Carnegie Mellon University

CARNEGIE INSTITUTE OF TECHNOLOGY

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF Doctor of Philosophy

TITLE Characterization and Monitoring of Geochemical and Geophysical Effects of CO_2 Leakage from Sequestration Reservoirs

PRESENTED BY Zan Wang

ACCEPTED BY THE DEPARTMENTS OF

Civil and Environmental Engineering

Mitchell J. Small — February 20, 2015
CO-ADVISOR, MAJOR PROFESSOR — DATE

Athanasios Karamalidis — February 20, 2015
CO-ADVISOR, MAJOR PROFESSOR — DATE

David A. Dzombak — February 20, 2015
DEPARTMENT HEAD — DATE

APPROVED BY THE COLLEGE COUNCIL

Vijayakumar Bhagavatula — February 24, 2015
DEAN — DATE

Characterization and Monitoring of Geochemical and Geophysical Effects of CO_2 Leakage from Sequestration Reservoirs

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Civil and Environmental Engineering

Zan Wang

B.S., Hydropower Engineering, Sichuan University
M.S., Civil and Environmental Engineering, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

February, 2015

Acknowledgements

First and foremost, I offer my sincerest gratitude to my advisor, Professor Mitchell Small, who shared his extensive knowledge with me and guided me throughout the research process. I am also very grateful to my co-advisor Professor Athanasios Karamalidis for his help and patience. This work would not have been possible without their support and guidance.

I would like to thank the CEE department head Professor David Dzombak for proving me the opportunity to conduct a study on CO_2 solubility with Professor Athanasios Karamalidis and Professor Mitchell Small when I was a Master's student at CMU.

I am very grateful to my committee members: Professor Mitchell Small at CMU (chair of the committee), Professor Athanasios Karamalidis at CMU, Professor Matteo Pozzi at CMU and Professor William Harbert at University of Pittsburgh for their constructive and valuable suggestions on this work. I would also like to thank Professor Grant Bromhal at National Energy Technology Laboratory for his review and instructions on this work.

Finally, I would like to acknowledge the National Energy Technology Laboratory for providing funding for this work. As part of the National Energy Technology Laboratory's Regional University Alliance (NETL-RUA), a collaborative initiative of the NETL, this technical effort was performed under the RES contract DE-FE0004000.

Abstract

Geological carbon sequestration is considered a promising initial option to slow or reduce global atmospheric CO_2 concentrations. To demonstrate that the implementation of carbon sequestration is safe and effective as a greenhouse gas control technology, characterization and monitoring of geochemical and geophysical effects of CO_2 leakage from sequestration reservoirs is crucial. A multimodel predictive system (MMoPS) has been developed to predict CO_2 solubility in brine more accurately, so that storage capacity and cost estimates could be improved. The CO_2 leakage level is characterized through an assessment of the integrity and permeability of the caprock inferred from pressure measurements in the injection zone using a Bayesian approach. The detection power of pressure monitoring is evaluated using the expected distribution of pressure increases in the injection zone for permeable vs. impermeable caprock cases. The distributions of detection power using seismic travel time measurements at different CO_2 leakage levels are obtained using a statistical analysis and test. The methodology and results in this work should improve our ability to understand the storage reservoir chemistry and the statistical performance of the sole pressure or seismic monitoring, which could be integrated into a monitoring network, combining multiple monitoring techniques for CO_2 leakage detection.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 MOTIVATION

1.1.1 Carbon Capture and Sequestration

In response to concerns about possible climate change and its associated economic and environmental impacts, significant efforts are likely to be made during the coming decades to reduce CO_2 and other greenhouse gas emissions. A particular technology for achieving these reductions in the relatively near future is carbon capture and sequestration (CCS). This technology entails the capture and pressurization of carbon dioxide emissions from electric-generating plants and other large point sources, and their injection and storage in deep geologic formations.

Geologic formations, such as saline aquifers, oil and gas deposits, unmineable coal seams, organic-rich shales and basalts have been proposed as long-term storage sites for CO_2. Among the options being considered in the United States, deep saline formations are believed to have the highest CO_2 storage potential, on the order of 2,000 – 20,000 Gt, equivalent to storage of 50 – 500 years of anticipated US emissions (USDOE, 2010; Goodman et al., 2011). For effective estimation of storage potential, formations must be characterized in terms of their available pore volume for liquid-phase (supercritical) CO_2 storage, as well as the potential dissolution of CO_2 into formation (e.g., brine) water. CO_2 dissolution into brine and other ground waters is also an important determinant of the fate and transport of CO_2 leakage (should it occur) as it migrates towards the ground surface and/or other subsurface locations.

1.1.2 CO_2 Leakage Pathway

Injected CO_2 could leak slowly and diffusively through a sealing caprock, or leak relatively quickly through high permeability conduits, such as faults and abandoned wells. Depending on the kind of storage formation, one leakage pathway might be more likely than the other. For example, as the sealing caprock in a depleted oil and gas reservoir is usually well defined and there might be abandoned wells present from former reservoir operations, the most likely pathway for leakage is through the boreholes of the wells. For saline aquifers, the sealing caprock is less well defined or lacks a structural

trap at some locations. The most likely pathway would then be through the caprock if its permeability is not low enough (Benson & Myer 2002). The high pressure buildups after CO_2 injection into the storage formation could potentially hydraulically fracture the caprock (Chiaramonte et al., 2007). This mechanism might trigger higher effective permeability and a greater resulting chance for CO_2 leakage through the caprock (Liu et al. 2012).

1.1.3 CO_2 Leakage Detection and Monitoring Techniques

To demonstrate that the implementation of carbon sequestration is safe and effective as a greenhouse gas control technology, monitoring CO_2 stored in deep saline aquifers is crucial (Benson & Myer 2002). Multiple monitoring techniques can be applied at different depths, including atmospheric monitoring, near-surface monitoring and deep subsurface monitoring. (Benson and Myer, 2002; USDOE, 2009). Studies to characterize the performance of near-surface monitoring techniques, including near-surface soil CO_2 flux measurements and tracer measurements, have been undertaken (Yang et al., 2011a; Yang et al., 2011b). A Bayesian belief network methodology, integrating multiple monitoring techniques for CO_2 leak detection, has also been demonstrated by Yang et al. using near-surface monitoring techniques (Yang et al. 2012). A review of alternative statistical methods and perspectives for leak detection monitoring and inference is provided by Jenkins (2013). The costs of deploying different monitoring techniques were evaluated by Benson et al. (2004).

One broadly recommended and cost-effective monitoring technique is pressure monitoring in the injection and above zone reservoirs (Meckel et al., 2008). Measurements of injection pressures at the wellhead and in the formation are common in oil field practice (Benson & Myer 2002). An above-zone monitoring technique that also measures pressure changes in a permeable zone overlying the injection zone has been deployed in the Cranfield CO_2 Storage Project (Meckel et al. 2008). The continuous pressure changes in the injection zone and those in the above zone during CO_2 loading can be measured simultaneously once downhole pressure gauges are equipped in the monitoring wells. Unanticipated pressure changes in both zones can also be used as indicators of a fluid migration out of the injection zone (Meckel et al. 2008).

Another important monitoring technique is the seismic monitoring, which has also been implemented in several pilot CO_2 sequestration sites, such as the Frio Formation near Houston and the Penn West CO_2-EOR site (Daley et al. 2007; Alshuhail & Lawton 2007; Silver et al. 2007). Different seismic monitoring methods, including time-lapse 2D or 3D surface seismic surveys, borehole vertical seismic profiling and cross-well seismic monitoring, have been employed at sequestration sites (Alshuhail & Lawton 2007; Daley et al. 2007; Silver et al. 2007; Couëslan et al. 2009). The seismic attributes that can be measured include seismic velocity, travel time, amplitude and impedance (Alshuhail & Lawton 2007).

1.2 Thesis Statement

Geological carbon sequestration is considered a promising strategy for mitigating potential climate change associated with increasing CO_2 concentration in the atmosphere (Hepple and Benson, 2005; IPCC, 2005). To have a significant effect on the atmospheric concentration of CO_2, large amounts of CO_2 need to be captured from electric power plants and industrial sources and injected into deep underground formations, such as saline aquifers, depleted oil and gas reservoirs and unmineable coal seams (Holloway 2005; Gale 2004). Deep saline aquifers are estimated to have the highest CO_2 storage potential among all the geological storage options (USDOE 2010). Estimating the storage capacity and monitoring CO_2 stored in sequestration reservoirs is crucial to demonstrate that the implementation of carbon sequestration is safe and effective as a greenhouse gas control technology. A few questions arise:

- Several models exist to predict CO_2 solubility in brine, which one is the best one for estimating CO_2 solubility at conditions relevant to carbon sequestration?
- How to characterize the CO_2 leakage level using pressure measurements and how large is the power of pressure-based leakage detection?
- Can we detect CO_2 leakage with high confidence using seismic wave travel time measurements?

My thesis is aimed at using statistical tools and analysis methodologies to address each of the issues above.

1.3 Chapter Overview

The discussion in the remainder of this thesis proceeds as follows:

Chapter 2: The prediction of carbon dioxide solubility in brine at conditions relevant to carbon sequestration, i.e. high temperature, pressure and salt concentration (T-P-X), is crucial when this technology is applied. Eleven mathematical models for predicting CO_2 solubility in brine are compared and considered for inclusion in a multi-model predictive system. Model goodness of fit is evaluated over a temperature range of 304-433 K, pressure 74-500 bar and salt concentration 0-7 molal (NaCl equivalent), using 173 published CO_2 solubility measurements, particularly selected for those conditions. The performance of each model is assessed using various statistical methods, including the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Different models emerge as best fits for different sub-ranges of the input conditions. A classification tree is generated using machine learning methods to predict the best-performing model under different T-P-X sub-ranges, allowing development of a Multi-Model Predictive System (MMoPS) that selects and applies the model expected to yield the most accurate CO_2 solubility prediction. Statistical analysis of the MMoPS predictions, including a stratified 5-fold cross validation, shows that MMoPS outperforms each individual model and increases the overall accuracy of CO_2 solubility prediction across the range of T-P-X conditions likely to be encountered in carbon sequestration applications.

Chapter 3: Diffusive leakage of CO_2 through the sealing caprock is likely to occur at saline sequestration sites if the permeability of the caprock is too high. Pressure monitoring is a widely used technique for CO_2 leakage detection. This chapter aims to characterize the CO_2 leakage level in an idealized CO_2 storage site through an assessment of the integrity and permeability of the caprock inferred from pressure measurements in the injection zone. Monte Carlo uncertainty analysis is conducted, allowing the detection power of pressure measurements to be estimated. Based on Bayesian classification theory, the probability of each caprock permeability class given pressure measurements with measurement error can be inferred. The influence of time and location of measurements, the relative magnitude of the measurement error, the CO_2 injection rate, and the assumed uncertainty in reservoir properties on the inferred caprock permeability class and the detection power of the pressure monitoring

is also evaluated. Pressure monitoring alone is not sufficient to detect a permeable caprock for the system considered in this study, unless the caprock permeability is very high. A significant improvement in the statistical detection power is evident in the case where the uncertainty in reservoir properties is reduced through further site characterization. Complementary monitoring techniques will likely be needed to improve the power of pressure-based CO_2 leakage detection to acceptable levels.

Chapter 4: This chapter evaluates the detection power of seismic wave travel time measurements and statistical tests at different CO_2 leakage rate levels. A simplified rock physics model is assumed for monitoring zones at sequestration sites and the effects of leakage-induced changes in pressure and CO_2 saturation on P-wave travel times are modeled. The empirical distributions of detection power using the P-wave travel time for four regions in the permeability-porosity input space at four leakage levels are obtained from the Monte Carlo uncertainty analysis with a stochastic response surface method. The detection power using the P-wave travel time measurements and test alone is generally not high enough, unless the porosity and the permeability of the monitoring zone are high, and/or a long period of time has elapsed since the leakage occurred. For monitoring layers with lower permeability and porosity, measurements from other monitoring techniques will likely be needed to increase the probability that leakage events are detected and addressed in a timely manner.

Chapter 5: The main contributions of the thesis and lessons learnt relevant to predicting CO_2 solubility and conducting CO_2 leakage detection for geological sequestration are discussed. Future research directions are given.

Chapter 2: Multimodel Predictive System for Carbon Dioxide Solubility in Saline Formation Waters

2.1 INTRODUCTION

In response to concerns about possible climate change and its associated economic and environmental impacts, significant efforts are likely to be made during the coming decades to reduce CO_2 and other greenhouse gas emissions. A particular technology for achieving these reductions in the relatively near future is carbon capture and sequestration (CCS). This technology entails the capture and pressurization of carbon dioxide emissions from electric-generating plants and other large point sources, and their injection and storage in deep geologic formations.

Geologic formations, such as saline aquifers, oil and gas deposits, unmineable coal seams, organic-rich shales and basalts have been proposed as long-term storage sites for CO_2. Among the options being considered in the United States, deep saline formations are believed to have the highest CO_2 storage potential, on the order of 2,000 – 20,000 Gt, equivalent to storage of 50 – 500 years of anticipated US emissions (USDOE 2010; Goodman et al. 2011). For effective estimation of storage potential, formations must be characterized in terms of their available pore volume for liquid-phase (supercritical) CO_2 storage, as well as the potential dissolution of CO_2 into formation (e.g., brine) water. CO_2 dissolution into brine and other ground waters is also an important determinant of the fate and transport of CO_2 leakage (should it occur) as it migrates towards the ground surface and/or other subsurface locations.

CO_2 dissolution into the aqueous phase is limited by its solubility, which in turn depends upon the water temperature, pressure, and solute concentrations. These parameters can vary significantly from one formation to another, as well as from location to location within a given reservoir. In particular, temperature, pressure and salt concentration (T-P-X) are considered the main factors controlling the solubility of CO_2 in saline formation waters. Salt composition in deep saline formation waters may thus vary substantially among different formations depending on their depth, thermal conditions, geomorphology, and reservoir minerals. The concentrations of multiple salt ions, such as Na^+, Ca^{2+}, K^+, Mg^{2+}, Cl^-, SO_4^{2-} are often present in high concentrations. The purpose of this paper is to develop an

integrated multi-model system able to predict CO_2 solubility across the very wide range of subsurface conditions likely to be of interest in CO_2 geologic storage and leak-detection applications.

Many models for predicting CO_2 aqueous solubility have been developed in studies over recent decades, using a combination of phase-equilibrium and empirical relationships, and a number of these model studies are reviewed in the following section. Most of the developers of these models compared their predictions to a particular experimental data set under certain T-P-X conditions, but studies comparing available mathematical models with a comprehensive experimental data set specifically selected for carbon sequestration conditions are scarce.

In this study, eleven candidate models are identified. Qualitative and quantitative comparisons are conducted of their formulations, assumptions, principal input variables, and prediction accuracy. Available experimental data from various sources are combined to compare the performance of each model in predicting CO_2 solubility across the entire dataset. While some models are shown to perform better than others, each of the models considered provides the best prediction for some subset of the dataset. This motivated our development of a Multi-Model Predictive System (MMoPS) able to identify and use the best-performing model under different T-P-X sub-ranges. The model selection is made using machine learning techniques to generate a classification tree that inputs the T-P-X value and outputs the best model for that sample and the corresponding CO_2 solubility prediction. Using a rigorous set of cross-validation and model complexity versus accuracy comparisons, the MMoPS is shown to provide superior predictive capability compared to any individual model.

2.2 Comparison of Available Models for the Prediction of CO_2 Solubility in Brine

Many mathematical expressions have been proposed in recent decades for the calculation of CO_2 solubility in aqueous media (Table 2.1 and S-1 of Appendix 1). We reviewed these models to identify those applicable to brine conditions, that is, at T-P-X values representative of those likely to be encountered in brine reservoirs at carbon sequestration sites (U.S. Brine Wells Database, 2003):

Reservoir brine Temperature range:	304-433 K
Pressure range:	74-500 bar

Salt concentration (NaCl equivalent): 0-7 molal

Table 2.1 indicates eleven models that have been applied at T-P-X points within these ranges. Most of the selected models span a broad range of the temperatures reported in geologic formations, though their ranges for pressure applicability are more variable. Most of these models are valid for pressures up to about 1000 bar, except for the Barta and Bradley model (Barta and Bradley, 1985), which was developed for pressures lower than 152 bar. Models applicable to very high pressure conditions include Duan and Sun (Duan & Sun 2003; Duan et al. 2006) and Darwish and Hilal (Darwish & Hilal 2010), spanning a broad range of pressure conditions from 0-2000 bar (Table 2.1).

The majority of the models in Table 2.1 consider salt concentrations up to 4.5 molal of electrolyte (mol/kg H_2O), except for the Harvey and Prausnitz (Harvey & Prausnitz 1989) and Zuo and Guo (Zuo and Guo, 1991) models, which were developed for more dilute solutions up to 1 molal of electrolyte solution. The unit commonly used to express salt concentration is molality (mol electrolyte/kg H_2O). However, the salinity (the total dissolved solids in mass percent), is used in the models of Enick and Klara (Enick and Klara, 1990) and Chang and Coats (Chang et al., 1998), thereby including all dissolved electrolytes on a mass percent basis. Coexisting multiple salt ions are considered in the Harvey and Prausnitz, Zuo and Guo, and Duan and Sun models, while the rest of the models are specifically developed for NaCl or NaCl/$CaCl_2$ solutions.

Table 2.1 also provides a brief summary of the modeling approach adopted in each of the eleven predictive models. As indicated, the models can be divided into two principal types, based on their use of mechanistic phase equilibrium equations vs. primarily empirical equations fit using regression analysis. Approximately half of the models are phase equilibrium models based on equations of state (EOS) in which phase equilibrium calculations are involved. The remaining models are empirical or semi-empirical in which CO_2 solubility can be explicitly expressed as a function of the temperature, pressure and salt concentration. However, the line separating these two types of models is often blurred, since all the models contain both theoretical and empirical fitting relationships and/or parameters. A more detailed summary of the individual equations and parameters used by each model can be found in section B of Appendix 1.

2.2.1 Empirical and Semi-empirical Fitting Models

The Barta and Bradley model is a semi-empirical model which was developed by the modification of the Pitzer specific interaction model (Pitzer 1973). The authors indicated that below a temperature of 338 K and a pressure of 202 bar, the effect of pressure on CO_2 equilibrium was negligible, and therefore, the interaction parameters of aqueous species in Pitzer's model were only treated as a function of the temperature. The calculation of CO_2 solubility in brine using the Barta and Bradley model involves the estimation of the CO_2 fugacity. Barta and Bradley used reported values for fugacity coefficient of CO_2. In order to expand the application range of Barta and Bradley model to conditions where no reported fugacity coefficient values can be found, the fugacity coefficient model of Spycher and Reed (Spycher & Reed 1988) is coupled with the Barta and Bradley model for the CO_2 solubility calculation in our study. Using a fugacity coefficient model instead of the reported values also make Barta and Bradley model easier to be programmed in computer. The chosen fugacity model can accurately represent pressure-volume-temperature data of CO_2 for pressure-temperature values up to 500 bars and 625 K (Spycher & Reed 1988).

Duan and Sun model also uses interaction parameters from the Pitzer model, however both temperature and pressure are considered in the CO_2 solubility calculations. The Duan and Sun (2003) model is based on a fifth-order virial equation of state derived by Duan et al. (Duan et al., 1992) for the vapor phase and the Pitzer specific particle interaction model (Pitzer 1973) for the liquid phase. Specifically, in Duan and Sun (2003), the interaction parameters (λ's and δ's) of ions with the same charge number have the same value. The values of interaction parameters λ's for CO_2-bivalent cations, (e.g. $\lambda_{CO2\text{-}Ca}$) are assumed to be twice those for CO_2-monovalent cations (e.g. $\lambda_{CO2\text{-}Na}$). All ternary parameters (δ's) are assumed to have the same value as $\delta_{CO2\text{-}Na\text{-}Cl}$. With these assumptions the model originally developed for aqueous NaCl solutions can be applied to other electrolyte solutions or solutions with multiple coexisting electrolytes. Although the Duan and Sun (2003) model is based on an equation of state, it does not involve phase equilibrium calculations when calculating CO_2 solubility in brine.

In the revised version of their 2003 model, Duan and Sun (2006) (Duan et al. 2006) directly estimate the fugacity coefficient of CO_2, φ_{CO2}, as a function of temperature and pressure. This is

accomplished using a continuous empirical fitting through six consecutive sections of the pressure and temperature (P-T) space. A set of parameters was fitted for each section of the partitioned P-T space. The model was optimized for prediction accuracy at temperatures below 288 K using experimental CO_2 solubility data available at that time. The Duan and Sun model (Duan & Sun 2003; Duan et al. 2006) may also be used to estimate CO_2 solubility in pure water.

Darwish and Hilal (Darwish & Hilal 2010) developed an empirical fitting model based on an extension of the Setschenow equation (Setschenow, 1889) by fitting the Setschenow constant (K_{gs}) as a function of temperature and pressure. For the calculation of CO_2 solubility in brine, the Darwish and Hilal model requires the input of accurate values of CO_2 solubility in pure water which, according to the authors, can be estimated by Duan and Sun (2006) model.

Another model that is based on the Setschenow equation is Akinfiev and Diamond's empirical fitting model (Akinfiev & Diamond 2010). Similar to the Darwish and Hilal model, a method for estimating the solubility of CO_2 in pure water is required; this value is then adjusted for solution molality using the Setschenow equation. The Akinfiev and Diamond model also uses Pitzer's equations to estimate additional thermodynamic properties of the CO_2-H_2O-NaCl system.

Table 2.1 Characteristics of CO_2 Solubility Predictive Models

Model (year)	Temperature Range (K)	Pressure Range (bar)	Salt Concentration (Molal, mol/kg H_2O)	Salt Composition	Model Type
Barta & Bradley (1985)	298-623	0-152	0-6.0	NaCl	Semi-empirical fitting based on Pitzer's specific interaction model and Setschenow equation
Li & Nghiem (1986)	≤473	0-1000	0-4.3	NaCl	Phase equilibrium model based on Peng-Robinson EOS[a] for the gas phase and Henry's law for the aqueous phase; Scaled-particle theory (SPT) was used to modify the Henry's law constant to account for the presence of salt
Harvey & Prausnitz (1989)	≤423	0-1400	≤1	Na, Li, K, Ca, Mg, Cl, Br	Phase equilibrium model based on a molecular-based EOS for non-electrolytes, combined with a modification of Born's equation accounting for ionic effects
Enick & Klara (1990)	298-523	30-850	0-7.3	Ca, Na, Cl	Empirical fitting
Zuo & Guo (1991)	≤392 for $CaCl_2$; ≤423 for NaCl	≤700 for $CaCl_2$; ≤1400 for NaCl	≤1	Na, Ca, K, Mg, Cl, NO_3, SO_4, Br, I.	Phase equilibrium model based on Patel-Teja EOS, combined with a Debye-Huckel electrostatic term
Soreide & Whitson (1992)	288-623	14-970	0-5.0	NaCl	Phase equilibrium model based on Peng-Robinson EOS
Chang & Coats (1998)	313-373	0-1000	0-6.0	NaCl	Empirical fitting
Duan & Sun (2003, 2006)	273-533	0-2000	0-4.5	Na, K, Ca, Mg, Cl, SO_4	Duan & Sun (2003) is based on a fifth-order EOS by Duan et al. for the vapor phase and Pitzer's specific interaction model for the aqueous phase. Duan & Sun (2006) is an improved model of Duan & Sun (2003) by developing a non-iterative equation to replace the previous fifth-order EOS in the calculation of CO_2 fugacity coefficients
Darwish & Hilal (2010)	300-500	50-2000	1-4.0	NaCl	Empirical fitting based on Setschenow equation
Akinfiev & Diamond (2010)	251-373	1-1000	Any salt concentration	NaCl	Empirical fitting based on Setschenow constant; Diamond-Akinfiev model (2003) is used for binary CO_2-H_2O subsystem
Spycher & Pruess (2010)	285-573	1-600	0-6.0	NaCl	Phase equilibrium model relying on fugacity coefficients for the CO_2-rich phase and activity coefficients for the aqueous phase

[a] EOS: Equation of state.

The empirical model of Enick and Klara is developed as a two-step process. The first step includes the calculation of Henry's law constant for estimating the solubility of CO_2 in pure water. Both the reference Henry's law constant, H_{CO2}^*, and the partial molar volume of CO_2 at infinite dilution were treated as adjustable parameters which were optimized to match the experimental CO_2 solubility data. Thus, the values of the partial molar volume of CO_2 in this model are not comparable to experimental measurements of molar volume of CO_2 at infinite dilution. In the second step, another input, the total dissolved solids concentration (on a mass basis) is incorporated to relate the solubility of CO_2 in brine with that in pure water, allowing for the consideration of a salting-out effect in the presence of multiple salts.

The Chang and Coats model is based on a series of empirical equations and parameters. Care is needed in applying these equations to ensure that the units of the inputs are specified properly (the temperature is in ^oF, pressure is in psia, and the salt concentration is measured as salinity in weight percent). The CO_2 solubility calculated from this model is in units of scf CO_2/stb brine. In order to maintain the consistency of the model comparison, a density model developed by Batzle and Wang (Batzle & Wang 1992) is used, in our study, to convert the solubility in units of scf CO_2/stb brine into molality.

2.2.2 Phase Equilibrium Models

The fugacity coefficient is a central variable in phase equilibrium models and is estimated using different equations of state in different models. Often, cubic equations of state, such as the Peng-Robinson EOS (Robinson et al., 1985) and the Patel-Teja EOS (Patel & Teja 1982), are chosen to estimate the fugacity coefficient of CO_2 in the gas phase. However, these EOS were found to be inadequate for describing the aqueous phase interactions needed to estimate CO_2 fugacity in the aqueous phase (Soreide and Whitson, 1992). Thus, several modifications to EOS equations have been made to allow the calculation of CO_2 solubility in brine.

In the Soreide and Whitson model (Soreide and Whitson, 1992), the Peng-Robinson EOS is used with two modifications. First, the energy parameter α in the Peng-Robinson EOS was modified for the water/brine component to include the effects of NaCl salinity (C_{sw}) and the pure water reduced temperature (T_r). Second, two sets of binary interaction parameters (BIPs) were developed for the aqueous and non-aqueous phase. The aqueous phase BIP (k_{ij}^{AQ}) is a function of the temperature and salinity of the brine. The non-aqueous phase BIP (k_{ij}^{NA}) is a constant.

In the model introduced by Zuo and Guo, the non-aqueous phase is modeled using the Patel-Teja EOS. For the aqueous phase, a Debye-Huckel electrostatic term (Li and Pitzer, 1986) is combined with the Patel-Teja EOS (Patel & Teja 1982), allowing for extension of the EOS to electrolyte solutions. The energy parameter α is calculated using a mixing rule proposed by Kurihara et al. (Kurihara, et al., 1987).

Li and Nghiem (Li and Ngheim, 1986) modify the Henry's law constant derived for pure water to estimate CO_2 solubility in brine by incorporating parameters of the scaled-particle theory (Reiss et al., 1960). The molar density of each species in the aqueous phase is a required input. These values are estimated from the aqueous solution density using the density model of Rowe and Chou (Rowe & Chou 1970). For the non-aqueous phase, the model uses the Peng-Robinson EOS in which the interaction coefficient k_{ij} between water and CO_2 is temperature-dependent.

Harvey and Prausnitz (Harvey & Prausnitz 1989) use a molecular-based EOS (rather than the Peng-Robinson or Patel-Teja EOS) to describe intermolecular forces for non-electrolytes. They also employ a modification of Born's equation (Born, 1920) to account for interactions arising from permanent electric charges. Binary interaction parameters k_{ij} and k_{ji}, for gas-water pairs, ion-water pairs and ion-gas pairs were respectively fitted to binary vapor-liquid equilibrium data, osmotic-coefficient data and Setschenow-constant data at room temperature. These binary interaction parameters are used in a cubic mixing rule for calculation of the attractive-energy parameter between molecules.

In the model developed by Spycher and Pruess (Spycher & Pruess 2010), fugacity coefficient for the gas phase (CO_2-rich phase) is calculated using a modified Redlich-Kong EOS (Spycher et al. 2003). Activity coefficients are used to quantify fugacity in the aqueous phase. A Margules expression for CO_2 in pure water (Carlson and Colburn, 1942) and a Pitzer expression for salting-out effects are combined to estimate activity coefficients for the aqueous phase.

In the models of Enick and Klara, Chang and Coats, Darwish and Hilal, and Akinfiev and Diamond, and Spycher and Pruess, the solubility of CO_2 in pure water is first calculated, with each model using a different equation for this purpose. Enick and Klara optimized the Krichevsky-Ilinskaya equation (Krichevsky and Illinskaya, 1945) and the Krichevsky-Kasarnovsky equation (Krichevsky and Kasarnovsky, 1935). Chang and Coats derived an equation by fitting experimental measurements, while Darwish and Hilal recommend use of the Duan and Sun (2006) model to calculate the solubility of CO_2 in pure water. Akinfiev and Diamond utilize a previous version of their model (2003) (Akinfiev & Diamond 2003) to estimate CO_2 solubility in pure water. Three complex sub-models, including the Span and Wagner EOS, the Akinfiev and Diamond EOS and the Hill EOS (Akinfiev and Diamond, 2003; Hill, 1990; Span and Wagner, 1996), were incorporated in the Diamond-Akinfiev (2003) model (Diamond & Akinfiev 2003), making the model difficult to apply in practice. For simplicity in our study, the Duan and Sun (2006) model is employed for estimating CO_2 solubility in pure water and combined with the Akinfiev and Diamond (2010) model to estimate the solubility of CO_2 in brine. Accordingly, the Spycher and Pruess model uses their original model

(Spycher et al. 2003) developed in 2003 for temperatures ≤ 100 ^{o}C, whereas for temperatures > 100 ^{o}C, an activity coefficient model using Margules expressions (Spycher et al. 2003) is implemented to calculate CO_2 solubility in pure water.

The frequently required variables for the calculation of CO_2 solubility in brine by the eleven models are summarized in Table 2.2. Most models use the fugacity coefficient of CO_2 when deriving CO_2 solubility in brine, especially for phase equilibrium models. The binary interaction parameters between gas-water pairs, ion-water pairs and ion-gas pairs, Henry's law constant and Setschenow constant are also required in most models. Setschenow equation correlates the activity coefficient of a non-electrolyte solute and the concentration of the electrolyte. It is valid for low electrolyte concentrations, but has been extensively used for moderate concentrations with sufficient results (Yasunishi & Yoshida 1979).

Table 2.2 Frequently required variables for the calculation of CO_2 solubility in brine by different models[a].

Calculated Variables	Barta & Bradley (1985)	Li & Nghiem (1986)	Harvey & Prausnitz (1989)	Enick & Klara (1990)	Zuo & Guo (1991)	Soreide & Whitson (1992)	Chang & Coats (1998)	Duan & Sun (2003, 2006)	Darwish & Hilal (2010)	Akinfiev & Diamond (2010)	Spycher & Pruess (2010)
Solubility of CO_2 in pure water				√			√		√[b]	√[b]	√
Fugacity coefficient of CO_2 (Φ_{CO2})	√[c]	√	√		√	√		√			√
Henry's law constant (H_{CO2})	√	√		√							
Setschenow constant (K_{gs})			√						√	√	
Binary interaction parameters[d]	√	√	√		√	√		√			√
Activity coefficient										√	√
Partial molar volume of CO_2				√		√					√
Parameters for species[e]		√	√		√						

[a] For detail description of the calculated variables for each model, see Appendix 1. [b] Solubility of CO_2 in pure water is estimated by Duan and Sun (2006) model. [c] Fugacity coefficient is estimated from the equations of Spycher and Reed. [d] Binary interaction parameters are for gas-water pairs, ion-water pairs and ion-gas pairs. [e] Parameters for species include intermolecular size parameter, σ; intermolecular potential-energy parameter, ε; and molecular polarizability, α. (Harvey & Prausnitz 1989).

2.3 SELECTION AND INTERPRETATION OF EXPERIMENTAL DATA

Koschel et al. (Koschel et al. 2006) provide an intensive review of available experimental measurements on carbon dioxide solubility in aqueous NaCl solution. Their study indicates about 1,000 CO_2 solubility observations available over a wide range of pressure, however, most measurements were made at relatively low pressures (<50 bars). Only 20% of the dataset are for measurements taken above 50 bar and measurements above 200 bar are extremely rare (only 2%). In this study, 173 experimental CO_2 solubility measurements are selected from various sources, with most of them being the data sets used to develop the sub-models. These 173 experimental measurements are used as the dataset to evaluate alternative models for conditions representative of carbon sequestration reservoirs, viz. at temperatures of 304-433 K, pressures of 74-500 bar and salt concentration of 0-7 molal (NaCl equivalent).

From the available experimental measurements of CO_2 solubility in brine, only those reported for brine composition of NaCl and $CaCl_2$ are used. This decision is based on the fact that most predictive models are developed for aqueous NaCl solutions and that the salting-out effects of Na^+ and Ca^{2+} ions on CO_2 solubility are similar in solutions containing the same weight percent of these two salts (Liu et al. 2011). Therefore, available experimental data of CO_2 solubility in solutions of NaCl and $CaCl_2$ are combined. Data collected from electrolyte solutions containing K^+ ions are excluded, because the solubility of CO_2 in KCl solution is much higher than that in solutions containing the same weight percent of NaCl or $CaCl_2$ (Liu et al. 2011).

The distribution of the selected experimental measurements across the targeted pressure-temperature and salt concentration-temperature domains are shown in Figures S-1A and S-1B of Appendix 1, respectively. While the distributions of the measurements across these dimensions are not uniform, most of the space of interest is covered by at least one observation, with the exception of low temperature (304-340 K) samples with high pressure (above 200 bar) and/or high salinity (above 3 molal).

2.4 EVALUATION OF MODELS

The goodness of fit of each individual model against the measurements in the dataset formed for carbon sequestration conditions is assessed using the root mean squared error (RMSE) and the average percent error (APE) of the predicted vs. observed CO_2 solubility (Appendix 1). Considering the differing complexity of the alternative models, they are also evaluated in terms of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Both the AIC and BIC metrics consider a trade-off between the goodness of fit and complexity of the model, with the complexity represented by the number of fitted parameters in the model (Appendix 1). In general, adding more parameters to a model will improve accuracy in model predictions, but that sacrifices the simplicity of the model and also adds more risk of over-fitting. In the AIC and BIC, models with a large number of parameters are penalized for their higher complexity. The preferred model is the one with the lowest AIC and/or BIC value (Table 2.3).

As indicated in Table 2.3, the best performing model on all measures considered (lowest RMSE, lowest APE, lowest AIC, lowest AIC adjusted for small sample size, and lowest BIC) is the modified Akinfiev and Diamond (2010) model in which Duan and Sun (2006) model is employed to calculate CO_2 solubility in pure water, instead of the original equations (Diamond-Akinfiev 2003 model) proposed in Akinfiev and Diamond's study (2010). The number of data points for which each model provided the most accurate prediction in the experimental dataset is shown as "n (in data set)", in the next-to-the-last column of Table 2.3. As indicated, the modified Akinfiev and Diamond (2010) model and Duan and Sun (2006) model provide the best prediction for the highest number of observations (for 30 of the 173 measurements), though all of the remaining models predict best for a significant subset (at least 9) of the observations.

To better identify where each model performs best, the model yielding the most accurate prediction is identified in Figure S-2 of Appendix 1 for each temperature, pressure and salt concentration (T-P-X) condition represented. Different models performed better at different conditions of T-P-X. In order to predict CO_2 solubility with higher accuracy over the full range of conditions, utilization of all of the available models is necessary. For example, the Darwish and Hilal

model according to the evaluation criteria performed well in general (Table 2.3), however it was rarely a best-performing model at specific T-P-X space. In contrast, the Li and Nghiem model performed poorly compared to other models (Table 2.3), but was very accurate in predicting the solubility of CO_2 at high salt concentrations. The Harvey and Prausnitz, and Zuo and Guo models were excluded from model evaluation as they were developed for salt concentrations lower than 1 molal.

Table 2.3 Evaluation of models based on different statistical analysis criteria

Model[a]	Number of parameters	RMSE[b] (mol/kg)	APE[c] (%)	AIC[d]	Adjusted AIC for small sample size[e]	BIC[f]	n^g (in data set)	m^h (in MMoPS)
Akinfiev and Diamond (modified)[i]	30	0.076	9.5	-831	-817	-736	30	32
Duan and Sun	35	0.082	9.7	-796	-777	-686	30	31
Spycher and Pruess	36	0.084	11.2	-787	-767	-673	18	15
Enick and Klara	18	0.122	14.2	-692	-687	-635	22	19
Darwish and Hilal	28	0.129	15.9	-654	-642	-565	9	8
Soreide and Whitson	17	0.155	21.2	-611	-607	-557	21	20
Chang and Coats	16	0.217	25.9	-497	-493	-446	25	27
Barta and Bradley	17	0.296	34.4	-387	-383	-333	9	12
Li and Nghiem	17	0.457	50.6	-237	-233	-184	9	9
MMoPS[j]	36[k]	0.061	4.9	-897	-878	-784		

[a]Models are arranged in ascending order of AIC and BIC; [b]RMSE: Root mean squared error; [c]APE: Average percent error; [d]AIC: Akaike Information Criteria; [e]Adjusted AIC: A second-order AIC in which a bias-adjusted term is added into the original AIC formula to account for the influence of small sample size; [f]BIC: Bayesian Information Criteria; [g]n (in data set): Number of data points for which each model provides the

most accurate prediction in the experimental data set; [h]*m* (in MMoPS): Number of data points where Multi-model predictive system suggests using model *i* to predict CO_2 solubility in the evaluation data set; [i]Akinfiev and Diamond (modified) model: A modification of the original Akinfiev and Diamond (2010) model by using the Duan and Sun (2006) model to calculate CO_2 solubility in pure water, instead of the original equations (Diamond-Akinfiev 2003 model) proposed in Akinfiev and Diamond's study (2010); [j]MMoPS: Multi-model predictive system; [k]36: The 36 parameters in MMoPS refers to the 36 splitting points in the generated classification tree in order to select the best-performing model for each condition.

2.5 Development of a Multi-Model Predictive System

A multi-model predictive system (MMoPS) designed to predict CO_2 solubility more accurately at specific T-P-X range representative of carbon storage reservoirs is developed by using a classification tree to combine the individual CO_2 solubility models of Table 2.3. The MMoPS is an overlying model that consists of a classification tree to select the best-performing model for specific T-P-X conditions, subsequently applying that model to predict CO_2 solubility for those conditions (Figure 2.1). Classification trees are developed using machine learning methods to find a set of sequential split-points in a multivariate input space (values of T-P-X) that best predict a discrete outcome (Hastle et al., 2001), in this case, the best-performing model. Classification and related regression trees (for continuous outputs) have been developed and applied for a number of environmental problems (Bennett et al., 2000; Eisenberg and McKone, 1998; Green et al., 2009; Hartman et al., 2009; Hughes et al., 2010; Ren, 2003; Worth and Cronin, 2003).

The classification tree developed in this study uses the T-P-X value for each of the 173 observations, along with the best-performing (lowest error) model determined for each. The tree is generated using the J48 algorithm in the Weka software system, a practical machine learning tool (Witten et al, 2011). The J48 algorithm selects optimal splitting points for the tree based on the information gain that is achieved (Witten et al., 2011). In deciding whether to add additional splits and branches, a tradeoff is made between the improved prediction obtained at each final node of the tree and the increased tree complexity that results (increased number of nodes, with fewer observations at each). The resulting classification tree chooses the model predicted to have the lowest error compared to the experimental data. Because of the variability across the dataset and the approximate nature of each of the models considered, the classification is not perfect – some classification errors do occur. This error is quantified, as is the prediction error across the 173 observations that results from a) choosing the wrong model (in some cases); and b) the error that remains when even the best-performing model is chosen. Bubble charts are provided in Figure S-3 (Appendix 1) to demonstrate the distribution of prediction error of MMoPS over T-P space at low,

moderate and high salt concentrations, respectively. The out-of-sample predictive capability of MMoPS is explored by a cross-validation method in which 20 percent of the data are randomly withheld when fitting the classification tree (i.e., a 5-fold cross validation) and predictions for the withheld data are compared to their observed values.

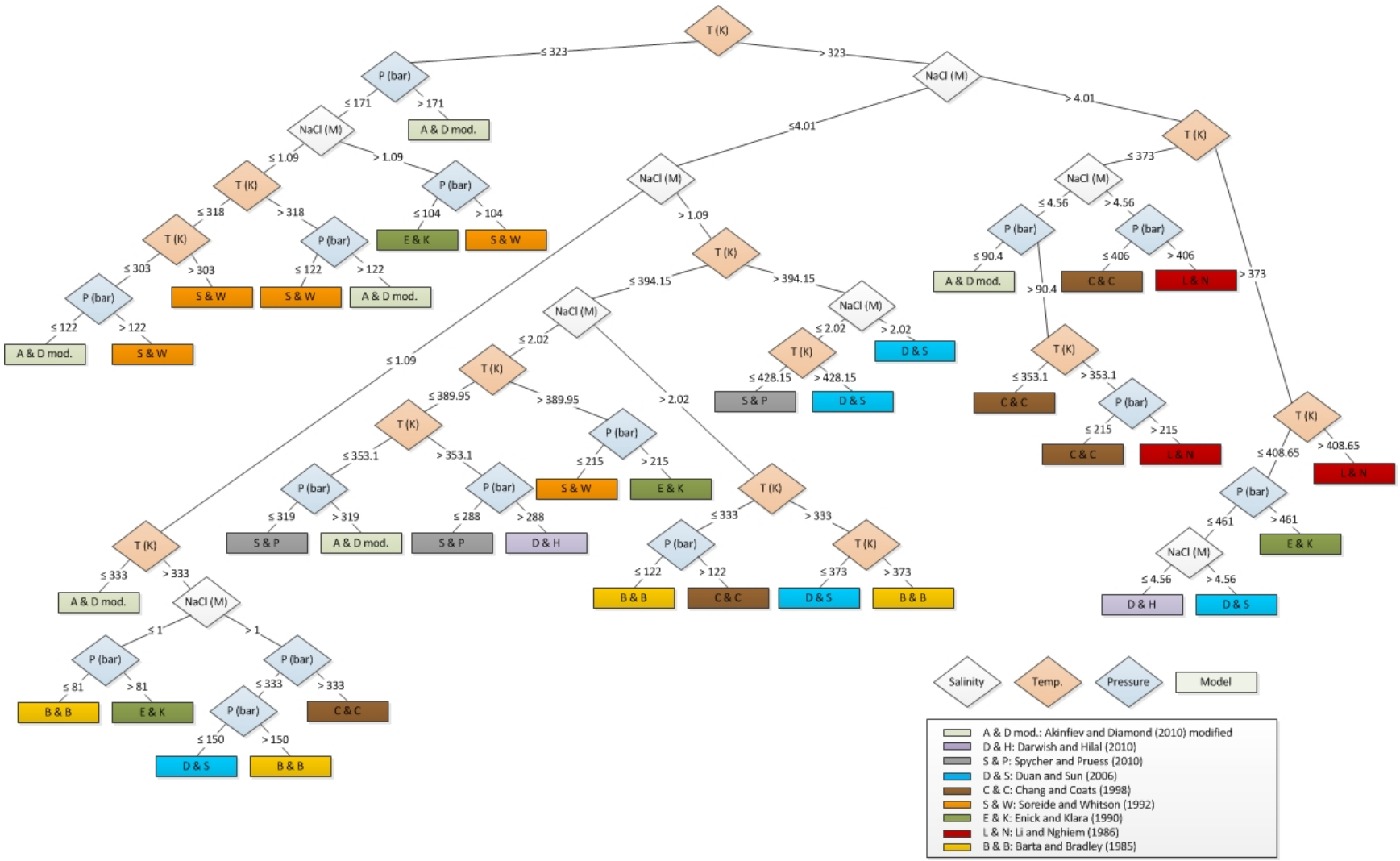


Figure 2.1 A multi-model predictive system for CO_2 solubility over T-P-X ranges encountered in carbon sequestration sites. Diamonds represent the splitting points in the T-P-X space. The colored rectangle at each leaf node is the selected best-performing sub-model for the specified range of T-P-X conditions.

2.6 RESULTS AND DISCUSSION

As indicated in Table 2.3, all of the individual models are chosen by the MMoPS to predict CO_2 solubility for at least eight of the original observations (last column of Table 2.3), and the number of times that a model is chosen correlates strongly with the number of times it provides the best fit (next to the last column of Table 2.3). The performance of MMoPS is evaluated using the same statistical analysis criteria as those used for the individual sub-models (Table 2.3). The MMoPS outperformed each individual model, achieving the lowest value of the RMSE, APE, AIC, adjusted AIC and BIC. The prediction error of CO_2 solubility can be greatly reduced by using MMoPS instead of any individual model for the whole T-P-X range of interest (see RMSE and APE of Table 2.3). The average prediction error of MMoPS (4.9%) is small, compared to the range of prediction error achieved by the individual models (9.5%-50.6%). This improved prediction is obtained by tapping results from each of the individual models, chosen over the range of conditions where each performs best. However, with 36 parameters, the MMoPS is a very complex overall model and it is more difficult to implement, requiring prior evaluation of the classification tree, followed by application of the chosen model. Is this additional complexity worth it in terms of the anticipated ability of the model to predict CO_2 solubility for observations not used to fit the model? Cross validation methods are designed to address this type of question.

In order to estimate the performance of MMoPS outside of the dataset used for fitting, a stratified 5-fold cross-validation was conducted. The experimental data set was divided randomly into 5 parts in which the distribution of preferred sub-models was approximately the same as that in the full dataset. Each part was withheld in turn and the classification tree was generated using the remaining four-fifths resulting in an RMSE of

0.07 for the full dataset. Comparing these results to those obtained for in-sample prediction using the entire dataset (RMSE = 0.061, see Table 2.3), some expected loss in predictive capability is indicated. However, the cross-validation errors are still lower than the in-sample error rates achieved by any of the individual sub-models, further supporting the predictive capability and value of the MMoPS.

In the classification tree, all the splits are binary splits. The classification tree sequentially splits T-P-X ranges into sub-ranges along each branch until the number of instances at a node reaches the minimum value set in the algorithm or further bifurcations no longer achieve sufficient differentiation to justify their inclusion in the tree. In general, salt concentration is the most informative attribute in distinguishing best-performing sub-models, followed by temperature and pressure, according to the evaluation of their chi-squared statistic with respect to the best-performing sub-model (Appendix 1). As shown in Figure 1, the resulting tree first splits into a small, low-temperature branch (with seven end-nodes) and a large high-temperature branch (with 27 end-nodes) at T = 323 K. The low-temperature branch includes four split-nodes for pressure, two for temperature and three for pressure; while the high-temperature branch has nine split-nodes for temperature, eight based on pressure, and six for salinity.

The generated classification tree model indicates that some models perform well for a specific region of T-P-X space, but cannot provide good estimation for other regions. All these sub-models, to some extent, involve fitting model parameters to experimental data, regardless of their type (empirical fitting or phase equilibrium models). The T-P-X conditions of the experimental data that each model chose to fit the parameters largely determine the suitable conditions for each model. For example, Duan and Sun, and Akinfiev and Diamond fitted their model parameters using many of the

experimental data included in this study over a wide range of T-P-X conditions. These models similarly exhibit good prediction over broad T-P-X ranges, providing the most accurate prediction for about 35% of the observations in the experimental dataset, and the most frequent use in MMoPS (Table 2.3). Extrapolation of models to conditions beyond fitted conditions may decrease the overall performance of these models. For example, the Li and Nghiem model was fitted using data primarily at high T-P-X conditions and appears only on the right portion of the tree (Figure 3). This indicates that this model is only most suitable for predicting CO_2 solubility at the high temperature, pressure and salt concentration conditions to which it was fit.

MMoPS performs better than any individual models tested in this study and increases the overall accuracy of CO_2 solubility prediction across carbon sequestration conditions, not by optimization of the individual model variables or constants, but because it chooses in the most effective manner the best-performing individual model for a specific region of the T-P-X space. That means that all individual models presented in Table 2.3 are utilized in MMoPS. For "less popular" sub-models, those with poor performance over the whole range of conditions having high RMSE, APE, AIC and BIC, such as Li & Nghiem model (Table 2.3), MMoPS identifies the suitable conditions for which this model should be used to predict CO_2 solubility. As such, the overall uncertainties of CO_2 solubility estimation for carbon sequestration conditions are substantially reduced. The predicted CO_2 solubility values from sub-models and MMoPS for all the 173 experimental data points are compared in Table S-7 (Appendix 1). In particular, a series of experimental data in which only one of the T-P-X conditions is varied (i.e. varying P at constant T and X, varying T at constant P and X, and varying X at constant T and P) are selected and plotted in Figure S-4 to graphically demonstrate the

accuracy of different sub-models and MMoPS. Note that these figures only represent the accuracy of each model at the selected conditions and the model performance would vary for other conditions. A discussion about the effect of using sub-sets of the 173 experimental data (i.e. the individual sub-set of NaCl data and $CaCl_2$ data) on MMoPS and evaluation of models is provided in Appendix 1.

Identifying further experimental studies, three CO_2 solubility measurements with high prediction errors in the MMoPS model are identified and listed in Table S-8 (Appendix 1). The prediction errors for the three measurements account for 66% of the total squared error for the whole evaluation set. For these observations the classification tree model indicated an alternative sub-model, instead of the model with the lowest error to predict CO_2 solubility. The three measurements are from the same experimental study, Tekenouchi and Kennedy's study (Tekenouchi and Kennedy, 1965). Another set of measurements from a different study, Gehrig's study (Gehrig, 1980) but at conditions close to those of Tekenouchi and Kennedy's study are also collected and listed in Table S-8. CO_2 solubility measurements from these two studies are not consistent with each other. Using the models selected in MMoPS more accurately predicts the CO_2 solubility measurements obtained from Gehrig. We conducted statistical analysis of the measurement error of CO_2 solubility between laboratories and different experimental practices using all the measurements we found that have parallel measurements (measured at identical T-P-X conditions). The mean of absolute measurement errors is 0.064 mol/kg which is of the same magnitude as the model root mean square errors (RMSE) shown in Table 2.3. Note that data from Tekenouchi and Kennedy's study and Gehrig's study are excluded from the statistical analysis because of the high discrepancy between these two studies. Uncertainties in CO_2 experimental measurements also transfer

to the uncertainties in the estimated performance of CO_2 solubility predictive models. More CO_2 solubility experimental measurements are necessary to verify the reliability of different experimental studies. The lack of experimental data in specific regions of the T-P-X space also increases the uncertainty of CO_2 solubility predictions. We plan to apply MMoPS to help identify further solubility experiments with the greatest potential to reduce the overall- and region-specific accuracy of the CO_2 solubility prediction. Predicting CO_2 solubility more accurately should improve our ability to understand and predict storage reservoir chemistry and improve storage capacity and cost estimates.

Chapter 3: A Bayesian Approach to CO_2 Leakage Detection at Saline Sequestration Sites using Pressure Measurements

3.1 INTRODUCTION

Geological carbon sequestration is considered a promising strategy for mitigating potential climate change associated with increasing CO_2 concentration in the atmosphere (Hepple and Benson, 2005; IPCC, 2005). To have a significant effect on the atmospheric concentration of CO_2, large amounts of CO_2 need to be captured from electric power plants and industrial sources and injected into deep underground formations, such as saline aquifers, depleted oil and gas reservoirs and unmineable coal seams (Holloway 2005; Gale 2004). Deep saline aquifers are estimated to have the highest CO_2 storage potential among all the geological storage options (USDOE 2010).

To demonstrate that the implementation of carbon sequestration is safe and effective as a greenhouse gas control technology, monitoring CO_2 stored in deep saline aquifers is crucial (Benson & Myer 2002). Multiple monitoring techniques can be applied at different depths, including atmospheric monitoring, near-surface monitoring and deep subsurface monitoring. (Benson and Myer, 2002; USDOE, 2009). Studies to characterize the performance of near-surface monitoring techniques, including near-surface soil CO_2 flux measurements and tracer measurements, have been undertaken (Yang et al., 2011a; Yang et al., 2011b). A Bayesian belief network methodology, integrating multiple monitoring techniques for CO_2 leak detection, has also been demonstrated by Yang et al. using near-surface monitoring techniques (Yang et al. 2012). A review of alternative statistical methods and perspectives for leak detection monitoring and inference is provided by Jenkins (2013). The costs of deploying different monitoring techniques were evaluated by Benson et al. (2004).

One broadly recommended and cost-effective monitoring technique is pressure monitoring in the injection and above zone reservoirs (Meckel et al., 2008). Measurements of injection pressures at the wellhead and in the formation are common in oil field practice (Benson & Myer 2002). An above-zone monitoring technique that also measures pressure changes in a permeable zone overlying the injection zone has been deployed in the Cranfield CO_2 Storage Project (Meckel et al. 2008). The continuous pressure changes in the injection zone and those in the above zone during CO_2 injection can be measured simultaneously once downhole pressure gauges are equipped in the monitoring wells. Unanticipated pressure changes in both zones can also be used as indicators of a fluid migration out of the injection zone (Meckel et al. 2008). The range of anticipated conditions under leak vs. no-leak scenarios can be characterized using reservoir flow and pressure models.

Semi-analytical solutions have been derived to describe the space-time evolution of a CO_2 plume during injection and CO_2 leakage through an abandoned well (Nordbotten et al. 2005). This semi-analytical model was further applied to a specific field site in Alberta, Canada to perform a risk analysis of CO_2 and brine leakage along old wells (Celia et al., 2011). The sensitivity and uncertainty analysis of pressure measurements has been studied by several authors (Birkholzer et al., 2009; Chabora and Benson, 2009; Sun et al., 2013a).

In the paper by Sun et al., a particular uncertainty-quantification technique especially well-suited for computationally complex models, the probabilistic collocation method, is used to assess leakage detectability based on the signal to noise ratio of pressure perturbations predicted to occur from leakage. In a related paper an optimization method is formulated for designing a pressure-based monitoring system at

carbon sequestration sites (Sun et al., 2013b). Azzolina et al. (2013) also propose a probabilistic framework for leak detection based on pressure monitoring. They utilize a multilayer reservoir pressure model to predict system response to CO_2 injection and evaluate its response using an approximate first-order uncertainty analysis method. Our analysis is similar in concept, but uses a more aggregate analytical model for the subsurface and more advanced methods for uncertainty analysis based on Monte Carlo simulation and Bayesian inference regarding caprock states and potential associated leakage rates.

Injected CO_2 could leak slowly and diffusively through a sealing caprock, or leak relatively quickly through high permeability conduits, such as faults and abandoned wells. Depending on the kind of storage formation, one leakage pathway might be more likely than the other. For example, as the sealing caprock in a depleted oil and gas reservoir is usually well defined and there might be abandoned wells present from former reservoir operations, the most likely pathway for leakage is through the boreholes of the wells. For saline aquifers, the sealing caprock is less well defined or lacks a structural trap at some locations. The most likely pathway would then be through the caprock if its permeability is not low enough (Benson & Myer 2002). The high pressure buildups after CO_2 injection into the storage formation could potentially hydraulically fracture the caprock (Chiaramonte et al., 2007). This mechanism might trigger higher effective permeability and a greater resulting chance for CO_2 leakage through the caprock (Liu et al. 2012).

This study focuses on diffusive CO_2 leakage through the sealing caprock and the resulting pressure changes in the injection zone. The objective of this study is to characterize the CO_2 leakage level through an assessment of the integrity and

permeability of the caprock inferred from pressure measurements and to evaluate the detection power of the pressure monitoring technique. The uncertainty of the pressure in the injection zone is estimated from Monte Carlo simulation of pressure buildups using a simplified geological reservoir model. Pressure measurements are assumed to be made with error, considering low, medium and high coefficients of variation for each measurement, conditioned on the modeled value. Based on Bayesian classification theory, the probability of each caprock permeability class given the observations of pressure in the injection zone can then be inferred. The power of pressure monitoring to detect a leaky, high-permeability caprock is also evaluated from the simulation results. Further, we evaluate the influence of time and location of measurements, the magnitude of the measurement error, the CO_2 injection rate, and the uncertainty in reservoir properties on the inferred caprock permeability class and the detection power of the pressure monitoring. The uncertain reservoir properties considered here include the thickness, permeability and porosity of the injection zone and the above zone.

3.2 Model Setup and Uncertainty Analysis

3.2.1 A Simplified Geological Reservoir Model

The CO_2 storage site is modeled as a confined system with two aquifers separated by a sealing caprock of thickness 120 m (Figure 3.1). The lower aquifer is the injection zone and the upper aquifer is the above zone. The injection zone is at a depth of 2000 m below the ground surface and the temperature is assumed to be 65 °C. The initial pressures in the injection zone and the above zone are stable at $2*10^7$ Pa and $1.97*10^7$ Pa, respectively. Each layer (the injection zone, the sealing caprock and the above zone) is assumed to be homogeneous, isotropic and with infinite radial extent. An injection well

completely penetrates the above zone and the caprock, and is screened at the top of the injection zone. CO_2 is continuously injected into the injection zone from the injection well at a fixed rate. A base-case observation well is located 500 m away from the injection well, with monitoring wells at 100 m also considered in a sensitivity analysis. Pressure buildups in the injection zone after continuous injection of CO_2 over 1, 2, 5, 10 and 20 years are calculated using the analytical solutions derived by Neuman and Witherspoon (Neuman & Witherspoon 1969).

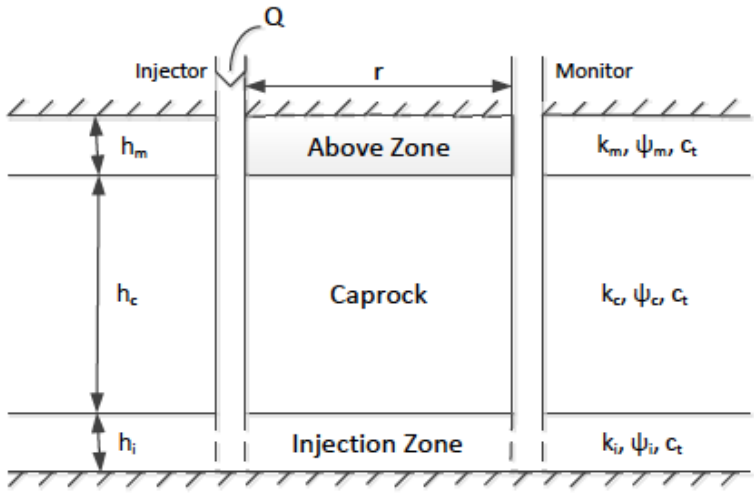


Figure 3.1 Schematic diagram of the simplified geological reservoir model, adapted from Neuman and Witherspoon (1969).

The Neuman-Witherspoon analytical solution is a single-phase solution, originally developed for the problem of transient flow to a well in a confined two-aquifer system. The equations have been previously used to calculate pressure buildups after CO_2 injection at carbon sequestration sites by Chabora and Benson (Chabora & Benson 2009). In the Neuman-Witherspoon solution, the pressure change is expressed as an infinite integral of a highly oscillating function. This integral is solved numerically using the adaptive Simpson quadrature in MATLAB (Gander & Gautschi 2000; Chabora & Benson

2009). The equations of the Neuman-Witherspoon analytical solutions and the MATLAB codes are listed in Appendix 2A. Solutions are obtained with relatively short computation time, allowing Monte Carlo uncertainty analysis (with 20 scenarios and a sample size of n=1000 replications for each scenario) to be conducted for the predicted pressure buildups.

3.2.2 Uncertainty in Pressure Evolution

The uncertainty in pressure evolution following initiation of the CO_2 injection is quantified using the Neuman-Witherspoon solution implemented with Monte Carlo sampling of uncertain model inputs and parameters. The presumptive conditions of a carbon sequestration site and CO_2 injection rate used in the Monte Carlo uncertainty analysis are summarized in Table 3.1. These parameters are obtained from geological properties at the Cranfield CO_2 storage site (Meckel et al. 2008) and the reservoir conditions used by Chabora and Benson (Chabora & Benson 2009), and are assumed to be fixed. The assumed uncertain inputs include the thickness, permeability and porosity of the injection zone and those in the above zone. The uncertain inputs are assumed to be independent variables with a specified distribution type. The geological properties in the above zone are assumed to follow the same distribution as those in the injection zone. Based on the permeability and porosity database developed by the National Petroleum Council (NPC, 1984), the distributions of uncertain inputs shown in Table 3.2 were selected to conduct the Monte Carlo uncertainty analysis. Scatter plots of permeability vs. porosity based on the NPC database can be found in Appendix 2A.3.

One assumption made in the Neuman-Witherspoon analytical solution is that the flow is vertical in the caprock and horizontal in the aquifers. As pointed out by Neuman and Witherspoon (1969), the permeability of the aquifers needs to be two or more orders-

of-magnitude greater than that of the caprock in order to reduce the errors introduced by this assumption. Therefore, the assumed lognormal distribution of permeability in reservoirs is truncated to guarantee that the reservoir permeabilities are at least four orders of magnitude greater than that of the caprock. As this study focuses on the diffusive leakage of CO_2 through the sealing caprock, the caprock permeability class, defined by intervals of sealing caprock permeability values, is assumed to serve as a surrogate for the magnitude of CO_2 leakage (Table 3.3).

Table 3.1 Fixed reservoir and injection conditions assumed in Monte Carlo uncertainty analysis

Parameter	Symbol	Value
Depth (m)	H	2000
Temperature (°C)	T	65
Initial pressure in the injection zone (Pa)	pI_initial	$2*10^7$
Initial pressure in the above zone (Pa)	pM_initial	$1.97*10^7$
Porosity of the caprock	ϕ_c	0.05
Thickness of the caprock (m)	h_c	120
Total formation compressibility (Pa^{-1})	c_t	$1*10^{-9}$
Injection rate (MT-CO_2/yr)	Q	1
Radial distance from injection well (m)	R	500

Table 3.2 Distributions of uncertain input parameters assumed in Monte Carlo uncertainty analysis (based on values found in the National Petroleum Council database (NPC, 1984))

Parameter	Distribution Type	Mean (μ)	Standard deviation (σ)	Lower limit	Upper limit

Formation thickness, $h_{I/M}$ (m)	Normal	10	2		
Permeability, $k_{I/M}$ (mD)	Truncated Log10 normal	2[a]	1[a]	$10^4 k_c$	
Porosity, $\varphi_{I/M}$	Truncated Normal	0.2[b]	0.1[b]	0.0005	0.6

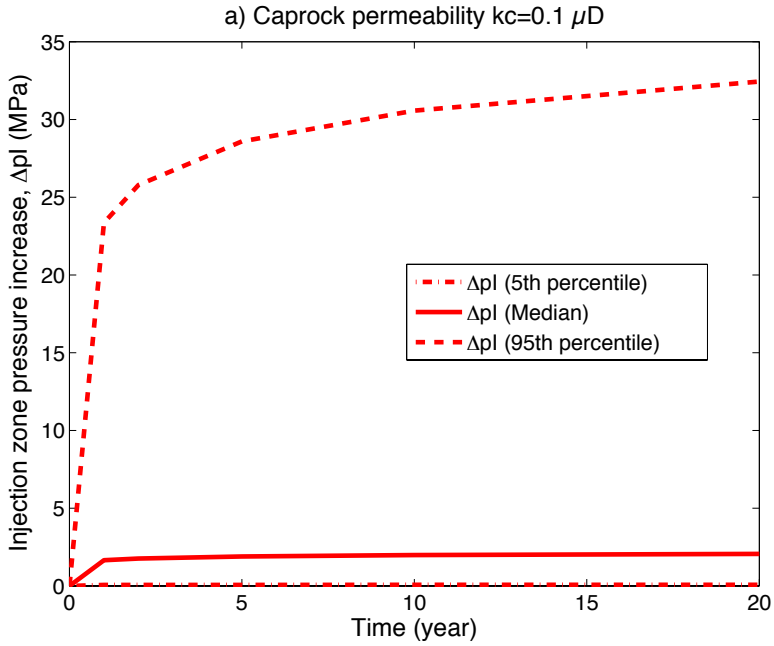
a: The mean and standard deviation values are parameters of the lognormal distribution in base 10, that is, the mean and standard deviation of the base 10 logarithm of the variable before truncation.
b: The mean and standard deviation values are parameters of the normal distribution before truncation.

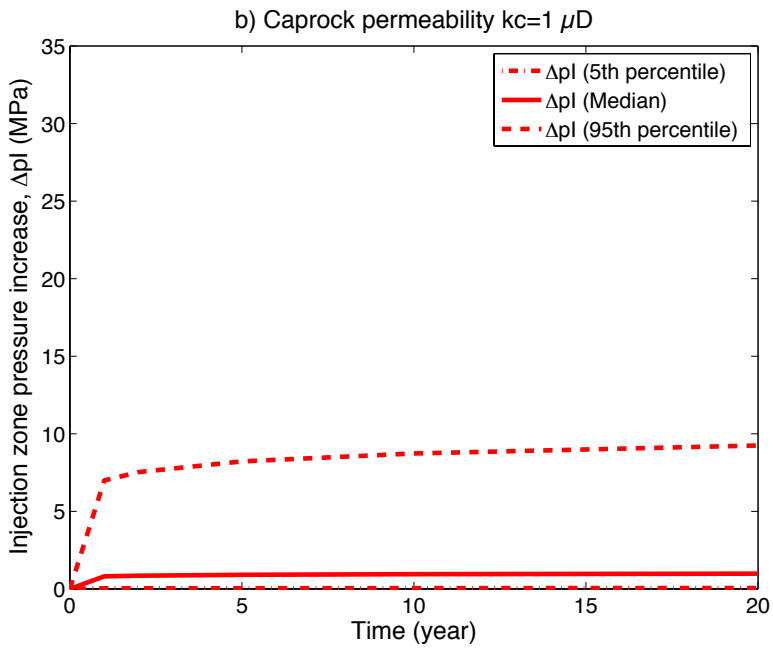
Table 3.3 Intervals of caprock permeability values representing different leakage levels

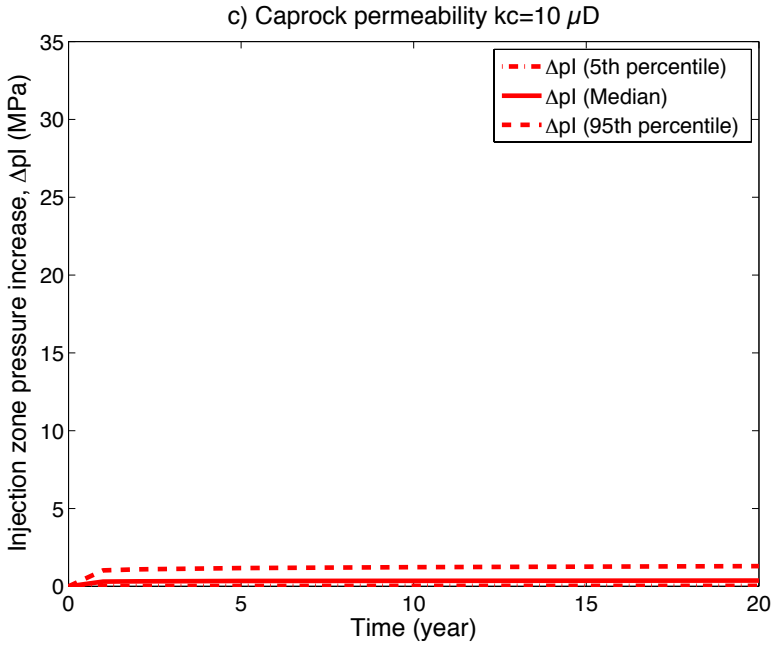
Caprock permeability, k_c (μD)	CPC = Caprock Permeability Class	CO_2 Leakage Magnitude	Assumed Prior Probability, $\pi(CPC)$[a]
0.1-1	1. Impermeable	Negligible	0.25
1-10	2. Moderate	Moderate	0.25
10-100	3. High	High	0.25
100-1000	4. Very High	Very High	0.25

Given the assumed distributions of uncertain inputs, Monte Carlo simulations were run for five caprock permeability values within each caprock permeability class, resulting in 20 caprock permeability scenarios in total. Each execution of the Monte Carlo model simulates (using the Neumann-Witherspoon analytical solution) 1000 different temporal profiles of pressure buildup in the injection zone at a location 500 m away from the injection well. Pressure buildups in the above zone can also be calculated using the Neuman-Witherspoon solution, but are not as responsive as those in the injection zone when the caprock permeability increases. The value of adding a second

measurement (i.e., the above-zone pressure) is further diminished by the presence of measurement error. Therefore, this study focuses solely on pressure buildups in the injection zone. Pressure values are noted and evaluated at t=1, 2, 5, 10 and 20 years following the initial injection. As indicated in Table 3.1, a continuous injection rate of 1 MT CO_2/yr is maintained over the 20-year simulation period. The simulated uncertainty in pressure buildups in the injection zone (i.e. pressure changes, ΔpI) after CO_2 injection at dividing points of caprock permeability classes is depicted in Figure 3.2, including the empirical 5th, 50th and 95th percentile values of the simulation results. As the caprock permeability increases, the uncertainty in the simulation results decreases markedly. Also, the pressure buildups in the injection zone (ΔpI) decrease as the caprock permeability increases. This is reasonable because, as the caprock permeability increases, more fluid from the injection zone is expected to migrate into the above zone through the caprock, resulting in a pressure decrease in the injection zone.




b) Caprock permeability kc=1 μD
Injection zone pressure increase, Δpl (MPa)
Time (year)
Δpl (5th percentile)
Δpl (Median)
Δpl (95th percentile)
0
5
10
15
20
25
30
35


c) Caprock permeability kc=10 μD
Injection zone pressure increase, Δpl (MPa)
Time (year)
Δpl (5th percentile)
Δpl (Median)
Δpl (95th percentile)
0
5
10
15
20
25
30
35

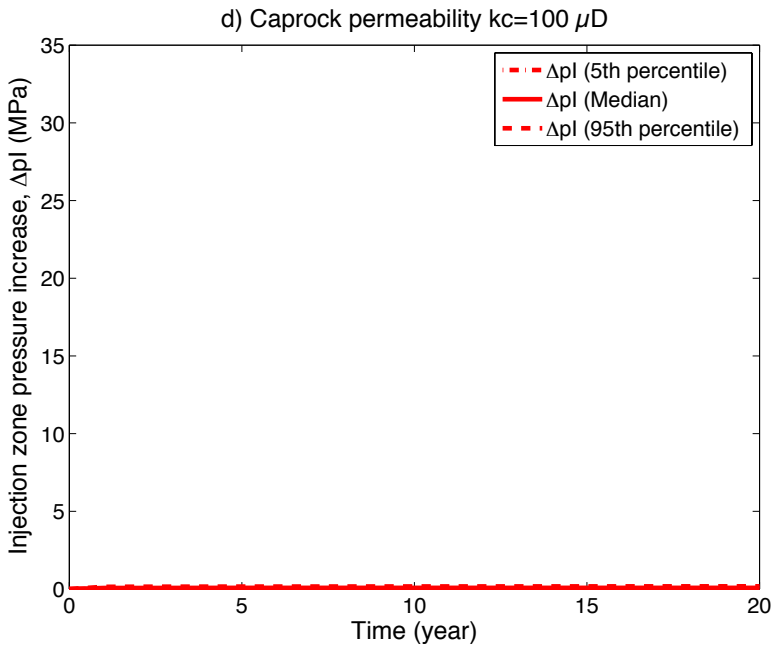


Figure 3.2 Simulated uncertainty in pressure buildups in the injection zone at 500 m from point of injection after injecting CO_2 with a) Caprock permeability kc = 0.1 md; b) Caprock permeability kc = 1 md; c) Caprock permeability kc = 10 md; and d) Caprock permeability kc = 100 md. Results shown are the median (solid line), 5th and 95th percentile values (dashed lines) of 1000 Monte Carlo simulations for the injection zone (*ΔpI*: pressure buildups in the injection zone, shown in red).

3.3 Bayesian Classification Methodology for CO_2 Leakage Detection

Bayesian theory provides a mathematical tool to combine an expert's belief regarding the probability of each caprock permeability class (CPC) with evidence obtained from field measurements. An expert's belief is represented by a prior distribution for the presence of each CPC at a site. In the Bayesian classification methodology, the prior distribution and the field measurements are combined using the likelihood function to derive the posterior distribution. If an uninformative prior

distribution is assumed, the results will be objective, that is, totally dependent upon the field measurements. In this case, the posterior probability of a caprock permeability class is proportional to the likelihood function for the field pressure measurements, given each CPC.

The likelihood function for the field pressure measurements is composed of two components. The first is the uncertainty in the true value of the pressure in the injection zone (at a given location and time) that results from the uncertain reservoir properties; this uncertainty is captured by the 1000 discrete Monte Carlo simulation results. The second is the uncertainty associated with measurement error when observing the pressure in the injection zone. This is captured by an assumed lognormal measurement error function that maps simulated true into simulated measured pressures. Combining these, the likelihood for each simulation is computed as the value of the measurement error probability density function evaluated at the observed value, assuming that the simulated value is true. The overall likelihood function for the full set of simulations (e.g., associated with a given caprock permeability class) is the total likelihood computed for each of the individual (e.g., 1000) simulations. This method is similar to that employed in Bayes Monte Carlo simulation (e.g., Dilks et al., 1992; Sohn et al., 2000; Pinder et al., 2004). The number of simulations allocated to each caprock permeability class is proportional to the prior probability for that class; in this case an equal number of simulations is allocated to each of the four CPC's.

3.3.1 Inferring Caprock Permeability Class

Equal prior probability (= 0.25) is assigned to each of the four caprock permeability classes, as shown in Table 3.3. The pressure buildup in the injection zone (ΔpI) is chosen to be the measured variable for inference of caprock permeability class in

this study. Generally, ΔpI would decrease as the caprock permeability increases. ΔpI ranges theoretically between zero and infinity. For the cases considered and simulated in our analysis, the value of ΔpI ranges from 10^{-4} MPa to 45 MPa. When measurement errors are also considered, observed values of ΔpI beyond these limits are possible.

To estimate the likelihood of observing ΔpI given each caprock permeability class, $f(\Delta pI|CPC_j)$, first, 1000 ΔpI values are computed for each of the caprock permeability classes, incorporating the fixed values in Table 3.1, the uncertain properties in Table 3.2 and a random sample of kc from a uniform distribution within each caprock permeability class (for example kc ~ U(1, 10) for the moderate caprock permeability class). Second, to simulate the effects of measurement error, the injection zone pressure buildups ($\Delta pI\_measured$) are assumed to be lognormally distributed about the true (model simulation) values:

$$\Delta pI\_measured \sim Lognormal(a_I, b_I) \qquad (3.1)$$

$$\text{with} \quad a_I = log(\Delta pI\_simulated)$$

$$b_I = [\log(c.v.^2 + 1)]^{1/2}$$

As indicated, the ΔpI results from each Monte Carlo simulation run specify the median values for the respective lognormal distributions of $\Delta pI\_measured$ (specifying the parameters a_I, which represent the logarithm of the median of the respective measurements), while the second parameter, b_I, corresponds to the standard deviation of the logarithm of $\Delta pI\_measured$. For ease of interpretation, these second parameters are specified by the coefficients of variation of $\Delta pI\_measured$: $c.v.$.

The lognormal distribution for $\Delta pI\_measured$ is thus fully specified by the computed value of ΔpI_i for each simulation i and the assumed error (coefficient of variation) for each pressure measurement (determined by b_I, which is the same for all

simulations). The pdf of the lognormal distribution serves as the likelihood function for the pressure observations given the simulation result:

Likelihood(simulation i) =

$$f(\Delta pI\_measured|a_{Ii}, b_I) = \frac{1}{\Delta pI\_measured * a_{Ii} * \sqrt{2\pi}} \exp\{\frac{-(\ln(\Delta pI\_measured) - a_{Ii})^2}{2{b_I}^2}\}$$

(3.2)

When a subset of the simulations is used to represent a particular case, (i.e., a particular CPC_j), then the case is represented by an "or" operation across the individual simulations in the subset, and the overall likelihood is given by the sum of the individual likelihoods. In the application that follows we utilize 1000 simulations for each case, so that:

Likelihood(case j) $=\sum_{i=1}^{1000} f(\Delta pI\_measured|a_{Ii}, b_I)$ (3.3)

The prior distribution of caprock permeability class and the likelihood function are combined using Bayes rule to calculate the posterior distribution of each caprock permeability class, $\pi(CPC_j|\Delta pI\_measured)$.

$$\pi(CPC_j|\Delta pI\_measured) = \frac{f(\Delta pI\_measured|CPC_j)*(CPC_j)}{\sum_{j=1}^{4} f(\Delta pI\_measured|CPC_j)*(CPC_j)} \quad (3.4)$$

3.3.2 Power Calculation

Leakage detection is assumed to occur whenever a permeable or a highly permeable caprock is inferred to be present from the results of the pressure measurements. This inference is made using a statistical test of hypothesis. The test may be set up to determine whether there is sufficient evidence in the data to very strongly support the argument that the caprock is impermeable (the null hypothesis is that the

caprock is permeable). Alternatively, a traditional approach used for regulatory compliance can be employed in which the caprock is assumed to be impermeable, unless the data provides very strong evidence that the caprock is permeable. While either approach can be implemented with the proposed methodology, here we demonstrate the latter (regulatory compliance) perspective in which the null and alternative hypotheses are given by:

H_o: CPC = Class 1 (Impermeable)

H_1: CPC = Class 2, 3, or 4 (Moderate, High, or Very High Permeability) (3.5)

We employ a Bayesian test of hypothesis in which the P-value for the test is computed as the probability of H_o. This probability is calculated directly in Equation 3.4 as a function of *ΔpI_measured*. If the P-value is low enough, this provides statistically significant evidence that the caprock is permeable, and H_o is rejected. As such, we can identify a critical value of the measured pressure buildup in the injection zone, $\Delta pI\_measured_{crit}$, beyond which the P-value falls below a, the significance level chosen for the test (e.g., a = 0.1, 0.05, or 0.01). The chosen significance level also corresponds to the false positive (Type 1 error) rate for the test of hypothesis. Power is calculated as the probability that H_o is rejected for simulations representing CPC Class 2, 3, or 4 (i.e. caprock permeability kc > 1 mD); for CPC Class 1 (i.e. kc ≤ 1 mD), rejecting H_o represents a false positive (Type 1 error), so the concept of power does not apply.

The probability that Ho is rejected (with $\Delta pI\_measured < \Delta pI\_measured_{crit}$) for a given caprock permeability value, kc_i (kc_i >1 mD) is the expected fraction of measurements below $\Delta pI\_measured_{crit}$, where this expectation is calculated across the 1000 simulations for that caprock permeability value, kc_i:

$$Power(kc_i) = Prob[\Delta pI\_measured < \Delta pI\_measured_{cri} | kc_i] = \frac{1}{1000}\sum_{i=1}^{1000} F_{\Delta pI\_measured}(\Delta pI\_measured_{cri} | a_{Ii}, b_I) \qquad (3.6)$$

where $F_{\Delta pI\_measured}(\Delta pI\_measured_{crit} \mid a_{Ii}, b_I)$ is the lognormal cdf of $\Delta pI\_measured$, evaluated at the value of $\Delta pI\_measured_{crit}$, with lognormal parameters given by a_{Ii} and b_I, as specified in Equation 3.1.

3.3.3 Evaluating the Influence of Input Parameters

The parameters listed in Table 3.1 and 2 serve as a baseline case, representing typical parameter values and a nominal level of uncertainty assumed to apply to a particular reservoir application. To illustrate the sensitivity of model predictions to variations in selected assumptions and parameters, we consider alternative monitoring time t, monitoring locations r, an increased CO_2 injection rate Q, and reductions in the magnitude of the assumed uncertainty of the reservoir properties in Table 3.2 (as might occur as a result of additional site characterization studies), computing how these alternative assumptions affect simulation results and the probability of each caprock permeability class that is inferred from an observation of ΔpI. The baseline monitoring time is at t = 5 year after the injection begins; ΔpI values at t = 1 year and t = 10 year are also simulated to explore the sensitivity of results to monitoring time. The baseline monitoring location is r = 500 m from the injection well, and we also present simulation results for locations at r = 100 m from the injection well. The injection rate is doubled from the baseline case Q = 1 MT-CO_2/yr to 2 MT-CO_2/yr. The uncertainty of the reservoir properties (i.e. the thickness, permeability and porosity of the injection zone and the above zone) is characterized by the variance of each distribution described in Table

3.2. Here, we evaluate the influence of reduced reservoir uncertainty by reducing the variance of each reservoir property (listed in Table 3.2) by half.

3.4 RESULTS

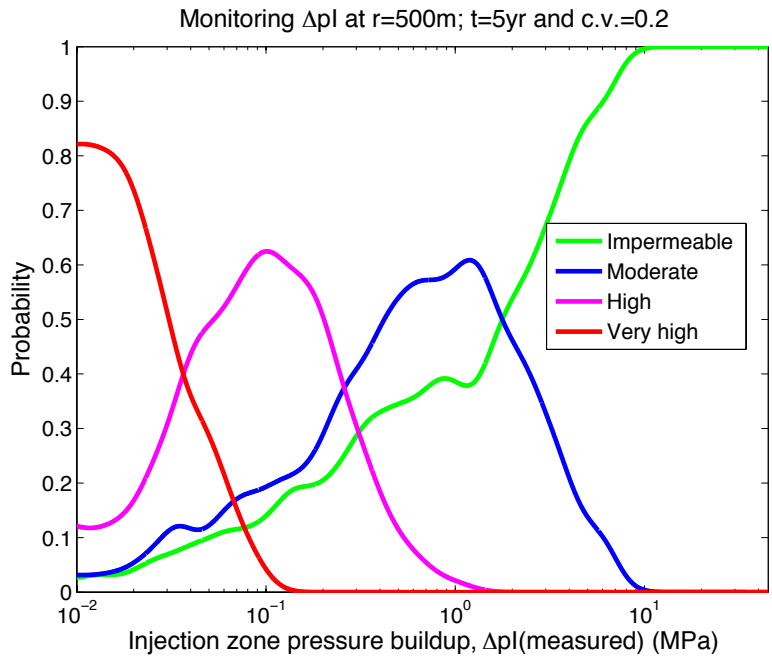
3.4.1 Inferred Caprock Permeability Class

The prior values of ΔpI across the four CPC simulations extend over the range ~ 10^{-4} Mpa– 45 MPa. Since the Bayesian update in Equation 3.4 assigns zero posterior probability wherever the prior probability is zero, this prior range for ΔpI determines the posterior domain as well. As the results below 0.01 MPa in the simulation runs are very inaccurate, the posterior domain of ΔpI is further chosen to be from 0.01 MPa to 45 MPa. Figure 3.3 shows the posterior probability of each CPC as a function of the observed *ΔpI_measured*, for monitoring conducted at a) low measurement error (c.v. = 0.2); b) medium measurement error (c.v. = 0.3); and c) high measurement error (c.v. = 0.5), all at a monitoring location r = 500 m away from the injection well and at t = 5 year.

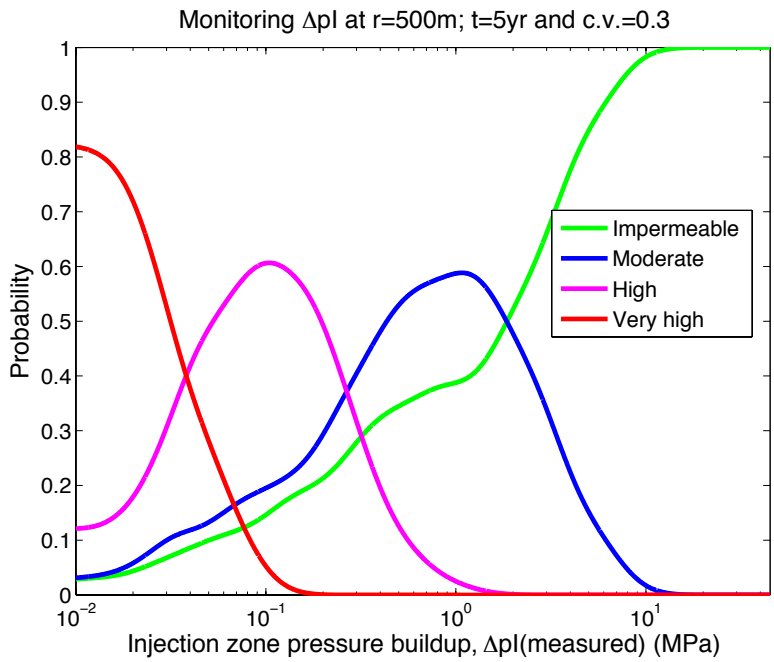
In Figure 3.3, the green line represents the posterior probability that an impermeable caprock is present. Other lines represent the probabilities that a permeable caprock (i.e. a leaky caprock) of varying degree are present. Generally, as the *ΔpI_measured* value decreases, the probability of an impermeable caprock decreases and it is more likely that an increasingly leaky caprock is present. For example, in the low measurement error case, with a *ΔpI_measured* value of 10, the probability of an impermeable caprock is 0.996; while with *ΔpI_measured* = 1, the probability that an impermeable caprock is present is only 0.386 and the probability of a caprock with moderate permeability has reached a value of about 0.6. At low measurement error, the posterior probability curves exhibit fluctuations at some *ΔpI_measured* values. This

occurs in part due to discrete (non-overlapping) ranges of permeability associated with each of the four CPC classes. Nonetheless, the local fluctuation in the predicted posterior probability is small enough that it does not affect the predicted dominant CPC class. As the measurement error increases, the local fluctuations are dampened and the posterior probability curves become both broader and smoother.

a)



b)


Monitoring ΔpI at r=500m; t=5yr and c.v.=0.3
Probability
Injection zone pressure buildup, ΔpI(measured) (MPa)
Impermeable
Moderate
High
Very high

c)

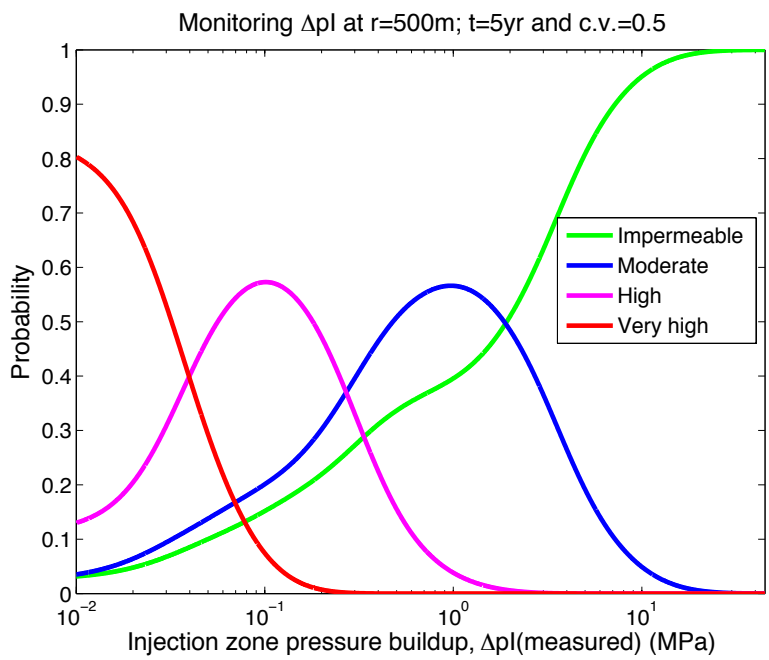

Monitoring ΔpI at r=500m; t=5yr and c.v.=0.5
Probability
Injection zone pressure buildup, ΔpI(measured) (MPa)
Impermeable
Moderate
High
Very high

Figure 3.3 Posterior distribution of caprock permeability class monitored at r = 500 m; t = 5 year and a) low measurement error (c.v. = 0.2); b) medium measurement error (c.v. = 0.3); c) high measurement error (c.v. = 0.5). *ΔpI(measured)* is the measured pressure buildup in the injection zone. Critical values of *ΔpI(measured)* at $\alpha = 0.05$ are computed from the impermeable (green curve) cases as: a) 0.0224 MPa; b) 0.0225 MPa; and c) 0.0211 MPa.

3.4.2 Results of Power Analysis

The power analysis is implemented using the method described in Section 3.3.2. The value of $\Delta pI\_measured_{critical}$ in Equation 3.6 depends on the value of a chosen for the test of hypothesis in Equation 3.5; lower values of a require stronger evidence that H_o is false (i.e., a lower P-value) before it is rejected, dictating a lower value of *ΔpI_measured*. As such, there is a lower probability of rejection of H_o as a decreases, and as a result, less power.

The results of the power calculation for the case with medium measurement error (c.v. = 0.3) and with a = 0.1, 0.05 and 0.01, at the location r = 500 m away from the injection well and at t = 5 year, are shown in Figure 3.4. For a moderately leaky caprock with permeability kc=10 mD and a moderate level of measurement error (c.v. = 0.3), the simulated statistical detection power of the pressure monitoring is still below 0.1 even at the significance level $\alpha = 0.1$. The detection power increases to values of 0.6 as kc reaches approximately 400 mD at $\alpha = 0.05$, with the power further increasing to 0.9 at kc ~ 800 mD. As such, very high caprock permeability is required before pressure monitoring as prescribed in this simulation will yield a high probability of detecting deviations from the baseline, impermeable assumption.

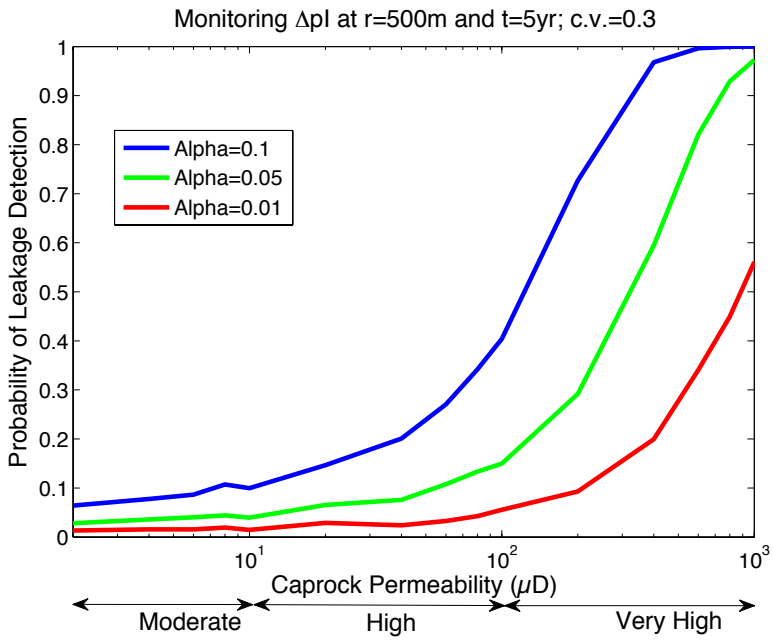


Figure 3.4 Power relationship indicating probability of leakage detection for permeable caprock, (i.e. caprock permeability value, kc > 1 mD) with medium measurement error (c.v. = 0.3), a = 0.1, 0.05 and 0.01 respectively, monitoring at a location r = 500 m away from the injection well, and t = 5 year.

3.4.3 Sensitivity Analysis of Assumed Input Parameters

The following sensitivity analyses were conducted to determine how model inferences are affected by changing the monitoring time (base case t = 5 year), the monitoring location (base case r = 500 m from injection well); the CO_2 injection rate (base case = 1 MT-CO_2/yr); and the initial uncertainty in reservoir properties (base case specified by distributions in Table 3.2):

- Simulating the monitoring of pressures after injection for one year (t = 1 year) and after injection for ten years (t = 10 year);
- Simulating the monitoring of pressures at locations closer to the injection well (r = 100 m);

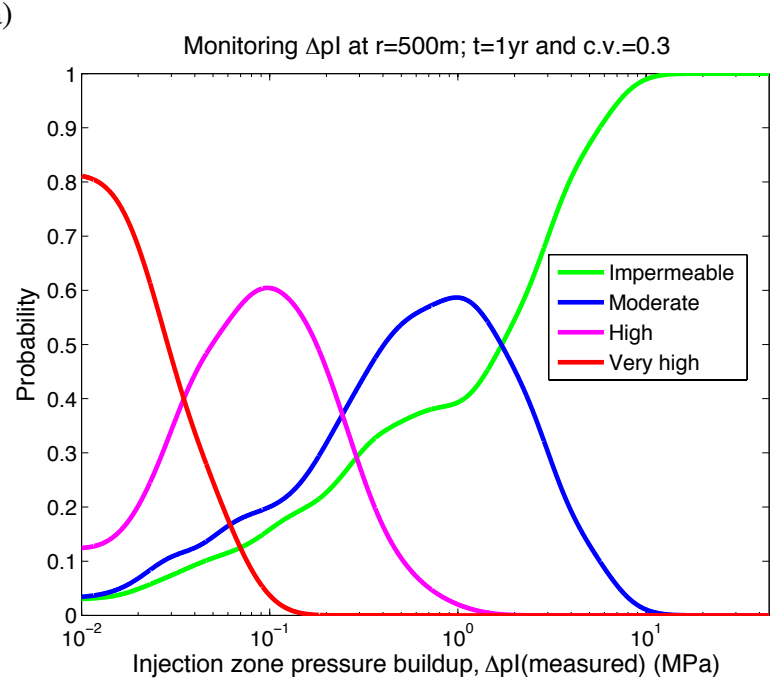
- Doubling the injection rate to 2 MT-CO_2/yr; and
- Reparameterizing the distributions in Table 3.2 so that their variances are reduced by a factor of two.

Figure 3.5 shows the posterior probability of each CPC as a function of the observed *ΔpI_measured*, for monitoring conducted with medium measurement error (c.v. = 0.3) a) at t = 1 year and b) at t = 10 year. Figures 3b, 5a and 5b are very similar, which indicates that beyond the first year of continuous CO_2 injection, the additional system response has little effect on the inferred caprock permeability class. The calculation results at r = 100 m are shown in Figure S2 of Appendix 2B. As expected, for all the cases considered, the lower the value of *ΔpI_measured*, the higher the probability that the caprock is permeable. Furthermore, for a given value of *ΔpI_measured* the predicted most likely (dominant) caprock permeability class remains the same across monitoring locations. However, the location can affect the predicted probability of different caprock permeability classes.

The effects of a doubled injection rate and reduced uncertainty are summarized in Figure 3.6a and b, respectively, for classification of CPC's as a function of *ΔpI_measured*, and in Figure 3.7a and b, respectively, for the statistical power to infer a permeable caprock (kc > 1 mD) at different caprock permeability values. The base case monitoring location (r = 500 m) and medium measurement error are assumed, with pressure measured at t = 5 year. Comparing the curves in Figure 3.6a to those in Figure 3.3b suggests that doubling the injection rate has an effect on the predicted probabilities of different CPC's as a function of *ΔpI_measured*. The green curve for the impermeable CPC (Figure 3.6a) starts to decrease at a higher value of *ΔpI_measured*. Likewise, a doubling of the injection rate yields an increase in the statistical power for inferring a

permeable caprock. This is as would be expected – increasing the injection rate increases the likelihood of leakage and a lower *ΔpI_measured*, yielding a more likely detection of a permeable caprock. Halving the uncertainty in reservoir properties in Table 3.2 yields significant differences in results, with tighter CPC curves exhibited in Figure 3.6b (compared to Figure 3.3b), implying greater resolution of the CPC for a given value of *ΔpI_measured*. As a result, the statistical power to properly detect a permeable caprock increases significantly for kc > 100 mD (see Figure 3.7b).

a)



b)

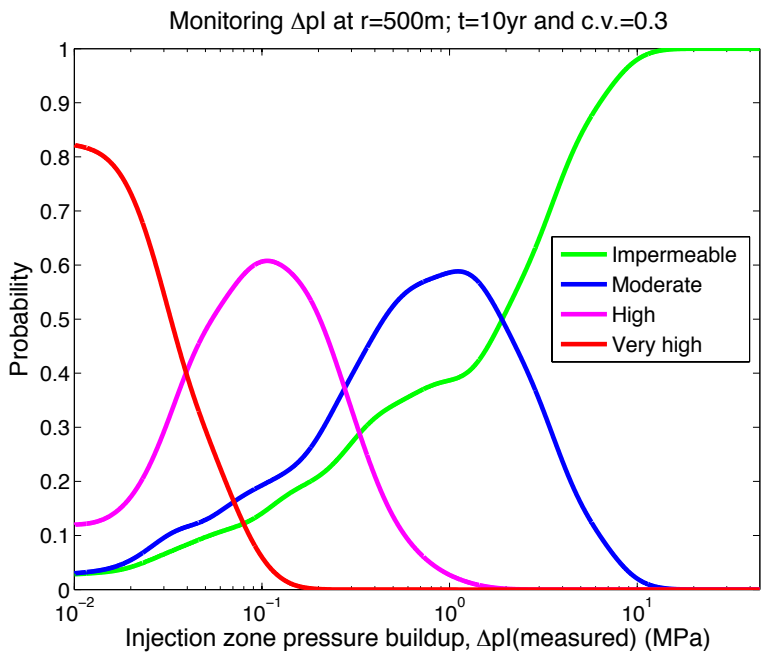
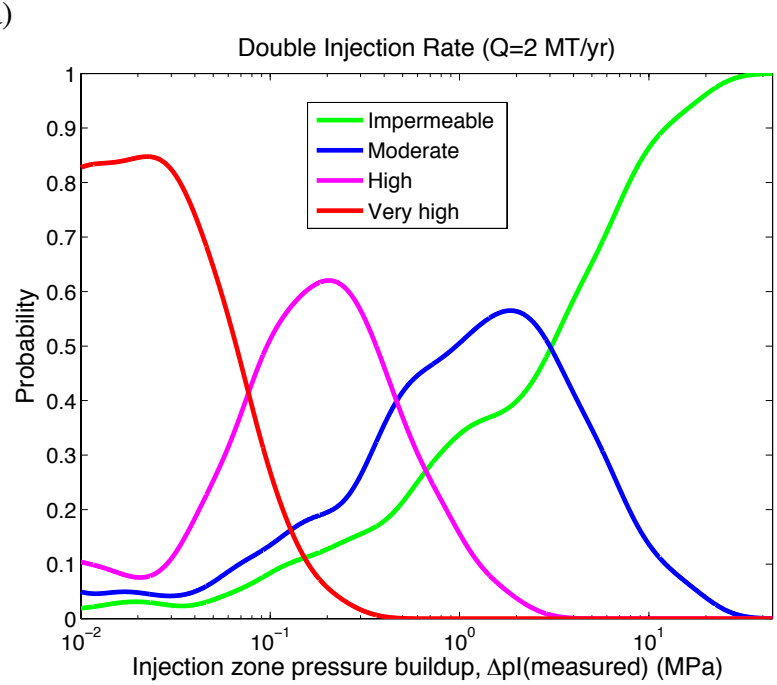


Figure 3.5 Posterior distribution of caprock permeability class monitored with medium measurement error (c.v. =0.3) a) at t = 1 year; b) at t = 10 year. *ΔpI(measured)* is the measured pressure buildup in the injection zone.

a)



b)

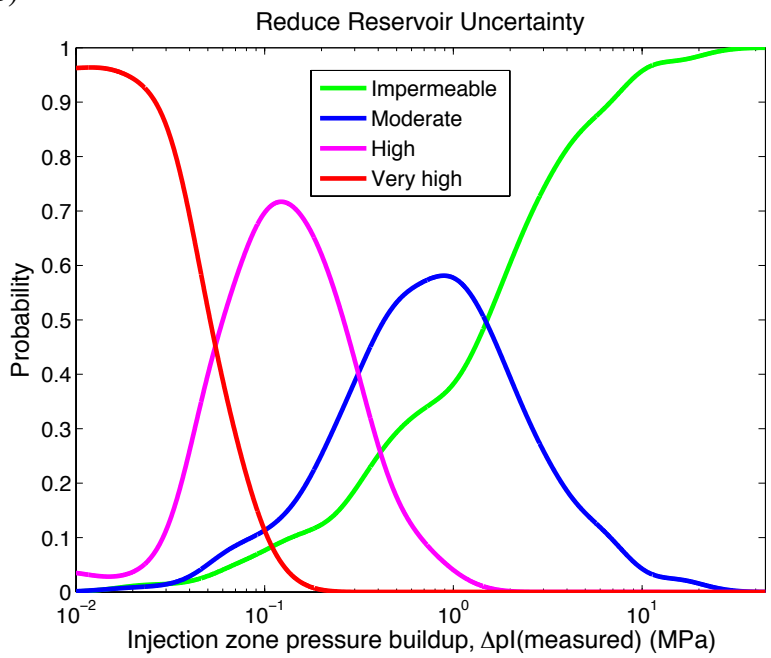
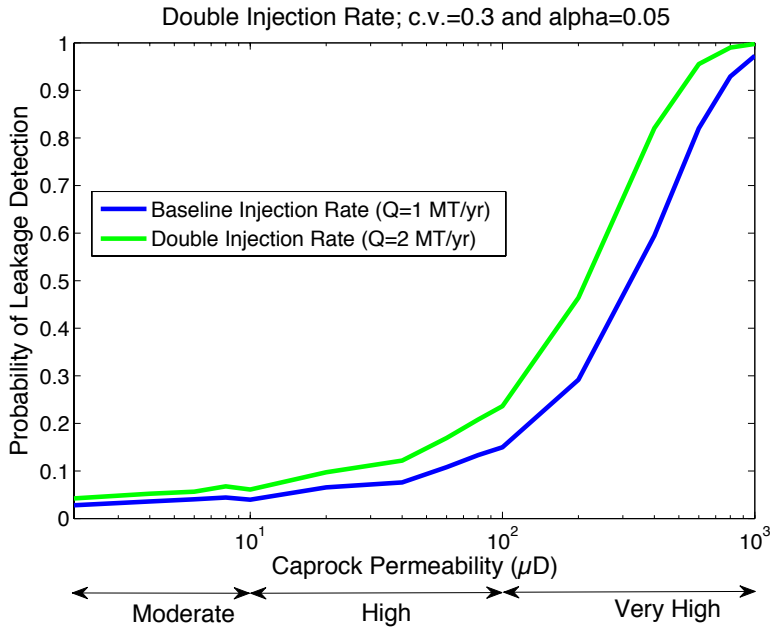


Figure 3.6 Posterior distribution of caprock permeability class at r = 500 m, t = 5 year and medium measurement error (c.v. = 0.3) a) at doubled injection rate (Q = 2 MT-CO_2/yr); b) at reduced reservoir uncertainty. *ΔpI(measured)* is the measured pressure buildup in the injection zone.

a)



b)

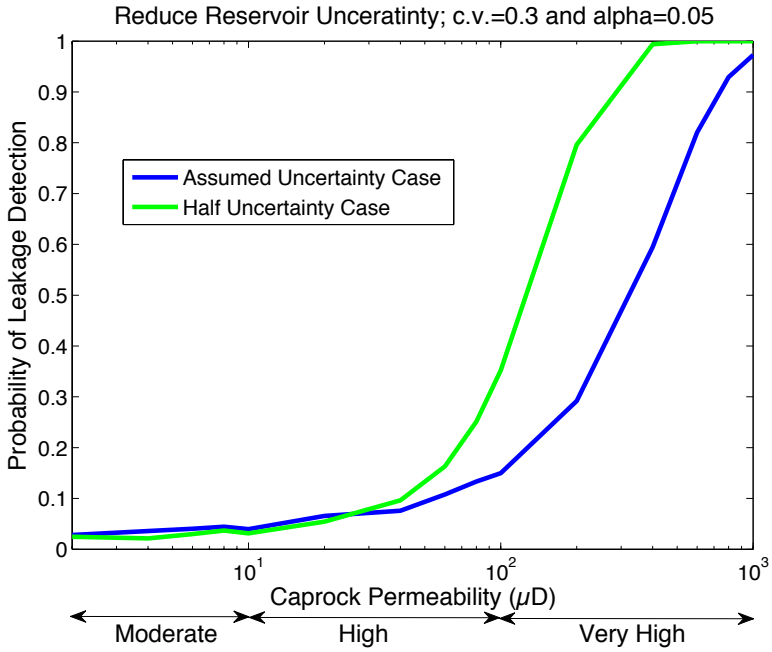


Figure 3.7 Comparison of the detection power for a) doubled injection rate case with baseline injection rate case; b) reduced half reservoir uncertainty case with assumed reservoir uncertainty case at r = 500 m, t = 5 year, a = 0.05, and medium measurement error (c.v. = 0.3).

3.5 DISCUSSION

The principal purpose of this paper is to provide a proof of concept for statistically-based CO_2 leak detection using pressure monitoring supported by reservoir modeling. In addition, we have elucidated the factors that affect the inferences that can be made using the method. Reducing the uncertainty in reservoir properties is shown to allow caprock permeability classification to be made with greater resolution and power. At CO_2 sequestration sites, reservoir properties are usually quantified by site characterization activities (Chadwick et al. 2004; Chiaramonte et al. 2007). In addition to traditional site characterization techniques, including geological mapping, geophysical imaging, core analyses and hydraulic well testing, data collected during the initial CO_2 injection from monitoring techniques, such as crosswell seismic imaging and sampling fluid at an observation well, can be employed to better characterize reservoir properties (Doughty et al. 2007). Combining analysis of the potential for these characterization efforts to reduce uncertainty in site properties with statistical evaluation of ongoing monitoring, such as those presented in this study, can help to better quantify the value of information achieved by better site characterization.

Whether these results indicate that the sole use of pressure monitoring for leak detection at actual sites will be insufficient remains to be seen, though further application of this method with more sophisticated reservoir models and site assumptions should allow this question to be explored with greater insight. Further analysis should also consider the joint application of multiple monitoring methods and tools, since a combination of methods will likely result in improved power. Complementary monitoring techniques, such as seismic monitoring, surface deformation monitoring,

tracer monitoring, and groundwater chemistry monitoring are promising candidates to be employed for this purpose.

3.6 CONCLUSIONS

A Bayesian classification methodology is developed to infer the caprock permeability class, as a surrogate for potential CO_2 leakage at a geologic sequestration site, using the expected distribution of pressure increases in the injection zone for permeable vs. impermeable caprock cases. The detection power of pressure monitoring is evaluated using a Bayesian formulation that considers uncertainty in site conditions, Monte Carlo simulation results for modeled injection zone pressure profiles, and measurement errors for the corresponding pressure measurements. The results indicate that pressure monitoring alone is not sufficient to detect a permeable caprock for the system considered in this study, except for very high caprock permeability values. Time has little impact on the inferred caprock permeability class and the detection power of pressure monitoring. The injection rate has an influence on the predicted caprock permeability class and the detection power of the simulated pressure monitoring. Higher confidence in the predicted caprock permeability class and a significant improvement in the statistical detection power for a leaky caprock can be achieved by reducing the uncertainty in reservoir properties through further site characterization. Other monitoring techniques, such as seismic monitoring, surface deformation monitoring and groundwater chemistry monitoring, may need to be combined with pressure monitoring to conduct CO_2 leakage detection monitoring with high confidence.

Chapter 4: Statistical Performance of CO_2 Leakage Detection using Seismic Travel Time Monitoring

4.1 INTRODUCTION

It is estimated that the atmospheric carbon has increased by 240±10 PgC between 1750 and 2011 (Ciais et al. 2013). Most of the increase is due to emissions from fossil fuel combustion and cement production (Ciais et al. 2013). Geological carbon sequestration in which CO_2 is injected and remains permanently stored in a subsurface formation is considered a promising initial option to slow or reduce global atmospheric CO_2 concentrations (USDOE 2010).

Monitoring for possible CO_2 leakage is an important part of a safe and effective geological sequestration program (USDOE 2009). An overview of available monitoring techniques is provided in the report of U.S. Department of Energy on Monitoring, Verification, and Accounting of CO_2 Stored in Deep Geologic Formations (USDOE 2009). Monitoring techniques can be categorized into atmospheric monitoring methods, near-surface monitoring methods and deep-subsurface monitoring methods based on the depth where these techniques are deployed (Benson & Myer 2002; USDOE 2009). The performance of near-surface monitoring techniques, including soil CO_2 flux measurements and tracer measurements, has been evaluated by Yang et al. using a Bayesian belief network approach (Yang, Mitchell J Small, et al. 2011; Yang, Mitchell J. Small, et al. 2011; Yang et al. 2012). Pressure monitoring, a widely used subsurface monitoring technique, has also been extensively studied (Wang & Small 2014; Azzolina et al. 2014; Birkholzer et al. 2009; Chabora & Benson 2009; Sun, Zeidouni, et al. 2013; Sun, Nicot, et al. 2013). A methodology for detecting CO_2 leakage through abandoned wells was developed using pressure and surface-deformation measurements (Jung et al.

2013). Song et al. (Song et al. 2014) studied the monitoring feasibility of a surface wave seismic method and concluded it to be a challenge to detect abnormalities using this method alone. It is generally found that the individual monitoring techniques studied are unlikely to provide sufficient statistical power for the CO_2 leakage detection task, so that multiple monitoring techniques may need to be utilized together at a site (Yang, Mitchell J Small, et al. 2011; Yang, Mitchell J. Small, et al. 2011; Wang & Small 2014; Azzolina et al. 2014). Our study focuses on the performance and detection power of a particular deep-subsurface method, seismic monitoring above the reservoir caprock.

Seismic monitoring has been commonly applied in oil and gas reservoir characterization and for monitoring changes in reservoir rock properties remotely (Khatiwada et al. 2012). It has also been implemented in several pilot CO_2 sequestration sites, such as the Frio Formation near Houston and the Penn West CO_2-EOR site (Daley et al. 2007; Alshuhail & Lawton 2007; Silver et al. 2007). Different seismic monitoring methods, including time-lapse 2D or 3D surface seismic surveys, borehole vertical seismic profiling and cross-well seismic monitoring, have been employed at sequestration sites (Alshuhail & Lawton 2007; Daley et al. 2007; Silver et al. 2007; Couëslan et al. 2009). The seismic attributes that can be measured include seismic velocity, travel time, amplitude and impedance (Alshuhail & Lawton 2007). In our study, predicted changes in seismic velocity and travel time after a CO_2 leakage event occurs are considered.

Extensive studies have been undertaken to model CO_2 leakage from sequestration sites. The governing equations controlling these processes are the mass and energy balance equations, which can be solved analytically under simplified assumptions or numerically by space and time discretization. Nordbotten et al. (Nordbotten et al. 2009) developed analytical solutions for modeling CO_2 leakage through multiple abandoned

wells. There have also been numerical simulators, such as TOUGH2 and STOMP, developed to model multi-phase flow in the subsurface (Pruess et al. 2012; White et al. 2013). The application of the TOUGH2 simulator in modeling CO_2 leakage has been demonstrated in several studies (Humez et al., 2011; Yang et al., 2011b; Zheng et al., 2013).

At geological sequestration sites, CO_2 is injected into formations usually more than one thousand meters deep. Under the conditions encountered in sequestration reservoirs, CO_2 is in the supercritical fluid phase. To model changes in seismic velocity and seismic travel time after fluid substitution, the Biot-Gassmann equation (Gassmann 1951) is commonly used (Kumar 2006; Khatiwada et al. 2012). Gutierrez et al. (Gutierrez et al. 2013) studied the effects of CO_2 injection on the seismic velocity of sandstone and confirmed that the Biot-Gassmann substitution theory can be used in modeling changes in the P-wave velocity of sandstone containing supercritical CO_2 and saline water when the distribution of the two fluids in the pore space is accounted for in the calculation.

Uncertainty quantification in modeling CO_2 leakage is crucial in regulatory compliance assessment (Isukapalli et al. 1998). The uncertainty can be attributed to natural variability within the physical system; incomplete knowledge of model input parameters caused by insufficient or inaccurate site characterization; and simplifications of the physical model (Isukapalli et al. 1998). Oladyshkin et al. (Oladyshkin et al. 2011) developed a methodology for data-driven uncertainty quantification using the polynomial chaos expansion (PCE) and applied it to modeling CO_2 storage in geological reservoirs. The PCE approach allows uncertainty analysis of complex and computationally intensive models to be conducted with relatively low computation time. In our study, the uncertainty of the calculated seismic travel time is analyzed using the arbitrary

polynomial chaos expansion method (Oladyshkin & Nowak 2012) combined with a traditional uncertainty analysis method, Monte Carlo simulation. The PCE model serves as a stochastic response surface for numerical simulation results from TOUGH2.

The objective of this study is to evaluate the detection power of seismic travel time monitoring at different CO_2 leakage rate levels. A simplified rock physics model is assumed for monitoring zones at sequestration sites. Changes in pore pressure and CO_2 saturation at different leakage levels are estimated using the TOUGH2 numerical simulator. The model results are then approximated by the polynomial chaos expansion (PCE). Monte Carlo uncertainty analysis is performed for seismic travel time using the Biot-Gassmann model and the estimated pressure and CO_2 saturation from the fitted PCE models. A methodology for conducting the power analysis is developed and demonstrated based on the ability to statistically distinguish between P-wave travel time with and without pressure and CO_2 saturation changes modeled to result from CO_2 leakage. The empirical distributions of detection power are obtained at four leakage levels from the Monte Carlo uncertainty analysis.

4.2 Modeling Changes in Seismic Travel Time

4.2.1 Model Setup

The monitoring zone where the seismic source and receiver are located is modeled as a homogenous layer at a depth of 1 km below the ground surface. The geometry of the simplified rock physics model for the monitoring zone is illustrated in Figure 4.1. The lithology type is sandstone with a density of 2600 kg/m^3. The lateral extent of the model domain is 10 km and the thickness of the monitoring zone is 50 m. The model domain is discretized with a mesh of *80×5 (400)* grids. The mesh is refined

around the leakage point with a grid width of 50 m. The grid width 1 km farther away from the leakage point is 200 m. The CO_2 leakage point is at the center bottom of the monitoring zone. The temperature is constant at 45 °C and the gradient of initial hydrostatic pressure is 0.1 bar/m. The initial CO_2 saturation in the pore space is assumed to be zero. The monitoring zone is initially filled with brine with a salinity of 0.15 in weight fraction. The point source of the seismic signal is located at the top of the monitoring zone and the receiver is directly below the source at the bottom of the monitoring zone. The source and the receiver thus form a vertical line perpendicular to the bottom. The signal recorded is the direct arrival from top to bottom of the monitoring zone. The horizontal distance between the receiver and the leakage point is x, which is assumed to be either 0 or 500 m, respectively in this study. The permeability, porosity and initial pressure at the top of the monitoring zone are uncertain input parameters with empirical distributions defined in section 4.3.1. The symbols and values for parameters of the modeled monitoring zone are summarized in Table 4.1.

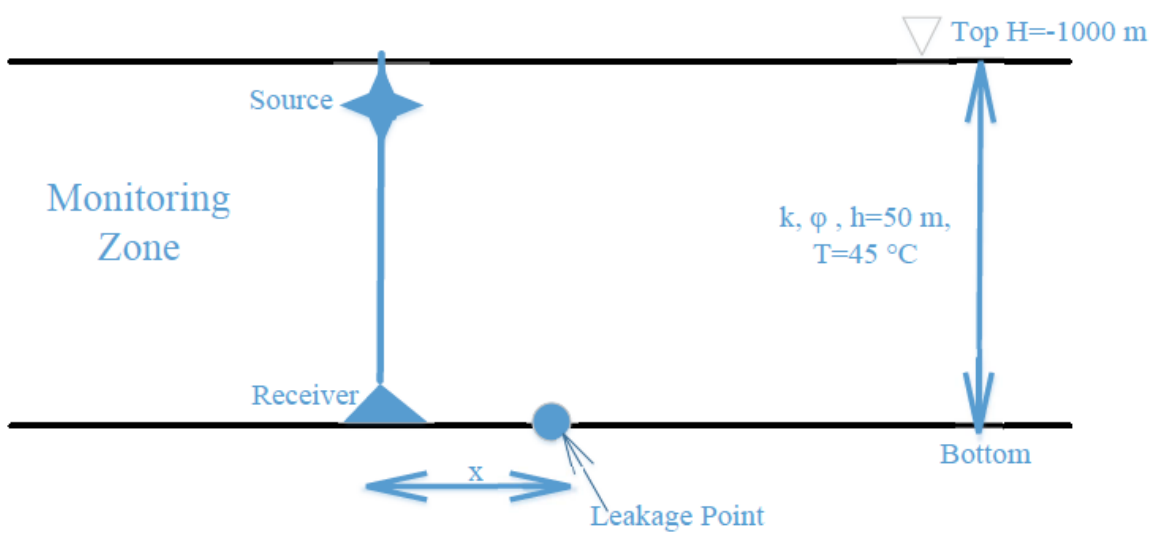


Figure 4.1 Geometry of the simplified rock physics model for the monitoring zone.

Table 4.1 Symbols and values for parameters of the modeled monitoring zone.

Parameter	Symbol	Value
Depth below the ground surface (m)	H	1000
Thickness (m)	h	50
Temperature (°C)	T	45
Initial pressure (MPa)	P_initial	Uncertain
Permeability (mD)	k	Uncertain
Porosity (%)	ϕ	Uncertain
Initial CO_2 saturation	Sg_initial	0
Density of rock (kg/m^3)	ϱ	2600
Salinity of brine (weight fraction)	S	0.15
CO_2 leakage rate (tonnes/year)	LR	10, 50, 100 and 150
Distance of monitor from leakage point (m)	x	0 or 500

4.2.2 Calculation of Seismic Velocity and Travel Time

Based on the laboratory study of rock samples conducted by Wang et al. (Wang et al. 1998), changes in seismic velocity are mainly due to changes in pressure and CO_2 saturation in the pore space. The pressure and CO_2 saturation for the no leakage case and the four leakage rate levels (i.e. small, moderate, large and very large) are estimated using the TOUGH2/ECO2N module (Pruess 2005). The CO_2 leakage rates that correspond to the four leakage levels are 10 tonnes/year, 50 tonnes/tear, 100 tonnes/year and 150 tonnes/year. The small leakage level (i.e. 10 tonnes/year) represents a scenario where 1% of injected CO_2 is leaking from a storage reservoir containing 10^5 tonnes of CO_2 over 100 years.

Seismic velocity can be estimated using the moduli and density of the rock:

$$V_p = \sqrt{\frac{K_{sat} + \frac{4}{3}\mu}{\rho_{sat}}} \qquad (4.1)$$

$$V_s = \sqrt{\frac{\mu}{\rho_{sat}}} \qquad (4.2)$$

where V_p and V_s are the P-wave and S-wave velocity; K_{sat} and μ are the bulk and shear moduli of the rock; and $\boldsymbol{\rho_{sat}}$ is the density of the rock.

The bulk modulus of a saturated rock is computed using the Gassmann equation (Gassmann, 1951):

$$K_{sat} = K_{frame} + \frac{\left(1 - \frac{K_{frame}}{K_{matrix}}\right)^2}{\frac{\varphi}{K_{fl}} + \frac{1-\varphi}{K_{matrix}} - \frac{K_{frame}}{K_{matrix}^2}} \qquad (4.3)$$

where $K_{sat}, K_{frame}, K_{matrix}$ and K_{fl} are the bulk moduli of the saturated rock, porous rock frame, mineral matrix and pore fluid, respectively. φ is the porosity of the rock as a fraction. The shear modulus is held constant during the fluid substitution in the Gassmann theory. The bulk and shear moduli of the rock frame are estimated using the Hertz-Mindlin contact theory for consolidated rock and the Hashin-Shtrikman bounds (Mavko et al. 2009; Carcione et al. 2006). The equations are included in Appendix 2. The density of the rock is calculated with the volume averaging equation:

$$\rho_{sat} = \varphi\rho_{fl} + (1-\varphi)\rho_{matrix} \qquad (4.4)$$

where ρ_{fl} and ρ_{matrix} are the density of the fluid phase and the density of the mineral matrix, respectively.

For a formulation and Matlab codes that implement the Gassmann fluid substitution theory, the readers are referred to Kumar (2006).

Once the gridded fields of seismic velocities are determined, the direct travel time along the vertical wave path from top to bottom is computed as:

$$t = \int_0^h \frac{1}{v}\, dz \qquad (4.5)$$

where v is the seismic velocity and h is the thickness of the monitoring zone.

4.3 UNCERTAINTY QUANTIFICATION AND POWER ANALYSIS

The uncertainty of the seismic travel time is estimated using the Monte Carlo (MC) method. As the computation time of the TOUGH2 simulation is too large to directly implement the MC method with a sufficient sample size, a stochastic response surface using the polynomial chaos expansion is fitted to the TOUGH2 outputs (i.e. pressure and CO_2 saturation), allowing more rapid MC evaluation of this surrogate model.

4.3.1 Stochastic Response Surface Method

The stochastic inputs in the uncertainty analysis are permeability, porosity and initial pressure at the top of the monitoring zone. The empirical distributions of permeability and porosity are obtained from the U.S. National Petroleum Council (NPC) database, which contains the raw data for over one thousand reservoirs (NPC 1984). The initial pressure is assumed to be independent of the permeability and porosity pairs. Scatter plots of permeability vs. initial pressure data and porosity vs. initial pressure data in the NPC database are provided in Appendix 3. Although the NPC database also contains initial pressure data for reservoirs, they are at various depth and lithology types. In our model setting, the top of the monitoring zone is fixed at a depth of 1 km and the lithology type of the monitoring zone is sandstone. Therefore, a synthetic dataset for initial pressure at a depth of 1 km and sandstone reservoirs is generated based on the initial pressure data in the NPC database. The distribution of initial pressure is assumed to be a truncated normal with a mean of 9.68 MPa and a standard deviation of 3.94 MPa before the truncation. The lower and upper limits of the truncated normal distribution are 1.93 MPa and 17.42 MPa, respectively. The procedure for generating the synthetic dataset for initial pressure from the truncated normal distribution is shown in Appendix 3. Figure 4.2 illustrates the empirical distributions of the three uncertain input parameters.

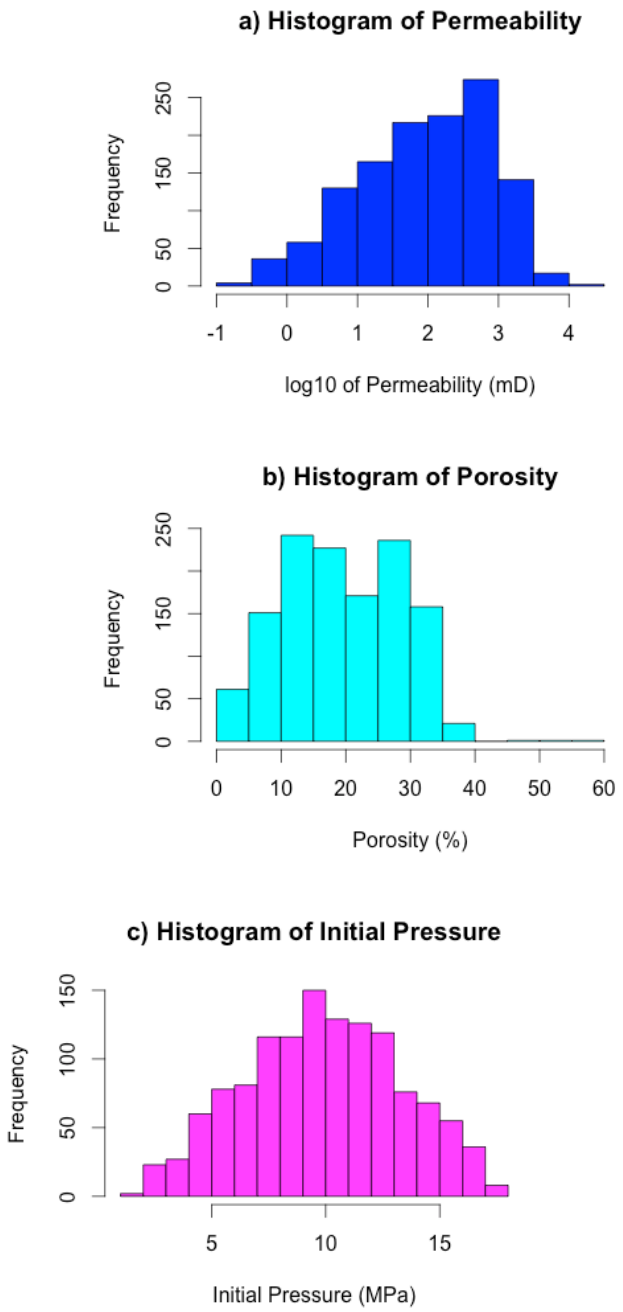


Figure 4.2 Histograms of stochastic inputs: a) Permeability; b) Porosity; and c) Initial pressure.

The pressure and CO_2 saturation outputs from the TOUGH2 simulation at each leakage level are approximated by the polynomial chaos expansion (PCE) on the set of uncertain input parameters $\{\omega_i\}_{i=1}^{n}$ given by:

$$y = a_0 + \sum_{i1=1}^{n} a_{i1}\,\Gamma_1(\omega_{i1}) + \sum_{i1=1}^{n}\sum_{i2=1}^{i1} a_{i1i2}\,\Gamma_2(\omega_{i1,}\omega_{i2}) + \sum_{i1=1}^{n}\sum_{i2=1}^{i1}\sum_{i3=1}^{i2} a_{i1i2i3}\,\Gamma_3(\omega_{i1,}\omega_{i2},\omega_{i3}) + \cdots \quad (4.6)$$

where y is an output (either pressure or CO_2 saturation) of the TOUGH2 model; $a_{i1\ldots}$are coefficients to be estimated; and the $\Gamma_p(\omega_{i1,},\ldots,\omega_{ip})$ are multidimensional polynomial basis functions that are orthogonal in the input space.

The multidimensional polynomial basis functions $\Gamma_p(\omega_{i1,},\ldots,\omega_{ip})$ are constructed using the arbitrary polynomial chaos expansion (aPC) approach of Oladyshkin and Nowak (2012). The set of aPC basis functions is constructed from the moments of uncertain input variables. The equations used to construct the aPC basis are provided in Appendix 3. The aPC approach allows the use of empirical datasets to represent input uncertainties, which avoids errors introduced by fitting standard parametric statistical distributions to empirical datasets. The coefficients $(a_{i1,\ldots})$ of the polynomial basis can be estimated by the probabilistic collocation method (Tatang et al. 1997; Li & Zhang 2007) or a statistical regression method (Isukapalli et al. 1998; Zhang & Sahinidis 2013).

In this study, a third-order polynomial expansion function is chosen to approximate the TOUGH2 outputs. As there are three uncertain input parameters (n=3) and the order of the expansion is three (d=3), the total number of terms (M) in the polynomial expansion function is 20, according to the combinatory formula:

$$M = \frac{(n+d)!}{n!\,d!} \quad (4.7)$$

The coefficients of the polynomial basis are estimated using the probabilistic collocation method (Tatang et al., 1997), which requires evaluating the TOUGH2 model with M different sets of uncertain input parameters $\{\omega_i\}_{i=1}^{n}$ that are called collocation points. The collocation points are selected from the high probability roots of the (d+1)th (i.e. fourth in this study) order orthogonal polynomial for each uncertain input parameter. At each leakage level, 20 runs of the TOUGH2 model are needed to solve the 20 coefficients in the polynomial expansion function. The PCE models are constructed at each element in the model domain for each output variable (i.e. pressure and CO_2 saturation), resulting in a total of $(80\times5)\times2 = 800$ polynomial functions at each specific time and leakage level.

4.3.2 Monte Carlo Uncertainty Analysis

Once the PCE models are constructed, the seismic velocity at each element in the model domain can be calculated at a relatively low computational cost, which facilitates Monte Carlo uncertainty analysis. For each set of uncertain input parameters introduced in section 4.3.1, combined with the fixed parameters in Table 4.1, the direct seismic travel time from the top to bottom of the formation at locations x=0 m and 500 m are calculated for the no leakage case and the four leakage levels, at t=5, 10 and 50 years, respectively. The empirical distributions of changes in seismic travel time (Dt) at four leakage levels are obtained by subtracting the travel time for the no leakage case (T_0) from the travel time at each leakage level (T_i, i=1,4).

$$Dt = T_i - T_0 \quad (4.8)$$

where i = 1 for small leak, 2 for moderate leak, 3 for large leak and 4 for very large leak

4.3.3 Power Analysis

To assess the power of CO_2 leakage detection using seismic travel time, a statistical hypothesis test is assumed to be performed on measured changes in travel time (Dt_m) at each leakage level. The null and alternative hypotheses are given by:

$$H_0: Dt_m \leq 0 \ (no\ leak)$$
$$H_1: Dt_m > 0 \ (a\ leakage\ event\ occurs) \qquad (4.9)$$

As travel time changes of 0.2 to 1.0 ms can be observed using a continuous active-source seismic monitoring technique, (Daley et al. 2007), the standard deviations of measurement errors of the seismic travel times (T_0 and T_i) are assumed to be 0.2 ms in the hypothesis test, corresponding to a coefficient of variation of 0.2 to 1 in the study by Daley et al.. The distribution of measured seismic travel time is thus assumed to be normal with a mean of the simulated value from the Monte Carlo uncertainty analysis and a standard deviation of 0.2 ms, which is the assumed standard deviation of measurement error (σ_e). Therefore, the measured changes in seismic travel time (Dt_m) at each leakage level also follow the normal distribution:

$$T_{i,m} \sim N(T_{i,s}, \sigma_e)$$
$$T_{0,m} \sim N(T_{0,s}, \sigma_e)$$
$$Dt_m = \left(T_{i,m} - T_{0,m}\right) \sim N(\mu, \sigma) \qquad (4.10)$$

where: $\mu = T_{i,s} - T_{0,s}$

$\sigma = \sqrt{2} \times \sigma_e$

The hypothesis test (Equation 4.9) tests whether the mean of the measured changes in travel time (Dt_m) is statistically significantly different from zero. If it is, the null hypothesis (i.e. the no leak case) is rejected and we conclude that the measured travel times provide sufficient evidence of a leakage event. The critical value of Dt_m for the hypothesis test is chosen to be $Dt_{cri}=2\sigma$, corresponding to a significance level of about 0.02 for the test of hypothesis. The power of the hypothesis test for CO_2 leakage detection is computed as the probability that Dt_m is larger than Dt_{cri} when a leakage event occurs (at

each of the four leakage levels). Note that a group of measurements of changes in seismic travel time is assumed to be made, after which the mean of the observed values will be compared to Dt_{cri}. However, since these measurements (and their errors with respect to leakage effects and inference) are expected to be highly correlated, the effective sample size is conservatively assumed to be one.

4.4 Results and Discussion

4.4.1 Predicted Seismic Velocity in Model Domain

Seismic velocities, including the P-wave velocity and the S-wave velocity, at each element in the model domain are calculated using the predicted pressure and CO_2 saturation values from the PCE models, combined with the Biot-Gassmann model for fluid substitution introduced in section 4.2.2. Figure 4.3 shows the model results for the no leak case and the moderate leakage level at t=10 year. In this figure, the values of the three uncertain input parameters are as follows: permeability = 100 mD, porosity = 15% and initial pressure = 9.5 MPa. The results at t=1, 5, and 50 year and at different leakage levels are shown in Appendix 3.

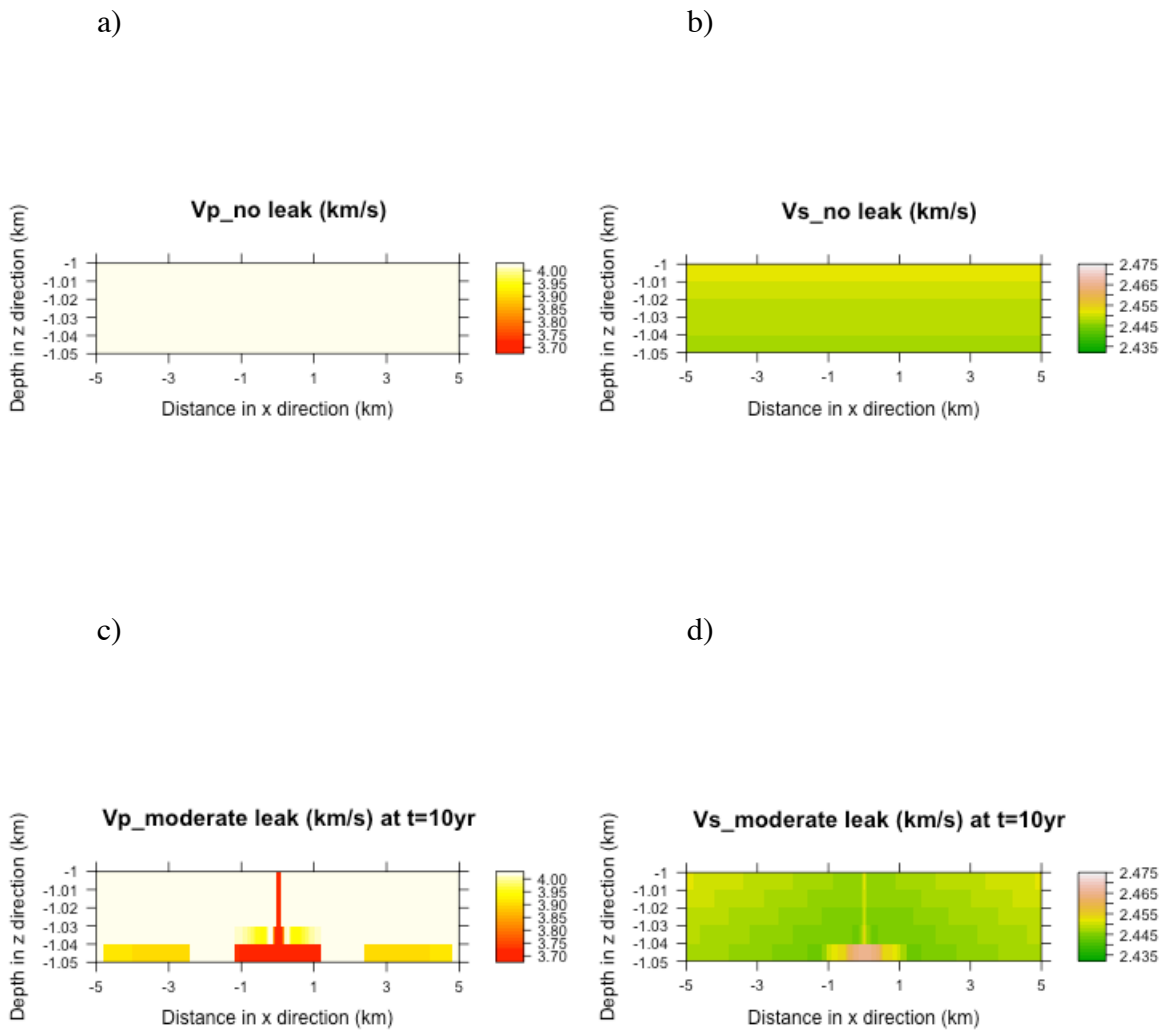


Figure 4.3 Sample model results of P-wave (a and c) and S-wave (b and d) seismic velocities for the no leak case (a and b) and the moderate leakage level (c and d) at t=10 year (*k=100 mD, φ=15%, P_initial=9.5 MPa)*.

The velocity of the P-wave decreases by about 7.5% around the leakage point at the moderate leakage level after 10 years of leakage. The effect of the decreased P-wave

velocity propagates upwards in the center towards the top of the monitoring zone. For the areas at the bottom of the monitoring zone but far away from the center, the P-wave velocity decreases by about 3.7%. Discontinuities in the P-wave velocity occur at distances beyond 1000 m from the point of leakage; however these do not influence the estimates that follow, as the estimates are made for monitors located at x=0 or x=500 m from the leakage point. In general, as the leakage level and time increases, the drop in the P-wave velocity gets larger and affects larger areas in the monitoring zone. The decrease in P-wave velocity is a combined effect of pore pressure and CO_2 saturation increase after CO_2 leakage occurs in the monitoring zone.

The S-wave velocity changes little, increasing slightly around the leakage point by about 0.8% at the moderate leakage level after 10 years of leakage. The maximum increase in S-wave velocity at all the leakage levels considered in this study is about 1% (at the very large leakage level). The effect of CO_2 saturation on the changes in S-wave velocity is caused mainly by the change in the density of the pore fluid. The shear modulus of the rock is not affected by CO_2 saturation because the shear modulus of fluids is zero. As the changes in S-wave velocities are too small to serve as a basis for leakage detection, the S-wave travel time is excluded from further analysis.

4.4.2 Detection Power using P-wave Travel Time

Monte Carlo uncertainty analysis is performed for the P-wave travel time along a vertical wave path from the top to the bottom of the formation (Figure 4.1) at locations x=0 and 500 m respectively, from the point of leakage. The input parameter space is divided into four regions based on permeability and porosity ranges, as in practice, the permeability and porosity ranges of the monitoring zone could be known in advance by site characterization. Four primary groups are identified in the input permeability-

porosity space from the NPC database, denoted by different colors in Figure 4.4. In addition, initial pore pressures are sampled independently for each case from the assumed truncated normal distribution shown in Figure 4.2c. The group of points in red represents inputs with low porosity. The groups of points in green, orange and purple represent inputs with relatively high porosity, and with low, moderate and high permeability, respectively. Figure5-7 show the resulting computed empirical cumulative distribution function (CDF) of detection power using the P-wave travel time at the location x=0 m (center of the monitoring zone) for the four regions in the input permeability-porosity space at t = 5, 10 and 50 year, respectively. The results for the location x=500 m are shown in Appendix 3.

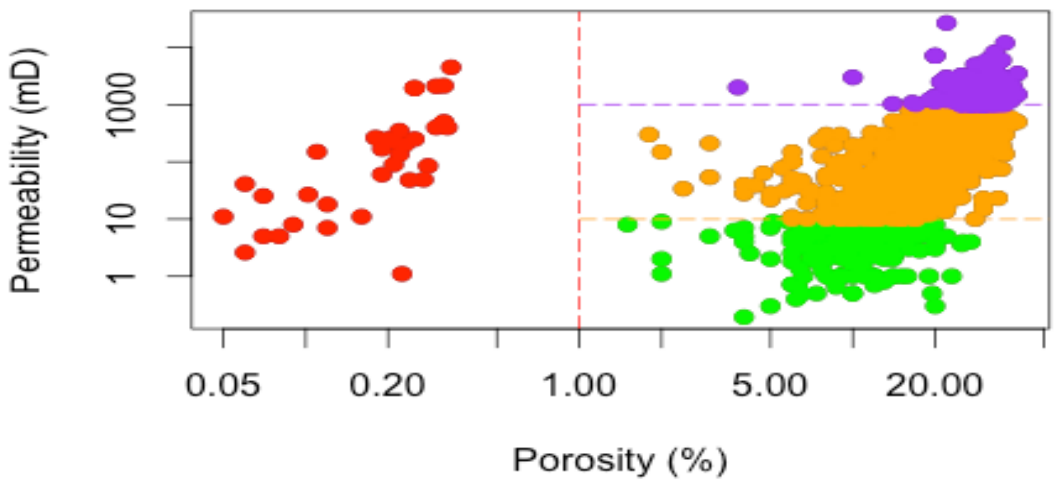


Figure 4.4 Scatter plot of permeability vs. porosity input parameters used in the Monte Carlo simulation. Each point represents an observation in the U.S. National Petroleum Council (NPC) database. The observations are classified into four regions of the parameter space as indicated by color.

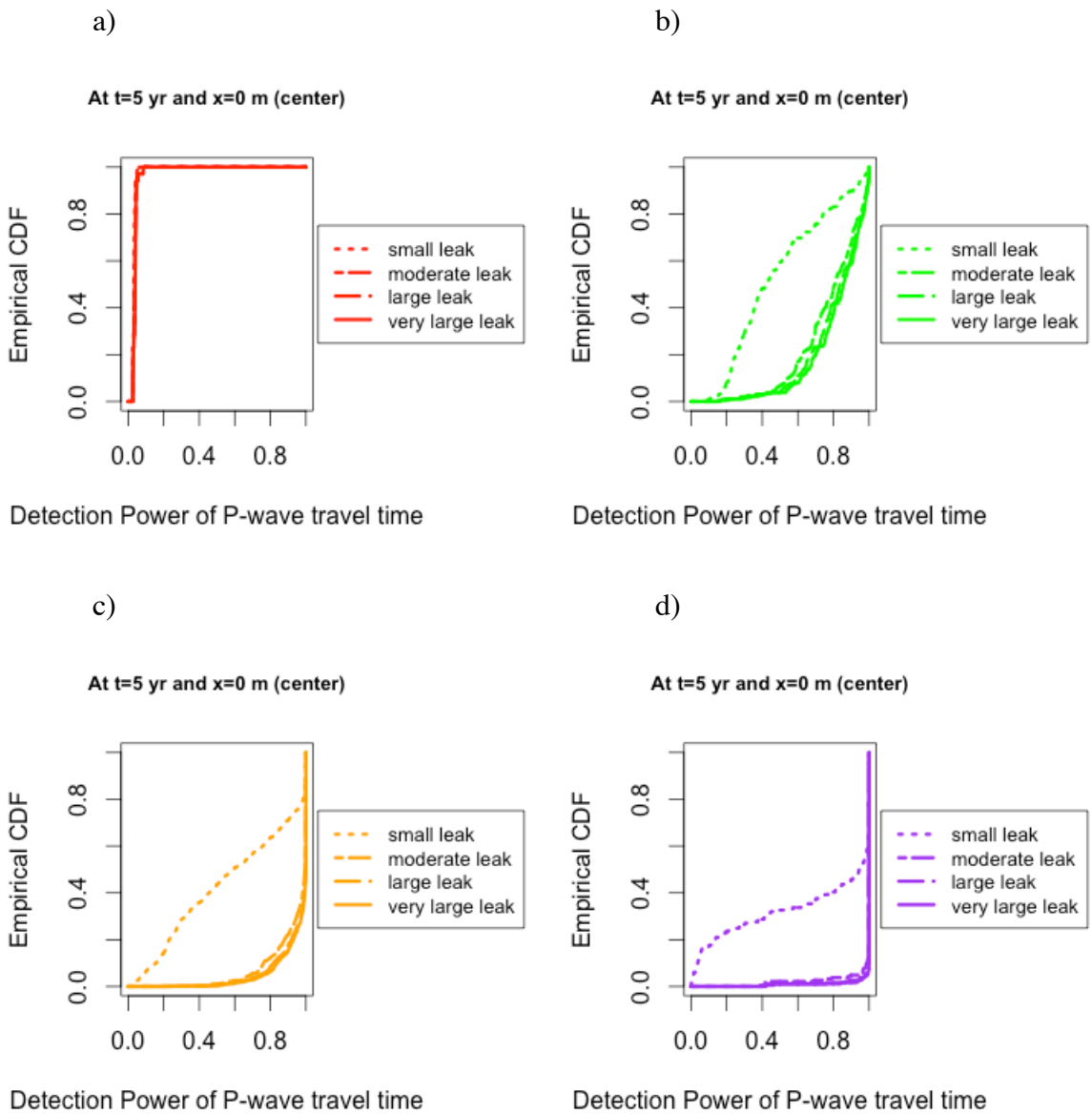


Figure 4.5 Empirical CDF of detection power of P-wave travel time test of hypothesis (Equation 4.9, at significance level $\alpha=0.02$) at four leakage levels, t=5 year, x=0 m and at a) low porosity region (red points in Figure 4.4); b) high porosity and low permeability region (green points in Figure 4.4); c) high porosity and moderate permeability region (orange points in Figure 4.4) and d) high porosity and high permeability region (purple points in Figure 4.4).

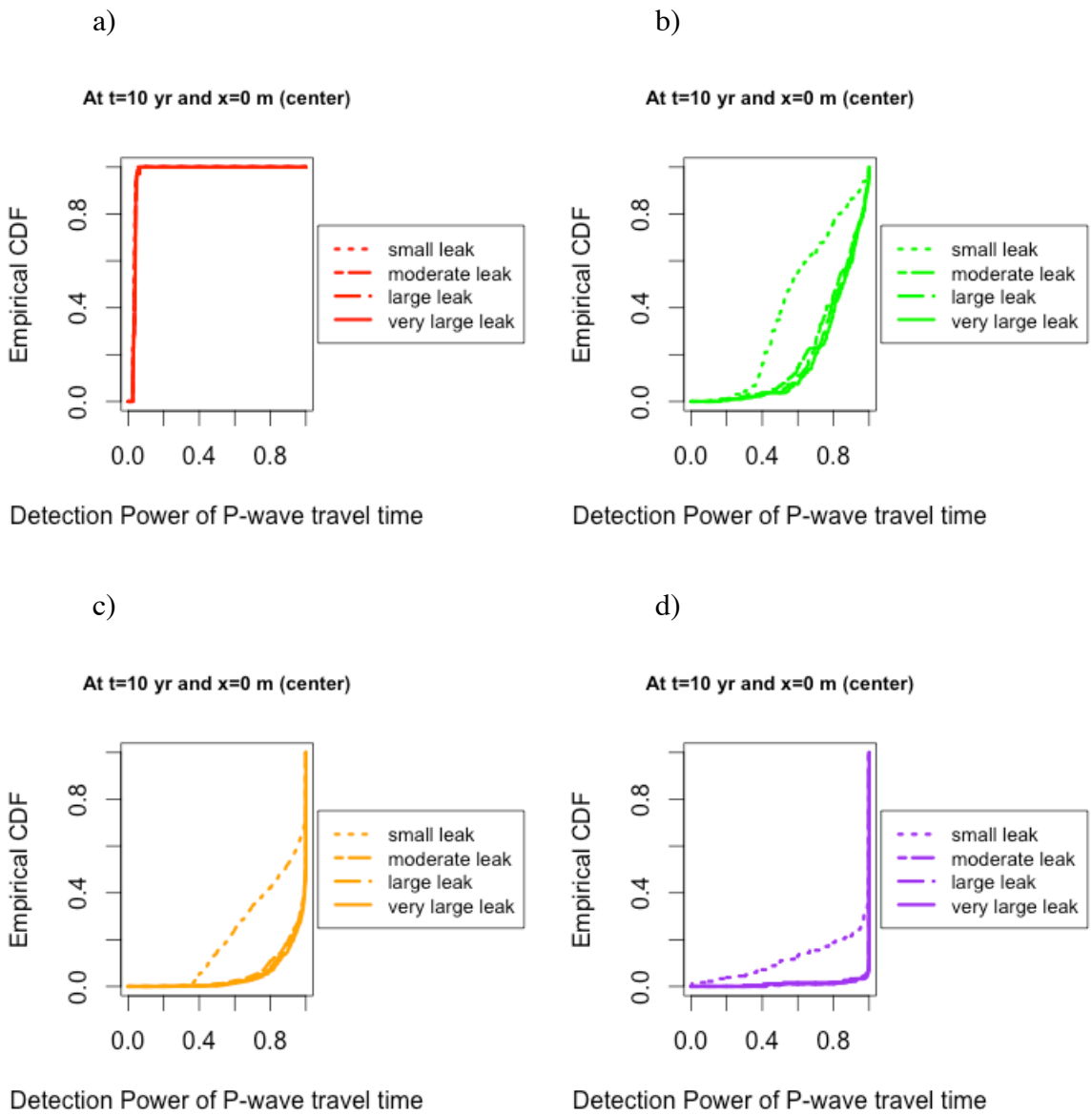


Figure 4.6 Empirical CDF of detection power of P-wave travel time test of hypothesis (Equation 4.9, at significance level $\alpha=0.02$) at four leakage levels, t=10 year, x=0 m and at a) low porosity region (red points in Figure 4.4); b) high porosity and low permeability region (green points in Figure 4.4); c) high porosity and moderate permeability region (orange points in Figure 4.4) and d) high porosity and high permeability region (purple points in Figure 4.4).

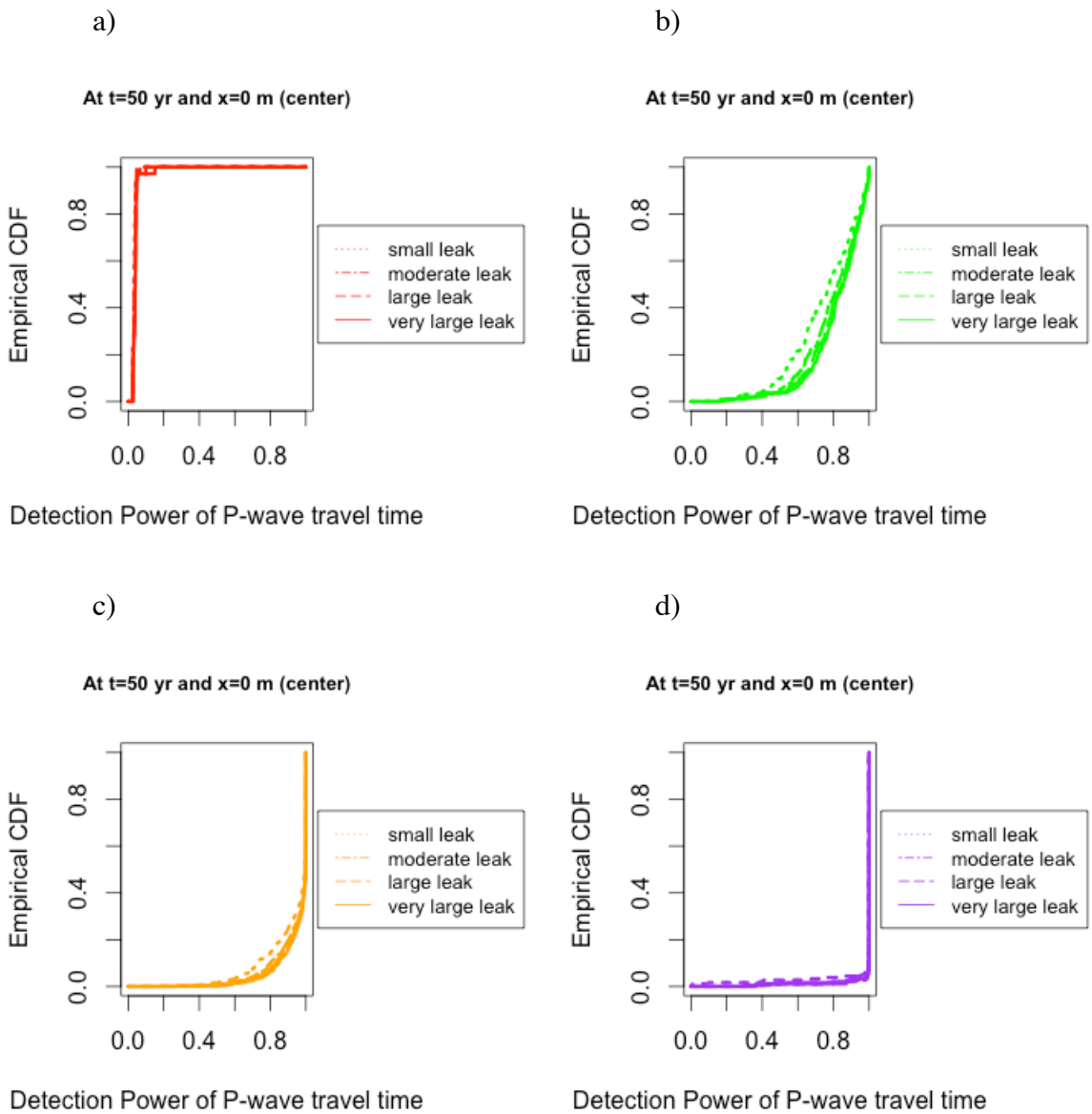


Figure 4.7 Empirical CDF of detection power of P-wave travel time test of hypothesis (Equation 4.9, at significance level $\alpha=0.02$) at four leakage levels, t=50 year, x=0 m and at a) low porosity region (red points in Figure 4.4); b) high porosity and low permeability region (green points in Figure 4.4); c) high porosity and moderate permeability region (orange points in Figure 4.4) and d) high porosity and high permeability region (purple points in Figure 4.4).

It is evident that the detection power of the test of hypothesis for decreases in the P-wave travel time differs for different regions of the permeability and porosity input space. The P-wave travel time test exhibits good power for CO_2 leakage detection if the permeability and porosity of the monitoring zone are relatively high (purple curves in Figure 4.5-7d). The detection power using the P-wave travel time test is very low for monitoring zones with low porosity, no matter the value of the permeability value (red curves in Figure 4.5-7a). For monitoring zones with porosity larger than 1% (green, orange, and purple regions in Figure 4.4), the detection power using the P-wave travel time test is generally higher for a high permeability formation than for a low permeability layer. The detection power increases with increasing time since the leakage occurs. For example, comparing Figure 4.5c to Figure 4.7c, the probability of achieving a detection power larger than 0.8 at the small leakage level increases from 0.4 (at t=5 year) to about 0.82 (at t=50 year) if the permeability and porosity ranges of the monitoring zone fall into the orange region in Figure 4.4. As shown in Figure 4.7d, high detection power using the P-wave travel time test can be achieved for monitoring zones with high permeability at t=50 year. Comparing the results at x=0 m to those at x=500 m, it is seen that the detection power is higher at the location closer to the leakage point (x=0 m) than at the location farther away (x=500 m). The detection power does not change much for different CO_2 leakage levels at the location x=0 m. However, at the location x=500 m, the detection power more notably increases as the CO_2 leakage level increases (see Appendix 3).

4.5 CONCLUSIONS

In this work, the power of CO_2 leakage detection using a statistical analysis and test of P-wave travel times is assessed with a simplified rock physics model for the

monitoring zone. The empirical distributions of detection power using the P-wave travel time for four regions in the permeability-porosity input space at four leakage levels are obtained from the Monte Carlo uncertainty analysis with a stochastic response surface method. The model results predict that the P-wave velocity decreases by about 7.5% around the leakage point at the moderate leakage level after 10 years of leakage. The detection power using the P-wave travel time measurements and test alone is generally not high enough for small leakage events, unless the porosity and the permeability of the monitoring zone are high, and/or a long period of time has elapsed since the leakage occurred. The results indicate that there is an advantage to choosing a monitoring zone with high porosity and high permeability for the purpose of monitoring CO_2 leakage with seismic wave velocity detection. The detection power is higher when the leak occurs near the location of the monitoring system, suggesting the possible need for multiple monitoring locations when the reservoir and the monitoring zone are large. As the likely pathways for concentrated CO_2 leakage at sequestration sites include abandoned wells and faults, the seismic monitoring equipment may be most effectively deployed along the abandoned wells or above the faults (if any are present). As the sole measurement of the P-wave travel time cannot provide sufficient power for CO_2 leakage detection in many cases, measurements from other monitoring techniques, such as pressure monitoring and near-surface monitoring will likely need to be combined at sequestration sites to increase the probability that leakage events are detected and addressed in a timely manner.

Chapter 5: Conclusions

5.1 Predicting CO_2 Solubility in Brine

The accuracy of CO_2 solubility prediction at the conditions of carbon storage reservoirs can be improved by a multimodel predictive system (MMoPS), which utilizes nine individual models in an effective manner.

The generated classification tree indicates that some sub-models perform well for a specific region of T-P-X space, but cannot provide good estimation for other regions.

Three CO_2 solubility measurements with high prediction errors in the MMoPS model are identified. Uncertainties in CO_2 experimental measurements also transfer to the uncertainties in the estimated performance of CO_2 solubility predictive models. More CO_2 solubility experimental measurements are necessary to verify the reliability of different experimental studies. The lack of experimental data in specific regions of the T-P-X space also increases the uncertainty of CO_2 solubility predictions. Further solubility experiments with the greatest potential to reduce the overall- and region-specific accuracy of the CO_2 solubility prediction could be identified from MMoPS. Predicting CO_2 solubility more accurately should improve our ability to estimate CO_2 storage potential and determine the fate and transport of CO_2 leakage.

5.2 CO_2 Leakage Detection using Pressure Measurements

CO_2 leakage level can be characterized through an assessment of the integrity and permeability of the caprock inferred from pressure measurements in the injection zone using a Bayesian approach.

The detection power of pressure monitoring is evaluated using the expected distribution of pressure increases in the injection zone for permeable vs. impermeable caprock cases. The results indicate that pressure monitoring alone is not sufficient to

detect a permeable caprock for the system considered in this study, except for very high caprock permeability values.

Time has little impact on the inferred caprock permeability class and the detection power of pressure monitoring. The injection rate has an influence on the predicted caprock permeability class and the detection power of the simulated pressure monitoring. Higher confidence in the predicted caprock permeability class and a significant improvement in the statistical detection power for a leaky caprock can be achieved by reducing the uncertainty in reservoir properties through further site characterization.

5.3 CO_2 Leakage Detection using Seismic Travel Time Measurements

The power of CO_2 leakage detection is assessed using a statistical analysis and test of P-wave travel times in the monitoring zone.

The detection power using the P-wave travel time measurements and test alone is generally not high enough for small leakage events, unless the porosity and the permeability of the monitoring zone are high, and/or a long period of time has elapsed since the leakage occurred. Therefore, there is an advantage to choosing a monitoring zone with high porosity and high permeability for the purpose of monitoring CO_2 leakage with seismic wave velocity detection.

The detection power is higher when the leak occurs near the location of the monitoring system, suggesting the possible need for multiple monitoring locations when the reservoir and the monitoring zone are large. As the likely pathways for concentrated CO_2 leakage at sequestration sites include abandoned wells and faults, the seismic monitoring equipment may be most effectively deployed along the abandoned wells or above the faults (if any are present).

5.4 FUTURE DIRECTIONS

As the sole measurement of the pressure or seismic travel time cannot provide sufficient power for CO_2 leakage detection in many cases, further investigations are needed for the performance of CO_2 leakage detection using a monitoring network which combines the pressure and seismic monitoring, and/or other monitoring techniques, such as soil CO_2 flux monitoring, surface deformation monitoring and groundwater chemistry monitoring. The optimization of the monitoring network also needs to be investigated to conduct CO_2 leakage detection with high confidence and in a timely and cost-effective manner.

Spatial heterogeneity at sequestration sites is not considered in this study. Further investigations are needed to model spatial variations in the input parameters, such as permeability and porosity. Candidate statistical methods include spatially autoregressive models, multilevel models, geographically weighted regression, switching regressions and spatial stochastic modeling.

For seismic monitoring, the change in P-wave travel time from the top to the bottom of the monitoring zone is chosen as the indicator for CO_2 leakage detection in this study. The detection power using other seismic attributes, such as changes in seismic velocities measured from acoustic-velocity logs, could be further investigated.

Appendices

APPENDIX 1: SUPPORTING INFORMATION FOR CHAPTER 2

A.

Table S-1: Models capable of calculating CO_2 solubility in water or low temperature, low pressure conditions

Model	Validity	Type
Li and Nghiem (1986)	Pure water and brine	Peng-Robinson EOS for the gas phase and Henry's law for the aqueous phase
Nighswander et al. (1989)	353-473K (80-200°C), up to 100 bar, Pure Water and 1 wt% NaCl solution	Peng-Robinson EOS for the gas phase and an empirical Henry's law constant correlation for the aqueous phase
Chang and Coats et al. (1998)	Pure water and Brine	Empirical fitting
Teng & Yamasaki (1998)	278-293K, 64.4-294.9 bar, Sea water	Modified Henry's law and Setchenow equation
Enick and Klara (1990)	298-523K, 30-850 bar, Pure water and brine	Empirical fitting
Soreide and Whitson (1992)	423-623K, 0-1000 bar, Pure water and brine	Modified Peng-Robinson EOS
Carroll and Mather (1992): 4 models	below 373K, up to 100 bar (solubility of 2 mol%)	EOS: Krichevsky-Kasarnovsky equation (based on Henry' law)
King et al. (1992)	1-500 bar, Pure water	Henry's law based on Redlich-Kwong equation
Pruess and Garcia (PG-CSM) (2002)	298-373K (25-100°C), 0-600 bar, Pure water	Empirical fitting
Spycher and Pruess (2003)	285-373K, 0-600 bar, Pure water	EOS[a]
Duan and Sun (2003)	273-533K, 0-2000 bar, Pure water and brine	EOS

[a] Equation of State.

B. Equations and Parameters in CO_2 Solubility Predictive Models

1. Duan and Sun (2006) Model

$$\ln m_{CO_2} = \ln y_{CO_2}\Phi_{CO_2}P - \frac{u_{CO_2}}{RT} - 2\lambda_{CO_2-Na}(m_{Na} + m_K + 2m_{Ca} + 2m_{Mg}) - \\ -\xi_{CO_2-Na-Cl}m_{Cl}(m_{Na} + m_K + m_{Ca} + m_{Mg}) + 0.07m_{SO_4} \quad (1.1)$$

In the above equation, λ_{CO_2-Na}, $\xi_{CO_2-Na-Cl}$ and the dimensionless standard chemical potential, $\frac{u_{CO_2}}{RT}$ are calculated using the following equation developed by Pitzer et al. (1984).

$$Par(T,P) = c_1 + c_2T + c_3/T + c_4T^2 + c_5/(630-T) + c_6P + \\ + c_7P\ln T + c_8P/T + c_9P/(630-T) + c_{10}P^2/(630-T)^2 + c_{11}T\ln P \quad (1.2)$$

The parameters c_1-c_{11} for λ_{CO_2-Na}, $\xi_{CO_2-Na-Cl}$ and $\frac{u_{CO_2}}{RT}$ are listed in Table 2 of Duan & Sun (2003) paper.

$$\Phi_{CO_2} = C_1 + \left[C_2 + C_3T + \frac{C_4}{T} + \frac{C_5}{T-150}\right]P + \left[C_6 + C_7T + \frac{C_8}{T}\right]P^2 + \\ + \left[C_9 + C_{10}T + \frac{C_{11}}{T}\right]\ln P + \frac{(C_{12}+C_{13}T)}{P} + \frac{C_{14}}{T} + C_{15}T^2 \quad (1.3)$$

The parameters C_1, C_2, C_3,…C_{14}, C_{15} are given in Table 1 of Duan et al. (2006) paper.

$$y_{CO_2} = (P - P_{H_2O})/P \quad (1.4)$$

$$P_{H_2O} = (P_cT/T_c)[1 + c_1(-t)^{1.9} + c_2t + c_3t^2 + c_4t^3 + c_5t^4] \quad (1.5)$$

Table S-2. Parameters for P_{H2O} calculation

Parameters for P_{H2O} calculation	value
c_1	-38.640844
c_2	5.8948420
c_3	59.876516
c_4	26.654627
c_5	10.637097

Nomenclature

m	molality of components dissolved in water
y_{CO_2}	mole fraction of CO_2 in vapor phase
Φ_{CO_2}	fugacity coefficient ofCO_2
P	total pressure of the system in bar

u_{CO2}	standard chemical potential of CO_2 in liquid phase
R	universal gas constant
T	absolute temperature in Kelvin
λ_{CO_2-Na}	interaction parameter between CO_2 and Na^+
$\xi_{CO_2-Na-Cl}$	interaction parameter between CO_2 and Na^+, Cl^-

2. Darwish and Hilal (2010) Model

$$\log\left(\frac{C_{g0}}{C_{gs}}\right) = K_{gs}m \tag{2.1}$$

$$\left(\frac{K_{gs}}{K_{g0}}\right) = \left(\frac{P}{P_0}\right)^{f(T_r,m)} \tag{2.2}$$

$$f(T_r,m) = (0.461 + 0.078m)\left(\frac{T_r}{(1.0 + 0.1267m)} - 1\right) \qquad 1 \le m \le 4 \tag{2.3}$$

$$T_r = \frac{T}{T_{c,CO_2}} = \frac{T}{304.2} \tag{2.4}$$

Nomenclature

P	pressure, bar
T	temperature, K
C_{g0}	Solubility of CO_2 in pure water
C_{gs}	Solubility of CO_2 in brine
K_{gs}	Setschenov constant in kg/mol
K_{g0}	0.086 mol CO_2 per kg H_2O
P_0	50 bar
m	molality of NaCl in aqueous phase

3. Enick & Klara (1990) Model

$$w_{CO_2,b} = w_{CO_2,w}(1.0 - 4.893414\cdot10^{-2}(TDS) + 0.1302838\cdot10^{-2}(TDS)^2 - \\ -0.1871199\cdot10^{-4}(TDS)^3) \tag{3.1}$$

$$\ln\left(\frac{f_{CO_2}}{y_{CO_2,w}}\right) = \ln H^*_{CO_2} + \frac{v^\infty_{CO_2}}{RT}(p) \tag{3.2}$$

$$H^*_{CO_2} = -5032.99 + 30.74113T - 0.052667T^2 + 2.630218\cdot 10^{-5}T^3 \tag{3.3}$$

$$v^\infty_{CO_2} = 1799.36 - 17.8218T + 0.0659297T^2 - 1.05786\cdot 10^{-4}T^3 + 6.200275\cdot 10^{-8}T^4 \tag{3.4}$$

$$w_{CO_2,w} = \frac{y_{CO_2,w}\cdot 44}{18 + y_{CO_2,w}\cdot 44} \tag{3.5}$$

Nomenclature

P	pressure, MPa
T	temperature, K
$w_{CO_2,b}$	solubility, weight fraction in brine
$w_{CO_2,w}$	solubility, weight fraction in pure water
TDS	weight %
f_{CO_2}	fugacity of CO_2, MPa
$y_{CO_2,w}$	mole fraction of CO_2 in the aqueous phase
$v^\infty_{CO_2}$	molar volume of CO_2 at infinite dilution, cm^3/mole
$H^*_{CO_2}$	reference Henry's constant

4. Barta and Bradley (1985) Model

$$K_H = \frac{f}{m_{CO_2}} = \frac{P\Phi_{T,P}}{m_{CO_2}} \tag{4.1}$$

$$\begin{aligned}\ln K_H &= \Delta^{(T)}{}_{Na,CO_2,Cl}\cdot m_{Na} + \frac{3}{2}\chi^{(T)}{}_{Na,CO_2,Cl} m_{Na}{}^2 + \\ &+ \frac{1}{2}[(2m_{Na}\Delta^{(T)}{}_{Na,CO_2,Cl} + 3\chi^{(T)}{}_{Na,CO_2,Cl} m_{Na}{}^2)^2 + 8\lambda^{(T)}{}_{CO_2,CO_2}]^{1/2}\end{aligned} \tag{4.2}$$

$$\lambda^{(T)}{}_{CO_2,CO_2} = (B_0 + B_1T + B_2/T)^2 \tag{4.3}$$

$$\Delta^{(T)}{}_{Na,CO_2,Cl} = B_3 + B_4 T + B_5 / T + B_6 T^2 \quad (4.4)$$

$$\chi^{(T)}{}_{Na,CO_2,Cl} = B_7 + B_8 T + B_9 / T + B_{10} T^2 \quad (4.5)$$

$$\ln \Phi_{T,P} = (a / T^2 + b / T + c)P + (d / T^2 + e / T + f)P^2 / 2 \quad (4.6)$$

(Equation 4.6 for estimating $\Phi_{T,P}$ is adopted from Spycher and Reed, 1988)

Nomenclature

P	pressure, bar
T	temperature, K
K_H	Henry's law ratio (fugacity to molality)
f	gas fugacity in vapor phase relative to ideal gas standard state
m_{CO2}	molality of CO_2
$\Phi_{T,P}$	fugacity coefficient at temperature T and total pressure P
$\Delta^{(T)}{}_{Na,CO_2,Cl}$	parameter for binary interaction between CO_2 and ions
$\chi^{(T)}{}_{Na,CO_2,Cl}$	parameter for interactions of two neutral species with one ion
$\lambda^{(T)}{}_{CO_2,CO_2}$	second virial coefficient for interaction of CO_2 species with itself

Table S-3.Regression Coefficients B_0-B_{10} in Barta and Bradley (1985) Model (Barta and Bradley, 1985)

Coefficient	Estimate	σ[a]
B_0	13.85682	6.300
B_1	-1.172399 x10^{-2}	0.0681 x10^{-2}
B_2	-2.349329 x10^{3}	0.1316 x10^{3}
B_3	9.639938	25.41
B_4	-2.431955 x10^{-2}	0.8018 x10^{-2}
B_5	-1.194304 x10^{3}	0.5071 x10^{3}
B_6	2.056847 x10^{-5}	0.5919 x10^{-5}
B_7	-1.183187	0.5759
B_8	2.911457 x10^{-3}	1.300 x10^{-3}
B_9	1.544037 x10^{2}	0.8286 x10^{2}
B_{10}	-2.362693 x10^{-6}	0.9536 x10^{-6}

[a] two times the standard error (95% confidence limit)

5. Zuo & Guo (1991) Model

$$\ln\phi = \ln\phi^{EOS} + \ln\phi^{DH} \tag{5.1}$$

$$P = \frac{RT}{v-b} - \frac{a[T]}{v(v+b)+c(v-b)} \tag{5.2}$$

$$a[T] = \Omega_a (RT_c)^2 / P_c \alpha(T_r) \tag{5.3}$$

$$b = \Omega_b (RT_c / P_c) \tag{5.4}$$

$$c = \Omega_c (RT_c / P_c) \tag{5.5}$$

Ω_b is the smallest positive root of the following cubic equation:

$$\Omega_b^{\ 3} + (2-3\delta)\Omega_b^{\ 2} + 3\delta^2\Omega_b - \delta^3 = 0 \tag{5.6}$$

$$\alpha(T_r) = [1 + F(1 - T_r^{0.5})]^2 \tag{5.7}$$

For molecular components (except water)

$$F = 0.452413 + 1.30982\omega - 0.295937\omega^2 \tag{5.8}$$

$$\delta = 0.329032 - 0.076799\omega + 0.0211947\omega^2 \tag{5.9}$$

For water: F=0.689803 δ=0.269

$$\Omega_a = 3\delta^2 + 3(1-2\delta)\Omega_b + \Omega_b^{\ 2} + 1 - 3\delta \tag{5.10}$$

$$\Omega_c = 1 - 3\delta \tag{5.11}$$

For ionic species:

$$b = (2/3)\pi N_a \sigma^3 \tag{5.12}$$

$$c = b \tag{5.13}$$

$$a = 2.57012\pi\varepsilon N_a^{\ 2}\sigma^3 f \qquad f = 6 \tag{5.14}$$

$$\varepsilon = k * (2.2789 * 10^{-8}\eta^{0.5}\alpha^{1.5}\sigma^{-6}) \tag{5.15}$$

$$b = \Sigma x_i b_i \tag{5.16}$$

$$c = \Sigma x_i c_i \tag{5.17}$$

$$a = a^{VDW} - 2dg^{E}{}_{\infty} / \ln(\frac{Q+d}{Q-d}) \tag{5.18}$$

$$a^{VDW} = \sum_i \sum_j x_i x_j (a_i a_j)^{0.5} (1 - k_{ij}) \tag{5.19}$$

$$d = [bc + (b+c)^2 / 4]^{0.5} \tag{5.20}$$

$$Q = b + (b+c)/2 \tag{5.21}$$

$$g^{E}{}_{\infty} = RT \sum_i \sum_j x_i x_j A_{ij} \tag{5.22}$$

$$k_{gw} = m_1 + m_2 / T \tag{5.23}$$

m_1= 0.19235, m_2= -93.953, A_{gw}= 0.04532, A_{sg}= -7.01
NaCl-H_2O interaction parameter k_{sw}= -0.1556
CO_2-NaCl A_{sg}= -7.01 (Li and Pitzer, 1986)

$$\ln \phi^{DH} = -A[\frac{2Z_i^2}{B} \ln(\frac{1+BI^{0.5}}{1+B/\sqrt{2}}) + (\frac{I^{0.5} Z_i^2 - 2I^{1.5}}{1+BI^{0.5}})] \tag{5.24}$$

$$I = 0.5 \sum_i x_i Z_i^2 \tag{5.25}$$

$$A = \frac{1}{3} (\frac{2\pi N_a d_0}{M_s})^{0.5} (\frac{e^2}{DkT})^{1.5} \tag{5.26}$$

$$B = 2150 (d_0 / DT)^{0.5} \tag{5.27}$$

$$D = 78.54[1 - 4.579 \cdot 10^{-3}(T - 298.15) + 1.19 \cdot 10^{-5}(T - 298.15)^2 - 2.8 \cdot 10^{-8}(T - 298.15)^3] \tag{5.28}$$

Nomenclature

P	pressure, bar
T	temperature, K
R	universal gas constant 8.314 J/mole-K
v	molar volume
N_a	Avogadro's number $6.02 \cdot 10^{23}$ /mole
σ	ionic diameter Na^+ 1.92 Å Cl^- 3.61 Å
η	number of electrons in an ion
α	polarizability of an ion Na^+ 0.18 $Å^3$ Cl^- 3.69 $Å^3$
$g^{E}{}_{\infty}$	excess Gibbs energy at infinite pressure which is calculated from Margules model in this work
x	liquid phase mole fraction

e unit electronic charge
Z number of charges
Ms molecular weight of solvent
d_0 solvent density
k Boltzmann constant

6. Soreide & Whitson (1992) Model

$$P = \frac{RT}{v-b} - \frac{a(T)}{v(v+b)+b(v-b)} \tag{6.1}$$

$$b = 0.0778\frac{RT_c}{P_c} \tag{6.2}$$

$$a(T_c) = 0.45724\frac{R^2T_c^{\,2}}{P_c} \tag{6.3}$$

$$a(T) = \alpha[a(T_c)] \tag{6.4}$$

For CO_2:
T_c= 304.2 K; P_c= 73.8 bar; T_b= 194.7 K; Acentric factor, w: 0.2273
For water:
T_c= 647.3 K; P_c= 221.2 bar; T_b= 373.2 K; Acentric factor, w: 0.3434
For pure compound (e.g. CO_2 and water), 0.2<w<2.0:

$$\alpha^{1/2} = 1 + k(1 - T_r^{1/2}) \tag{6.5}$$

$$k = 0.3746 + 1.5423w - 0.2699w^2 \tag{6.6}$$

For brine:

$$\alpha^{1/2} = 1 + 0.4530[1 - T_r(1 - 0.0103c_{sw}^{\;1.1})] + 0.0034(T_r^{-3} - 1) \tag{6.7}$$

$$b_m = \sum_{i=1}^{n} x_i b_i \tag{6.8}$$

$$a_m = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j a_{ij} \tag{6.9}$$

$$a_{ij}^{\;AQ} = \sum_i \sum_j x_i x_j (a_i a_j)^{0.5}(1 - k_{ij}^{\;AQ}) \tag{6.10}$$

$$k_{ij}^{\;AQ} = -0.31092(1 + 0.15587c_{sw}^{\;0.7505}) + 0.2358(1 + 0.17837c_{sw}^{\;0.979})T_{ri} \\ -21.2566\exp(-6.7222T_{ri} - c_{sw}) \tag{6.11}$$

$$a_{ij}^{NA} = \sum_i \sum_j y_i y_j (a_i a_j)^{0.5} (1 - k_{ij}^{NA}) \qquad k_{ij}^{NA} = 0.1896 \tag{6.12}$$

Nomenclature

P	pressure, bar
T	temperature, K
R	gas constant, 83.14472 cm^3 bar/(mole K)
V	molar volume, cm^3/mole
c_{sw}	brine salinity, mole/kg (molality)
T_r	reduced temperature ($T_r = T / T_c$)
X_i	normalized mole fraction of aqueous phase
Y_i	normalized mole fraction of non-aqueous phase

7. Chang & Coats (1998) Model

$$R_w = aP[1 - b\sin(\frac{\pi}{2}\frac{cP}{cP+1})] \qquad \text{for } P < P^0 \tag{7.1}$$

$$a = \sum_{i=0}^{4} (a_i \cdot 10^{-3i})(\frac{9}{5} \cdot T + 32)^i \tag{7.2}$$

$$b = \sum_{i=0}^{4} (b_i \cdot 10^{-3i})(\frac{9}{5} \cdot T + 32)^i \tag{7.3}$$

$$c = 10^{-3} \sum_{i=0}^{4} (c_i \cdot 10^{-3i})(\frac{9}{5} \cdot T + 32)^i \tag{7.4}$$

$$p^0 = \frac{2}{\pi} \cdot \frac{\sin^{-1}(b^2)}{c[1 - \frac{2}{\pi}\sin^{-1}(b^2)]} \tag{7.5}$$

$$R_b = 10^{-0.028S/[(9/5)T+32]^{0.12}} R_w \tag{7.6}$$

The values of coefficients ai, bi and ci are shown in the table below.

Table S-4. Parameters a_i, b_i and c_i in Chang and Coats Model (Chang et al., 1998)

i	a_i	b_i	c_i
0	1.163	0.965	1.28
1	-16.63	-0.272	-10.757
2	111.073	0.0923	52.696

3	-376.859	-0.1008	-222.395
4	524.889	0.0998	462.672

Nomenclature

P total pressure of CO_2 and water, psia

T temperature, ºC

S salinity of brine which is defined as the total dissolved salts in the solution in mass percent

R_w solubility of CO_2 in pure water, scf CO_2/stb H_2O

R_b solubility of CO_2 in brine, scf CO_2/stb H_2O

Another density model is needed if the desired solubility of CO_2 is in the unit of molality (mole/kg). A density model presented by Batzle and Wang is shown here.

$$\rho_w = 1 + 10^{-6}(-80T - 3.3T^2 + 0.00175T^3 + 489P - 2TP + 0.016T^2P$$
$$-1.3\cdot 10^{-5}T^3P - 0.333P^2 - 0.002TP^2) \quad (7.7)$$

$$\rho_b = \rho_w + Sfrac\{0.668 + 0.44Sfrac + 10^{-6}[300P - 2400PSfrac$$
$$+T(80 - 3T - 3300Sfrac - 13P + 47PSfrac)]\} \quad (7.8)$$

Where, T is temperature in ºC, P is pressure in MPa. *Sfrac* is salinity in mass fraction (ppm/10^6) and ϱ_w and ϱ_b are pure water and brine density values, respectively, in g/cm^3.

8. Li & Nghiem (1986) Model

Gas solubility in pure water

$$\ln H_i = \ln H_i^* + v_i^\infty p/(RT) \quad (8.1)$$

$$\ln \frac{H_i^*}{f_w^s} = -11.3021 + 10.6030(\frac{10^3}{T}) - 1.20696(\frac{10^6}{T^2}) \quad (8.2)$$

$$(\frac{p_{ci}v_i^\infty}{RT_{ci}}) = 0.095 + 2.35(\frac{Tp_{ci}}{CT_{ci}}) \quad (8.3)$$

$$C = (h_w^0 - h_w^s + p_w^s v_w^s - RT)/v_w^s \quad (8.4)$$

p_w^s and ($h_w^0 - h_w^s$) are calculated respectively from equations in Reid et al. (1977).

for T'>90 ºF:

$$f_w^s = \phi_w^s \cdot p_w^s \quad (8.5)$$

$$\phi_w^s = 0.9958 + 9.6833\cdot 10^{-5}T' - 6.175\cdot 10^{-7}T'^2 - 3.08333\cdot 10^{-10}T'^3 \quad (8.6)$$

for $T' \leq 90\ ^{o}F$:

$$\phi_w^{\ s} = 1 \tag{8.7}$$

$$T' = 1.8T - 459.67 \tag{8.8}$$

Where T is the temperature in K, and T' is the temperature in ^{o}F.
The gas phase is modeled with the Peng-Robinson EOS. The EOS interaction coefficient d_{ij} between water and CO_2 was found to be temperature-dependent.
d_{ij}=0.2 at T≤373K; d_{ij}=0.49852-0.008(T), at T>373K.
Gas solubility in brines

$$\ln \frac{H_i v_s}{RT} = (\frac{G_{bi}}{RT}) + (\frac{G_{ai}}{RT}) \tag{8.9}$$

Where G_{bi} is the Gibbs energy required to create a cavity in the solvent. G_{ai} is the molar Gibbs energy resulting from the interaction of the solute and the solvent molecules. v_s is the molar volume of pure water.

$$\frac{G_{bi}}{RT} = -\ln(1-\theta_3) + \frac{3\theta_2}{1-\theta_3}\sigma_i^* + [\frac{3\theta_1}{1-\theta_3} + \frac{9}{2}\frac{\theta_2^{\ 2}}{(1-\theta_3)^2}](\sigma_i^*)^2 + \frac{p}{RT}\frac{\pi}{6}\eta_w(\sigma_i^*)^3 \tag{8.10}$$

$$\theta_n = \frac{\pi}{6}\eta_w \sum_j \rho_j (\sigma_i^*)^n \qquad n=1,2,3 \tag{8.11}$$

$$\sigma_k^{\ *} = \sigma_k / \sigma_w \qquad k=\text{i, j} \tag{8.12}$$

$$\eta_w = N\sigma_w^{\ 3} \tag{8.13}$$

$$\frac{G_{ai}}{RT} = -\frac{32}{9}\frac{\pi}{T}\eta_w \sum_j \frac{\varepsilon_{ij}}{k}\rho_j(\sigma_{ij}^{\ *})^3 - \frac{4}{3}\frac{\pi}{kT}\frac{\rho_w N u_w^{\ 2}(N\alpha_i)}{(\sigma_{iw}^{\ *})^3\eta_w} \tag{8.14}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \tag{8.15}$$

$$\sigma_{ij}^{\ *} = F(\sigma_i^* + \sigma_i^*)/2 \tag{8.16}$$

$$F = 16280/T^2 - 141.75/T + 1.2978 \tag{8.17}$$

$$\rho_w = (1-\phi)\rho_s / M_w \tag{8.18}$$

$$\rho_{Na} = \rho_{Cl} = \phi\rho_s / M_{NaCl} \tag{8.19}$$

Where M_w and M_{NaCl} are the molecular weight of water and NaCl, respectively; ρ_s is the mass density of NaCl solution and ω is the weight fraction of NaCl in water. An accurate value ρ_s is required. In this paper, ρ_s is estimated by Rowe and Chou model (1970).

Table S-5. Parameters in Li and Nghiem Model (Li and Nghiem, 1986)

	σ, nm	ε/k, K	10^{24}α, cm^3/molecule
Water	0.275	85.3	1.590
Carbon Dioxide	0.332	300.0	2.590
Methane	0.413	160.0	2.700
Nitrogen	0.373	95.0	1.730
Na^+	0.190	147.4	0.210
Cl^-	0.362	225.5	3.020

Note: $N = 6.02283 \times 10^{23}$; $\mu_w = 1.84 \times 10^{-18}$ (erg·cm)$^{1/2}$; cm = 1 nm = 10^{-9}m.

Nomenclature

H_i^*	Reference Henry's law constant
v_i^∞	Molar volume at infinite dilution
C	cohesive energy density of water
p_w^s	Saturated vapor pressure of water at the temperature T
v_w^s	Molar volume of water at p_w^s and T
$h_w^0 - h_w^s$	Enthalpy departure of liquid water at p_w^s and T
σ_i and σ_j	molecular diameter of solute I and of species j in the solvent
σ_w	molecular diameter of pure water
N	Avogadro's number
ρ_j	molar density of species j in the solvent
u_w	dipole moment of water
α_i	polarizability of solute i
k	Boltzmann's constant
ε_i	Parameter in equation
T	Temperature, K
F	unity

9. Akinfiev and Diamond (2010) Model

$$k_{sbb} = \frac{1}{m_{NaCl}} \ln \frac{m_{CO_2}^{\ 0}}{m_{CO_2}} \tag{9.1}$$

$$k_{sbb} = 0.116791 + \frac{6.0616}{T - 228} - 0.0394175 \cdot m_{NaCl}{}^{0.5} + 0.227168 \cdot T \cdot 10^{-3} - 0.001971 \cdot m_{CO_2} \qquad (9.2)$$

Where, T is temperature in K.

m_{NaCl} and m_{CO_2} are in the unit of mole/kg H_2O.

Duan & Sun (2006) model is employed to calculate solubility of CO_2 in pure water ($m_{CO_2}{}^0$).

C. Figures Referred to in Section 2.3 and 2.4 of the Main Text

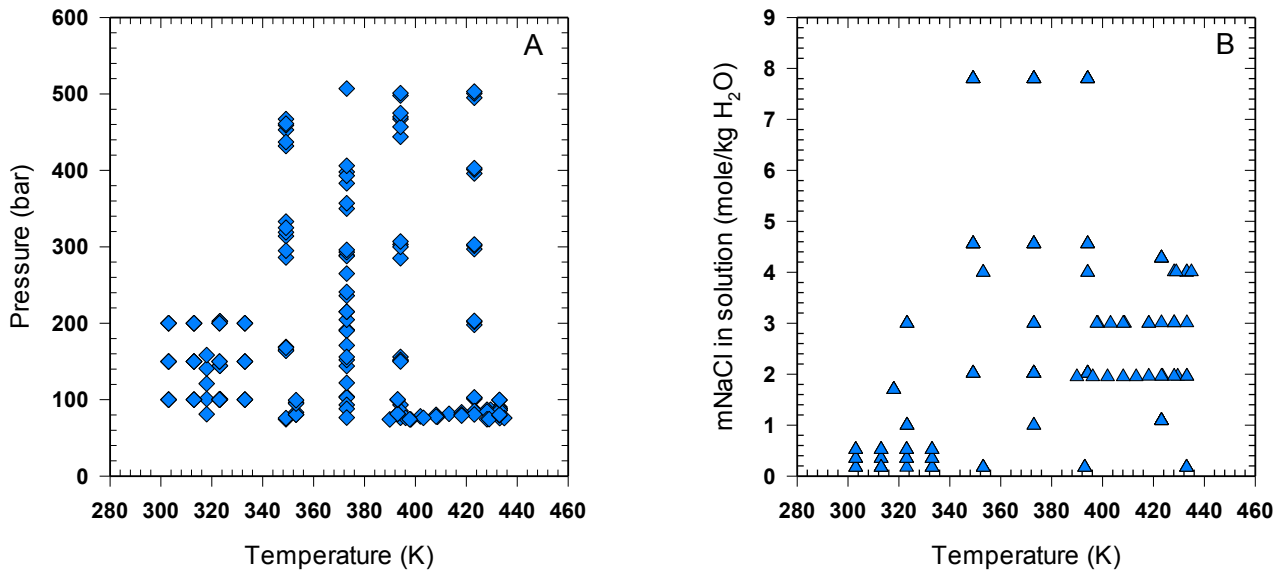


Figure S-1. Distribution of 173 experimental CO_2 solubility observations in A) pressure-temperature space and B) salt concentration – temperature space.

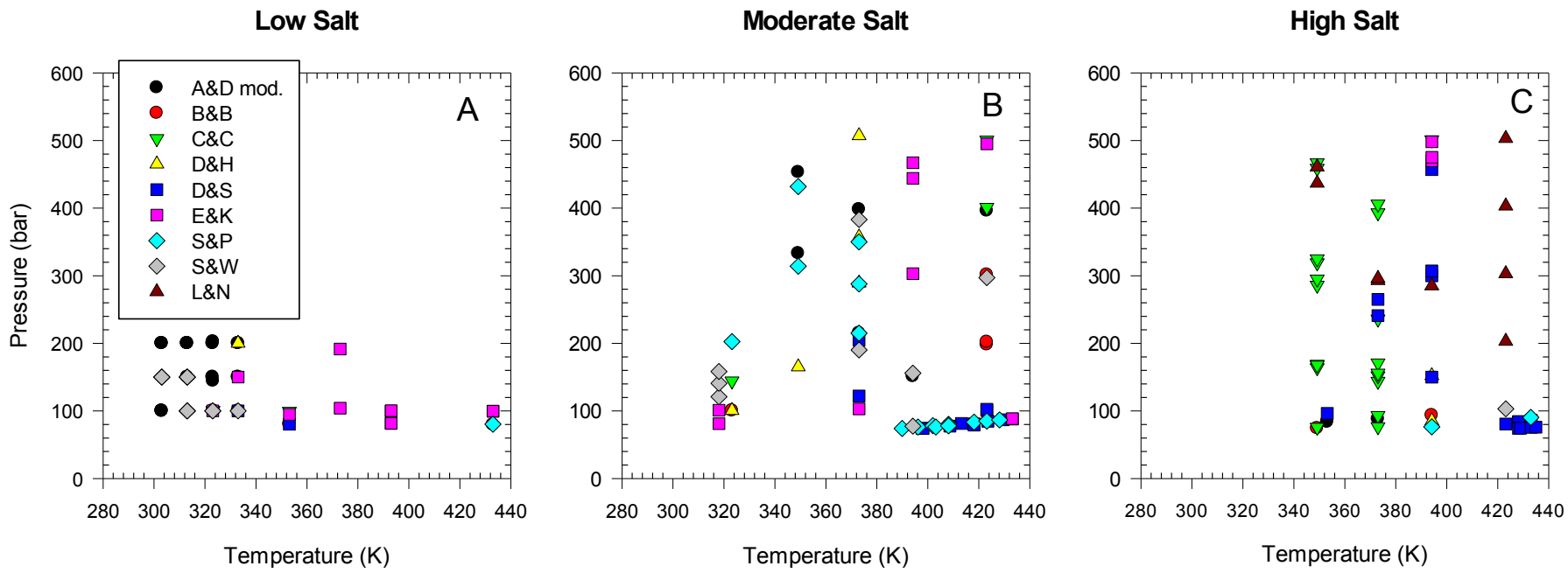


Figure S-2. Distribution of best-performing models in pressure-temperature space at various salt concentrations: A) Low (≤1 Molal), B) Moderate (1-3 Molal), and C) High (3-8 Molal). Models: A&D (modified): Akinfiev and Diamond coupled with Duan and Sun; B&B: Barta and Bradley; C&C: Chang and Coats; D&H: Darwish and Hilal; D&S: Duan and Sun; E&K: Enick and Klara; S&P: Spycher and Pruess; S&W: Soreide and Whitson; L&N: Li and Nghiem. The symbols for each model for all graphs are presented in the figure label of graph A. Y axis is pressure (bar) for all graphs.

D. RMSE, APE, AIC, BIC and Chi-squared Statistics

Root Mean Square Error for each model was calculated using:

$$RSS = \sum_{j=1}^{n} \varepsilon_j^{\,2}$$
$$RMSE = \sqrt{\frac{RSS}{n}} \qquad \text{(D.1)}$$

Where, RSS is Residual Sum of Squares. ε_j is the error, defined as the difference between the predicted value by each model and the experimental value. The errors are assumed to be normally distributed. n is the number of data points in the evaluation data set which is 173 in our case.
The average percent error (APE) was calculated by:

$$APE = \frac{\sum_{j=1}^{n} abs(\frac{prediction_j - observation_j}{observation_j})}{n} \qquad \text{(D.2)}$$

Akaike Information Criteria, AIC was calculated using the equation below (Burnham & Anderson, 2002).

$$AIC = 2K + n\ln(\frac{RSS}{n}) \qquad \text{(D.3)}$$

Where K is the number of parameters in the model
We also evaluated these models using adjusted AIC, AIC_C equation as follows:

$$AIC_C = AIC + (2K\frac{K+1}{n-K-1}) \qquad \text{(D.4)}$$

The added term $(2 \cdot K \cdot (K+1)/(n-K-1)$, is a small sample adjustment. As mentioned in Burnhan & Anderson (2002), if the ratio of n/K<40, which is the case for our evaluation data set with 173 data points, the small sample adjustment is recommended.

Bayesian Information Criteria:

$$BIC = K\ln n + n\ln(\frac{RSS}{n}) \qquad \text{(D.5)}$$

Table S-6. Chi-squared statistics of Temperature, Pressure and Salt concentration with respect to best-performing model.

Condition	Chi-squared statistic[a]
Temperature	75.11
Pressure	52.95
Salt Concentration	143.28

[a]Chi-squared statistic is calculated using the Weka software, the machine learning tool used to develop MMoPS. The value of the chi-squared statistic indicates the worth of an attribute (i.e. T, P or X) in predicting the class value (i.e. the best-performing model). The higher the value is, the more informative the attribute is in distinguishing best-performing models.

E. Table S-7. CO_2 Solubility Experimental Data and Predicted Values from Different Models

T, K	P, bar	mNaCl, M	mCO_2(ob)[a]	mCO_2(pr)[b] B&B	mCO_2(pr) E&K	mCO_2(pr) D&H	mCO_2(pr) D&S	mCO_2(pr) S&W	mCO_2(pr) C&C	mCO_2(pr) A&D mod.	mCO_2(pr) L&N	mCO_2(pr) S&P
303	100	0.17	1.28	1.50	1.52	1.31	1.31	1.34	1.36	1.30	0.41	1.31
303	150	0.17	1.33	1.79	1.71	1.38	1.38	1.35	1.46	1.36	0.43	1.38
303	200	0.17	1.39	1.97	1.76	1.44	1.45	1.16	1.51	1.43	0.41	1.43
303	100	0.35	1.20	1.44	1.45	1.26	1.27	1.29	1.30	1.24	0.41	1.27
303	150	0.35	1.25	1.72	1.62	1.33	1.34	1.29	1.40	1.31	0.42	1.33
303	200	0.35	1.30	1.89	1.67	1.39	1.40	1.12	1.44	1.37	0.41	1.38
303	100	0.52	1.16	1.38	1.37	1.22	1.23	1.23	1.24	1.19	0.40	1.23
303	150	0.52	1.22	1.65	1.54	1.29	1.30	1.24	1.34	1.26	0.42	1.28
303	200	0.52	1.26	1.81	1.59	1.35	1.36	1.07	1.38	1.32	0.40	1.33
313	100	0.17	1.14	1.31	1.22	1.19	1.20	1.14	1.22	1.19	0.40	1.20
313	150	0.17	1.22	1.60	1.41	1.26	1.27	1.18	1.35	1.26	0.42	1.28
313	200	0.17	1.28	1.79	1.48	1.33	1.34	1.08	1.41	1.32	0.41	1.34
313	100	0.35	1.12	1.26	1.16	1.15	1.16	1.11	1.17	1.14	0.39	1.16
313	150	0.35	1.19	1.54	1.34	1.22	1.23	1.15	1.29	1.21	0.41	1.23
313	200	0.35	1.23	1.72	1.41	1.28	1.30	1.05	1.35	1.27	0.40	1.29
313	100	0.52	1.07	1.20	1.10	1.12	1.13	1.07	1.12	1.10	0.39	1.12
313	150	0.52	1.13	1.48	1.27	1.18	1.20	1.11	1.24	1.16	0.41	1.19
313	200	0.52	1.19	1.65	1.34	1.24	1.26	1.02	1.29	1.22	0.40	1.25
313	150	0.17	1.13	1.60	1.41	1.26	1.27	1.18	1.35	1.26	0.42	1.28
313	200	0.17	1.23	1.79	1.48	1.33	1.34	1.08	1.41	1.32	0.41	1.34
318	101	1.70	0.74	0.88	0.76	0.87	0.89	0.77	0.83	0.84	0.36	0.86
318	121	1.70	0.77	0.97	0.82	0.89	0.92	0.79	0.87	0.86	0.37	0.89
318	141	1.70	0.78	1.05	0.86	0.92	0.94	0.80	0.91	0.88	0.37	0.91
318	158	1.70	0.80	1.11	0.89	0.94	0.96	0.79	0.93	0.90	0.38	0.93
318	81	1.70	0.69	0.76	0.68	0.81	0.83	0.72	0.76	0.79	0.33	0.81
323	100	0.17	1.06	1.17	1.05	1.11	1.12	1.02	1.12	1.10	0.39	1.10

323	100	0.35	1.02	1.12	1.00	1.07	1.09	1.00	1.07	1.06	0.38	1.06
323	150	0.35	1.10	1.40	1.18	1.14	1.16	1.05	1.21	1.13	0.41	1.15
323	200	0.35	1.19	1.59	1.26	1.20	1.22	0.99	1.28	1.19	0.40	1.21
323	100	0.52	0.98	1.07	0.95	1.04	1.05	0.97	1.03	1.02	0.38	1.03
323	150	0.52	1.06	1.34	1.12	1.10	1.12	1.01	1.16	1.09	0.40	1.11
323	200	0.52	1.13	1.52	1.20	1.16	1.19	0.96	1.23	1.15	0.40	1.17
323.15	101	1.00	0.91	0.96	0.84	0.95	0.97	0.87	0.92	0.93	0.37	0.93
323.15	144	1.00	0.98	1.17	0.97	1.00	1.03	0.91	1.03	0.98	0.39	1.01
323.15	203	1.00	1.00	1.37	1.06	1.07	1.10	0.86	1.10	1.05	0.38	1.07
323.15	100	1.00	0.96	0.96	0.84	0.95	0.97	0.87	0.92	0.93	0.37	0.93
323.15	101	3.00	0.63	0.65	0.54	0.68	0.70	0.53	0.61	0.66	0.35	0.66
323.15	144	3.00	0.69	0.80	0.62	0.74	0.75	0.55	0.69	0.70	0.36	0.71
323.15	203	3.00	0.76	0.93	0.68	0.80	0.80	0.53	0.74	0.74	0.36	0.76
323.15	101	3.00	0.68	0.65	0.54	0.68	0.70	0.53	0.61	0.66	0.35	0.66
333	100	0.17	1.00	1.05	0.95	1.00	1.00	0.95	1.03	0.99	0.39	0.99
333	150	0.17	1.11	1.34	1.14	1.12	1.13	1.00	1.20	1.12	0.41	1.12
333	200	0.17	1.19	1.55	1.24	1.19	1.20	0.96	1.29	1.19	0.41	1.19
333	100	0.35	0.92	1.01	0.90	0.96	0.98	0.93	0.99	0.95	0.38	0.96
333	150	0.35	1.07	1.28	1.08	1.09	1.10	0.98	1.15	1.08	0.40	1.08
333	200	0.35	1.14	1.48	1.18	1.15	1.17	0.95	1.24	1.14	0.41	1.15
333	100	0.52	0.91	0.97	0.86	0.93	0.95	0.90	0.95	0.92	0.38	0.93
333	150	0.52	1.01	1.23	1.03	1.05	1.07	0.95	1.10	1.04	0.40	1.05
333	200	0.52	1.06	1.42	1.12	1.11	1.14	0.92	1.19	1.10	0.40	1.11
349.15	165	2.02	0.77	0.85	0.68	0.76	0.81	0.63	0.78	0.75	0.38	0.75
349.15	314	2.02	0.89	1.19	0.80	0.92	0.98	0.53	0.94	0.90	0.35	0.90
349.15	333	2.02	0.91	1.23	0.81	0.93	0.99	0.51	0.95	0.91	0.34	0.91
349.15	432	2.02	0.98	1.45	0.85	1.00	1.07	0.40	1.00	0.98	0.28	0.98
349.15	453	2.02	0.97	1.51	0.86	1.01	1.08	0.37	1.01	0.99	0.27	1.00
349.15	169	4.56	0.51	0.61	0.41	0.56	0.56	0.31	0.51	0.53	0.39	0.53

349.15	286	4.56	0.57	0.81	0.48	0.70	0.66	0.28	0.60	0.62	0.37	0.61
349.15	295	4.56	0.58	0.83	0.48	0.71	0.66	0.28	0.60	0.62	0.36	0.62
349.15	459	4.56	0.63	1.09	0.53	0.84	0.76	0.19	0.66	0.70	0.28	0.70
349.15	467	4.56	0.64	1.10	0.53	0.84	0.76	0.19	0.66	0.70	0.27	0.70
349.15	164	7.80	0.31	0.61	0.19	0.45	0.34	0.12	0.33	0.38	0.52	0.40
349.15	168	7.80	0.30	0.62	0.19	0.45	0.34	0.12	0.33	0.39	0.52	0.41
349.15	319	7.80	0.36	0.87	0.22	0.68	0.42	0.10	0.40	0.46	0.48	0.48
349.15	325	7.80	0.36	0.88	0.22	0.68	0.42	0.10	0.40	0.46	0.48	0.49
349.15	437	7.80	0.41	1.06	0.23	0.80	0.47	0.08	0.43	0.50	0.40	0.53
349.15	461	7.80	0.40	1.11	0.24	0.83	0.48	0.07	0.43	0.51	0.38	0.54
349.15	74	4.56	0.35	0.35	0.27	0.33	0.37	0.24	0.33	0.36	0.29	0.36
349.15	76	7.80	0.22	0.36	0.12	0.21	0.23	0.09	0.22	0.26	0.39	0.28
353.1	83	4.00	0.41	0.39	0.31	0.38	0.43	0.30	0.38	0.40	0.30	0.40
353.1	96	4.00	0.46	0.44	0.34	0.42	0.47	0.33	0.42	0.44	0.33	0.44
353.1	80	0.17	0.76	0.76	0.72	0.74	0.74	0.70	0.78	0.73	0.33	0.73
353.1	81	0.17	0.75	0.76	0.72	0.74	0.74	0.70	0.78	0.73	0.33	0.74
353.1	95	0.17	0.79	0.86	0.80	0.82	0.83	0.78	0.86	0.81	0.36	0.82
353.1	99	0.17	0.87	0.89	0.82	0.84	0.85	0.81	0.89	0.84	0.37	0.84
373	104	1.00	0.61	0.68	0.63	0.65	0.69	0.66	0.68	0.65	0.35	0.67
373	191	1.00	0.87	1.04	0.87	0.88	0.94	0.94	0.97	0.89	0.49	0.91
373	103	3.00	0.40	0.47	0.40	0.45	0.51	0.40	0.46	0.47	0.32	0.48
373	190	3.00	0.56	0.72	0.56	0.64	0.70	0.56	0.66	0.65	0.46	0.65
373	205	3.00	0.73	0.76	0.57	0.66	0.72	0.58	0.69	0.67	0.47	0.67
373	122	2.02	0.65	0.63	0.55	0.60	0.66	0.58	0.62	0.61	0.37	0.61
373	215	2.02	0.78	0.92	0.72	0.77	0.85	0.77	0.84	0.79	0.49	0.80
373	215	2.02	0.81	0.92	0.72	0.77	0.85	0.77	0.84	0.79	0.49	0.80
373	289	2.02	0.85	1.10	0.79	0.86	0.95	0.87	0.94	0.88	0.56	0.88
373	288	2.02	0.88	1.10	0.79	0.86	0.95	0.86	0.94	0.87	0.56	0.88
373	350	2.02	0.94	1.24	0.83	0.92	1.01	0.93	1.00	0.93	0.61	0.94

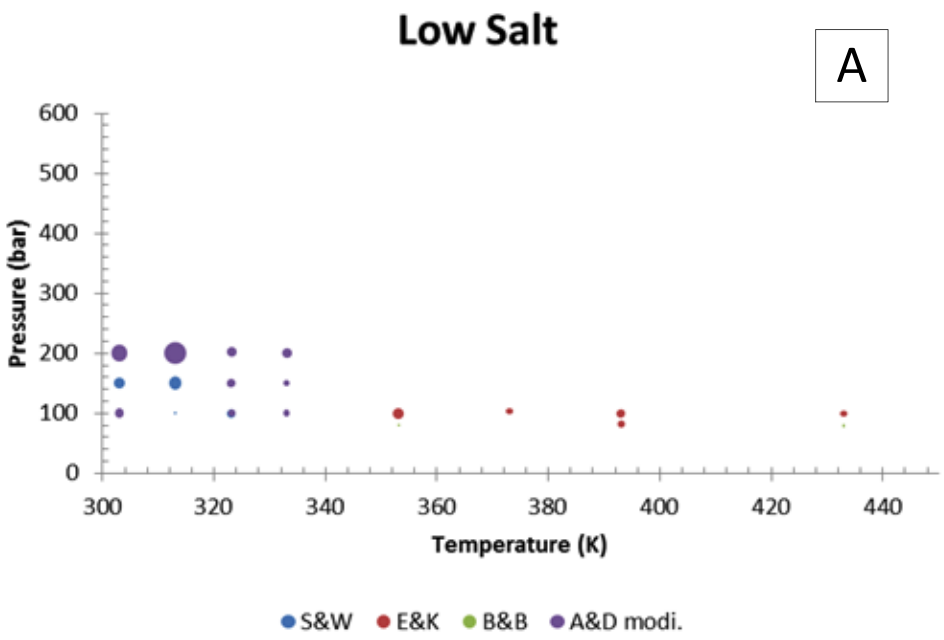
373	357	2.02	0.92	1.26	0.84	0.93	1.02	0.93	1.01	0.94	0.61	0.95
373	383	2.02	0.95	1.32	0.85	0.95	1.04	0.95	1.03	0.96	0.63	0.97
373	398	2.02	0.97	1.36	0.86	0.96	1.06	0.97	1.04	0.97	0.65	0.98
373	507	2.02	1.00	1.65	0.92	1.04	1.14	1.04	1.10	1.05	0.30	1.06
373	93	4.56	0.34	0.37	0.28	0.33	0.38	0.24	0.34	0.36	0.31	0.37
373	144	4.56	0.43	0.51	0.36	0.46	0.49	0.32	0.45	0.47	0.41	0.47
373	171	4.56	0.49	0.57	0.40	0.51	0.53	0.34	0.49	0.51	0.45	0.51
373	293	4.56	0.57	0.80	0.48	0.67	0.66	0.42	0.62	0.62	0.57	0.63
373	296	4.56	0.58	0.81	0.48	0.67	0.66	0.42	0.62	0.63	0.57	0.63
373	152	7.80	0.28	0.53	0.17	0.37	0.31	0.12	0.31	0.35	0.56	0.39
373	156	7.80	0.30	0.54	0.17	0.38	0.32	0.12	0.31	0.36	0.56	0.39
373	236	7.80	0.32	0.71	0.20	0.52	0.38	0.14	0.38	0.42	0.69	0.46
373	241	7.80	0.31	0.71	0.21	0.53	0.38	0.14	0.38	0.43	0.69	0.47
373	265	7.80	0.35	0.76	0.21	0.57	0.40	0.15	0.40	0.44	0.72	0.48
373	393	7.80	0.39	0.98	0.23	0.72	0.46	0.09	0.45	0.50	0.87	0.55
373	406	7.80	0.38	1.00	0.24	0.74	0.46	0.09	0.45	0.51	0.88	0.56
373	88	4.56	0.35	0.35	0.27	0.32	0.36	0.23	0.33	0.35	0.30	0.35
373	77	7.80	0.19	0.32	0.11	0.18	0.20	0.08	0.20	0.23	0.35	0.25
389.95	74	1.95	0.42	0.40	0.39	0.40	0.44	0.38	0.46	0.41	0.25	0.42
393	82	0.17	0.67	0.63	0.64	0.60	0.61	0.58	0.74	0.60	0.31	0.63
393	82	0.17	0.66	0.63	0.64	0.60	0.61	0.58	0.74	0.60	0.31	0.63
393	100	0.17	0.78	0.74	0.74	0.70	0.71	0.68	0.84	0.70	0.36	0.74
393	100	0.17	0.78	0.75	0.75	0.71	0.71	0.68	0.85	0.70	0.36	0.74
394.15	151	2.02	0.65	0.70	0.61	0.64	0.71	0.63	0.74	0.66	0.42	0.68
394.15	156	2.02	0.64	0.71	0.62	0.65	0.72	0.65	0.75	0.67	0.42	0.69
394.15	303	2.02	0.85	1.14	0.83	0.88	0.99	0.90	0.99	0.92	0.61	0.94
394.15	444	2.02	0.97	1.51	0.94	1.02	1.14	1.05	1.10	1.06	0.73	1.09
394.15	467	2.02	0.96	1.57	0.96	1.03	1.16	1.07	1.11	1.08	0.75	1.11
394.15	152	4.56	0.43	0.50	0.37	0.45	0.49	0.31	0.49	0.47	0.41	0.48

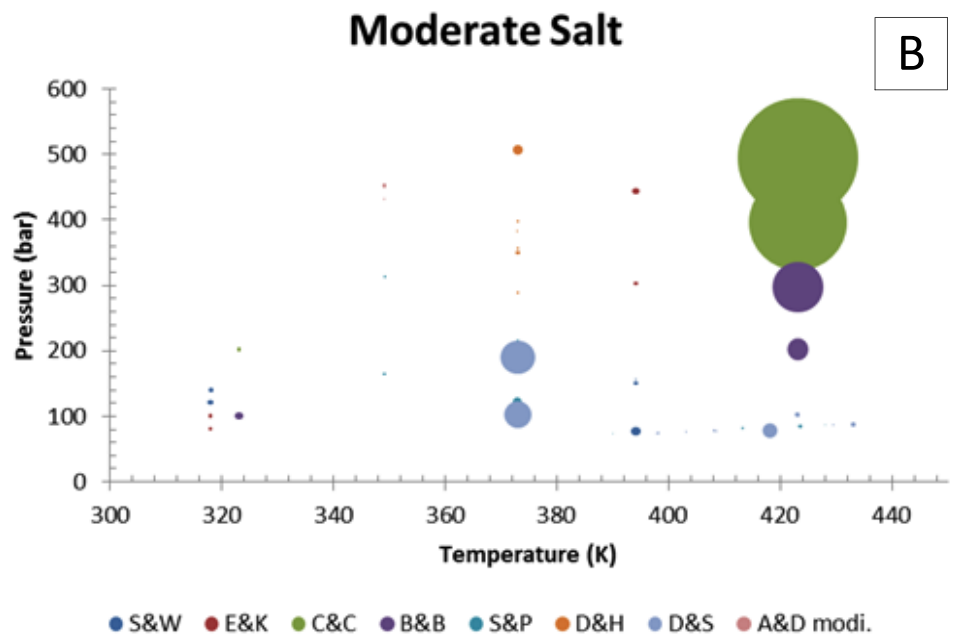
394.15	285	4.56	0.55	0.78	0.49	0.64	0.66	0.42	0.64	0.63	0.58	0.65
394.15	470	4.56	0.64	1.14	0.58	0.82	0.80	0.51	0.74	0.77	0.75	0.78
394.15	498	4.56	0.64	1.20	0.59	0.84	0.81	0.52	0.75	0.78	0.77	0.80
394.15	150	7.80	0.27	0.48	0.17	0.34	0.30	0.11	0.33	0.34	0.53	0.39
394.15	300	7.80	0.35	0.79	0.23	0.60	0.42	0.16	0.44	0.47	0.77	0.53
394.15	307	7.80	0.35	0.80	0.23	0.61	0.42	0.16	0.44	0.48	0.78	0.54
394.15	457	7.80	0.38	1.07	0.26	0.79	0.49	0.18	0.49	0.56	0.96	0.63
394.15	475	7.80	0.37	1.11	0.26	0.81	0.49	0.18	0.49	0.57	0.98	0.64
394.15	501	7.80	0.39	1.17	0.27	0.84	0.50	0.19	0.50	0.58	1.00	0.65
394.15	77	2.02	0.35	0.41	0.39	0.40	0.45	0.39	0.49	0.41	0.26	0.43
394.15	77	2.02	0.40	0.41	0.39	0.40	0.44	0.39	0.49	0.41	0.26	0.42
394.15	76	4.00	0.31	0.31	0.26	0.28	0.33	0.23	0.35	0.31	0.25	0.32
394.15	93	4.00	0.37	0.36	0.30	0.33	0.38	0.26	0.40	0.36	0.29	0.37
394.15	85	7.80	0.19	0.31	0.12	0.18	0.21	0.08	0.23	0.23	0.35	0.26
396.15	76	1.95	0.42	0.41	0.39	0.40	0.44	0.39	0.51	0.41	0.25	0.42
397.75	74	3.00	0.36	0.34	0.31	0.32	0.37	0.29	0.42	0.34	0.24	0.35
398.15	74	3.00	0.38	0.33	0.31	0.32	0.37	0.29	0.42	0.34	0.24	0.35
401.95	78	1.95	0.43	0.41	0.40	0.40	0.44	0.39	0.56	0.41	0.26	0.43
403.15	76	3.00	0.36	0.34	0.31	0.32	0.37	0.29	0.47	0.34	0.24	0.35
408.15	80	1.95	0.43	0.42	0.40	0.40	0.45	0.39	0.64	0.41	0.26	0.43
408.15	78	3.00	0.36	0.34	0.32	0.32	0.37	0.29	0.52	0.35	0.24	0.36
408.65	78	3.00	0.38	0.34	0.32	0.32	0.37	0.29	0.52	0.34	0.24	0.35
413.25	82	1.95	0.44	0.42	0.41	0.40	0.45	0.39	0.70	0.42	0.26	0.43
418.15	83	1.96	0.44	0.42	0.42	0.41	0.45	0.39	0.78	0.42	0.26	0.43
418.15	81	3.00	0.38	0.35	0.33	0.33	0.38	0.29	0.64	0.35	0.25	0.36
418.15	79	3.00	0.43	0.34	0.32	0.32	0.37	0.29	0.63	0.34	0.24	0.35
423.15	101	1.09	0.62	0.59	0.60	0.56	0.60	0.56	1.08	0.57	0.33	0.60
423.15	198	1.09	1.01	1.02	0.93	0.89	0.98	0.94	1.34	0.93	0.54	0.97
423.15	297	1.09	1.19	1.39	1.14	1.10	1.21	1.20	1.52	1.16	0.68	1.22

423.15	396	1.09	1.32	1.74	1.29	1.25	1.39	1.39	1.71	1.33	0.80	1.40
423.15	495	1.09	1.43	2.11	1.41	1.37	1.53	1.55	1.91	1.47	0.90	1.55
423.15	102	1.09	0.63	0.60	0.60	0.57	0.61	0.57	1.09	0.58	0.33	0.60
423.15	202	1.09	1.12	1.04	0.94	0.91	0.99	0.96	1.35	0.94	0.54	0.98
423.15	302	1.09	1.43	1.41	1.15	1.11	1.22	1.21	1.53	1.16	0.69	1.23
423.15	401	1.09	1.65	1.76	1.30	1.26	1.40	1.40	1.72	1.34	0.80	1.41
423.15	501	1.09	1.87	2.13	1.42	1.38	1.53	1.55	1.92	1.47	0.91	1.56
423.15	103	4.28	0.26	0.37	0.31	0.32	0.38	0.25	0.64	0.36	0.29	0.37
423.15	203	4.28	0.47	0.63	0.48	0.55	0.61	0.40	0.79	0.59	0.48	0.59
423.15	303	4.28	0.64	0.86	0.59	0.70	0.74	0.50	0.90	0.73	0.62	0.73
423.15	403	4.28	0.73	1.08	0.66	0.81	0.84	0.57	1.01	0.84	0.72	0.84
423.15	503	4.28	0.82	1.31	0.72	0.90	0.92	0.62	1.13	0.92	0.82	0.92
423.15	85	1.96	0.44	0.43	0.43	0.41	0.46	0.40	0.86	0.43	0.27	0.44
423.15	81	3.01	0.44	0.35	0.33	0.32	0.37	0.29	0.70	0.35	0.24	0.35
423.55	85	1.96	0.45	0.43	0.43	0.41	0.46	0.40	0.86	0.42	0.27	0.44
428.15	87	1.96	0.45	0.44	0.44	0.42	0.46	0.40	0.94	0.43	0.27	0.44
428.15	84	3.01	0.40	0.36	0.34	0.33	0.38	0.30	0.78	0.36	0.25	0.36
428.15	84	3.01	0.38	0.36	0.34	0.33	0.39	0.30	0.78	0.36	0.25	0.36
428.15	74	4.01	0.32	0.28	0.26	0.25	0.30	0.20	0.64	0.28	0.22	0.28
428.95	74	4.01	0.33	0.28	0.26	0.25	0.30	0.20	0.64	0.28	0.22	0.28
429.55	87	1.96	0.46	0.44	0.44	0.42	0.46	0.40	0.96	0.43	0.27	0.44
433	90	4.00	0.33	0.33	0.30	0.30	0.35	0.24	0.74	0.33	0.26	0.33
433	80	0.17	0.60	0.58	0.64	0.55	0.56	0.52	1.41	0.55	0.29	0.58
433	80	0.17	0.58	0.59	0.65	0.56	0.57	0.52	1.41	0.56	0.29	0.58
433	99	0.17	0.74	0.71	0.77	0.67	0.68	0.64	1.52	0.67	0.35	0.71
433	100	0.17	0.77	0.71	0.77	0.68	0.69	0.64	1.52	0.68	0.36	0.71
433.15	88	1.96	0.46	0.44	0.44	0.42	0.47	0.40	1.02	0.44	0.27	0.44
433.15	88	1.96	0.45	0.44	0.45	0.42	0.47	0.40	1.03	0.44	0.27	0.44
433.15	86	3.01	0.38	0.37	0.35	0.34	0.39	0.30	0.85	0.36	0.25	0.36

433.15	76	4.01	0.32	0.29	0.26	0.25	0.30	0.21	0.70	0.28	0.22	0.28
434.95	76	4.01	0.33	0.29	0.26	0.25	0.30	0.21	0.72	0.28	0.22	0.28

[a](ob): observed value,
[b](pr): predicted value





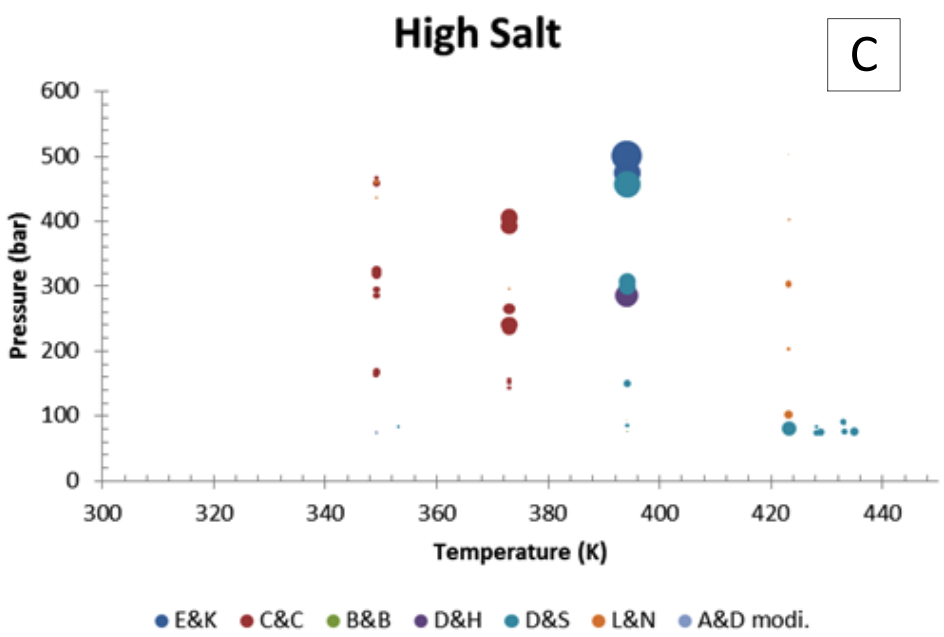


Figure S-3. Distribution of MMoPS Prediction Error over T-P space at A) low salt concentrations (<1 molal); B) moderate salt concentrations (1-3 molal); and C) high salt concentrations (3-8 molal). The size of the bubble represents the magnitude of the prediction error.
Models: A&D (modified): Akinfiev and Diamond coupled with Duan and Sun; B&B: Barta and Bradley; C&C: Chang and Coats; D&H: Darwish and Hilal; D&S: Duan and Sun; E&K: Enick and Klara; S&P: Spycher and Pruess; S&W: Soreide and Whitson; L&N: Li and Nghiem.

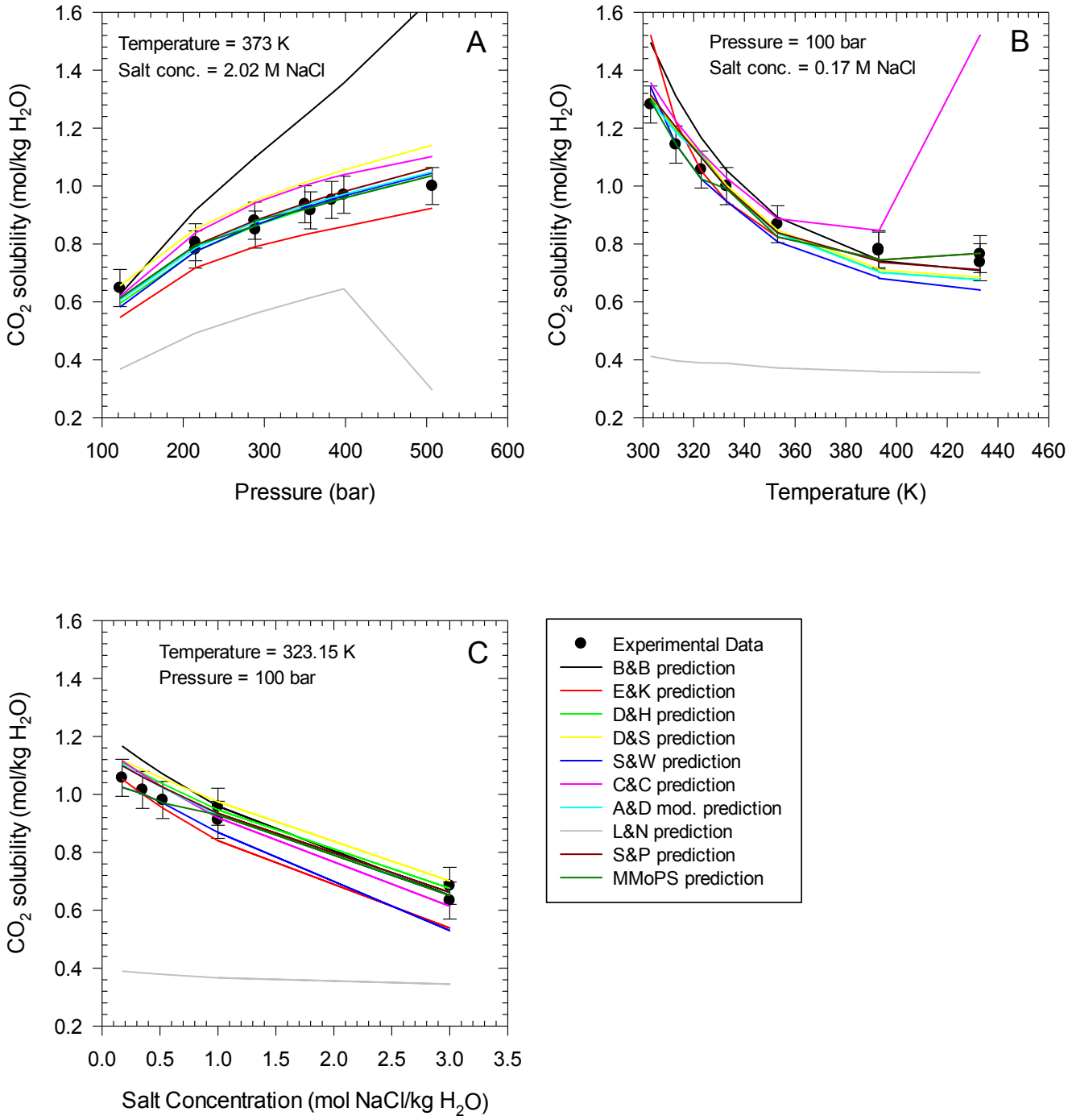


Figure S-4. Experimental and predicted CO_2 solubility values (mol/kg H_2O): A) as a function of pressure at constant temperature of 373 K and salt concentration of 2.02 mol NaCl/ kg H_2O; B) as a function of temperature at constant pressure of 100 bar and salt concentration of 0.17 mol NaCl/kg H_2O; and C) as a function of salt concentration at constant temperature of 323.15 K and pressure of 100 bar. The error bars on experimental data represent the mean of absolute measurement errors between laboratories and different experimental studies, which is 0.064 mol/kg.

Table S-8. Outlier Analysis for MMoPS Predictions

	T (K)	P (bar)	mNaCl (mol/kg)	Experimental Study	Observed Value (mol/kg)	MMoPS Model	MMoPS Predicted Value (mol/kg)	MMoPS Error (mol/kg)	Best-performing Model	Predicted Value using Best-performing Model (mol/kg)	Best-performing Model Error (mol/kg)
Outlier 1	423.15	495	1.09	Tekenouchi & Kennedy, 1964	1.43	C&C	1.91	0.48	E&K	1.41	-0.02
Related Sample 1	423.15	501	1.09	Gehrig, 1980	1.87	C&C	1.92	0.05	C&C	1.92	0.05
Outlier 2	423.15	396	1.09	Tekenouchi & Kennedy, 1964	1.32	C&C	1.71	0.39	A&D mod.	1.33	0.01
Related Sample 2	423.15	401	1.09	Gehrig, 1980	1.65	C&C	1.72	0.07	C&C	1.72	0.07
Outlier 3	423.15	297	1.09	Tekenouchi & Kennedy, 1964	1.19	B&B	1.39	0.20	S&W	1.20	0.01
Related Sample 3	423.15	302	1.09	Gehrig, 1980	1.43	B&B	1.41	-0.02	B&B	1.41	-0.02

F. Discussion of Experimental Data Set and Use of Both NaCl and $CaCl_2$ Data

As described in the main body of the paper, a combined NaCl and $CaCl_2$ data set is used to evaluate different CO_2 solubility models and to develop the MMoPS. In the combined data set, there are 113 NaCl and 60 $CaCl_2$ observations. Here, we also evaluate models using the NaCl and $CaCl_2$ data sets separately, and compare how these selections affect model performance.

Table S-9 summarizes differences in the RMSE, APE, the number of times a model provides the best estimate, and the number of times that MMoPS predicts the model will provide the best estimate, across the three subsets of observations: i) combined NaCl and $CaCl_2$ data set (n = 173, as evaluated in the main body of the paper); ii NaCl data only (n =113); and iii) $CaCl_2$ data only (n = 60). For each individual model, the performance does vary with the experimental data set. Some models perform better when only the NaCl data are used to evaluate them (e.g., the models of Spycher and Pruess; Darwish and Hilal; Soreide and Whitson; and Barta and Bradley), while for others adding the $CaCl_2$ data leads to an improvement in the assessed performance (e.g., the models of Enick and Klara; Chang and Coats; and Li and Nghiem). No matter which data set is used to evaluate the models, MMoPS outperforms the other models, achieving the lowest values in terms of RMSE and APE.

Insight is also gained by comparing the number of misclassifications of NaCl observations in MMoPS using the model fitted to the combined data set (MMoPS in the paper) vs. one fit using only the NaCl data. The original MMoPS failed to select the best model for 27 (24%) of the 113 NaCl observations, while refitting MMoPS using only the NaCl data resulted in misclassification (choosing the wrong model) for 28 (25%) of these 113 observations. Therefore, the addition of the $CaCl_2$ observations when fitting MMoPS does *not* degrade its ability to select the best fitting model for predicting the NaCl observations. Indeed, a very small (though not statistically significant) improvement is achieved.

Table S-9. Evaluation of models using different experimental data sets

Model	RMSE[a] (combined NaCl and $CaCl_2$ Data)	RMSE[a] (NaCl Data Only)	RMSE[a] ($CaCl_2$ Data Only)	APE[b] (combined NaCl and $CaCl_2$ Data)	APE[b] (NaCl Data Only)	APE[b] ($CaCl_2$ Data Only)	n[c] (in combined data set)	m[d] (in MMoPS)	n[e] (in NaCl data set)	m[f] (in refitted MMoPS)
Akinfiev and Diamond (modified)	0.076	0.075	0.078	9.5	6.7	14.9	30	32	29	26
Duan and Sun	0.082	0.079	0.087	9.7	7.3	14.2	30	31	7	8
Spycher and Pruess	0.084	0.070	0.105	11.2	6.4	20.1	18	15	7	10
Enick and Klara	0.122	0.133	0.099	14.2	10.5	21.2	22	19	15	14
Darwish and Hilal	0.129	0.092	0.178	15.9	8.8	29.4	9	8	7	4
Soreide and Whitson	0.155	0.108	0.218	21.2	12.7	37.3	21	20	18	16
Chang and Coats	0.217	0.263	0.070	25.9	33.5	11.4	25	27	22	27
Barta and Bradley	0.296	0.236	0.385	34.4	16.8	67.6	9	12	4	4
Li and Nghiem	0.457	0.511	0.330	50.6	48.3	54.8	9	9	4	4
						sum=	173	173	113	113
MMoPS[g]	0.061	0.069	0.041	4.9	4.3	6.8				

[a]RMSE: Root mean squared error; [b]APE: Average percent error; [c]*n* (in combined data set): Number of data points for which each model provides the most accurate prediction in the combined NaCl and $CaCl_2$ data set; [d]*m* (in MMoPS): Number of data points where Multi-model predictive system suggests using model *i* to predict CO_2 solubility in the combined data set; [e]*n* (in NaCl data set): Number of data points for which each model provides the most accurate prediction in NaCl data set; [f]*m* (in refitted MMoPS): The classification tree is regenerated using NaCl data only and m is the number of data points where refitted MMoPS (using NaCl data only) suggests using model *i* to predict CO_2 solubility in the NaCl data set; [g]MMoPS: This is the multi-model predictive system developed using the combined NaCl and $CaCl_2$ data set (same as the one in Table 3 of the body of the paper).

References in Appendix 1

Akinfiev, N. N.; Diamond, L. W. Thermodynamic model of aqueous CO_2 –H_2O-NaCl solutions from -22 to 100 ºC and from 0.1 to 100 Mpa. *Fluid Phase Equilibria* **2010**, 295, 104-124.

Barta, L.; Bradley, D. J. Extension of the specific interaction model to include gas solubilities in high temperature brines, *Geochim. Cosmochim. Acta* **1985**, **49,** 195–203.

Batzle, M.; Wang, Z. Seismic properties of pore fluids. *Geophysics* **1992**, 57(11), 1396–408.

Burnham, K. P.; Anderson, D. R. *Model selection and inference: a practical information-theoretic approach* ,2nd edition; Springer Press: New York, **2002**.

Chang, Yih-Bor; Coats, B. K.; Nolen, J. S. A compositional model for CO_2 floods including CO_2 solubility in water. *SPE Reservoir Evaluation & Engineering* **1998**, 1, 155-160.

Darwish, N. A.; Hilal N. A simple model for the prediction of CO_2 solubility in H_2O–NaCl system at geological sequestration conditions. *Desalination* **2010**, 260 (1-3), 114–118.

Duan, Z.; Sun, R. An improved model calculating CO_2 solubility in pure water and aqueous NaCl solutions from 273 to 533 K and from 0 to 2000 bar. *Chem. Geol.* **2003**, 193, 257–271.

Duan, Z.; Sun, R.; Zhu, C.; Chou, I-M. An improved model for the calculation of CO_2 solubility in aqueous solutions containing Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, and SO_4^{2-}. *Marine Chemistry*, **2006**, 98, 131–139.

Enick, R. M.; Klara, S. M. CO_2 solubility in water and brine under reservoir conditions. *Chem. Eng. Comm.* **1990**, 90, 23–33.

Li, Z.; Dong, M.; Li, S.; Dai, L. Densities and solubilities for binary systems of carbon dioxide + water and carbon dioxide + brine at 59 ºC and pressures to 29 MPa. *J.Chem. Eng. Data* **2004**, 49, 1026-1031.

Li, Y.; Ngheim, L. X. Phase equilibria of oil, gas and water/brine mixtures from a cubic equation of state and Henry's law. *Can. J. Chem. Eng.* **1986**, 64, 486-496.

Li, Y.-G.; Pitzer, K. S. Thermodynamics of aqueous sodium chloride solutions at high temperatures and pressures (I): thermodynamic properties over 373-573 K and 0.1-100 MPa. J. *Chem. Ind. Eng.* China **1986**, 1, 40-50.

Patel, N. C.; Teja, A. S. A new cubic equation of state for fluids and fluid mixtures. *Chem. Enq. Sci.* **1982**, 37, 463-473.

Pitzer, K. S.; Peiper, J. C., Busey, R. H. Thermodynamic properties of aqueous sodium chloride solutions. *J. Phys. Chem. Ref. Data* **1984**, 13, 1–102.

Robinson, D. B.; Peng, D. Y.; Chung, S. Y. Development of the the Peng-Robinson equation and its application to phase equilibrium in a system containing methanol. *Fluid Phase Equilibria* **1985**, 24, 25-41.

Soreide, I.; Whitson, C. H. Peng-Robinson predictions for hydrocarbons, CO_2, N_2, and H_2S with pure water and NaCl brine. *Fluid Phase Equilibria*, **1992**, 77, 217-240.

Spycher, N. F.; Reed, M. H. Fugacity coefficients of H_2, CO_2, CH_4, H_2O and of H_2O-CO_2-CH_4 mixtures: A virial equation treatment for moderate pressures and temperatures applicable to calculations of hydrothermal boiling, *Geochimica et Cosmochimica Acta* **1988**, 52, 739-749.

Zuo, Y.; Guo, T. M. Extension of the Patel–Teja equation of state to the prediction of the solubility of natural gas in formation water, *Chem. Eng. Sci.* **1991**, 46, 3251–3258.

APPENDIX 2: SUPPORTING INFORMATION FOR CHAPTER 3

A.

A.1 Governing equations and boundary conditions describing the confined two-aquifer system depicted in Figure 3.1.

For the injection zone:

$$\frac{\partial^2 s_i}{\partial r^2} + \frac{1}{r}\cdot\frac{\partial s_i}{\partial r} + \frac{K_c}{K_i h_i}\cdot\left.\frac{\partial s_c}{\partial z}\right|_{z=0} = \frac{S_i}{K_i}\cdot\frac{\partial s_i}{\partial t} \quad \text{(1a)}$$

$$s_i(r,0) = 0 \quad \text{(1b)}$$

$$s_i(\infty,t) = 0 \quad \text{(1c)}$$

$$\lim_{r\to 0} r\frac{\partial s_i}{\partial r} = -\frac{Q}{2\pi K_i h_i} \quad \text{(1d)}$$

For the monitoring zone:

$$\frac{\partial^2 s_m}{\partial r^2} + \frac{1}{r}\cdot\frac{\partial s_m}{\partial r} + \frac{K_c}{K_m h_m}\cdot\left.\frac{\partial s_c}{\partial z}\right|_{z=h_c} = \frac{S_m}{K_m}\cdot\frac{\partial s_m}{\partial t} \quad \text{(2a)}$$

$$s_m(r,0) = 0 \quad \text{(2b)}$$

$$s_m(\infty,t) = 0 \quad \text{(2c)}$$

$$\lim_{r\to 0}\frac{\partial s_m}{\partial r} = 0 \quad \text{(2d)}$$

For the sealing caprock:

$$\frac{\partial^2 s_c}{\partial z^2} = \frac{S_c}{K_c}\cdot\frac{\partial s_c}{\partial t} \quad \text{(3a)}$$

$$s_c(r,z,0) = 0 \quad \text{(3b)}$$

$$s_c(r,0,t) = s_i(r,t) \quad \text{(3c)}$$

$$s_c(r,h_c,t) = s_m(r,t) \quad \text{(3d)}$$

where $S_j = \rho_{fl}\cdot g\cdot\varphi_j\cdot c_t,\ (j = i,m,c)$ and $K_j = \frac{k_j\cdot\rho_{fl}\cdot g}{\mu_{fl}},\ (j = i,m,c)$

Based on Neuman-Witherspoon solution (Neuman & Witherspoon 1969; Chabora & Benson 2009), pressure buildups in the injection zone and the monitoring zone can be calculated by the following equations.

$$s_{Di} = \int_0^\infty (1 - e^{-y^2 r_{Di}^2 t_D})\{[1+G(y)]J_0[\omega_1(y)] + [1-G(y)]J_0[\omega_2(y)]\}\frac{dy}{y} \quad \text{(4)}$$

$$s_{Dm} = 2r_{Tm}^2 \int_0^\infty (1 - e^{-y^2 r_{Di}^2 t_D})\{J_0[\omega_1(y)] - J_0[\omega_2(y)]\} \frac{dy}{F(y)\sin y} \quad (5)$$

$$G(y) = M(y) / F(y)$$

$$\omega_1^2(y) = 0.5[N(y) + F(y)]$$

$$\omega_2^2(y) = 0.5[N(y) - F(y)]$$

$$F(y) = \sqrt{M^2(y) + 4(r_{Ti} \cdot r_{Tm} \cdot y / \sin y)^2}$$

$$M(y) = (r_{Di}^2 - r_{Dm}^2)y^2 - (r_{Ti}^2 - r_{Tm}^2)y \cdot \cot y$$

$$N(y) = (r_{Di}^2 + r_{Dm}^2)y^2 - (r_{Ti}^2 + r_{Tm}^2)y \cdot \cot y$$

$$t_D = \frac{K_i t}{S_i r^2}$$

$$s_{Dj} = \frac{4\pi K_i h_i s_j}{Q}$$

$$r_{Dj} = \frac{r}{h_c}\sqrt{\frac{K_c S_j}{S_c K_j}}$$

$$r_{Tj} = \frac{r}{h_c}\sqrt{\frac{K_c h_c}{K_j h_j}}$$

$$j = i, m$$

Symbols used in the equations above are defined in Table S1.

Table S1 Symbols used in equations

Property	Symbol
Formation thickness	h_i
Formation permeability	k_i
Formation porosity	φ_i
Total formation compressibility	c_t
Volumetric injection rate	Q
Injection duration	t
Radial distance from the injection well	r
Pressure	s_i
Dimensionless pressure	s_{Dj}
Specific storage	S_i
Hydraulic conductivity	K_i
Fluid density	ρ_{fl}
Fluid viscosity	μ_{fl}

Note: The subscript j represents either the injection zone (i), the above zone (m), or the caprock (c).

A.2 MATLAB codes implementing the Neuman-Witherspoon analytical solution

```
function [sDs,sDm]=neumanrevised(h,por,k,t)
g=9.81; %m/s2
%Fluid properties
rhofl=691.7; %kg/m3; density of CO2
visfl=5.61*10^-5; %Pa-s; viscosity of CO2
%Geologic Settings
ct=[1*10^-9 1*10^-9 1*10^-9]; %total compressibility of injection zone, cap rock
    and monitoring zone
S=rhofl*g*(por.*ct); %specific storage of injection zone, cap rock and monitoring
    zone
K=k*rhofl*g/visfl; %hydraulic conductivity of injection zone, cap rock and
    monitoring zone
r=0.5*10^3; %m distance between injection well and monitoring well
%Dimensionless parameters
tD=K(1)*t/(S(1)*r^2); % dimensionless time for injection zone
rD(1)=r/h(2)*(K(2)*S(1)/(K(1)*S(2)))^0.5;
rD(2)=r/h(2)*(K(2)*S(2)/(K(2)*S(2)))^0.5;
rD(3)=r/h(2)*(K(2)*S(3)/(K(3)*S(2)))^0.5;
rT(1)=r/h(2)*(K(2)*h(2)/(K(1)*h(1)))^0.5;
rT(2)=r/h(2)*(K(2)*h(2)/(K(2)*h(2)))^0.5;
rT(3)=r/h(2)*(K(2)*h(2)/(K(3)*h(3)))^0.5;
%Define functions
function f=integrand1(y)
M_y=(rD(1)^2-rD(3)^2)*y.^2-(rT(1)^2-rT(3)^2)*y.*cot(y);
N_y=(rD(1)^2+rD(3)^2)*y.^2-(rT(1)^2+rT(3)^2)*y.*cot(y);
F_y=(M_y.^2+4*(rT(1)*rT(3)*y./sin(y)).^2).^0.5;
G_y=M_y./F_y;
if 0.5*(N_y+F_y)<0
    if 0.5*(N_y-F_y)<0
        f=0;
    else
        w2_y=sqrt(0.5*(N_y-F_y));
        f=(1-exp(-y.^2*rD(1)^2*tD)).*((1-G_y).*besselj(0,w2_y))./y;
    end
else
    if 0.5*(N_y-F_y)<0
       w1_y=sqrt(0.5*(N_y+F_y));
       f=(1-exp(-y.^2*rD(1)^2*tD)).*((1+G_y).*besselj(0,w1_y))./y;
    else
        w1_y=sqrt(0.5*(N_y+F_y));
        w2_y=sqrt(0.5*(N_y-F_y));
        f=(1-exp(-y.^2*rD(1)^2*tD)).*((1+G_y).*besselj(0,w1_y)+(1-G_y).*besselj(0
            ,w2_y))./y;
    end
end
end
function f=integrand2(y)
M_y=(rD(1)^2-rD(3)^2)*y.^2-(rT(1)^2-rT(3)^2)*y.*cot(y);
N_y=(rD(1)^2+rD(3)^2)*y.^2-(rT(1)^2+rT(3)^2)*y.*cot(y);
F_y=(M_y.^2+4*(rT(1)*rT(3)*y./sin(y)).^2).^0.5;
if 0.5*(N_y+F_y)<0
    if 0.5*(N_y-F_y)<0
        f=0;
    else
        w2_y=sqrt(0.5*(N_y-F_y));
        f=(1-exp(-y.^2*rD(1)^2*tD)).*(0-besselj(0,w2_y))./(F_y.*sin(y));
```

```
    end
else
    if 0.5*(N_y-F_y)<0
       w1_y=sqrt(0.5*(N_y+F_y));
       f=(1-exp(-y.^2*rD(1)^2*tD)).*(besselj(0,w1_y)-0)./(F_y.*sin(y));
    else
        w1_y=sqrt(0.5*(N_y+F_y));
        w2_y=sqrt(0.5*(N_y-F_y));
        f=(1-exp(-y.^2*rD(1)^2*tD)).*(besselj(0,w1_y)-besselj(0,w2_y))./(F_y.*sin
            (y));
    end
end
end
sDs_sub=zeros(1,1000);
sDm_sub=zeros(1,1000);
tol1=1e-5;
tol2=1e-4;
for i=1:1000
    sDs_sub(i)=quad(@integrand1,(i-1)*pi,i*pi,tol1);
    sDm_sub(i)=2*rT(3)^2*quad(@integrand2,(i-1)*pi,i*pi,tol2);
end
    sDs=sum(sDs_sub);
    sDm=sum(sDm_sub);
end
```

A.3 Scatter plots of permeability vs. porosity based on the NPC database

The scatter plots of permeability vs. porosity and ln(Permeability) vs. ln(Porosity/(1-Porosity) based on the NPC database are shown in Figure S1. The red line is a linear regression between the transformed permeability and the transformed porosity variables. The correlation coefficient between permeability and porosity is 0.33 (R^2 = 0.11). The correlation coefficient between ln(Permeability) and ln(Porosity/(1-Porosity) is 0.43 (R^2 = 0.185) – in both cases less than 20% of the variance in permeability is explained by the relationship with porosity. From the scatter plots, we can see that for a certain range of porosity values, permeability is somewhat correlated with porosity. But as a whole, permeability is not strongly correlated with porosity. So rather than adding another parameter to the model (a weak correlation coefficient between permeability and porosity, or their transforms), we believe it is preferable to assume permeability and porosity are independent variables.

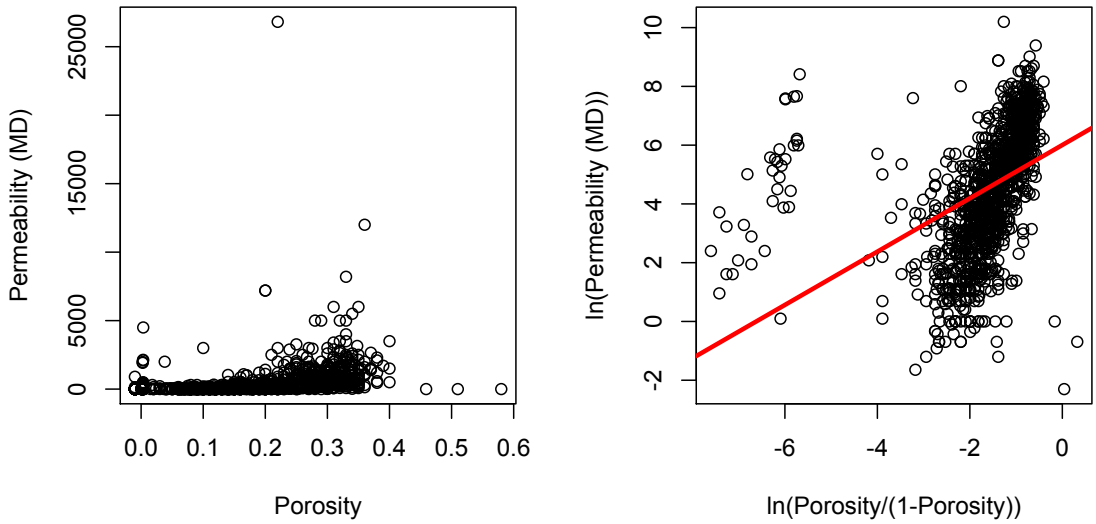
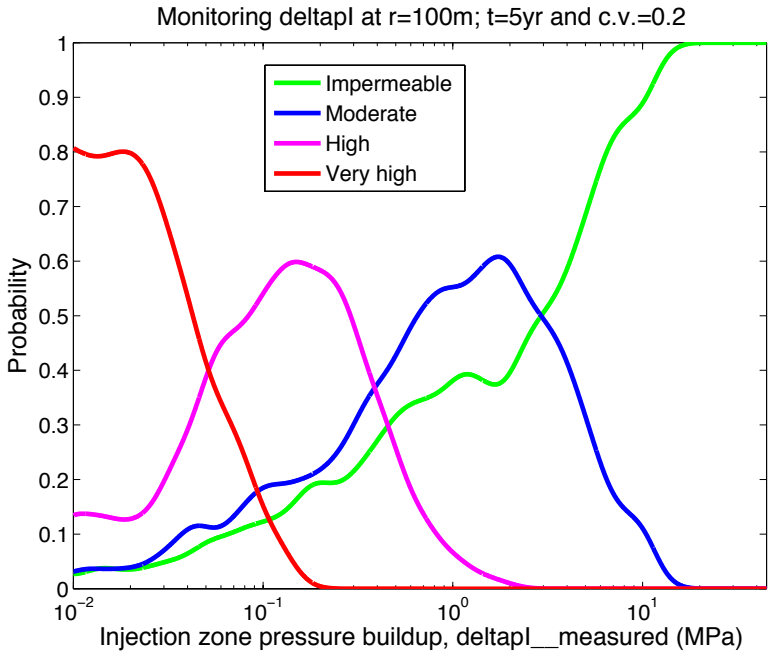


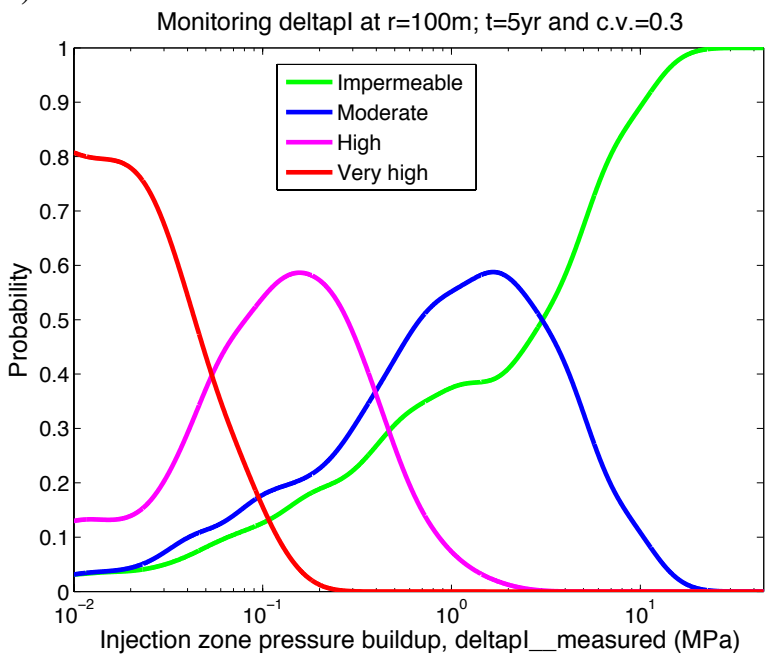
Figure S1 Scatter plots of permeability vs. porosity based on the NPC database

B. Calculation results for the location 100 m away from the injection well

a)



b)


Monitoring deltapI at r=100m; t=5yr and c.v.=0.3
Impermeable
Moderate
High
Very high
Probability
Injection zone pressure buildup, deltapI__measured (MPa)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
10^{-2}
10^{-1}
10^{0}
10^{1}

c)

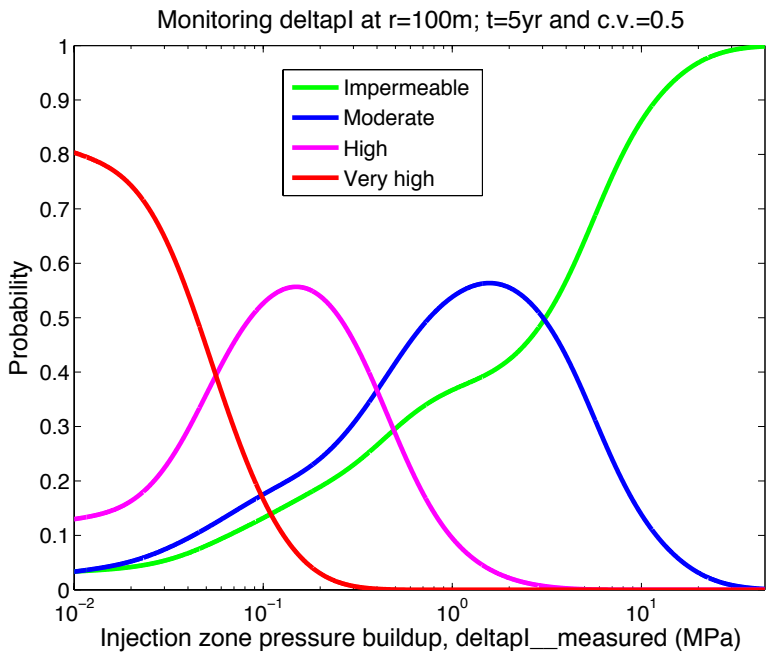


Figure S2 Posterior distribution of caprock permeability class monitored at r = 100 m; t = 5 year and a) low measurement error (c.v.=0.2); b) medium measurement error (c.v.=0.3); and c) high measurement error (c.v.=0.5). *deltapI_measured* is the measured pressure buildup in the injection zone.

References in Appendix 2

Chabora, E.R., Benson, S.M., 2009. Brine Displacement and Leakage Detection Using Pressure Measurements in Aquifers Overlying CO2 Storage Reservoirs. Energy Procedia 1(1), 2405–2412.

Neuman, S.P., Witherspoon, P.A., 1969. Theory of flow in a confined two aquifer system. Water Resources Research 5(4), 803–816.

APPENDIX 3: SUPPORTING INFORMATION FOR CHAPTER 4

A. Equations for estimating the bulk and shear moduli of rock frame

1) Estimate the shear modulus of the mineral matrix using the Hashin-Shtrikman bounds (Mavko et al., 2009)**:**
The Hashin-Shtrikman upper (+) and lower (-) bounds for the shear modulus of the mineral matrix is:

$$u^{HS\pm} = \Gamma[\xi(K_{\pm}, \mu_{\pm})]$$

where

$$\Gamma(\xi) = \langle \frac{1}{\mu + \xi} \rangle^{-1} - \xi$$

$$\xi(K_{\pm}, \mu_{\pm}) = \frac{\mu_{\pm}}{6}\left(\frac{9K_{\pm} + 8\mu_{\pm}}{K_{\pm} + 2\mu_{\pm}}\right)$$

The subscripts + and – denote the maximum and the minimum moduli of the individual mineral. The bracket $\langle \quad \rangle$ denotes an average over the minerals weighted by their volume fractions. The shear modulus of the mineral matrix is calculated as:

$$u_{matrix} = \frac{u^{HS+} + u^{HS-}}{2}$$

2) Estimate the bulk and shear moduli of the frame using the Hertz-Mindlin contact theory for consolidated rock (Mavko et al., 2009)**:**
The bulk and shear moduli at the critical porosity are given by:

$$K_{mc} = \left(\frac{C^2(1-\varphi_c)^2\mu_{matrix}^2 p_d}{18\pi^2(1-v_s)^2}\right)^{1/3}$$

$$\mu_{mc} = \frac{5-4v_s}{5(2-v_s)}\left(\frac{3C^2(1-\varphi_c)^2\mu_{matrix}^2 p_d}{2\pi^2(1-v_s)^2}\right)^{1/3}$$

where p_d is the differential pressure; v_s is the Poisson ratio of the minerals; φ_c is the critical porosity and C is the average number of contacts per spherical mineral. Based on data by Murphy (Carcione et al., 2006; Murphy, 1982), $C = 2.8/\varphi_c$.
The bulk and shear moduli of the frame are calculated as an arithmetic average of the Voigt and Wood moduli:

$$K_{frame} = \frac{K_V + K_W}{2}$$

$$\mu_{frame} = \frac{\mu_V + \mu_W}{2}$$

where

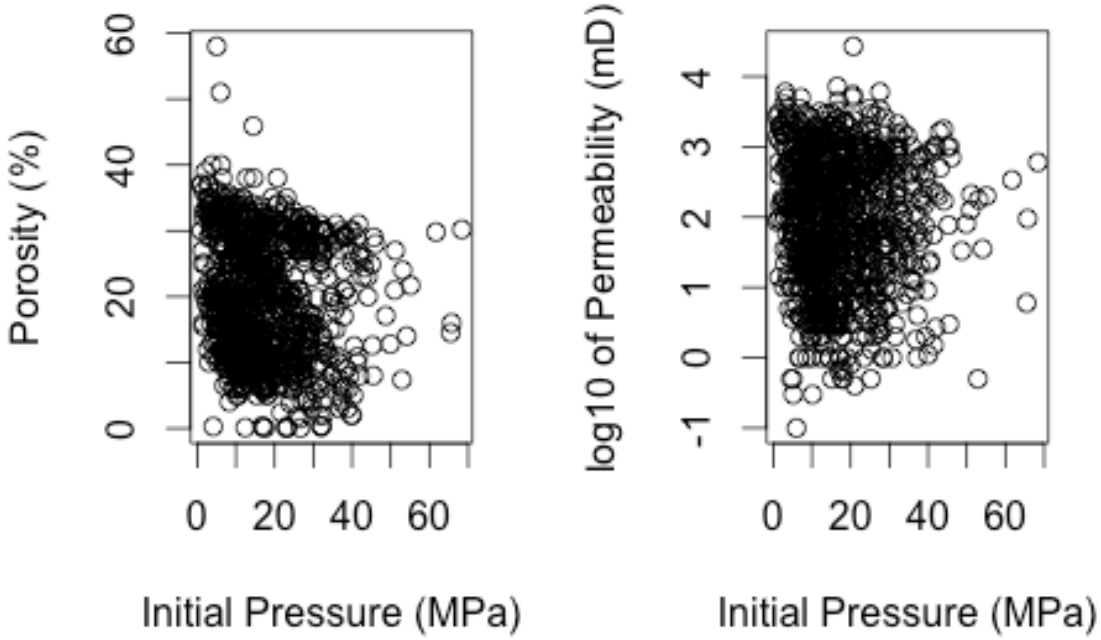
$$K_V = \left(1 - \frac{\varphi}{\varphi_c}\right) K_{matrix} + \frac{\varphi}{\varphi_C} K_{mc}$$

$$\mu_V = \left(1 - \frac{\varphi}{\varphi_c}\right) \mu_{matrix} + \frac{\varphi}{\varphi_C} \mu_{mc}$$

$$\frac{1}{K_W} = \frac{1 - \frac{\varphi}{\varphi_c}}{K_{matrix}} + \frac{\frac{\varphi}{\varphi_c}}{K_{mc}}$$

$$\frac{1}{\mu_W} = \frac{1 - \frac{\varphi}{\varphi_c}}{\mu_{matrix}} + \frac{\frac{\varphi}{\varphi_c}}{\mu_{mc}}$$

B. Scatter Plots of Porosity vs. Initial Pressure Data and Permeability vs. Initial Pressure Data in the NPC Database

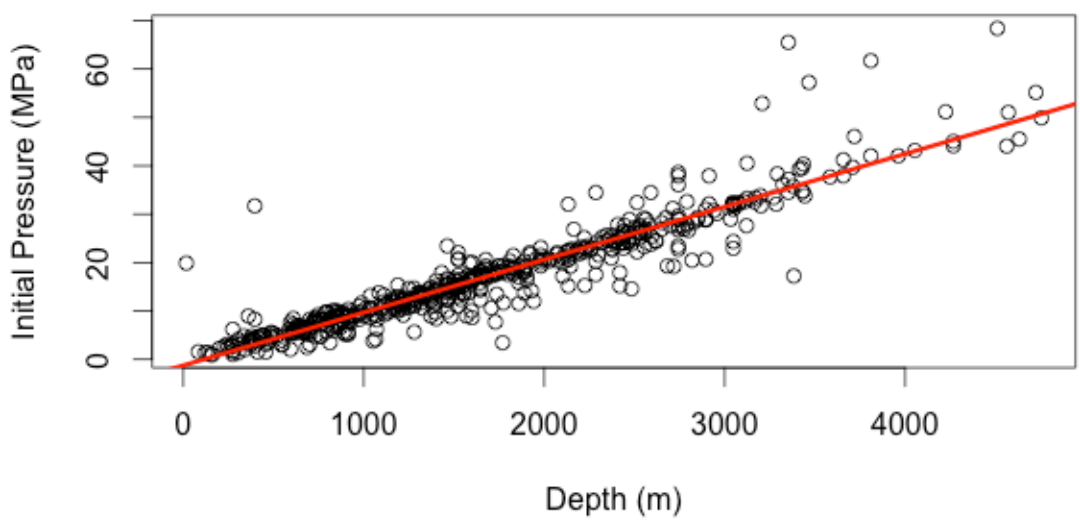


The correlation coefficient between porosity and initial pressure data in the NPC database is -0.13; while the correlation coefficient between permeability and initial pressure data is -0.04. As the porosity and permeability are only weakly correlated with the initial pressure based on the NPC database, the initial pressures are sampled independently for each case in this study.

C. Synthetic Dataset for Initial Pressure

The U.S. National Petroleum Council (NPC) database contains the initial pressure data for 1381 oil and gas reservoirs at various depth and lithology types. In our model, the top of the monitoring zone is fixed at a depth of 1000 m and the lithology type of the monitoring zone is sandstone. Therefore, a synthetic dataset for initial pressure at a depth of 1000 m and sandstone reservoirs has been generated based on the NPC database. The procedures of generating the synthetic dataset are as follows:

Step 1: Select a subset from the NPC database only for sandstone reservoirs. After removing "-1" values, this results in a dataset of 542 observations.
Step 2: Linear regression of initial pressure versus depth.



```
Call:
lm(formula = p_ini ~ depth)

Residuals:
     Min       1Q   Median       3Q      Max 
-18.4991  -1.1914   0.0173   1.0417  30.0975 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.2572571  0.3423904  -3.672 0.000265 ***
depth        0.0109341  0.0001776  61.554  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.936 on 540 degrees of freedom
Multiple R-squared:  0.8753,	Adjusted R-squared:  0.875 
F-statistic:  3789 on 1 and 540 DF,  p-value: < 2.2e-16
```

Step 3: Calculate the fitted value for initial pressure at a depth of 1000 m, the standard error of the prediction and the 95% prediction interval. The fitted initial pressure at 1000 m is 9.68 MPa. The standard error of the prediction is 3.94 MPa. The 95% prediction interval is (1.93, 17.42).

Step 4: Assume the initial pressure at 1000 m is normally distributed with mean of 9.68 MPa and standard deviation of 3.94 MPa. Generate 2000 normal random variables ~ N(9.68, 3.94). Remove numbers out of the 95% prediction interval (1.93, 17.42) MPa. The remaining random variables generated from the normal distribution form the synthetic dataset for initial pressure.

D. Equations for constructing the arbitrary polynomial chaos (aPC) basis

The polynomial of degree k (k=0,…,d) in the random variable $\omega \in \Omega$ is defined as (Oladyshkin and Nowak, 2012):

$$P^{(k)}(\omega) = \sum_{i=0}^{k} p_i^{(k)} \omega^i \qquad k = 0, \dots, d$$

where $p_i^{(k)}$ are the coefficients of ω^i in $P^{(k)}(\omega)$; d is the degree of expansion.

The coefficients $p_i^{(k)}$ in the aPC basis can be estimated by solving the system of linear equations:

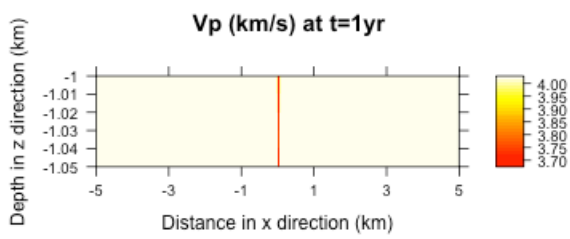
$$\begin{bmatrix} \mu_0 & \mu_1 & \dots & \mu_k \\ \mu_1 & \mu_2 & \dots & \mu_{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{k-1} & \mu_k & \dots & \mu_{2k-1} \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} p_0^{(k)} \\ p_1^{(k)} \\ \vdots \\ p_{k-1}^{(k)} \\ p_k^{(k)} \end{bmatrix} = \begin{bmatrix}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

where μ_k is the kth raw moment of the random variable ω.

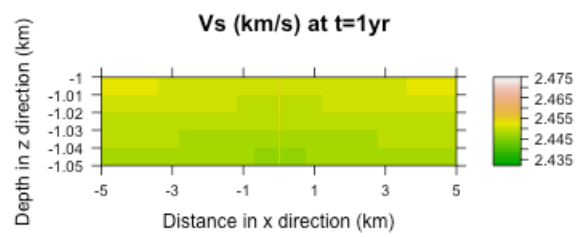
For detailed description of the aPC approach, the readers are referred to the paper by Oladyshkin and Nowak (Oladyshkin and Nowak, 2012).

E. Simulated seismic velocity in model domain

a)

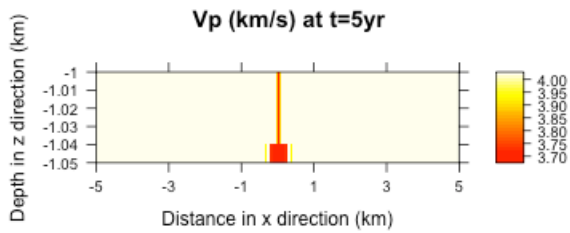


b)

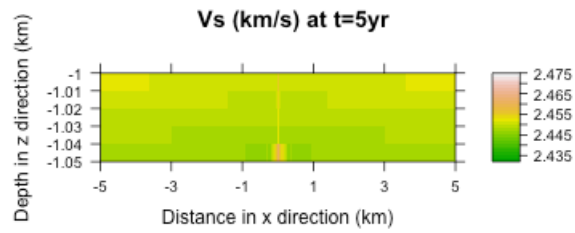


c)

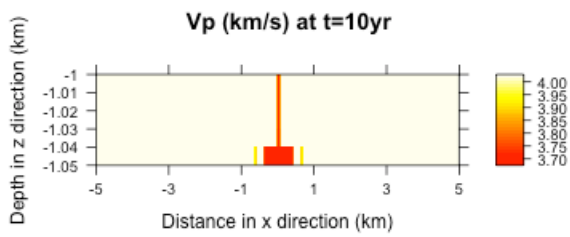
d)


Vp (km/s) at t=5yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
4.00
3.95
3.90
3.85
3.80
3.75
3.70

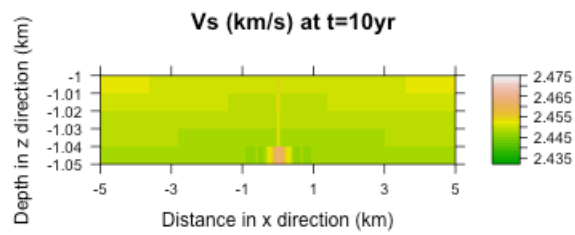
e)


Vs (km/s) at t=5yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
2.475
2.465
2.455
2.445
2.435

f)


Vp (km/s) at t=10yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
4.00
3.95
3.90
3.85
3.80
3.75
3.70

g)


Vs (km/s) at t=10yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
2.475
2.465
2.455
2.445
2.435

h)

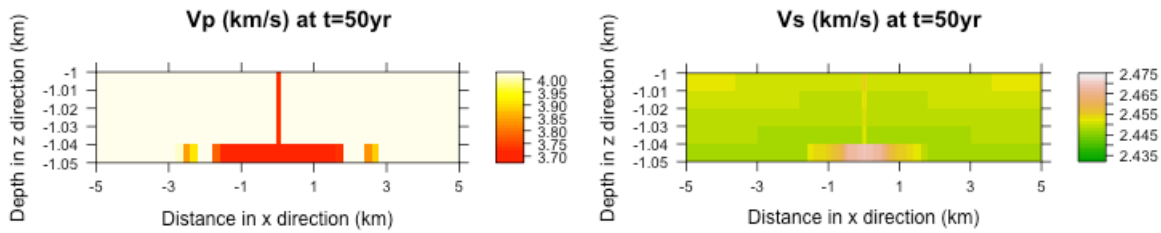
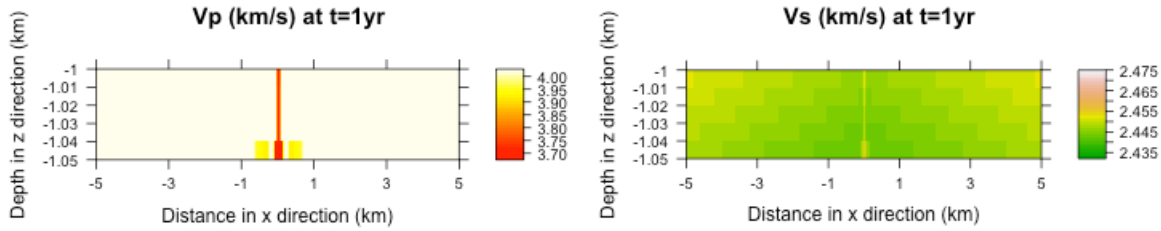
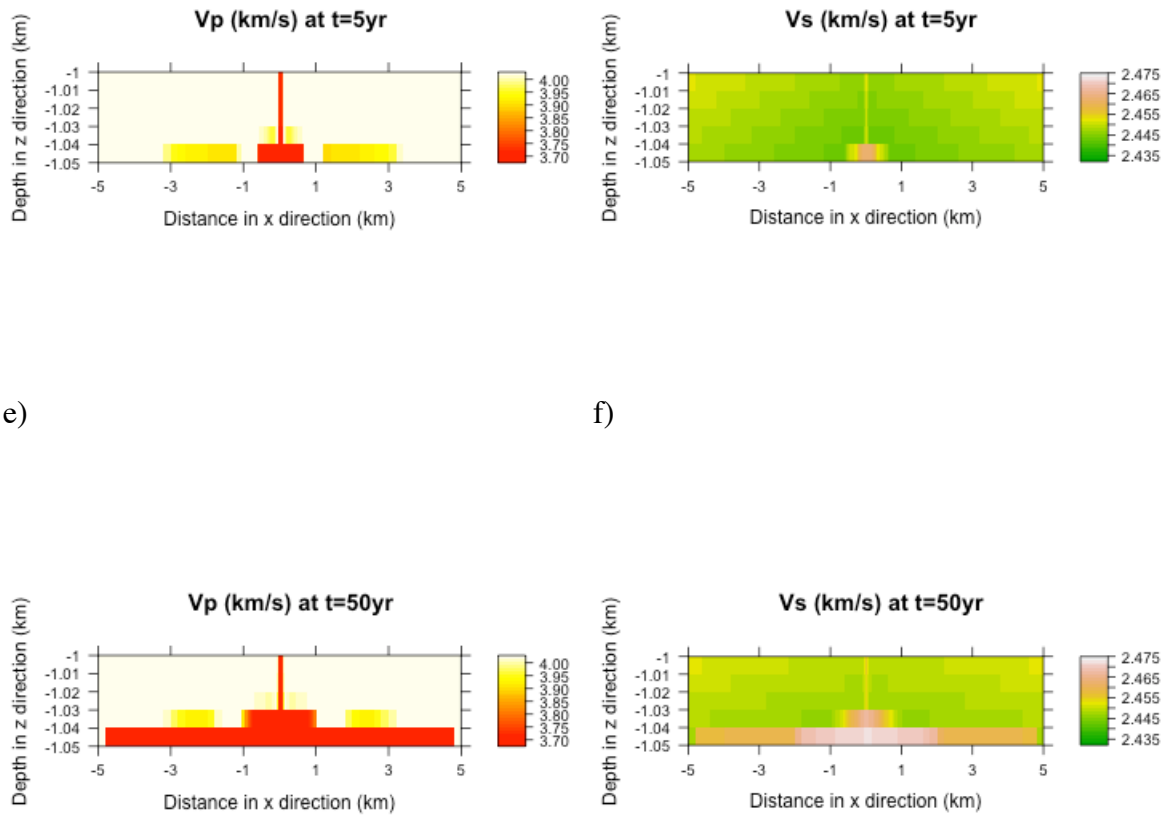


Figure E1 Sample modeling results of seismic velocity at the small leakage level at t=1, 5, 10 and 50 year

a) b)



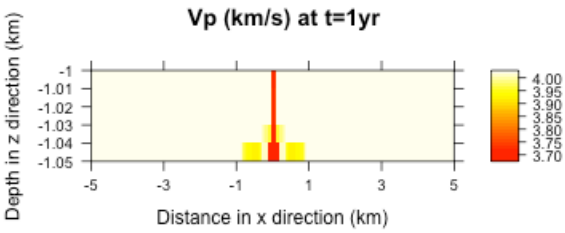
c) d)



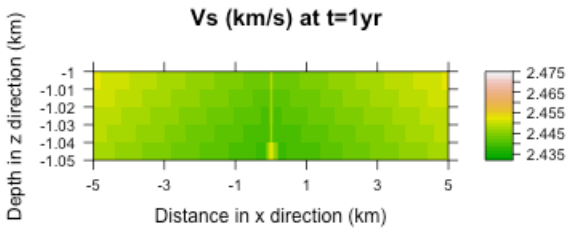
e) f)

Figure E2 Sample modeling results of seismic velocity at the moderate leakage level at t=1, 5, and 50 year (results at t=10 year are shown in the main text of the paper)

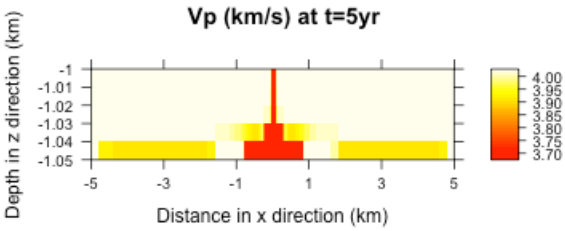
a)


Vp (km/s) at t=1yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
4.00
3.95
3.90
3.85
3.80
3.75
3.70

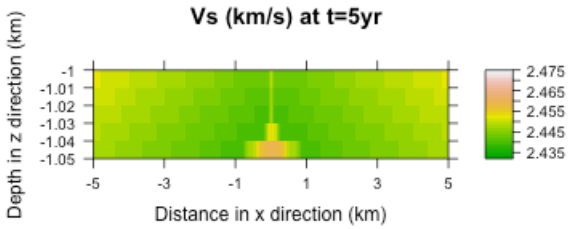
b)


Vs (km/s) at t=1yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
2.475
2.465
2.455
2.445
2.435

c)

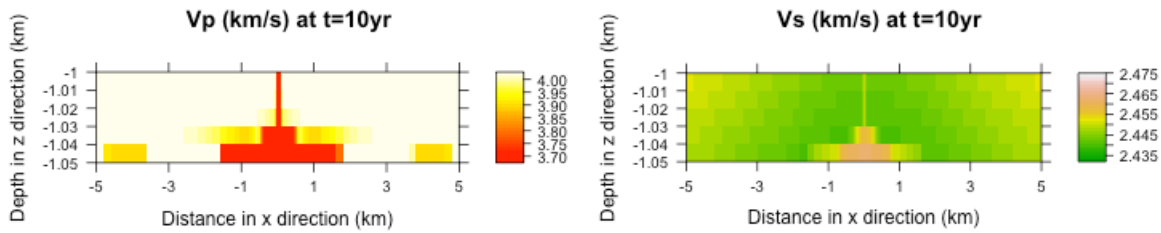

Vp (km/s) at t=5yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
4.00
3.95
3.90
3.85
3.80
3.75
3.70

d)


Vs (km/s) at t=5yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
2.475
2.465
2.455
2.445
2.435

e)

f)



g) h)

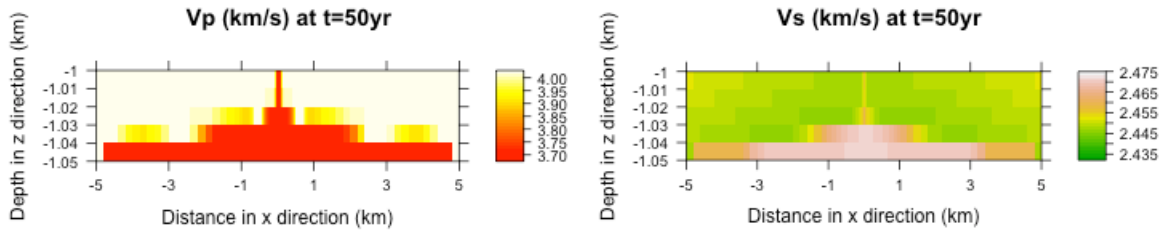
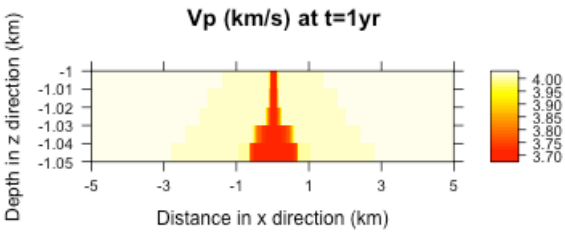
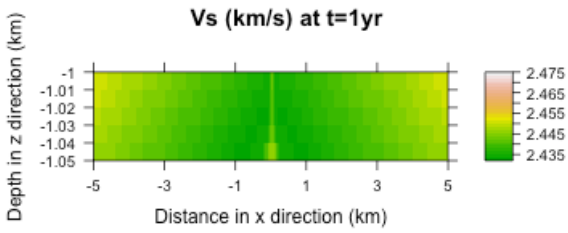


Figure E3 Sample modeling results of seismic velocity at the large leakage level at t=1, 5, 10 and 50 year

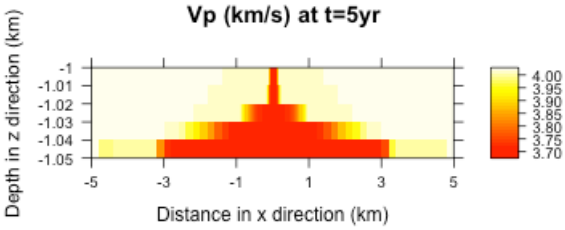
a)


Vp (km/s) at t=1yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
4.00
3.95
3.90
3.85
3.80
3.75
3.70

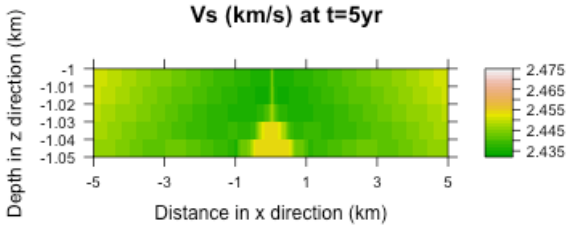
b)


Vs (km/s) at t=1yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
2.475
2.465
2.455
2.445
2.435

c)

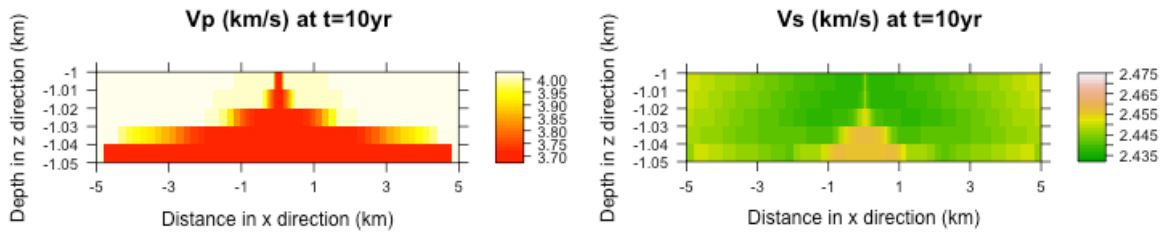

Vp (km/s) at t=5yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
4.00
3.95
3.90
3.85
3.80
3.75
3.70

d)


Vs (km/s) at t=5yr
Depth in z direction (km)
-1
-1.01
-1.02
-1.03
-1.04
-1.05
-5
-3
-1
1
3
5
Distance in x direction (km)
2.475
2.465
2.455
2.445
2.435

e)

f)



g) h)

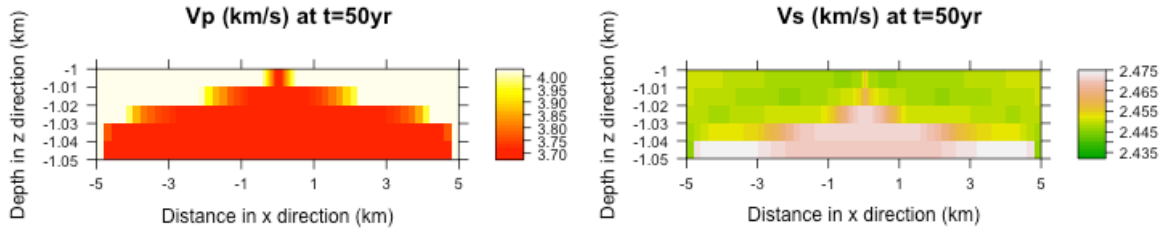


Figure E4 Sample modeling results of seismic velocity at the very large leakage level at t=1, 5, 10 and 50 year

F. Empirical CDF of detection power at x=500m

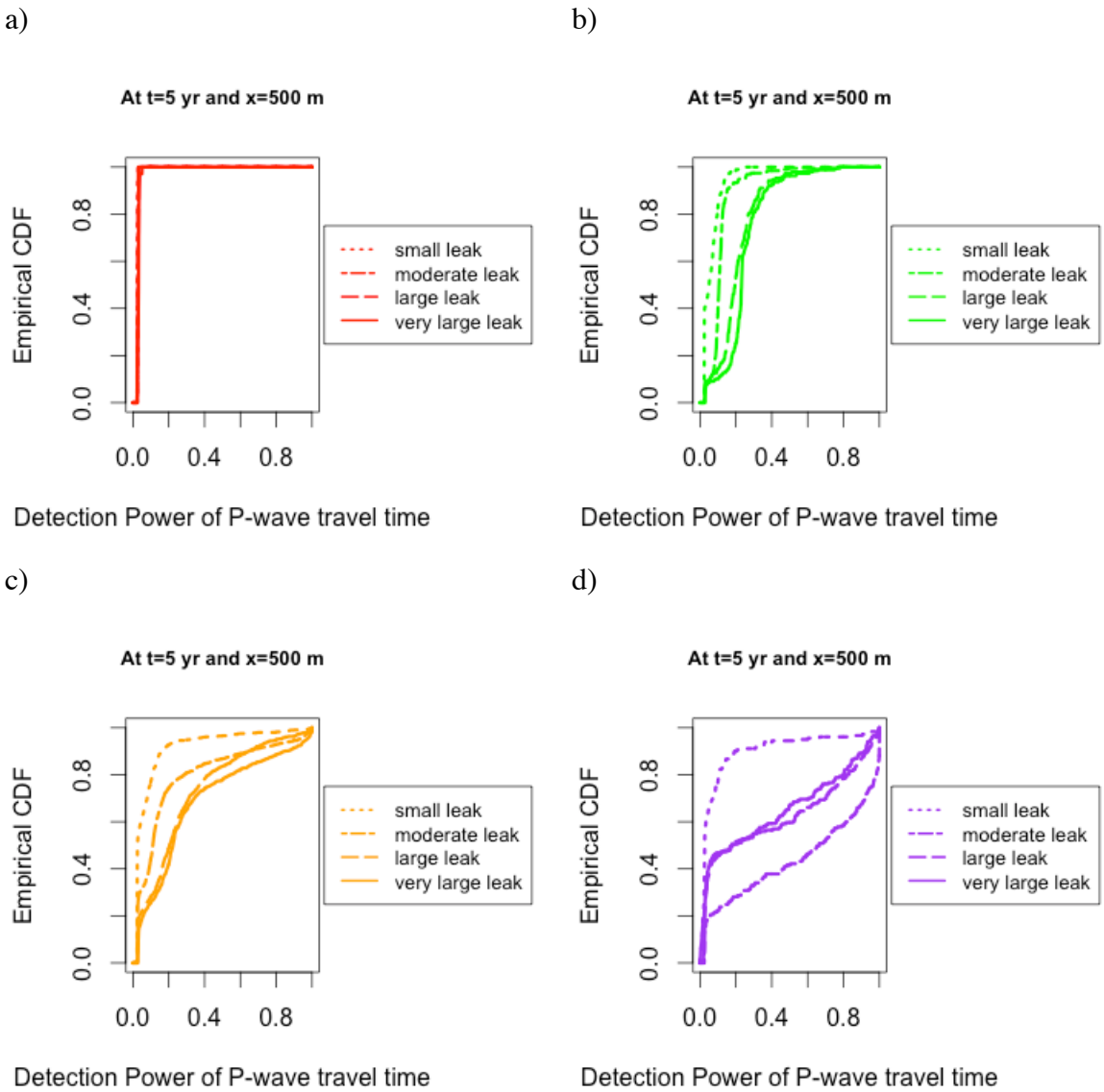


Figure F1 Empirical CDF of detection power of P-wave travel time at four leakage levels, t=5 year, x=500 m and at a) low porosity region (red points in Figure 4.4); b) high porosity and low permeability region (green points in Figure 3); c) high porosity and moderate permeability region (orange points in Figure 3) and d) high porosity and high permeability region (purple points in Figure 3).

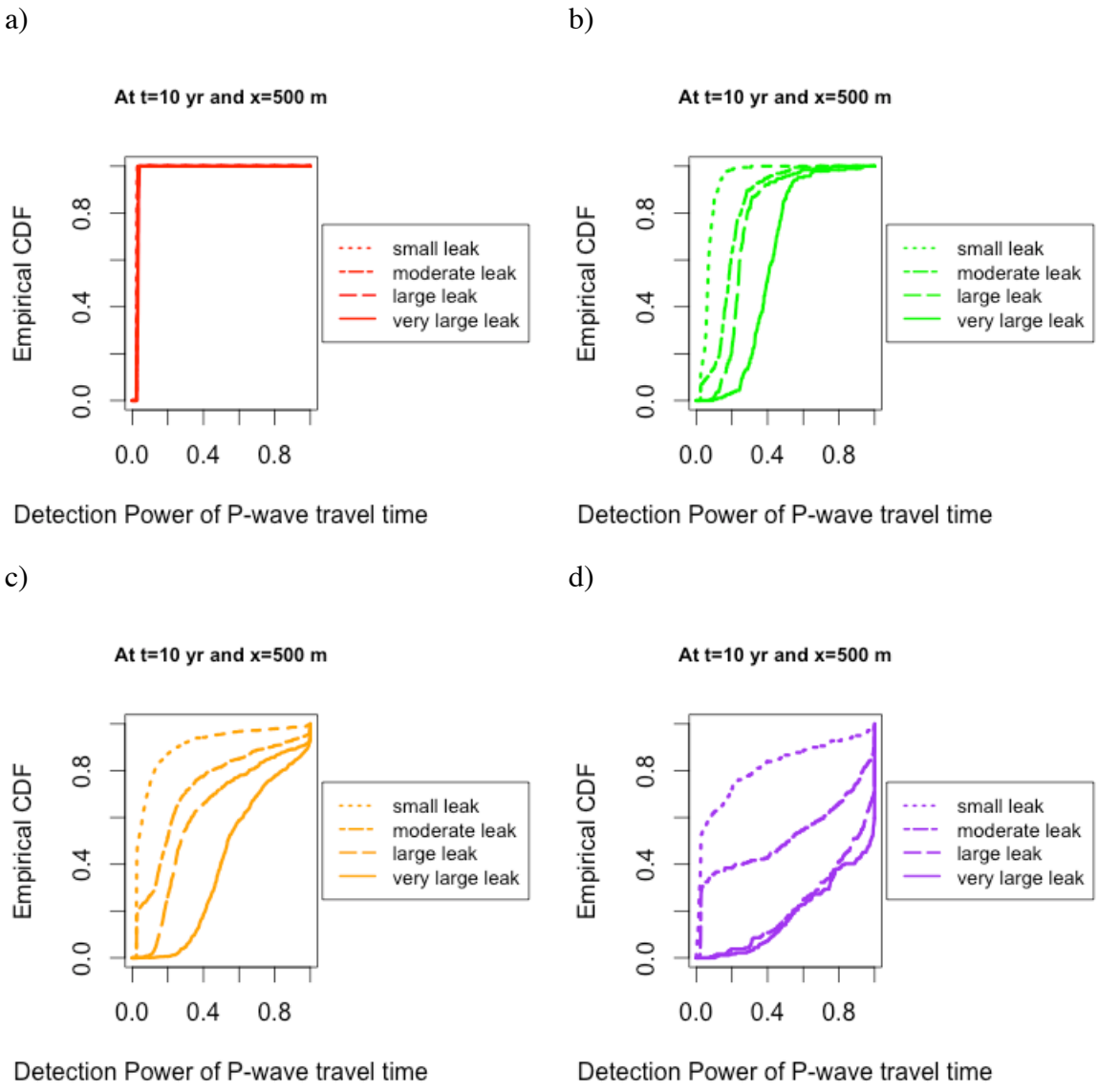


Figure F2 Empirical CDF of detection power of P-wave travel time at four leakage levels, t=10 year, x=500 m and at a) low porosity region (red points in Figure 4.4); b) high porosity and low permeability region (green points in Figure 4.4); c) high porosity and moderate permeability region (orange points in Figure 4.4) and d) high porosity and high permeability region (purple points in Figure 4.4).

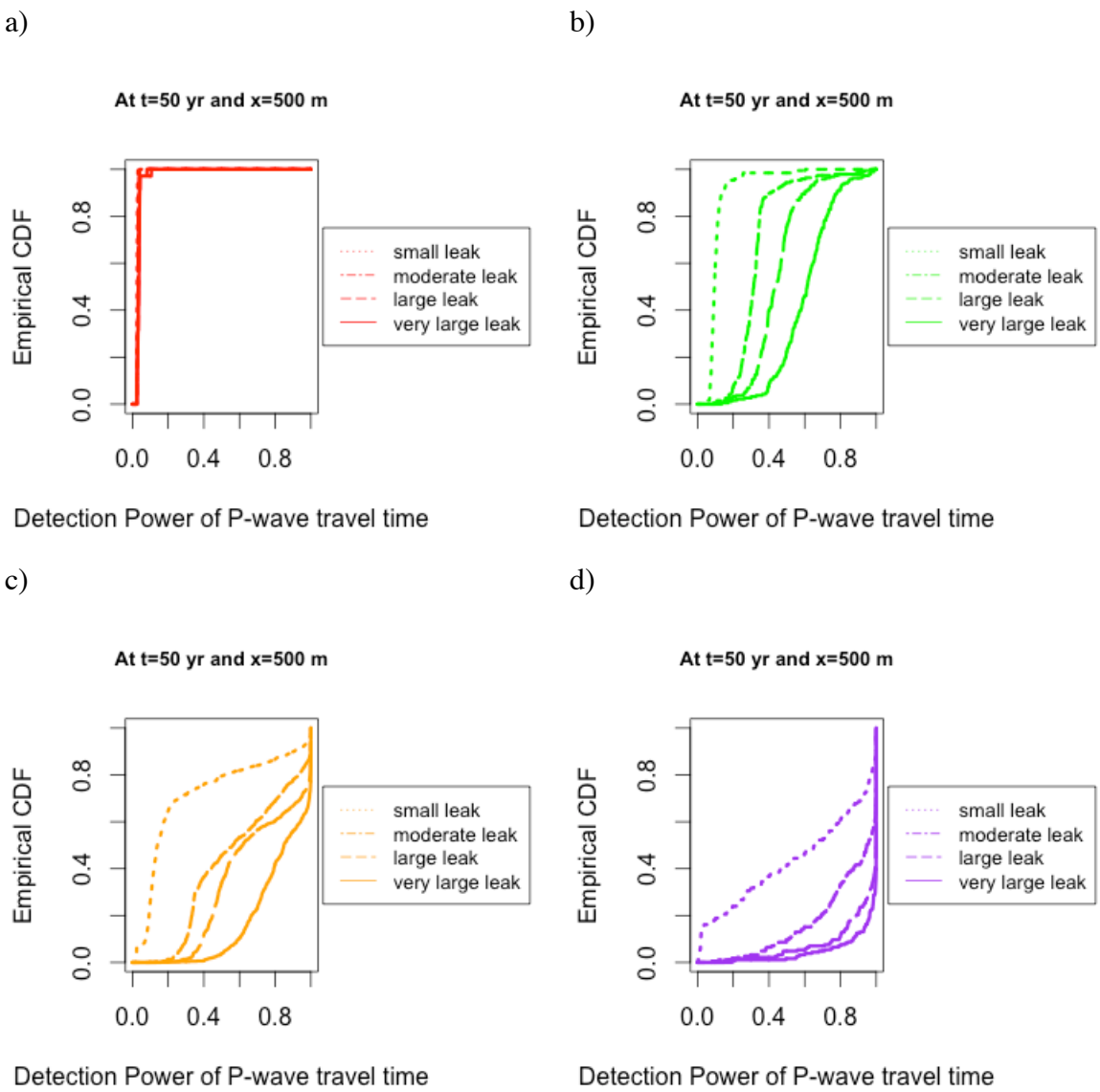


Figure F3 Empirical CDF of detection power of P-wave travel time at four leakage levels, t=50 year, x=500 m and at a) low porosity region (red points in Figure 4.4); b) high porosity and low permeability region (green points in Figure 4.4); c) high porosity and moderate permeability region (orange points in Figure 4.4) and d) high porosity and high permeability region (purple points in Figure 4.4).

References in Appendix 3

Carcione, J.M., Picotti, S., Gei, D., Rossi, G., 2006. Physics and Seismic Modeling for Monitoring CO2 Storage. Pure Appl. Geophys. 163, 175–207. doi:10.1007/s00024-005-0002-1

Mavko, G., Mukerji, T., Dvorkin, J., 2009. The Rock Physics Handbook: Tools for Seismic Analysis in Porous Media. Cambridge University Press, Cambridge, UK.

Murphy, W.F., 1982. Effects of microstructure and pore fluids on the acoustic properties of granular sedimentary materials. Stanford University.

Oladyshkin, S., Nowak, W., 2012. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. Reliab. Eng. Syst. Saf. 106, 179–190. doi:10.1016/j.ress.2012.05.002

References

Akinfiev, N.N., Diamond, L.W., 2003. Thermodynamic description of aqueous nonelectrolytes at infinite dilution over a wide range of state parameters. Geochim. Cosmochim. Acta 67, 613–627. doi:10.1016/S0016-7037(02)01141-9

Akinfiev, N.N., Diamond, L.W., 2010. Thermodynamic model of aqueous CO2-H2O-NaCl solutions from -22 to 100 ˚C and from 0.1 to 100MPa. Fluid Phase Equilib. 295, 104–124. doi:10.1016/j.fluid.2010.04.007

Alshuhail, A., Lawton, D.C., 2007. Time-lapse surface seismic monitoring of injected CO2 at the Penn West CO2-EOR site, Violet Grove, Alberta. CREWES Res. Rep. 19, 1–19.

Azzolina, N.A., Small, M.J., Nakles, D. V., Bromhal, G.S., 2014. Effectiveness of subsurface pressure monitoring for brine leakage detection in an uncertain CO2 sequestration system. Stoch. Environ. Res. Risk Assess. 28, 895–909.

Barta, L., Bradley, D. J., 1985. Extension of the specific interaction model to include gas solubilities in high temperature brines. Geochim. Cosmochim. Acta 49**,** 195–203.

Batzle, M., Wang, Z., 1992. Seismic properties of pore fluids. Geophysics 57, 1396–1408. doi:10.1190/1.1443207

Bennett, D. H., McKone, T. E., Kastenberg, W.E., 2000. Evaluating multimedia chemical persistence: Classification and regression tree analysis. Environ. Toxicol. Chem. 19, 810–819.

Benson, S.M., Myer, L., 2002. Monitoring to ensure safe and effective geologic storage of carbon dioxide. In: Proceedings of the Intergovernmental Panel on Climate Change (IPCC) Wokshop on Carbon Dioxide Capture and Storage. Regina, Saskatchewan, Canada.

Benson, S.M., Gasperikova, E. and Hoversten, G.M., 2004. Overview of monitoring techniques and protocols for geologic storage projects.

Birkholzer, J., Zhou, Q., Tsang, C., 2009. Large-scale impact of CO2 storage in deep saline aquifers: A sensitivity study on pressure response in stratified systems. Int. J. Greenh. Gas Control 3, 181–194. doi:10.1016/j.ijggc.2008.08.002

Born, M., 1920. Volumen und Hydratationswärme der Ionen. Zeitschrift für Physik. 1, 45–48.

Carcione, J.M., Picotti, S., Gei, D., Rossi, G., 2006. Physics and seismic modeling for monitoring CO2 storage. Pure Appl. Geophys. 163, 175–207. doi:10.1007/s00024-005-0002-1

Carlson, H., Colburn, A.P., 1942. Vapor-liquid equilibria of nonideal solutions: utilization of theoretical methods to extend data. Ind. Eng. Chem. 34, 581-589.

Celia, M.A., Nordbotten, J.M., Court, B., Dobossy, M., Bachu, S., 2011. Field-scale appli-cation of a semi-analytical model for estimation of CO_2 and brine leakage alongold wells. Int. J. Greenh. Gas Con. 5, 257–269.

Chabora, E.R., Benson, S.M., 2009. Brine displacement and leakage detection using pressure measurements in aquifers overlying CO2 storage reservoirs. Energy Procedia 1, 2405–2412. doi:10.1016/j.egypro.2009.01.313

Chadwick, R.., Zweigel, P., Gregersen, U., Kirby, G.., Holloway, S., Johannessen, P.., 2004. Geological reservoir characterization of a CO2 storage site: The Utsira Sand, Sleipner, northern North Sea. Energy 29, 1371–1381. doi:10.1016/j.energy.2004.03.071

Chang, Yih-Bor, Coats, B. K., Nolen, J. S., 1998. A compositional model for CO_2 floods including CO_2 solubility in water. SPE Reservoir Evaluation & Engineering 1, 155-160.

Chiaramonte, L., Zoback, M.D., Friedmann, J., Stamp, V., 2007. Seal integrity and feasibility of CO2 sequestration in the Teapot Dome EOR pilot: geomechanical site characterization. Environ. Geol. 54, 1667–1675. doi:10.1007/s00254-007-0948-7

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Quéré, C. Le, Myneni, R.B., Piao, S., Thornton, P., France, P.C., Willem, J., Friedlingstein, P., Munhoven, G., 2013. Carbon and Other Biogeochemical Cycles, Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. doi:10.1017/CBO9781107415324.015

Couëslan, M.L., Leetaru, H.E., Brice, T., Scott Leaney, W., McBride, J.H., 2009. Designing a seismic program for an industrial CCS site: Trials and tribulations. Energy Procedia 1, 2193–2200. doi:10.1016/j.egypro.2009.01.285

Daley, T.M., Solbau, R.D., Ajo-Franklin, J.B., Benson, S.M., 2007. Continuous active-source seismic monitoring of injection in a brine aquifer. Geophysics 72, A57–A61. doi:10.1190/1.2754716

Darwish, N. a., Hilal, N., 2010. A simple model for the prediction of CO2 solubility in H2O-NaCl system at geological sequestration conditions. Desalination 260, 114–118. doi:10.1016/j.desal.2010.04.056

Diamond, L.W., Akinfiev, N.N., 2003. Solubility of CO2 in water from -1.5 to 100°C and from 0.1 to 100 MPa: Evaluation of literature data and thermodynamic modelling. Fluid Phase Equilib. 208, 265–290. doi:10.1016/S0378-3812(03)00041-4

Dilks, D. W., Canale, R. P., Meier, P. G., 1992. Development of Bayesian Monte Carlo techniques for water quality model uncertainty. Ecological Modelling, 62(1), 149-162.

Doughty, C., Freifeld, B.M., Trautz, R.C., 2007. Site characterization for CO2 geologic storage and vice versa: the Frio brine pilot, Texas, USA as a case study. Environ. Geol. 54, 1635–1656. doi:10.1007/s00254-007-0942-0

Duan, Z., Moller, N., Weare, J. H., 1992. An equation of state for the CH_4–CO_2–H_2O system: I. Pure systems for 0 to 1000 ºC and 0 to 8000 bar. Geochim. Cosmochim. Acta 56, 2605– 2617.

Duan, Z., Sun, R., 2003. An improved model calculating CO2 solubility in pure water and aqueous NaCl solutions from 273 to 5333 K and from 0 to 2000 bar. Chem Geol 193, 257–271.

Duan, Z., Sun, R., Zhu, C., Chou, I.M., 2006. An improved model for the calculation of CO2 solubility in aqueous solutions containing Na+, K+, Ca2+, Mg2+, Cl-, and SO4 2-. Mar. Chem. 98, 131–139. doi:10.1016/j.marchem.2005.09.001

Eisenberg, J. N. S., Mckone, T. E., 1998. Decision tree method for the classification of chemical pollutants: Incorporation of across-chemical variability and within-chemical uncertainty. Environ. Sci. Technol. 32, 3396–3404.

Enick, R. M., Klara, S. M., 1990. CO_2 solubility in water and brine under reservoir conditions. Chem. Eng. Comm. 90, 23–33.

Gale, J., 2004. Geological storage of CO2: What do we know, where are the gaps and what more needs to be done? Energy 29, 1329–1338. doi:10.1016/j.energy.2004.03.068

Gander, W., Gautschi, W., 2000. Adaptive quadrature—revisited. BIT 40(1), 84–101.

Gassmann, F., 1951. Über die elastizität poröser medien. Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich 96, 1–23.

Gehrig, M., 1980. Phasengleichgewichte und pVT-daten ternarer Mischungen aus Wasser, Kohlendioxid und Natriumchlorid bis 3 kbar und 550 °C. Ph.D. Dissertation, Univ. Karlsrushe, Freiburg.

Goodman, A., Hakala, A., Bromhal, G., Deel, D., Rodosta, T., Frailey, S., Small, M., Allen, D., Romanov, V., Fazio, J., Huerta, N., McIntyre, D., Kutchko, B., Guthrie, G., 2011. U.S. DOE methodology for the development of geologic storage potential for carbon dioxide at the national and regional scale. Int. J. Greenh. Gas Control 5, 952–965. doi:10.1016/j.ijggc.2011.03.010

Green, S. T., Small, M. J., Casman, E. A., 2009. Determinants of national diarrheal disease burden. Environ. Sci. Technol. 43, 993-999.

Gutierrez, M., Katsuki, D., Almrabat, A., 2013. Use of the Biot-Gassmann equation in modeling of the seismic velocity changes during supercritical CO2 injection in sandstone. Poromechanics V 1–10.

Hartman, J. S., Weisberg, P. J., Pillai, R., Ericksen, J. A., Kuiken, T., Lindberg, S. E., Zhang, H., Rytuba, J. J., Gustin, M. S., 2009. Application of a Rule-Based Model to Estimate Mercury Exchange for Three Background Biomes in the Continental United States. Environ. Sci. Technol. 43, 4989–4994.

Harvey, A.H., Prausnitz, J.M., 1989. Thermodynamics of high-pressure aqueous systems containing gases and salts. AIChE J. 35, 635–644. doi:10.1002/aic.690350413

Hastle, T., Tibshirani, R., Friedman, J., 2001. *The Elements of Statistical Learning*: *Data Mining, Inference, and Prediction*, 1st edition; Springer Press: New York.

Hepple, R.P., Benson, S.M., 2005. Geologic storage of carbon dioxide as a climate change mitigation strategy: performance requirements and the implications of surface seepage. Environ. Geol. 47, 576–585. doi:10.1007/s00254-004-1181-2

Hill, P.G., 1990. A unified fundamental equation for the thermodynamic properties of H_2O. J. Phys. Chem. Ref. Data 19, 1233-1274.

Holloway, S., 2005. Underground sequestration of carbon dioxide—a viable greenhouse gas mitigation option. Energy 30, 2318–2333. doi:10.1016/j.energy.2003.10.023

Hughes, A. S., VanBriesen, J. M., Small, M. J., 2010. Identification of structural properties associated with polychlorinated biphenyl dechlorination processes. Environ. Sci. Technol. 44, 2842-2848.

Humez, P., Audigane, P., Lions, J., Chiaberge, C., Bellenfant, G., 2011. Modeling of CO2 leakage up through an abandoned well from deep saline aquifer to shallow fresh groundwaters. Transp. Porous Media 90, 153–181.

IPCC (Intergovernmental Panel on Climate Change), 2005. IPCC Special Report on Carbon Dioxide Capture and Storage. Cambridge University Press, New York.

Isukapalli, S.S., Roy, A., Georgopoulosl, P.G., 1998. Stochastic response surface methods (SRSMs) for uncertainty propagation: Application to environmental and biological systems. Risk Anal. 18.

Jenkins, C., 2013. Statistical aspects of monitoring and verification. International Journal of Greenhouse Gas Control 13, 215–229.

Jung, Y., Zhou, Q., Birkholzer, J.T., 2013. Early detection of brine and CO2 leakage through abandoned wells using pressure and surface-deformation monitoring data: Concept and demonstration. Adv. Water Resour. 62, 555–569. doi:10.1016/j.advwatres.2013.06.008

Khatiwada, M., Adam, L., Morrison, M., van Wijk, K., 2012. A feasibility study of time-lapse seismic monitoring of CO2 sequestration in a layered basalt reservoir. J. Appl. Geophys. 82, 145–152. doi:10.1016/j.jappgeo.2012.03.005

Koschel, D., Coxam, J.Y., Rodier, L., Majer, V., 2006. Enthalpy and solubility data of CO2 in water and NaCl(aq) at conditions of interest for geological sequestration. Fluid Phase Equilib. 247, 107–120. doi:10.1016/j.fluid.2006.06.006

Kumar, D., 2006. A Tutorial on Gassmann Fluid Substitution: Formulation, Algorithm and Matlab Code. GEOHORIZONS 4–12.

Li, H., Zhang, D., 2007. Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods. Water Resour. Res. 43, 1–13. doi:10.1029/2006WR005673

Li, Y., Ngheim, L. X., 1986. Phase equilibria of oil, gas and water/brine mixtures from a cubic equation of state and Henry's law. Can. J. Chem. Eng. 64, 486-496.

Li, Y.-G., Pitzer, K. S., 1986. Thermodynamics of aqueous sodium chloride solutions at high temperatures and pressures (I): thermodynamic properties over 373-573 K and 0.1-100 MPa. J. Chem. Ind. Eng. China 1, 40-50.

Liu, F., Lu, P., Griffith, C., Hedges, S.W., Soong, Y., Hellevang, H., Zhu, C., 2012. CO2–brine–caprock interaction: Reactivity experiments on Eau Claire shale and a review of relevant literature. Int. J. Greenh. Gas Control 7, 153–167. doi:10.1016/j.ijggc.2012.01.012

Liu, Y., Hou, M., Yang, G., Han, B., 2011. Solubility of CO2 in aqueous solutions of NaCl, KCl, CaCl 2 and their mixed salts at different temperatures and pressures. J. Supercrit. Fluids 56, 125–129. doi:10.1016/j.supflu.2010.12.003

Mavko, G., Mukerji, T., Dvorkin, J., 1998. The Rock Physics Handbook: Tools for Seismic Analysis in Porous Media. Cambridge University Press, Cambridge, UK.

Meckel, T.A., Hovorka, S.D., Kalyanaraman, N., 2008. Continuous pressure monitoring for large volume CO 2 injections GCCC Digital Publication Series # 08-03a C.

Neuman, S.P., Witherspoon, P.A., 1969. Theory of flow in a confined two aquifer system. Water Resources Research 5(4), 803–816.

Nordbotten, J.A.N.M., Kavetski, D., Celia, M.A., 2009. Model for CO2 leakage including multiple geological layers and multiple leaky wells 43, 743–749.

Nordbotten, J.M., Celia, M.A., Bachu, S., 2005a. Injection and storage of CO_2 in deep saline aquifers: analytical solution for CO_2 plume evolution during injection. Transport Porous Med 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.

NPC, 1984. U.S. NPC Public Database [WWW Document]. URL http://www.netl.doe.gov/research/oil-and-gas/software/databases (accessed 5.5.14).

Oladyshkin, S., Class, H., Helmig, R., Nowak, W., 2011. A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations. Adv. Water Resour. 34, 1508–1518. doi:10.1016/j.advwatres.2011.08.005

Oladyshkin, S., Nowak, W., 2012. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. Reliab. Eng. Syst. Saf. 106, 179–190. doi:10.1016/j.ress.2012.05.002

Patel, N.C., Teja, A.S., 1982. A new cubic equation of state for fluids and fluid mixtures. Chem. Eng. Sci. 37, 463–473. doi:10.1016/0009-2509(82)80099-7

Pinder, R. W., Pekney, N. J., Davidson, C. I., Adams, P. J., 2004. A process-based model of ammonia emissions from dairy cows: improved temporal and spatial resolution. Atmospheric Environment, 38(9), 1357-1365.

Pitzer, K.S., 1973. Thermodytriarnics of Electrolytes. I. Theoretical Basis and General Equations. J. Phys. Chem. 77, 268–277. doi:10.1021/j100621a026

Pruess, K., 2005. ECO2N: A TOUGH2 fluid property module for mixtures of water, NaCl, and CO2. University of California, Berkeley, CA.

Pruess, K., Oldenburg, C., Moridis, G., 2012. TOUGH2 User's Guide, Version 2. University of California, Berkeley, CA.

Reiss, H., Frisch, H.L., Helfand, E., Lebowitz, J. L., 1960. Aspects of the Statistical Thermodynamics of Real Fluids, J. Chem. Phys. 32, 119-124.

Ren, S., 2003. Phenol Mechanism of Toxic Action Classification and Prediction: A Decision Tree Approach. Toxicol. Lett. 144, 313–323.

Robinson, D. B., Peng, D. Y., Chung, S. Y., 1985. Development of the the Peng-Robinson equation and its application to phase equilibrium in a system containing methanol. Fluid Phase Equilib. 24, 25-41.

Rowe, A.M., Chou, J.C.S., 1970. Pressure-Volume-Temperature-Concentration Relation of Aqueous NaCl Solutions. J. Chem. Eng. Data 15, 61–66. doi:10.1021/je60044a016

Setschenow, M., 1889. Uberdiekonstitution der Salzlosungen auf Grund ihres Verhaltens zu Kohlensaure. Zeitschrift fiir Physikalische Chemie. 4, 117-128.

Silver, P.G., Daley, T.M., Niu, F., Majer, E.L., 2007. Active source monitoring of cross-well seismic travel time for stress-induced changes. Bull. Seismol. Soc. Am. 97, 281–293. doi:10.1785/0120060120

Sohn, M. D., Small, M. J., Pantazidou, M, 2000. Reducing uncertainty in site characterization using Bayes Monte Carlo methods. Journal of Environmental Engineering, 126(10), 893-902.

Song, X., Dai, K., Chen, G., Pan, Y., Zhong, Z., 2014. Sensitivity study of surface waves for CO2 storage monitoring. Int. J. Greenh. Gas Control 20, 180–188. doi:10.1016/j.ijggc.2013.11.008

Soreide, I., Whitson, C. H., 1992. Peng-Robinson predictions for hydrocarbons, CO_2, N_2, and H_2S with pure water and NaCl brine. Fluid Phase Equilib. 77, 217-240.

Span, R., Wagner, W., 1996. A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100 K at pressures up to 800 MPa. J. Phys. Chem. Ref. Data 25, 1509-1596.

Spycher, N., Pruess, K., 2010. A Phase-partitioning model for CO2-brine mixtures at elevated temperatures and pressures: Application to CO2-enhanced geothermal systems. Transp. Porous Media 82, 173–196. doi:10.1007/s11242-009-9425-y

Spycher, N., Pruess, K., Ennis-King, J., 2003. CO 2-H 2O mixtures in the geological sequestration of CO 2. I. Assessment and calculation of mutual solubilities from 12 to 100??C and up to 600 bar. Geochim. Cosmochim. Acta 67, 3015–3031. doi:10.1016/S0016-7037(03)00273-4

Spycher, N.F., Reed, M.H., 1988. Fugacity coefficients of H2, CO2, CH4, H2O and of H2O- CO2-CH4 mixtures: A virial equation treatment for moderate pressures and temperatures applicable to calculations of hydrothermal boiling. Geochim. Cosmochim. Acta 52, 739–749. doi:10.1016/0016-7037(88)90334-1

Sun, A.Y., Nicot, J.-P., Zhang, X., 2013a. Optimal design of pressure-based, leakage detection monitoring networks for geologic carbon sequestration repositories. Int. J. Greenh. Gas Control 19, 251–261. doi:10.1016/j.ijggc.2013.09.005

Sun, A.Y., Zeidouni, M., Nicot, J.-P., Lu, Z., Zhang, D., 2013b. Assessing leakage detectability at geologic CO2 sequestration sites using the probabilistic collocation method. Adv. Water Resour. 56, 49–60. doi:10.1016/j.advwatres.2012.11.017

Takenouchi, S., Kennedy, G. C., 1965. The solubility of carbon dioxide in NaCl solutions at high temperatures and pressures. Am. J. Sci. 263, 445–454.

Tatang, M.A., Pan, W., Prinn, R.G., McRae, G.J., 1997. An efficient method for parametric uncertainty analysis of numerical geophysical models. J. Geophys. Res. 102, 21925–21932.

U.S. Brine Wells Database, DOE/NETL-2003/119, 2003. U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV.

USDOE, 2009. Monitoring, Verification, and Accounting of CO2 Stored in Deep Geologic Formations. Morgantown, WV, USA.

USDOE, 2010. Carbon Sequestration Atlas of the United Sates and Canada. Morgantown, WV, USA.

Wang, Z., Cates, M.E., Langan, R.T., 1998. Seismc monitoring of a CO2 flood in a carbonate resevoir: A rock physics study. Geophysics 63, 1604–1617.

Wang, Z., Small, M.J., 2014. A Bayesian approach to CO2 leakage detection at saline sequestration sites using pressure measurements. Int. J. Greenh. Gas Control 30, 188–196. doi:10.1016/j.ijggc.2014.09.011

White, M.D., Watson, D.J., Bacon, D.H., White, S.K., McGrail, B.P., Zhang, Z.F., 2013. STOMP: Subsurface Transport Over Multiple Phases. STOMP-CO2 and -CO2e Guide. Richland, Washington.

Witten, I. H., Frank, E., Hall, M., 2011. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd edition; Elsevier Press: San Francisco.

Worth, A. P., Cronin, M. T. D., 2003. The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. J. Molecular Struct. 622, 97–111.

Yang, Y.-M., Small, M.J., Junker, B., Bromhal, G.S., Strazisar, B., Wells, A., 2011a. Bayesian hierarchical models for soil CO2 flux and leak detection at geologic sequestration sites. Environ. Earth Sci. 64, 787–798. doi:10.1007/s12665-011-0903-5

Yang, Y.-M., Small, M.J., Ogretim, E.O., Gray, D.D., Bromhal, G.S., Strazisar, B.R., Wells, A.W., 2011b. Probabilistic design of a near-surface CO2 leak detection system. Environ. Sci. Technol. 45, 6380–7. doi:10.1021/es104379m

Yang, Y.-M., Small, M.J., Ogretim, E.O., Gray, D.D., Wells, A.W., Bromhal, G.S., Strazisar, B.R., 2012. A Bayesian Belief Network (BBN) for combining evidence

from multiple CO2 leak detection technologies. Greenh. Gases Sci. Technol. 2, 185–199.

Yasunishi, a, Yoshida, F., 1979. Solubility of carbon dioxide in aqueous electrolyte solutions. J. Chem. Eng. 24, 13–16.

Zhang, Y., Sahinidis, N. V, 2013. Uncertainty quantification in CO2 sequestration using surrogate models from polynomial chaos expansion. Ind. Eng. Chem. Res. 52, 3121–3132.

Zheng, L., Spycher, N., Birkholzer, J., Xu, T., Apps, J., Kharaka, Y., 2013. On modeling the potential impacts of CO2 sequestration on shallow groundwater: Transport of organics and co-injected H2S by supercritical CO2 to shallow aquifers. Int. J. Greenh. Gas Control 14, 113–127. doi:10.1016/j.ijggc.2013.01.014

Zuo, Y., Guo, T. M., 1991. Extension of the Patel–Teja equation of state to the prediction of the solubility of natural gas in formation water. Chem. Eng. Sci. 46, 3251–3258.