolution in the reference field at t= 2 years for: a) Scenario 1 and b) Scenario 2.

3.2 Model forecast

In this stage, the forecast ensemble $\boldsymbol{\phi} = [\boldsymbol{p}_1, \dots, \boldsymbol{p}_{N_{MC}}]$ of fluid pressure distributions in the model domain is computed with a Monte Carlo multiphase flow ECLIPSE (Schlumberger, 2010) simulation using the generated ensemble $\boldsymbol{P} = [\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_{N_{MC}}]$ of realizations of the Y field. The results of this stochastic simulation are used to assemble the forecast state-parameter matrix $\boldsymbol{X}_f = [\boldsymbol{\phi}\ \boldsymbol{P}]^{\mathrm{T}}$.

As previously commented, ensemble-based KF algorithms (Equations (4-17)-(4-18)) provide an optimal solution to Equation (4-11) under the hypotheses of linear systems with state variables conforming to multi-Gaussian joint distributions, and measurement errors unbiased, uncorrelated and normal-distributed. Since in the considered case the multiphase flow model (Section 2.1) is non-linear, the ES and EnKS solutions can be sub-optimal. This sub-optimality can however be reduced if all requirements of the KF other than model linearity are met. It is thus suitable to require that: (i) the state and parameter ensembles in the matrix $\boldsymbol{X}_f$ be Gaussian distributed; (ii) all fluid-pressure measurement errors be unbiased and normal distributed (see Section 2.2.3). To meet these conditions, adequate transformations may be applied to the parameter and the state variables that are not Gaussian.

Since the caprock permeability is simulated as a bimodal random process, a NST algorithm is employed to transform permeability values to fit to a Gaussian PDF. A modified NST based on the works of Zhou et al. (2011) and Crestani et al. (2013) is here proposed to transform caprock Y values. The "local" NST used by Zhou et al. (2011) builds a CDF for each model cell and each observation time. While Crestani et al. (2013) show that a "global" NST, with a CDF composed by model results at all cells and observation times, may provide better

results, since it avoids the corruption of the parameter-state cross-correlation structure occurring when a different NST is applied for each cell.

In this study, we use a global NST for the current observation time when the EnKS is applied and a global NST for all the N_t when ES is applied. Both global NSTs are based on a sample CDF defined by all simulated $Y_{i,j}$ values of each cell (i=1,2,…,n) and all realizations of the ensemble (j=1,2,…,N_{MC}). The size of the sample used to build this CDF is nN_{MC}. After ordering the $Y_{i,j}$ values in ascending order, $Y_1 < Y_2 < \cdots < Y_{nN_{MC}}$, the corresponding CDF values are calculated as $CDF(Y_i) = (i - 0.5)/(nN_{MC})$ (Hahn, 1967) , where i=1,2,…,nN_{MC}.

For both the ES and the EnKS, after the update of state and parameters, the inverse function of the adopted NST has to be applied to the parameter ensemble in order to retrieve the spatial distribution of k in the caprock.

In the considered problem, the state variables of interest, that is, the Δp (Equation (4-10)) distribution within the model domain, require no transformation since a fit analysis indicates that this variable is normal-distributed. This fit analysis is based on a linear regression of the dataset $\left[f_\Phi(\Delta p); \Phi^{-1}\left(\frac{i-0.5}{N_{MC}}\right)\right]$ ($i = 1,2,\dots,N_{MC}$), where Δp represents fluid-overpressure values sampled from Aquifer 2 by stochastic simulation, Φ^{-1} is the inverse of the theoretical CDF, numerically computed, and f_Φ is a function of Δp that depends on the type of CDF under consideration. The fit analysis is conducted by testing different types of CDF. The "goodness" of fit for the sampled distribution is assessed based on the accuracy with which the regression line approximates the sample, as quantified by the coefficient of determination R^2 and the Kolmogorov-Smirnov statistics KS (Massey, 1951).Values of R^2 close to 1 and low values of KS are indicators of a good fit. Figure 4-8 shows the results of a fit test for Δp values sampled at a cell of Aquifer 2 at observation time t=0.25 years when Φ is chosen as a Gaussian CDF. For this

example, values of R^2= 0.975 and KS= 0.063 are obtained, indicating Gaussian behavior. One must be aware, however, that, whenever possible, transformations of parameters and/or states can at most meet conditions of "marginal" Gaussianity for each of the state variables, but do not guarantee that states and parameters fit into a multi-Gaussian joint distribution as the KF hypotheses would require. Therefore applying transformations to states and parameters does not necessarily improve the optimality of the ES and the EnKS solutions.

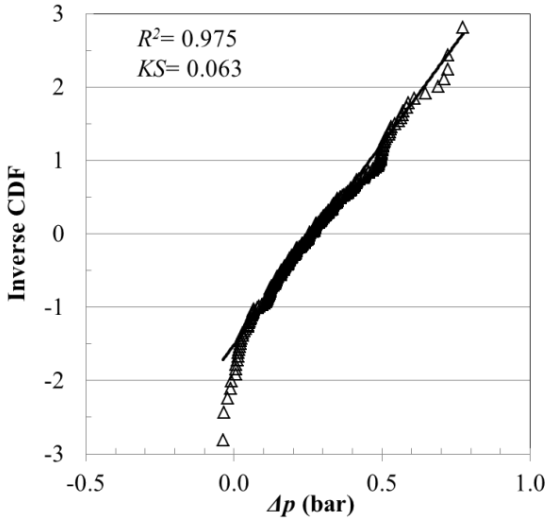

Figure 4-8. Inverse CDF of the fluid pressure change in one cell of Aquifer 2 at observation time t=0.25 years, showing its Gaussian distributed behavior.

Table 4-3 summarizes the numerical tests that are considered in this work. The ES and the EnKS are applied both with and without NST of the forecast parameter ensemble $\boldsymbol{P}$. For both Scenarios 1 and 2, four sub-scenarios, termed "a", "b", "c", and "d", are investigated. Preliminary tests are run to evaluate an optimal ensemble size. Based on these tests, a sample size N_{MC}=1,000 is adopted for the ES, whereas a sample size N_{MC}=200 is used for the EnKS. These sizes of N_{MC} are big enough to have a good representation of the covariance and to process the number of measurements that each algorithm uses.

Table 4-3. Scenario definition.

Scenario	Aquifers 1 and 2	Algorithm	Scenario	Aquifers 1 and 2	Algorithm	N_{MC}
1-a	Homogeneous	ES	2-a	Heterogeneous	ES	1,000
1-b	Homogeneous	ES-NST	2-b	Heterogeneous	ES-NST	
1-c	Homogeneous	EnKS	2-c	Heterogeneous	EnKS	200
1-d	Homogeneous	EnKS-NST	2-d	Heterogeneous	EnKS-NST	

3.3 Model update: Scenario 1

In Scenario 1, both Aquifers 1 and 2 are considered homogeneous and isotropic. In scenarios 1a and 1b, the ES is applied without and with NST (Section 3.2) of the caprock Y values, respectively. In scenarios 1c and 1d, the EnKS is applied without and with NST of the caprock Y values.

Figure 4-9 and Figure 4-10 summarize results of these four scenarios. In Figure 4-9 for each scenario a-d, the subpanel on the left column shows the maps of the updated median value of the caprock Y, whereas the subpanel on the right column shows the map of the ensemble spread (Section 2.3) at time t=2 years.

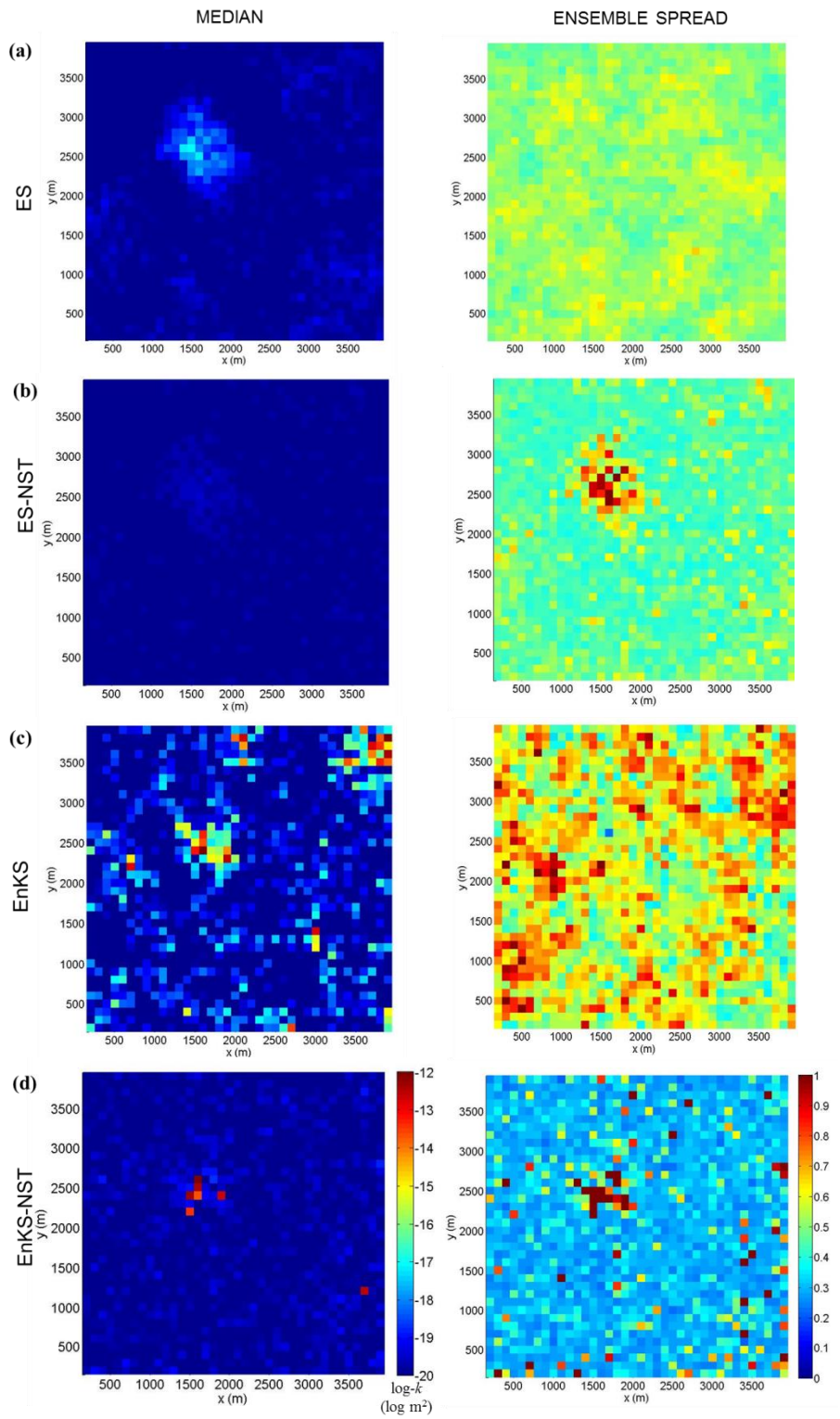

Figure 4-9. Scenario 1. Spatial distribution of updated caprock log-k at t=2 years (left) and ensemble spread maps of caprock log-k at t=2 years (right) for the Scenarios 1a, 1b, 1c, and 1d defined in Table 4-3.

Figure 4-9a illustrates results obtained with the ES with no parameter transformation. The ES is able to locate the caprock discontinuity in the proximity of the CO_2 injection well but cannot quantify its permeability value with good accuracy. Figure 4-9b displays results obtained by the ES with NST of the caprock Y. In this case, no high permeability zone is detected in the caprock.

Figure 4-9c indicates that the EnKS is able to locate the caprock discontinuity in proximity of the CO_2 injection well. However, it also points out to the existence of other caprock discontinuities at the top and top-right corner of the domain that do not truly exist in the reference field. Figure 4-9d presents the results of the combined EnKS-NST update. In this case, the EnKS with parameter NST locates the leakage area and quantifies its permeability with fairly good accuracy. In addition, the combined EnKS-NST is able to identify the leakage area located in proximity of the lower-right corner of the domain.

Figure 4-9 also shows maps of the ensemble spread of the caprock Y at t=2 years for scenarios 1a, 1b, 1c and 1d. The ES (Figure 4-9a) and the ES with NST (Figure 4-9b) produce ensemble spread values between 0.5 and 0.6, and between 0.4 and 1, respectively. Globally, the ES seems to perform better than the combined ES-NST, since the caprock discontinuity closer to the injection well is detected quite effectively. Interestingly, however, while the map of the median caprock Y obtained using the ES with NST shows no potential leakage areas, high values of the ensemble spread are noticed in proximity of the weak caprock area nearby the injection well, which reveals that the ES-NST results are quite uncertain around it.

Figure 4-9c shows a pronounced spread distribution with values between 0.5 and 1, which further highlights the low effectiveness of the EnKS in narrowing the uncertainty on the presence of weak caprock areas. Although the EnKS updates the parameter ensemble of caprock

Y sequentially, over N_t= 8 assimilation times, it presents the highest ensemble spread values. This is expected, since the EnKS would require the parameter statistical distribution to be Gaussian, when the actual one is truly bimodal. On the other hand, the combined EnKS-NST in Scenario 1d (Figure 4-9d) exhibits a spatial distribution of the ensemble spread that is quite contained, with values that are typically between 0.2 and 0.3, but can be as high as 1 over a few areas of the domain. These results demonstrate the beneficial effect of the NST when combined with the EnKS.

Figure 4-10 shows results of the 10th percentile of the caprock Y on the left column at time t=2 years, whereas the results of the 90th percentile of the caprock Y are shown on the right column, also at time t=2 years. Each row of this figure corresponds to one sub-scenario of Scenario 1 ("a", "b", "c", and "d" respectively). The 10th percentile results of ES (Figure 4-10a) show an area with a greater value of Y equal to -18, that has a high probability of being exceed. However, 90th percentile results show that this area has a maximum probable value of Y=-16.

ES-NST results, Figure 4-10b, do not display the detection of any leakage at the 10th percentile, but the 90th percentile results show that there are two points that cannot be defined as sealing formation with a probability of the 90%.

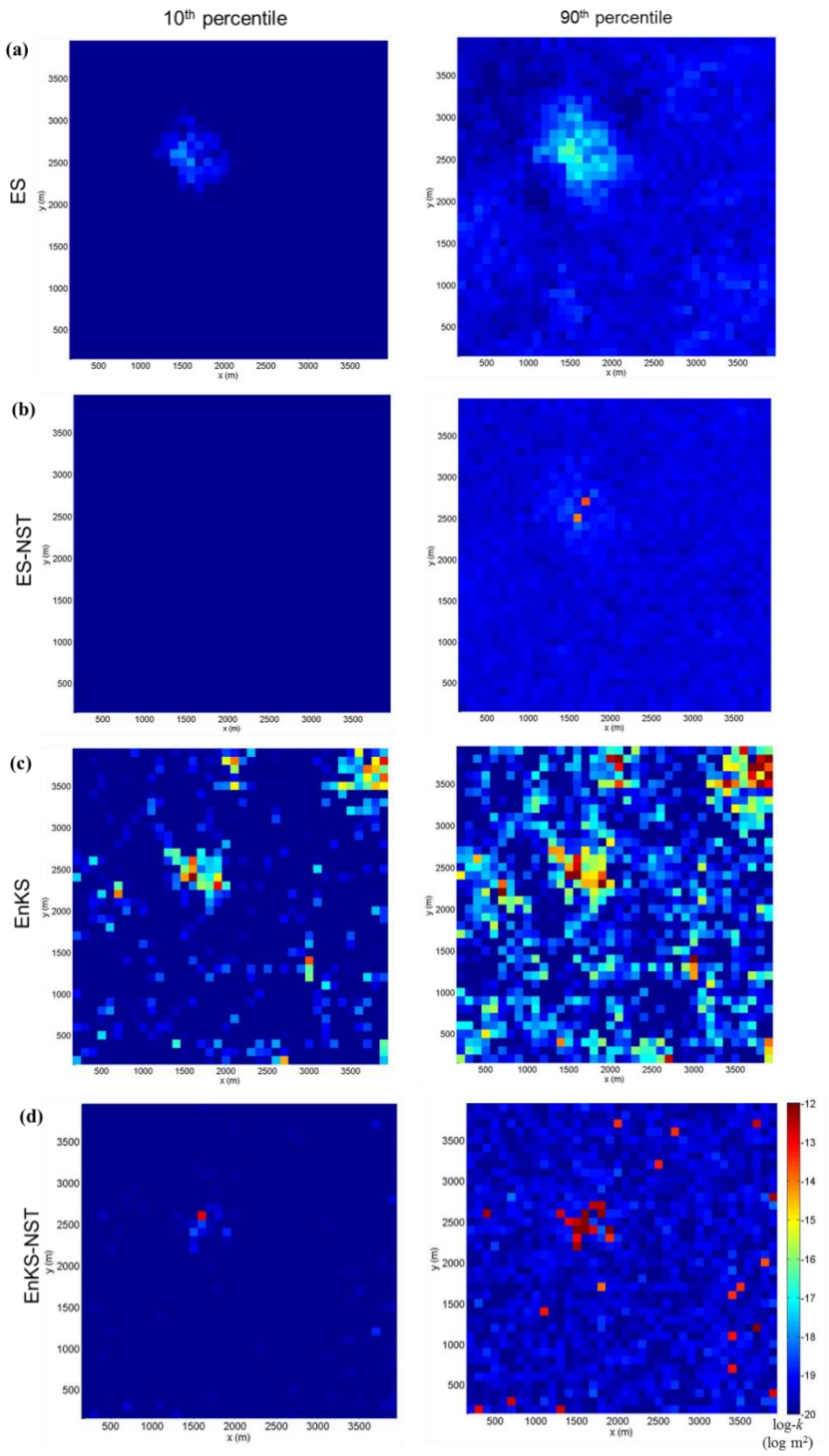

Figure 4-10. Scenario 1. Spatial distribution of the 10th percentile maps of caprock log-k at t=2 years (left) and 90th percentile maps of caprock log-k at t=2 years (right) for the Scenarios 1a, 1b, 1c, and 1d defined in Table 4-3.

Figure 4-10c points out that the EnKS locates the caprock discontinuity closer to the CO_2 injection well with high values of Y at the 10^{th} percentile. However, it also indicates the presence of other discontinuities at the top right corner and spread over the domain that do not exist at the reference field. On the other hand, in Figure 4-10d the EnKS-NST, at the 10^{th} percentile, detects there is a spot with a 90% of probability of being leaky. This spot corresponds to the caprock discontinuity in the proximity of the CO_2 injection well. The EnKS-NST results of the 90^{th} percentile report that there are spots that the code does not reproduce as caprock with a 90% of probability. As Figure 4-10d shows, these spots are spread over the entire domain.

3.4 Model update: Scenario 2

In Scenario 2 the capabilities of the ES and the EnKS to detect leakage areas of the caprock are investigated by assuming Aquifer 1 and Aquifer 2 as heterogeneous. Results from Scenarios 2b and 2c are not shown here since ES-NST and EnKS fail in determining log-k distributions and detect any of the leakage areas in Scenario 1.

Figure 4-11 reports results of the ES on the left column and results of the EnKS on the right one. Each row of Figure 4-11 displays a different map of Scenarios 2a and 2d evaluated at time t=2 years: map of the median value of the updated ensemble of the caprock Y, map of the ensemble spread, map of the 10^{th} percentile, and last, the map of the 90^{th} percentile.

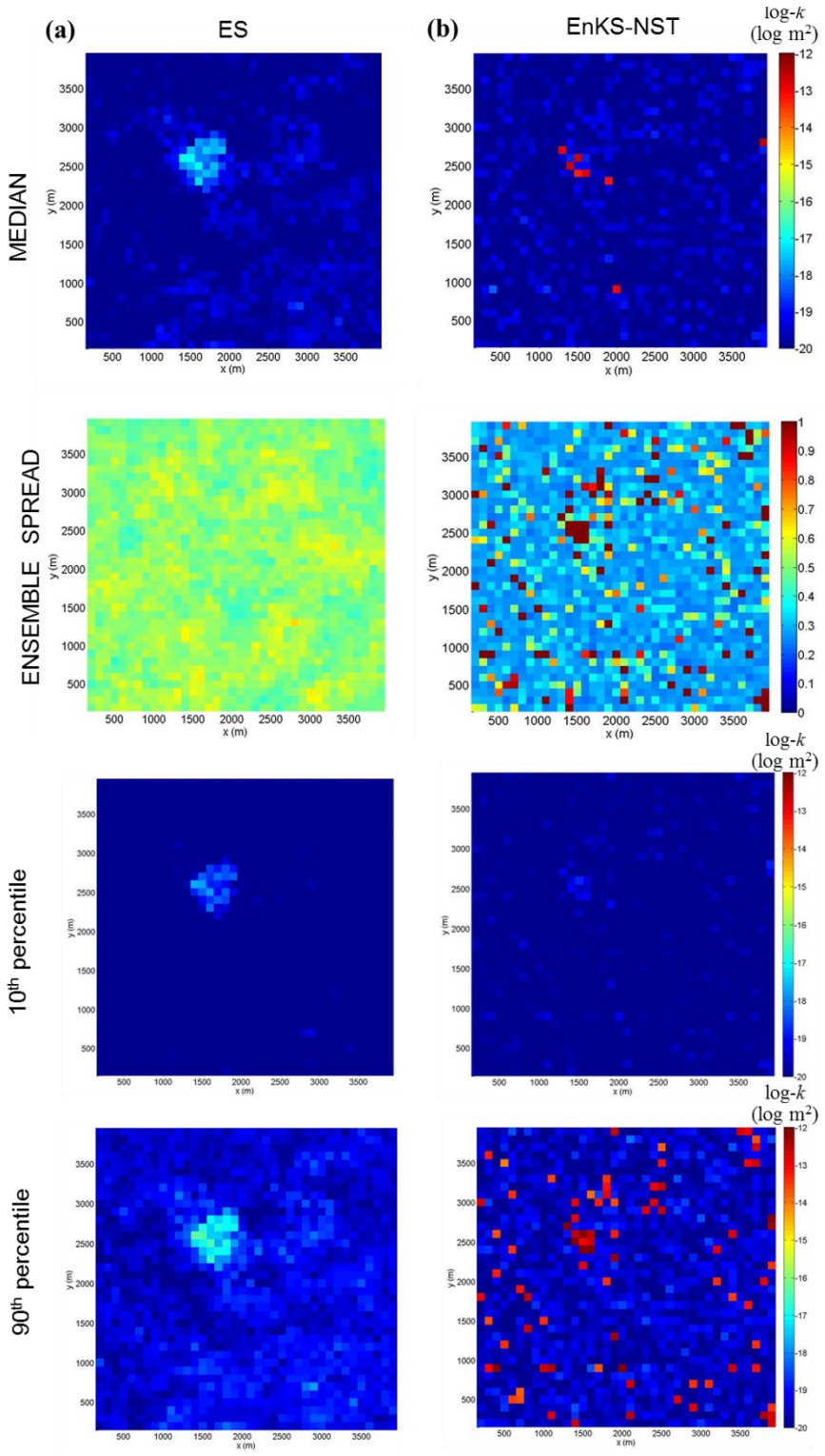

Figure 4-11. Scenario 2. Spatial distribution of updated caprock log-k maps (first row), ensemble spread maps (second row), 10th percentile maps (third row), and 90th percentile maps (fourth row) of caprock log-k at t=2 years for (a) Scenario 2a (left), and (b) Scenario 2d (right) defined in Table 4-3.

Similarly to what observed in Scenario 1a, Figure 4-11a shows that ES is able to locate the caprock discontinuity nearby the injection well, but it does not to quantify its Y value with good accuracy. Figure 4-11b shows that also the combined EnKS-NST can detect the same caprock discontinuity, although it identifies two other regions of high permeability that do not exist in the reference system of Figure 4-5. Further numerical tests reveal that these inaccuracies are due to the particular model setting adopted and the fluid pressure data assimilated here. At time t=2 years, the algorithm is not able to distinguish for between fluid pressure changes produced by heterogeneities of Aquifer 2 or by caprock leakages. So the algorithm assigns a high permeability at that spot of the caprock.

If the time frame t is increased, the model is able to remove one of the inexistent leakages. Figure 4-12 shows the spatial distribution of the median of the caprock Y ensemble updated with the combined EnKS-NST at time t=2.75 years in Scenario 2d. This map shows that one of the non-realistic caprock discontinuities has now disappeared. This outcome is evidence that continued data collection is instrumental towards improving the identification of leakage areas and removing update inconsistencies.

Second row of Figure 4-11, with results from updates of the ensemble spread maps of caprock Y updated at t=2 years for Scenario 2, displays the same tendency observed in Scenario 1. Scenario 2a produces the lowest confidence results with high ensemble spread values (around 0.5 and 0.6). Whereas Scenario 2d, in which EnKS-NST is applied, obtains the lowest values of the ensemble spread (around 0.2). The comparison of these two maps suggests that the combined EnKS-NST is globally more effective than the ES in identifying potential caprock leakage pathways.

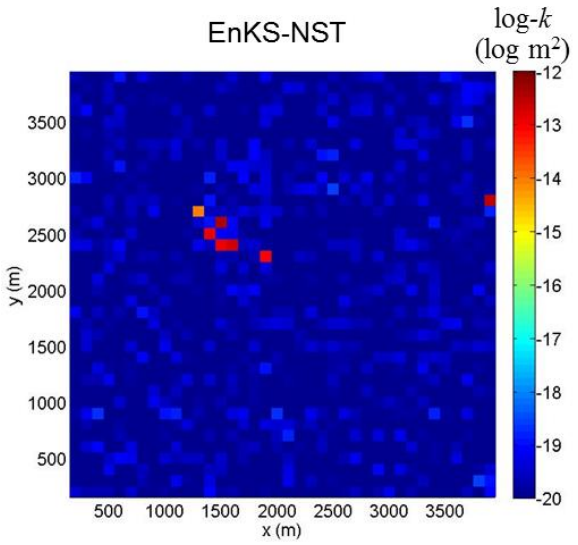

Figure 4-12. Scenario 2d. Spatial distribution of updated caprock log-k at t=2.75 years applying EnKS-NST.

Third and fourth row of Figure 4-11 display results of the 10th and 90th percentile of caprock Y maps respectively for (a) Scenario 2a and (b) Scenario 2d. The 10th percentile Y maps of both scenarios 2a and 2d are not able to detect any leakage with a 90% of probability. The 90th percentile caprock Y map of Scenario 2a show 90% of probability of sealing formation for the entire domain, except on the leakage area closer to the injection well. The 90th percentile caprock Y map of Scenario 2d presents a significant number of spots with a lower probability than 90% of being a sealing spot.

3.5 Discussion

In Scenarios 1a (Figure 4-9a and Figure 4-10a) and 2a (Figure 4-11a), the ES is applied under the assumption that the distribution of caprock Y is Gaussian and no modification is applied to the corresponding PDF. As expected, results obtained are not satisfactory and high values of Y in the caprock are not matched since the hypothesized PDF is bimodal rather than normal. In Scenarios 1b (Figure 4-9b and Figure 4-10b) and 2b, Gaussianity is introduced for caprock Y values by applying a NST before application of the ES. No improvement is observed

in these instances, and the performance of the combined ES-NST in detecting potential areas of CO_2 leakage is lower than the ES. The lower-quality results achieved with the ES and the ES-NST can be attributed to the fact that both algorithms update the parameter ensemble in a single comprehensive data assimilation step, thus not letting the parameter ensemble evolve gradually at each of the assimilation time.

The features of sequential data assimilation are better exploited when using the EnKS. The test conducted in Scenarios 1c, 1d, 2c, and 2d indicate that EnKS does not perform as effectively as the EnKS-NST. In scenario 1c (Figure 4-9c and Figure 4-10c), the EnKS can detect the leakage area nearby the injection well, yet it cannot quantify its permeability value with good accuracy. In addition, fictitious leakage zones are identified. In scenario 2c, the fictitious leakage spots that are generated increase in number at each update, as well as their caprock Y value. This results in a deterioration of the update with time and creating numerical errors in the multiphase flow code.

Figure 4-13 reports updates of the median caprock Y of Scenario 2c at two different times. For t=0.25 years (Figure 4-13a), EnKS detects the closest leakage to the injection well and one additional fictitious leakage area as the right top corner. At a later assimilation time t=1 year (Figure 4-13b) the algorithm detects more leakages than those present in the "true" reference field. This figure shows that the updates of caprock Y are deteriorated when increasing time of assimilation or more information is added.

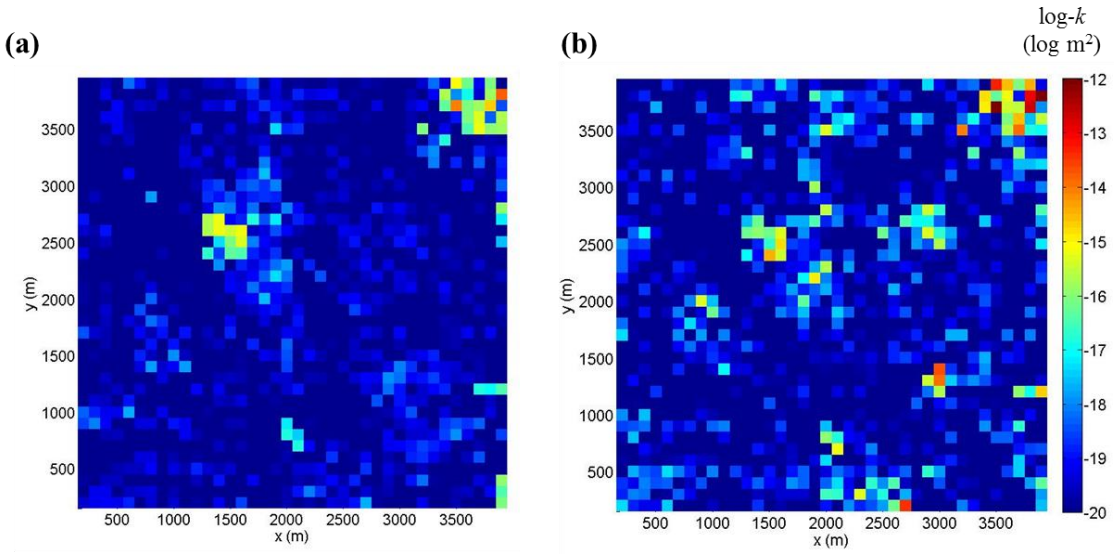

Figure 4-13. Scenario 2c. Spatial distribution of updated caprock log-k at (a) t=0.25 years and (b) t=1 year applying EnKS.

In scenarios 1d and 2d, the main leakage area close to the injection well is detected by the combined EnKS-NST early on at t=0.5 years (caprock Y maps are not shown here). In Scenario 1d, the combined EnKS-NST is able to detect also the caprock discontinuity nearby the lower-right corner of the domain at time t=2 years since an observable Δp reaches this region at this time. In Scenario 2d, however, when Aquifers 1 and 2 are considered heterogeneous, the EnKS-NST cannot detect this leakage spot, whereas it seems to identify fictitious caprock discontinuities in other regions of the domain. However, as reported in Figure 4-12, increasing the number N_t of assimilation times seems to improve the precision of the leakage identification and effectively removes possibly spurious caprock discontinuities. A similar result was observed in Nævdal et al. (2002; 2005).

Figure 4-14 and Figure 4-15 summarize the performance assessment for the investigated data assimilation algorithms in Scenario 1 and Scenario 2 respectively. Both figures show AE (Equation (4-21)) and AEP (Equation (4-22)) of the caprock Y as a function of time. Note that $AE(\boldsymbol{Y})$ and $AEP(\boldsymbol{Y})$ profiles of ES and ES-NST are formed by only two points: initial time t=0

(values of the forecast ensemble) and time t=2 years. In contrast, $AE(\boldsymbol{Y})$ and $AEP(\boldsymbol{Y})$ profiles of EnKS and EnKS-NST are constituted by 9 points: initial time t=0 and N_t=8 assimilation times.

In Figure 4-14a $AE(\boldsymbol{Y})$ of ES increases slightly with time. On the other hand, the AE of the combined ES-NST decreases with time. AE of EnKS increases significantly at each assimilation time indicating a global loss of accuracy in the characterization of the caprock Y. Similar to the ES-NST, the combined EnKS-NST shows an improvement of AE over time. The EnKS-NST is able to direct the solution to the reference one through the updating process.

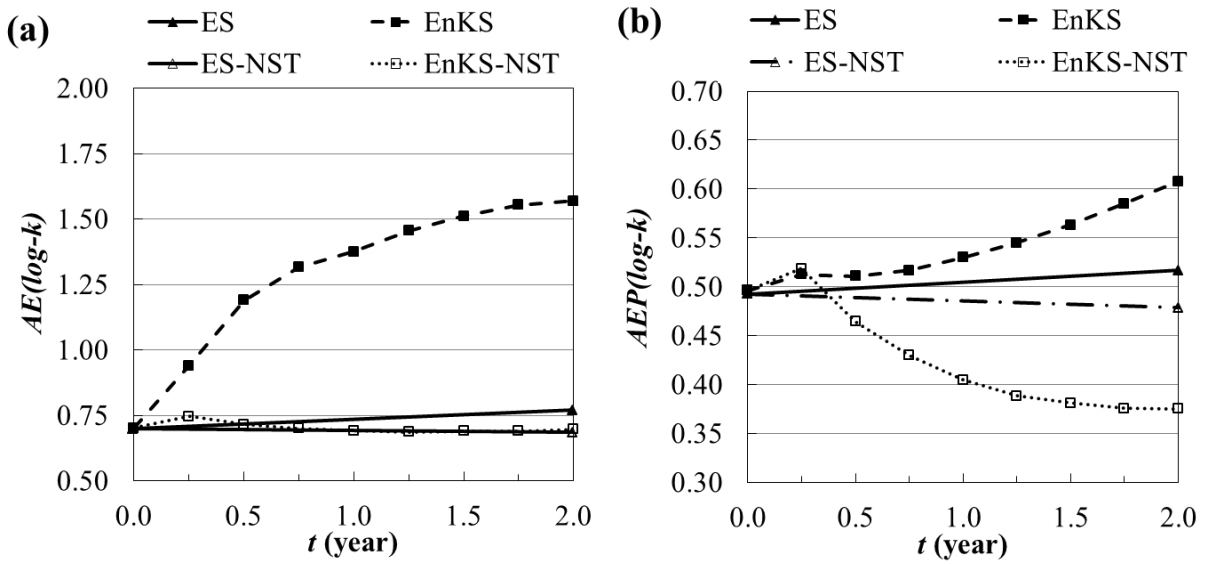

Figure 4-14. Performance assessment of Scenario 1: (a) AE and (b) AEP of caprock log-*k*. See Table 4-3 for description of Scenarios 1.

In Figure 4-14b $AEP(\boldsymbol{Y})$ profiles of ES and EnKS increase with time. However, AEP increase of the EnKS is greater, meaning a deterioration of the updated ensemble spread and a loss of confidence on the updated ensemble caprock Y values. AEP of ES-NST has an slight improvement of the ensemble spread. In contrast, EnKS-NST shows a significant reduction of the AEP profile, reporting a decrease on the uncertainty of the updated caprock Y values.

In Figure 4-15a none of the $AE(\boldsymbol{Y})$ profiles decreases. However, profile of the EnKS-NST first increases its value at the first time of assimilation and decreases with the following

updates. We point out how $AE(\boldsymbol{Y})$ of the EnKS increases excessively, indicating that updated log-*k* values are further away from the true reference state at each assimilation time.

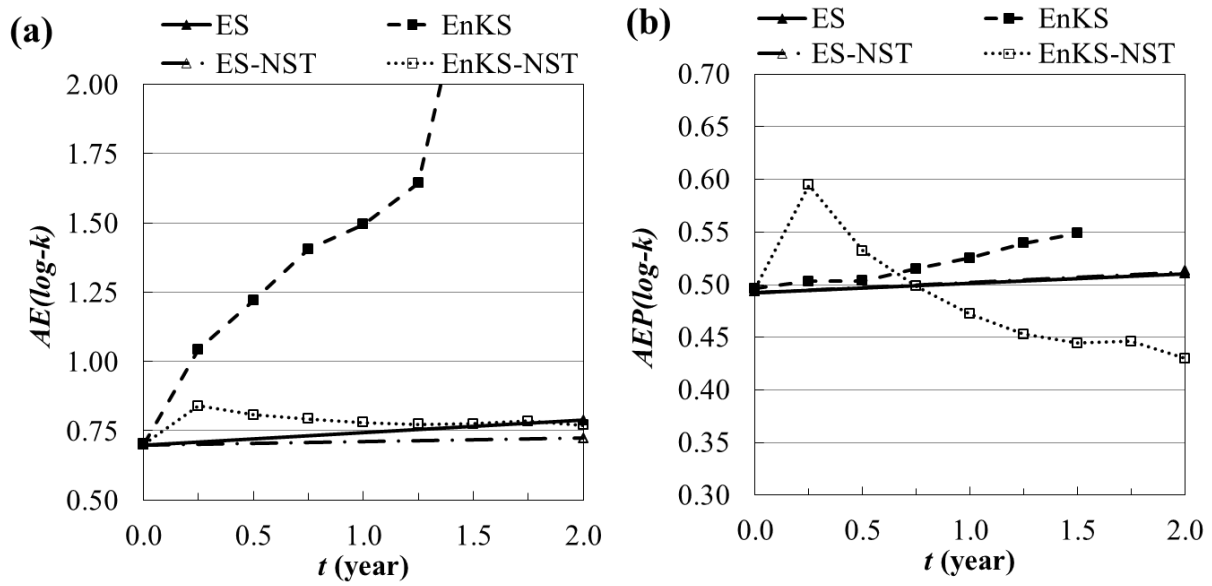

Figure 4-15. Performance assessment of Scenario 2: (a) AE and (b) AEP of caprock log-k. See Table 4-3 for description of Scenarios 2.

The AEP plot of Scenario 2 in Figure 4-15b follows the same tendency observed in Figure 4-14b. AEP profiles of ES, ES-NST, and EnKS increase. Only AEP of EnKS-NST indicates a decrease of the ensemble spread.

Scenario 2 presents the challenge that Aquifer 1 and Aquifer 2 are heterogeneous and each realization of the ensemble has different permeability fields for both aquifers. Jafarpour and McLaughlin (2009) remarked the importance of the geology used for the replicates of the ensemble, that in order to obtain a good characterization realistic reproductions of the geology are needed. In Scenario 2, the fact that both aquifers are heterogeneous and unknown hinders the characterization of the caprock through the data assimilation updating process, as it is observed in Section 0.

In summary, ES is able to locate the leakage, but not to quantify a good approximation of the log-*k* value at this area. Although ES performs better than ES-NST and EnKS, EnKS-NST presents the best performance. EnKS-NST is able to locate leaky areas and to quantify their log-*k* values. Both scenarios 1d and 2d provide better fits to the reference system. These results are in accordance to the conclusion reported by Evensen and van Leewen (2000), who found that EnKS generally provides a better performance than ES for strongly nonlinear models.

4 Summary and Conclusions

The goal of this paper is to identify and locate possible leaky areas of brine and CO_2 that could affect GCS operations. We use a hypothetical three-dimensional geological setup consisting of two deep saline aquifers separated by a sealing formation. Supercritical CO_2 is injected into the deepest aquifer. The caprock includes areas of low permeability (sealing formation) and high permeability (leaky pathways) generated with a sequential indicator Kriging simulator, following a bimodal distribution. Taking advantage of the fact that fluid pressure travels and propagates faster than CO_2 plume, measurements of fluid pressure at the overlying aquifer are used to identify and locate these discontinuities.

A subsurface multiphase flow model and ensemble-based data assimilation algorithms are combined to characterize the caprock discontinuities. We compare the capabilities of the ES and EnKS algorithms to identify the leaky areas. The assimilated data is fluid pressure change measured at several locations of the aquifer above the caprock and taken for a period of time of two years and every three months. A NST technique is applied to caprock log-*k* data to transform them into a Gaussian distribution. The ES and EnKS methods are applied to two different scenarios: 1) scenario in which permeability of both aquifers are considered known and

homogeneous, and 2) scenario in which aquifers' permeability are considered uncertain and heterogeneous.

Results show that ES-NST performs poorly since results rely on only one update of the system, so the fluid pressure pulse advances freely without modifications from the inversion procedure. On the other hand, EnKS-NST may constitute an effective tool for inverting pressure data in order to detect potential leakage pathway in the geological storage of supercritical CO_2, since it seems to provide a better performance. In the case of scenario with heterogeneous aquifers, more assimilation times seem to be necessary to improve the detection of caprock discontinuities and distinguish them from spurious leakages caused by fluid pressure changes from the heterogeneities of the aquifers.

Successful identification of caprock discontinuities through fluid pressure data assimilation is demonstrated, which improves the prediction of carbon storage models. This can be used to make decisions regarding injection rates and injection/production schemes and strategies to improve CO_2 storage and reduce brine/CO_2 leakages.

5 References

Apps, J., Zheng, L., Zhang, Y., Xu, T., Birkholzer, J., 2010. Evaluation of potential changes in groundwater quality in response to CO_2 leakage from deep geologic storage. Transport in Porous Media 82, 215-246.

Bailey, R.T., Bau, D., 2012. Estimating geostatistical parameters and spatially-variable hydraulic conductivity within a catchment system using an ensemble smoother. Hydrol. Earth Syst. Sci. 16, 287-304.

Baù, D., 2012. Multi-Objective Optimization Approaches for the Design of Carbon Geological Sequestration Systems. DE-FE0001830 Progress Report III, DOE, National Energy Technology Laboratory.

Baù, D.A., Mayer, A.S., 2008. Optimal design of pump-and-treat systems under uncertain hydraulic conductivity and plume distribution. J Contam Hydrol 100, 30-46.

Béal, D., Brasseur, P., Brankart, J.-M., Ourmières, Y., Verron, J., 2010. Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis. Ocean Science 6, 247-262.

Birkholzer, J., Zhou, Q., Tsang, C., 2009. Large-scale impact of CO_2 storage in deep saline aquifers: A sensitivity study on pressure response in stratified systems. International Journal of Greenhouse Gas Control 3, 181-194.

Bishop, A., 1959. The principle of effective stress. Teknisk Ukeblad 106, 859-863.

Brooks, R., Corey, A., 1964. 1964, Hydraulic properties of porous media, Hydrology Paper No. 3. Colorado State University, Fort Collins, CO.

Brusdal, K., Brankart, J., Halberstadt, G., Evensen, G., Brasseur, P., Van Leeuwen, P., Dombrowsky, E., Verron, J., 2003. A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. Journal of Marine Systems 40, 253-289.

Chabora, E.R., Benson, S.M., 2009. Brine Displacement and Leakage Detection Using Pressure Measurements in Aquifers Overlying CO_2 Storage Reservoirs. Energy Procedia 1, 2405-2412.

Chen, Y., Zhang, D., 2006. Data assimilation for transient flow in geologic formations via ensemble Kalman filter. Adv Water Resour 29, 1107-1122.

Cole, S., Lumley, D., Meadows, M., Tura, A., 2002. Pressure and saturation inversion of 4D seismic data by rock physics forward modeling, 2002 SEG Annual Meeting.

Cole, S.P., Lumley, D.E., Meadows, M.A., 2006. Method of obtaining pore pressure and fluid saturation changes in subterranean reservoirs by forward modeling. Google Patents.

Crestani, E., Camporese, M., Baú, D., Salandin, P., 2013. Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation. Hydrol. Earth Syst. Sci. 17, 1517-1531.

DOE/NETL, 2009. Monitoring, Verification and Accounting of CO_2 Stored in Deep Geologic Formations. DOE/NETL-311/081508.

Dunne, S., Entekhabi, D., 2006. Land surface state and flux estimation using the ensemble Kalman smoother during the Southern Great Plains 1997 field experiment. Water Resour Res 42.

Durand, M., Molotch, N.P., Margulis, S.A., 2008. A Bayesian approach to snow water equivalent reconstruction. Journal of Geophysical Research: Atmospheres (1984–2012) 113.

Evensen, G., 1994. Inverse methods and data assimilation in nonlinear ocean models. Physica D. Nonlinear phenomena 77, 108-129.

Evensen, G., 2009. Data assimilation: the ensemble Kalman filter. Springer.

Evensen, G., Van Leeuwen, P.J., 2000. An ensemble Kalman smoother for nonlinear dynamics. Mon Weather Rev 128, 1852-1867.

Fenghour, A., Wakeham, W.A., Vesovic, V., 1998. The viscosity of carbon dioxide. Journal of Physical and Chemical Reference Data 27, 31-44.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford university press.

Gray, W., Miller, C., 2007. Consistent thermodynamic formulations for multiscale hydrologic systems: Fluid pressures. Water Resour Res 43.

Gu, Y., Oliver, D.S., 2006. The ensemble Kalman filter for continuous updating of reservoir simulation models. Transactions-American society of mechanical engineers journal of energy resources technology 128, 79.

Hahn, G.J., 1967. Statistical models in engineering. Wiley, New York.

Hendricks Franssen, H., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resour Res 44.

Hendricks Franssen, H., Kinzelbach, W., 2009. Ensemble Kalman filtering versus sequential self-calibration for inverse modelling of dynamic groundwater flow systems. Journal of Hydrology 365, 261-274.

Herrera, G.S., Simuta-Champo, R., 2012. Optimal Design of Groundwater-Quality Sampling-Networks with 3D Selection of Sampling Locations Using an Ensemble Smoother. Journal of water resources planning and management 1, 169.

Jafarpour, B., McLaughlin, D., 2009. Estimating channelized-reservoir permeabilities with the ensemble Kalman filter: the importance of ensemble design. SPE Journal 14, 374-388.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82, 35-45.

Keppenne, C.L., Rienecker, M.M., 2003. Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter. Journal of Marine Systems 40, 363-380.

Kharaka, Y., Cole, D., Hovorka, S., Gunter, W., Knauss, K., Freifeld, B., 2006. Gas-water-rock interactions in Frio Formation following CO2 injection: Implications for the storage of greenhouse gases in sedimentary basins. Geology 34, 577-580.

Li, L., Zhou, H., Gómez-Hernández, J.J., Hendricks Franssen, H.-J., 2012. Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble Kalman filter. Journal of Hydrology 428–429, 152-169.

MacBeth, C., Floricich, M., Soldo, J., 2006. Going quantitative with 4D seismic analysis. Geophysical Prospecting 54, 303-317.

Marsily, G.d., 1986. Quantitative hydrogeology groundwater hydrology for engineers. Academic Press, Orlando, FL.

Massey, F.J., Jr., 1951. The Kolmogorov-Smirnov Test for Goodness of Fit. Journal of the American Statistical Association 46, 68-78.

Metz, B., Intergovernmental Panel on Climate Change. Working Group III., 2005. IPCC Special Report on Carbon Dioxide Capture and Storage. Cambridge University Press, for the Intergovernmental Panel on Climate Change, Cambridge.

Nævdal, G., Johnsen, L.M., Aanonsen, S.I., Vefring, E.H., 2005. Reservoir monitoring and continuous model updating using ensemble Kalman filter. SPE Journal 10, 66-74.

Nævdal, G., Mannseth, T., Vefring, E.H., 2002. Near-well reservoir monitoring through ensemble Kalman filter, SPE/DOE Improved Oil Recovery Symposium, pp. 13-17.

Ngodock, H.E., Jacobs, G.A., Chen, M.S., 2006. The representer method, the ensemble Kalman filter and the ensemble Kalman smoother: A comparison study using a nonlinear reduced gravity ocean model. Ocean Model 12, 378-400.

Nogues, J.P., Nordbotten, J.M., Celia, M.A., 2011. Detecting leakage of brine or CO_2 through abandoned wells in a geological sequestration operation using pressure monitoring wells. Energy Procedia 4, 3620-3627.

Park, Y.C., Huh, D.G., Park, C.H., 2012. A pressure-monitoring method to warn CO_2 leakage in geological storage sites. Environ Earth Sci 67, 425-433.

Redlich, O., Kwong, J.N.S., 1949. On the Thermodynamics of Solutions. V. An Equation of State. Fugacities of Gaseous Solutions. Chemical Reviews 44, 233-244.

Schlumberger, 2010. Eclipse technical description, v. 2010.1, Report, Abingdon, U.K.

Schöniger, A., Nowak, W., Hendricks Franssen, H.J., 2012. Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography. Water Resour Res 48, W04502.

Seto, C.J., McRae, G.J., 2011. Reducing Risk in Basin Scale CO_2 Sequestration: A Framework for Integrated Monitoring Design. Environ Sci Technol 45, 845-859.

Simon, E., Bertino, L., 2009. Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. Ocean Science 5, 495-510.

Skjervheim, J., Evensen, G., 2011. An ensemble smoother for assisted history matching, SPE Reservoir Simulation Symposium.

Spycher, N., Pruess, K., 2005. CO2-H2O mixtures in the geological sequestration of CO2. II. Partitioning in chloride brines at 12–100°C and up to 600 bar. Geochimica et Cosmochimica Acta 69, 3309-3320.

Spycher, N., Pruess, K., 2010. A phase-partitioning model for CO2–brine mixtures at elevated temperatures and pressures: application to CO2-enhanced geothermal systems. Transport in Porous Media 82, 173-196.

Sun, A.Y., Nicot, J.-P., 2012. Inversion of pressure anomaly data for detecting leakage at geologic carbon sequestration sites. Adv Water Resour 44, 20-29.

Van Genuchten, M.T., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils1. Soil Science Society of America Journal 44, 892.

van Leeuwen, P.J., Evensen, G., 1996. Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation. Mon Weather Rev 124, 2898-2913.

Vesovic, V., Wakeham, W., Olchowy, G., Sengers, J., Watson, J., Millat, J., 1990. The transport properties of carbon dioxide. Journal of Physical and Chemical Reference Data 19, 763-808.

Vinsome, P., 1976. Orthomin an Iterative Method for Solving Sparse Sets of Simultaneous Linear Equations, SPE Symposium on Numerical Simulation of Reservoir Performance. Society of Petroleum Engineers.

Zeidouni, M., Pooladi-Darvish, M., Keith, D.W., 2011. Leakage detection and characterization through pressure monitoring. Energy Procedia 4, 3534-3541.

Zhou, H., Gómez-Hernández, J.J., Hendricks Franssen, H.-J., Li, L., 2011. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. Adv Water Resour 34, 844-864.

Zhou, H., Li, L., Hendricks Franssen, H.-J., Gómez-Hernández, J.J., 2012. Pattern Recognition in a Bimodal Aquifer Using the Normal-Score Ensemble Kalman Filter. Math Geosci 44, 169-185.

Zhou, Q., Birkholzer, J.T., Mehnert, E., Lin, Y.-F., Zhang, K., 2009. Modeling Basin- and Plume-Scale Processes of CO2 Storage for Full-Scale Deployment. Ground Water 48, 494-514.