Carbon Sequestration: Uncertainty and Parameter Estimation

by

Mirhamed Sarkarfarshi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2015

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

I hereby declare that I am the sole author of this thesis, except for Sections 3.3.1 and 3.4.3 which are co-authored contributions.

Abstract

Carbon Capture and Sequestration (CCS) appears to be a practical technology for large-scale storage of CO_2 to reduce anthropogenic CO_2 emissions. Risk is an inevitable component of any geological project with the aim of storing CO_2, and thus, is a concern to the public, policy makers, and scientists. Uncertainty that arises in the application of mathematical Carbon Sequestration (CS) models has a negative impact on the quality of risk assessment. Parameter uncertainty is believed to play a dominant role in the uncertainty of the outputs of the CS system models. However, reducing parameter uncertainty in CS models involves a trade-off between accuracy and computational efficiency of the model calibration methodology.

The goal of this thesis is to reduce the trade-off between accuracy and computational efficiency when calibrating CS models. This is accomplished by, one, reducing the dimensionality of the parameter space; two, developing efficient calibration algorithms; and, three, reducing the computational cost of model simulation during calibration. The primary contributions of this thesis are:

1. The development of a sensitivity analysis to identify which parameters contribute the most to the uncertainty of the CS system model output, accounting for both parameter uncertainty and model structure.
2. The development of a computationally efficient and flexible Bayesian Importance Sampling (IS) method for continuous calibration of CS models using noisy monitoring data collected during the injection phase.
3. The development of the Response Surface Methodology (RSM) in a novel adaptive way to mitigate the computational demand of CS model calibration with negligible effect on the accuracy of the results.

The methodologies and results presented in this thesis contribute to efficient calibration of CS models by identifying the most influential parameters in uncertainty of CS model outputs and calibrating those models accurately and efficiently.

Acknowledgement

I would like to specially thank my PhD advisors, Professor Robert Gracie and Professor Mahesh Pandey for guiding me through all steps of this research. Through their support and instructions it was that I found my way to commence, continue and conclude this research.

I would also like to thank my PhD defense committee members, Professor John Harrison from University of Toronto and Professor Maurice Dusseault, Professor Bryan Tolson and Professor Kumaraswamy Ponnambalam from University of Waterloo.

Finally, I want to thank Negar, my wonderful wife and the love of my life for her patience and support. To her, I dedicate this thesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

The following table is a list of abbreviations used in the thesis.

Abbreviation	Meaning
ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Networks
ARSM	Adaptive Response Surface Method
BARSM	Bayesian Adaptive Response Surface Method
CCS	Carbon Capture and Sequestration
CDF	Cumulative Density Function
CMC	Carbon Management Canada
COV	Coefficient of Variation
CS	Carbon Sequestration
DOE	Design Of Experiment
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
FE	Finite Element
FEM	Finite Element Method
IID	Independent and Identically Distributed
IS	Importance Sampling
KF	Kalman Filter
LHS	Latin Hypercube Sampling
MARS	Multivariate Adaptive Regression Splines
MCMC	Markov-Chain Monte-Carlo
MCS	Monte-Carlo Simulation
MECE	Mutually Exclusive and Collectively Exhaustive
MG	Multi-Grid
MVFOSM	Mean Value First Order Second Moment
OAT	One-At-a-Time
PDF	Probability Distribution/Density Function
PEPM	Point Estimate method for Probability Moments
PMC	Population Monte-Carlo
PMF	Probability Mass Function
RMSE	Root Mean Square Error

RSM	Response Surface Method
RD	Relative Deviation
RS	Rejection Sampling
SA	Sensitivity Analysis
SSE	Sum of Squared Errors
UIS	Unscented transformation Importance Sampling
UKF	Unscented Kalman Filter
UPF	Unscented Particle Filter
UT	Unscented Transformation

Nomenclature

The following table is a list of symbols and notations used in the thesis. Local indices are described where they are introduced. Vectors and matrices are shown with a bold font.

M = Mass
L = Length
T = Time

Symbol	**Description**	**Dimensions**
A	Constant in Corey-Brooks relationship	$[ML^{-1}T^{-2}]$
$\boldsymbol{A}_m$	Matrix of parameters in a sample pool	-
B	Aquifer thickness	[L]
$\boldsymbol{B}^n$	Control input at time t_n in standard Kalman Filter (KF) notation	-
$\boldsymbol{B}_\beta$	Vector of coefficients in a linear regression model	-
$\mathbf{C}_m$	Covariance matrix of a vector of random variables $\boldsymbol{m}$	-
$\mathbf{C}_{a,b}$	Covariance matrix of two vectors of random variables $\boldsymbol{a}$ and $\boldsymbol{b}$	-
$\mathbf{C}_m^a$	Posterior covariance matrix of a vector of random variables $\boldsymbol{m}$	-
$\mathbf{C}_m^b$	Prior covariance matrix of a vector of random variables $\boldsymbol{m}$	-
$\mathbf{C}_m^n$	Posterior covariance matrix of a vector of random variables $\boldsymbol{m}$ at time t_n	-
$\mathbf{C}_d^{n\|n-1}$	Forecast covariance matrix at time t_n in standard KF notation	-
$\mathbf{C}_d^{n\|n}$	Updated covariance matrix at time t_n in standard KF notation	-
$\mathbf{C}_m^{n,n^*}$	Forecast of covariance matrix of a vector of random variables $\boldsymbol{m}$ at time t_{n^*} when the posterior distribution of $\boldsymbol{m}$ is available at time t_n	-
$\mathbf{C}_v^n$	Covariance matrix of measurement noise at time t_n in standard KF notation	-
$\mathbf{C}_w^n$	Covariance matrix of Gaussian model noise at time t_n in standard KF notation	-
$CI_{m,\alpha}^n$	The α^{th} quantile of the distribution of a vector of random variables $\boldsymbol{m}$ at time t_n	-
$\boldsymbol{D}^{n\|n-1}$	Prior ensemble matrix at time t_n in Ensemble Kalman Filter (EnKF)	-
$\boldsymbol{D}^{n\|n}$	Posterior ensemble matrix at time t_n in EnKF	-
$\boldsymbol{D}_{obs}^n$	Measurement matrix at time t_n in EnKF	-
$\boldsymbol{D}_i^{x,n}$	A set of model evaluations at time t_n and location $\boldsymbol{x}$ while the i^{th} model parameter is sampled from its probability distribution and other parameters are held constant at their mean value	-
$E[\boldsymbol{m}]$	Expected value of a vector of random variables $\boldsymbol{m}$	-
E_i^n	The average discrepancy between the full model and a meta-model for sample i at time t_n	-

$\boldsymbol{F}^n$	State transition model at time t_n in standard KF notation	-
$\boldsymbol{G}(.)$	A linear system model	-
$\boldsymbol{G}_m$	Matrix of model outputs in a linear regression model	-
$\boldsymbol{H}^n$	Measurement function at time t_n in standard KF notation	-
$\boldsymbol{I}$	Identity matrix	-
$I(t)$	Moment of inertia of a CO_2 plume at time t	-
$\boldsymbol{K}$	Kalman gain	-
$\boldsymbol{K}^n$	Kalman gain at time t_n in standard KF notation	-
$L(.)$	Likelihood function	-
N_b	Burn-in period in Markov Chain Monte Carlo (MCMC)	-
N_{eff}	Effective number of samples in Importance Sampling (IS)	-
N_m	Number of uncertain parameters	-
N_{ph}	Number of phases in a multiphase flow	-
N_s	Number of samples in a sampling-based statistical method	-
N_{sp}	Number of samples in a sample pool	-
N_t	Final calibration time of a model	-
N_{th}	Threshold of number of samples in escape mechanisms of Unscented transformation Importance Sampling (UIS)	-
N_x	Number of locations in a discretized system	-
$N(\boldsymbol{\mu}, \mathbf{C})$	A Normal (Gaussian) distribution with a mean $\boldsymbol{\mu}$ and a covariance matrix $\mathbf{C}$	-
O_j^i	i^{th} cost function for the j^{th} scenario in a multi-objective optimization	-
O_j^*	Combined cost functions for the j^{th} scenario in a multi-objective optimization	-
$\bar{O}_j^i$	Normalized i^{th} cost function for the j^{th} scenario in a multi-objective optimization	-
$P(A)$	Probability of an event A	-
$P(A\|B)$	Conditional probability of an event A given another event B has happened	-
Q	CO_2 injection rate	$[MT^{-1}]$
Q_i	Source/sink term for i^{th} phase in a multiphase flow	$[MT^{-1}]$
R^{BARSM}	Ratio of reduction in computational cost of full model run of the Bayesian Adaptive Response Surface Method (BARSM)	-
R_{eff}	Effective sample ratio in IS	-
$RD_i^{x,n}$	Relative deviation of the i^{th} parameter at time t_n and location $\boldsymbol{x}$ in the Relative Deviation (RD) method	-
$RMSE^n$	Root Mean Square Error of a predictive posterior distribution at time t_n	-
S	Sample space	-

S_e	Effective saturation	-
S_{front}	Minimum saturation of a CO_2 plume near boundaries of the plume	-
S_i	Saturation of phase i in a multiphase flow	-
S_r	Residual saturation	-
$U(a,b)$	Uniform probability distribution between a and b	-
V	Volume	[L^3]
W_i	i^{th} normalized weights in IS and Point Estimate Method	-
W_i^0	Probability masses of discrete points in a Probability Mass Function (PMF)	-
W_i^n	i^{th} normalized weights in UIS at time t_n	-
$\boldsymbol{X}$	Space grid	-
X_i	i^{th} subset of events in a sample space	-
X_β^i	the mass fraction of component β in phase i in a multiphase multi-component flow	-
b	Height of boundary of a CO_2 plume from the bottom of the reservoir	[L]
c	Normalizing constant	-
d	Evidence in Bayes' theorem	-
$\boldsymbol{d}^n$	System state at time t_n in standard KF notation	-
$\boldsymbol{d}^{n\|n-1}$	State forecast at time t_n in standard KF notation	-
$\boldsymbol{d}^{n\|n}$	State update at time t_n in standard KF notation	-
$\boldsymbol{d}_i^{n\|n-1}$	i^{th} column of the prior ensemble matrix at time t_n in EnKF	-
$\boldsymbol{d}_i^{n\|n}$	i^{th} column of the posterior ensemble matrix at time t_n in EnKF	-
$\boldsymbol{d}_i^n$	i^{th} column of the measurement matrix at time t_n in EnKF	-
$\boldsymbol{d}_{obs}^n$	Measurement of a quantifiable behaviour of a system in vector form at time t_n	-
$\boldsymbol{d}_{obs}(\boldsymbol{x},t_n)$	Measurement of a quantifiable behaviour of a system at location $\boldsymbol{x}$ and time t_n	-
$\boldsymbol{d}_{tru}^n$	True behaviour of a system in vector form at timet_n	-
$\boldsymbol{d}_{tru}(\boldsymbol{x},t_n)$	True behaviour of a system at location $\boldsymbol{x}$ and time t_n	-
$\boldsymbol{g}$	Acceleration of gravity	[LT^{-2}]
$\boldsymbol{g}(.)$	A system model (function)	-
$\hat{\boldsymbol{g}}(.)$	A meta-model	-
k	Intrinsic permeability	[L^2]
k_{ri}	Relative permeability of phase i in a multiphase flow in a porous medium	-
$\boldsymbol{m}$	A vector of uncertain model parameters	-
m_i	i^{th} uncertain model parameter	-
$\boldsymbol{m}^*$	Candidate sample in MCMC	-

$\hat{\boldsymbol{m}}_c$	A sample around which an IS algorithm has collapsed	-
$\hat{\boldsymbol{m}}_i$	i^{th} sample in IS, Multi-Grid (MG) method and MCMC	-
$\hat{m}_{i,j}$	j^{th} sample of the marginal distribution of the i^{th} parameter in RD method, or i^{th} parameter of the j^{th} sample in the sample pool in linear regression	-
$\hat{\boldsymbol{m}}_i^n$	i^{th} sample from the posterior distribution of UIS at time t_n	-
n	Time or cycle index in a cyclic algorithm of process	-
n_m	Number of meta-cycles between each two full cycles BARSM	-
n^*	Final forecast cycle of a model	-
p_c	Capillary pressure	$[ML^{-1}T^{-2}]$
p_i	Pressure of phase i in a multiphase flow	$[ML^{-1}T^{-2}]$
p_{ref}	Reference pressure in a multiphase flow	$[ML^{-1}T^{-2}]$
$\boldsymbol{q}_i$	Volumetric flux of i^{th} phase in a multiphase flow	$[L^3T^{-1}]$
$q(\boldsymbol{m})$	Proposal distribution of a parameter $\boldsymbol{m}$ in IS	-
$q(\boldsymbol{m}^i \vert \boldsymbol{m}^j)$	Probability of a Markov chain moving from $\boldsymbol{m}^i$ to $\boldsymbol{m}^j$	-
$\hat{q}(\boldsymbol{m})$	Defensive mixture distribution of a parameter $\boldsymbol{m}$ in defensive IS	-
$q^*(\boldsymbol{m})$	Combined proposal distribution of a parameter $\boldsymbol{m}$ in defensive IS	-
r	Radius of a CO_2 plume	[L]
$\boldsymbol{t}$	Time grid	[T]
t	Time value	[T]
t_n	Time value at cycle n in a cyclic algorithm or process	[T]
$u(.)$	Autoregressive error function	-
$\boldsymbol{u}^n$	Model control inputs at time t_n in standard Kalman Filter notation	-
$\boldsymbol{u}^*$	Linear autoregressive error coefficient	-
u_i	i^{th} random variable drawn from a uniform distribution	-
ν	Skewness coefficient of a probability distribution	-
$\boldsymbol{v}^t$	Gaussian measurement noise at time t_n in standard KF notation	-
w_i	i^{th} un-normalized weights in IS	-
$\boldsymbol{w}^t$	Gaussian modelling noise at time t_n in standard KF notation	-
$\boldsymbol{x}$	Location	[L]
$\boldsymbol{x}_i$	Location of the i^{th} measurement point	[L]
$\boldsymbol{y}^n(\boldsymbol{m})$	Output of a system model in vector form at time t_n	-
$\hat{\boldsymbol{y}}^n(.)$	A set of meta-models at time t_n	-
z	Height	[L]
$\Pi(\mathbf{m})$	Cumulative distribution function of a vector of random variables $\boldsymbol{m}$	-

$\boldsymbol{\Upsilon}_i$	i^{th} propagated Sigma point in Unscented Transformation (UT)	-
Ω_m	A feasible parameter space in which a vector of random variables $\boldsymbol{m}$ is defined	-
α	Constant in Corey-Brooks relationship for phase i in a multiphase flow	-
α_i	Constant in Corey-Brooks relationship for phase i in a multiphase flow	-
α_{MCMC}	Acceptance probability in MCMC	-
β	Constant in Corey-Brooks relationship	-
β_i	i^{th} coefficient in linear regression	-
$\delta(.)$	Dirac delta function	-
$\boldsymbol{\varepsilon}_{mod}$	Model error, time-independent	-
$\boldsymbol{\varepsilon}_{obs}$	Monitoring error	-
$\boldsymbol{\varepsilon}_{tot}$	Combined model and monitoring error	-
$\boldsymbol{\varepsilon}_{mod}^{0}$	Random noise in an autoregressive error model	-
$\boldsymbol{\varepsilon}_{tot}^{0}$	Random noise plus monitoring noise in an autoregressive error model	-
$\boldsymbol{\varepsilon}_{mod}^{n}$	Model error at time t_n	-
$\hat{\boldsymbol{\varepsilon}}_{mod}$	Meta-model error	-
$\hat{\boldsymbol{\varepsilon}}_{mod,i}$	Meta-model error for the i^{th} sample in linear regression	-
$\hat{\boldsymbol{\varepsilon}}_{tot}$	Combined meta-model, model and monitoring error	-
$\hat{\boldsymbol{\varepsilon}}_{mod}^{n}$	Meta-model error at time t_n	-
η	Defensive mixture ratio in defensive IS	-
η_{max}	Maximum defensive mixture ratio in escape mechanisms of UIS	-
η_{min}	Minimum defensive mixture ratio in escape mechanisms of UIS	-
μ_i	Dynamic viscosity of phase i in a multiphase flow	[ML^{-1}T^{-1}]
$\boldsymbol{\mu}_m$	Mean value of a vector of random variables $\boldsymbol{m}$	-
$\boldsymbol{\mu}_m^{a}$	Posterior mean value of a vector of random variables $\boldsymbol{m}$	-
$\boldsymbol{\mu}_m^{b}$	Prior mean value of a vector of random variables $\boldsymbol{m}$	-
$\boldsymbol{\mu}_m^{n}$	Mean of a random variable $\boldsymbol{m}$ at time t_n	-
$\boldsymbol{\mu}_g^{n,n^*}$	Mean of forecast of a function of random variables, $\boldsymbol{g}(\boldsymbol{m})$ at time t_{n^*} when the posterior distribution of $\boldsymbol{m}$ is available at time t_n	-
ϑ_i	Weight of i^{th} Sigma point in UT	-
$\pi_{mod}(.)$	Probability distribution of time-independent model error	-
$\hat{\pi}_{mod}^{n}(.)$	Probability distribution of meta-model error at time t_n	-
$\pi_{obs}(.)$	Probability distribution of the monitoring error	-
$\pi_{tot}(.)$	Probability distribution of combined model and monitoring error	-

$\hat{\pi}_{tot}^{n}(.)$	Probability distribution of combined meta-model, model and monitoring error at time t_n	-
$\pi(\boldsymbol{m})$	Probability distribution of a vector of random variables $\boldsymbol{m}$	-
$\pi_i(\boldsymbol{m}_i)$	Marginal probability distribution of a random variable m_i	-
$\pi^n(\boldsymbol{m})$	Posterior distribution of a vector of random variables $\boldsymbol{m}$ accounting for all monitoring data up to time t_n	-
$\pi_b^n(\boldsymbol{m})$	Prior Gaussian distribution of a vector of random variables $\boldsymbol{m}$ at time t_n	-
ρ_i	Density of phase i in a multiphase flow	[ML^{-3}]
σ	Constant in Corey-Brooks relationship	-
σ_m	Standard deviation of a random variable m	-
σ_m^n	Standard deviation of a random variable m at time t_n	-
ϕ	Porosity	-
$\boldsymbol{\chi}_i$	ith Sigma point in UT	-
ψ_i	ith term quadratic linear regression	-
ψ_i	ith term quadratic linear regression	-

Chapter 1.
Introduction

1.1. Carbon Capture and Sequestration

1.1.1. Global Warming

Global warming poses a significant challenge to society in the 21st century, with potentially drastic environmental consequences (IPCC, 2013). The average surface temperature has increased worldwide since the 19th century, as shown in Figure 1-1.

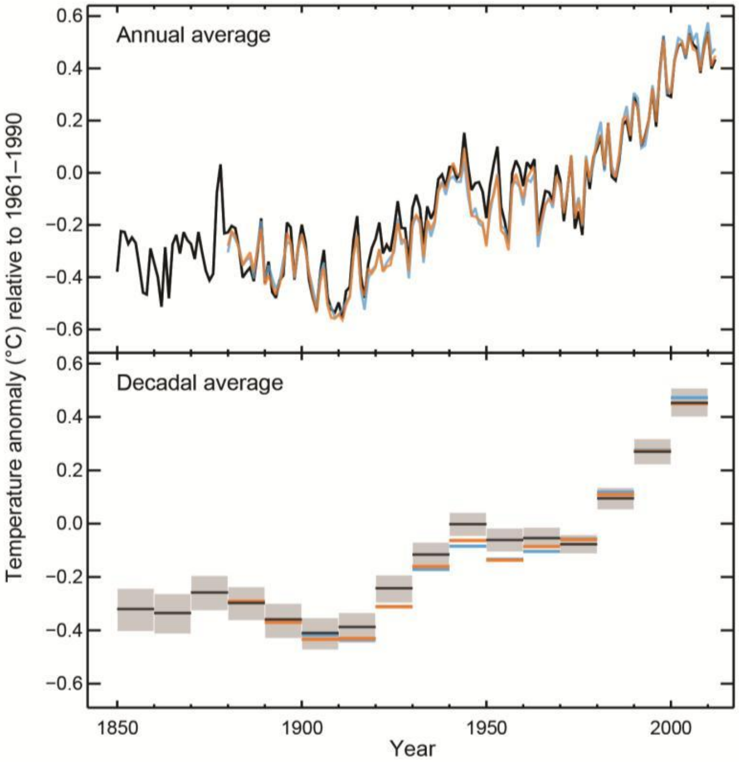


Figure 1-1: The average surface temperature anomaly between 1850 and 2012 relative to 1961 to1990, according to three different analyses. Grey shades in the bottom plot indicate uncertainty ranges. From: (IPCC, 2013).

The increase in global temperature is largely attributed to the concentration of anthropogenic greenhouse gases in the Earth's atmosphere, gases such as carbon dioxide (CO_2) and methane. The concentrations of theses have increased significantly since the beginning of the industrial age (IPCC, 2013). CO_2 is responsible for the majority of the anthropogenic green-house gas emissions. Atmospheric concentration of CO_2 was 391 ppm in 2011 and has increased by 40% since 1750 (IEA, 2012; IPCC, 2013), more than half of which has occurred in the past 45 years (The Royal Society, 2014). This sharp increase is significantly larger than what can be attributed to natural fluctuations over the past 650,000 years (Siegenthaler *et al*., 2005; Gasda, 2008) and there is a general agreement that this increase is directly correlated with the global warming (IPCC, 2013).

Phenomena such as more warmer days and fewer cooler days in most land areas, more frequent heat waves, heavier and more frequent precipitations, heavier and longer droughts, more tropical cyclonic activities and extreme increases in sea levels are likely correlated with the global warming (IPCC, 2013). Among the scientific community, it is generally agreed that significant action must be taken to drastically reduce the CO_2 emissions to mitigate global warming. Several solutions have been proposed, including using less carbon-intensive fossil fuels (e.g., natural gas instead of coal), improving the efficiency of current energy production methods, using renewable energy sources (e.g., solar, wind, hydro, and geothermal energy), using nuclear power plants, and Carbon Capture and Sequestration (CCS) in order to mitigate the increasing concentration of CO_2 in the Earth's atmosphere.

1.1.2. Carbon Capture and Sequestration Overview

CCS is the process of capturing CO_2 from anthropogenic point sources, i.e., the sources with large emission rates such as fossil fuel-based power plants, cement production plants and refineries, and transporting the CO_2 to a safe and long-term storage site, in order to isolate it from the atmosphere (IPCC, 2005). CCS is an attractive option to reduce CO_2 emissions because CCS can be utilized immediately with minor disruptions in current energy production methods and with minor adjustments made to the related infrastructure

(Seto & McRae, 2011). While CCS does not reduce the amount of CO_2 produced from the point sources, it can be considered a solution in relation to climate change, as CCS can mitigate CO_2 emissions over the next decades. By capturing CO_2 emissions in the present, it allows for the continued use of current energy production methods with fewer immediate environmental impacts, affording researchers time to effectively replace current methods with more sustainable solutions in the future (Ebigbo *et al*., 2007).

As shown in Figure 1-2, a CCS project consists of four stages: capture, transportation, injection, and storage. In the capture stage, CO_2 is separated from other emissions at the point source. Then, in the transportation stage, it is transported to a proper storage site. In the injection stage, CO_2 is injected into a reservoir. Finally, in the storage stage, it is stored separately from the atmosphere. This thesis will only address Carbon Sequestration strategies. Thus, the abbreviation “CS” will be employed for the rest of the thesis, as “CCS” also incorporates processes of Carbon Capture.

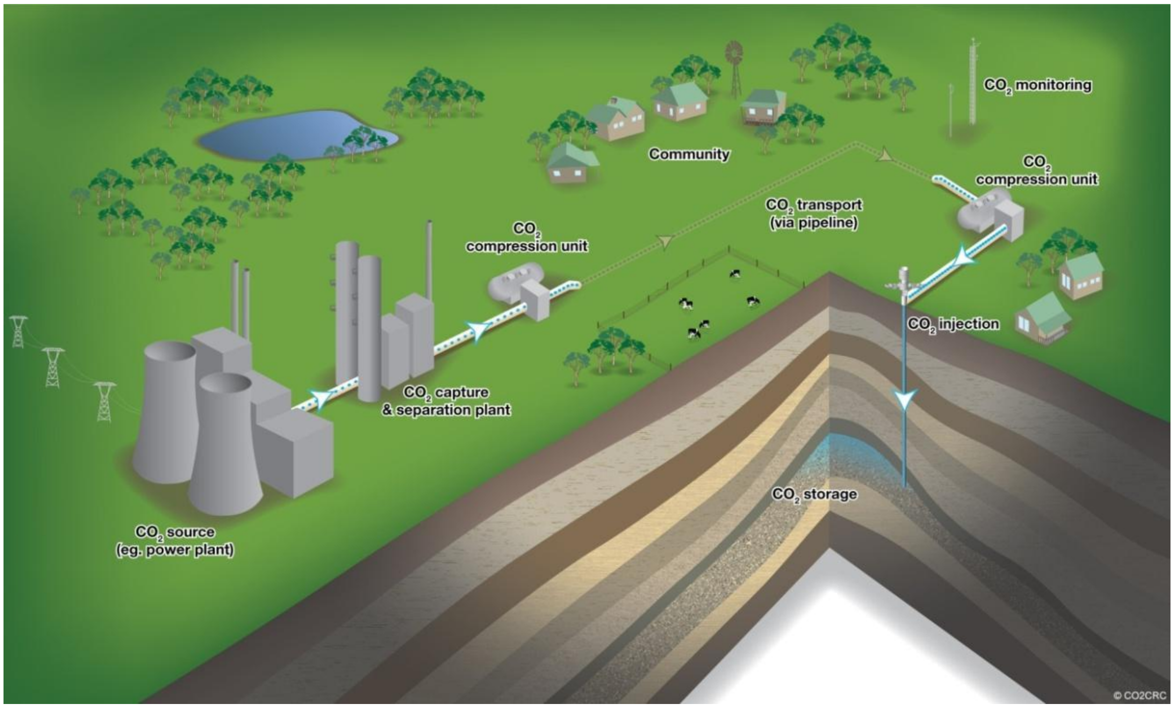


Figure 1-2: The process of Carbon Capture and Sequestration (CCS), including the four stages of capture, transportation, injection and storage. From: (CO2CRC, 2013).

Deep geological formations, as shown in Figure 1-2, are among the most practical storage options for CO_2. CO_2 in its supercritical form (above 304.25 K and 7.39 MPa) has a critical density of 467 kg/m^3 and occupies over 200 times less volume than CO_2 in the atmospheric conditions (IPCC, 2005). CO_2 will remain in a supercritical state if it is injected deeper than 800 m underground where the pressure and temperature exceed the critical pressure and temperature of CO_2. The idea behind geological storage of CO_2 is that by injecting supercritical CO_2 in a deep enough subsurface porous formation that is covered by an overlaying impermeable layer, CO_2 can be potentially stored for hundreds to thousands of years. Such storage times allow for the development of efficient long-term immobilizing mechanisms such as the dissolution of CO_2 into formation fluid (solubility trapping) and chemical reactions of dissolved CO_2 with minerals (mineral trapping), as they could isolate CO_2 indefinitely from the atmosphere (IPCC, 2005). The target subsurface porous formation is usually referred to as the "reservoir" and the impermeable overlaying layer is referred to as the "caprock". Deep saline aquifers are commonly considered the most practical CS reservoir because of their large capacity, the integrity of the caprock, the worldwide distribution, and the low economic value (Ebigbo *et al*., 2007; Zhao *et al*., 2010; Seto & Mcrae, 2011).

The density of supercritical CO_2 is less than the density of the resident brine in the saline aquifers (Ehlig-Economides & Economides, 2010; IPCC, 2005; Nordbotton *et al*., 2005). Thus, the supercritical CO_2 forms a buoyant phase beneath the caprock. As shown in Figure 1-3, this buoyant phase is usually referred to as "CO_2 plume". Due to the buoyancy forces, any geological or man-made opening in the caprock can provide a pathway for leakage of the supercritical CO_2. Pressure build-up due to injection in the reservoir might also cause the resident fluid of the reservoir to migrate to other formations. Figure 1-3 illustrates three of the potential leakage pathways in a CS system, i.e., caprock fractures, caprock faults and wells. Leakage risk is commonly considered as the most significant risk of a CS system and the primary factor measuring success or failure of a CS project (LeNeveu, 2008).

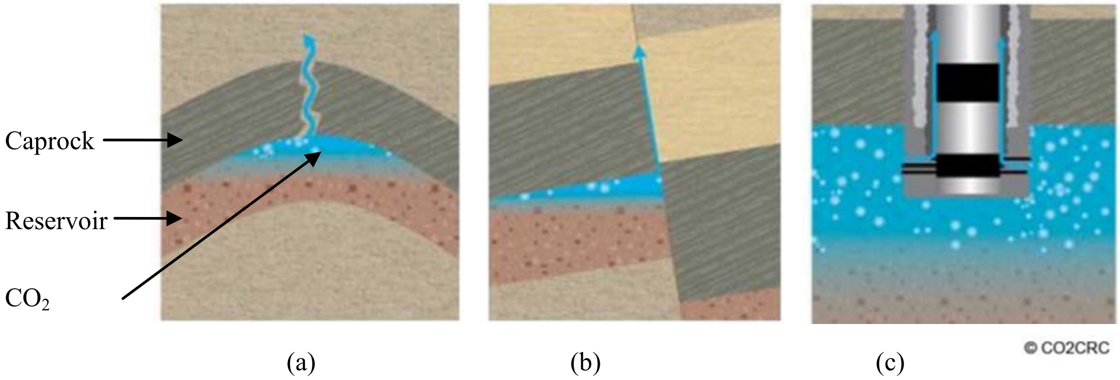


Figure 1-3: Three of potential leakage/migration pathways for the supercritical CO_2 or the resident fluid of the reservoir: (a) caprock fractures, (b) caprock faults and (c) wells. Modified from: (CO2CRC, 2013).

Quantitative risk assessment is a promising tool to forecast probability and the consequences of incidents (e.g., leakage) and to make informed risk-based decisions before, during, and/or after the injection of supercritical CO_2 in a CS project. A quantitative risk assessment requires a knowledge of probability distributions of system parameters and/or system behaviour forecasts. Accurate and precise knowledge of these distributions results in a more accurate and meaningful risk assessment, generating more effective risk management and better decision making abilities. As will be described in the following sub-section, uncertainties are the primary complicating factors in producing an accurate risk assessment, and uncertainties are especially prominent during the initial stages of a CS project (Wilson & Monea, 2004).

1.1.3. Carbon Sequestration Uncertainties

Uncertainties in our knowledge of the properties and behaviour of a CS system leads to uncertainty in the prediction of the CS system behaviour, such as plume migration, pressure buildup and flow rate through a leakage pathway. These uncertainties pose significant problems to generating accurate risk assessments, effectively managing risk, and making informed decisions. Uncertainty in a quantity (e.g., porosity of the target formation or flow rate through a leakage pathway) is due to a lack of knowledge (*epistemic* uncertainty), inherent randomness of that quantity (*aleatory* uncertainty) or both (Ang &

Tang, 2006). Unlike *aleatory* uncertainty, *epistemic* uncertainty is reducible through data collection and calibration efforts.

The primary components of *epistemic* uncertainty are, one, system models; two, parameters; and, three, measurements (Bedi & Harrison, 2012). Since this thesis primarily addresses *epistemic* uncertainty in CS, these three components of *epistemic* uncertainty are briefly described below in a CS context.

Model Uncertainty: A typical CS system can be described by a deterministic mathematical model, $\boldsymbol{g}(\boldsymbol{x}, t, \boldsymbol{m})$, where $\boldsymbol{x}$ is location, t is time and $\boldsymbol{m}$ is the vector of model parameters. It can be related to the *true* behaviour of the CS system by

$$\boldsymbol{d}_{tru}(\boldsymbol{x}, t) = \boldsymbol{g}(\boldsymbol{x}, t, \boldsymbol{m}) + \boldsymbol{\varepsilon}_{mod} \tag{1.1}$$

where $\boldsymbol{d}_{tru}(\boldsymbol{x}, t)$ is the true system behaviour (e.g., pressure or saturation) at time t and location $\boldsymbol{x}$, and $\boldsymbol{\varepsilon}_{mod}$ is the model error that represents the *epistemic* uncertainty about the outputs of the system model. Various analytical, semi-analytical, and numerical models have been developed to predict different aspects of the behaviour of a CS system such as CO_2 plume evolution (Nordbotton *et al.*, 2005; Malekzadeh & Dusseault, 2013), leakage to the surface through possible pathways (Gracie & Craig, 2011), leakage to ground water formations (Humez *et al.*, 2011), CO_2 dissolution into the brine (LeNeveu, 2008), and long-term mineralization of CO_2 (LeNeveu, 2008). Figure 1-4 shows a sample of a simulation of the distributions of CO_2 saturation and pressure in the reservoir in a CS project (Ghaderi & Leonenko, 2009). In Figure 1-4, model outputs are the average saturation of CO_2 and the average pressure, in each grid block of the reservoir domain (i.e., $\boldsymbol{x}$) where t is 50 years, and $\boldsymbol{d}_{tru}(\boldsymbol{x}, t)$ is the real average saturation of CO_2 and pressure, in the same grid block of the CS system and in the same time.

All of the models discussed above are subject to uncertainty due to the complexity of the physics of the multiphase, multi-component flow. Furthermore, uncertainty is increased by simplifying assumptions such as idealization of parameter correlations (i.e., use of empirical relationships), and/or by using approximate numerical solutions. Moreover, lack of knowledge about site characteristics such as undetected fractures and faults and unidentified heterogeneities decrease the predictive ability of the models. In a recent study, (Nordbotton *et al.*, 2012) investigated uncertainty in several CO_2 plume propagation models using a benchmark problem and highlighted the role of model uncertainty in CS systems.

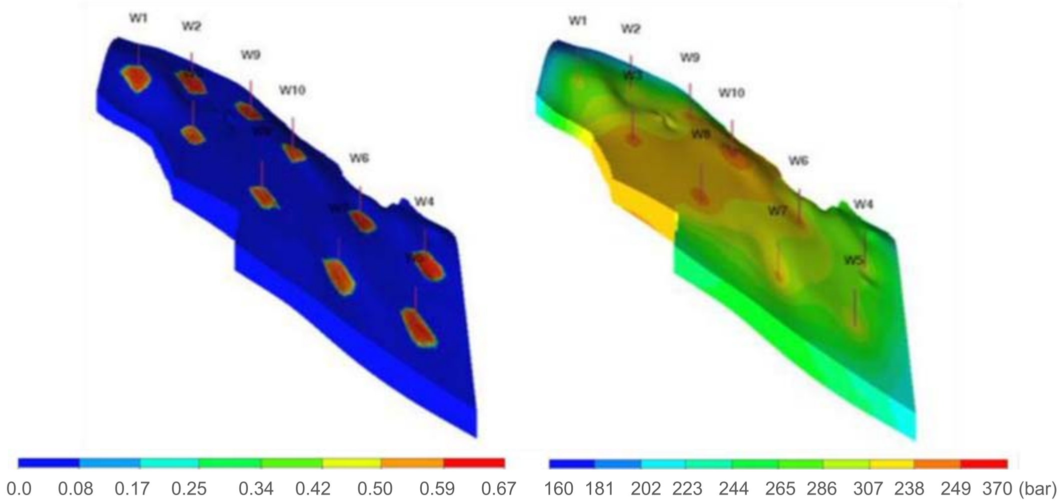


Figure 1-4: Simulation of saturation (left) and pressure (right) fields in the Wabamun Area Sequestration Project (WASP) with 10 injector wells and after 50 years of injection. Modified from: (Ghaderi & Leonenko, 2009).

Measurement Uncertainty: Several techniques are available for monitoring a CS system. Some of the currently available techniques for monitoring CS systems are:

- 3D and 4D seismic (Chadwick *et al.*, 2004) that can be used for 3D and time-lapse 3D estimation of subsurface conditions (e.g., estimation of CO_2 saturation and reservoir geometry, based on reflection of seismic waves).
- Gravity (Nooner *et al.*, 2007) that can be used for the estimation of CO_2 distribution in the reservoir based on subsurface mass redistribution, that results in gravity variation (Glegola, 2012).
- Electromagnetic (Hoversten *et al.*, 2004) that estimates properties of the CS system based on measurements of electrical properties of the reservoir (e.g., based on borehole-to-surface measurements of electrical resistivity of the reservoir) (Zhdanov *et al.*, 2013).

- Time lapse satellite imaging (Raikes *et al.*, 2008) that uses satellite measurements to find surface deformation due to CO_2 injection and pressure elevation.
- Fluid sampling from observation wells (Freifeld *et al.*, 2005) that can determine presence and/or saturation of CO_2, or other materials such brine or trackers, in a formation by sampling from the formation fluid.

A CS monitoring system can be described in a mathematical relationship with the true system behaviour in which the monitoring system provides an estimate of the system behaviour,

$$\boldsymbol{d}_{obs}(\boldsymbol{x}, t) = \boldsymbol{d}_{tru}(\boldsymbol{x}, t) + \boldsymbol{\varepsilon}_{obs} \tag{1.2}$$

where $\boldsymbol{d}_{obs}(\boldsymbol{x}, t)$ is the observed (monitored) value of the true system behaviour, $\boldsymbol{d}_{tru}(\boldsymbol{x}, t)$, at time t and location $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}_{obs}$ is the observation/monitoring error which represents the *epistemic* uncertainty about the monitoring data. The observation error, $\boldsymbol{\varepsilon}_{obs}$, indicates the accuracy of the monitoring and depends on both monitoring technique and project characteristics.

In CS projects, remote measurements of physical parameters of the subsurface are subject to large errors and considerable uncertainly (Prestona *et al.*, 2005). Direct measurements of system characteristics (such as porosity and permeability of the reservoir) or system behaviour (such as pressure and saturation) are more accurate than remote measurements, but they are very limited compared to reservoir extents because drilling monitoring wells penetrates the caprock of the target formation and creates potential leakage pathways for supercritical CO_2.

<u>Parameter Uncertainty</u>: Models require a set of parameters to describe a system, i.e., $\boldsymbol{m}$ in equation (1.1). The parameters describing a CS system can be physical characteristics of the system (e.g., porosity, injection depth, and reservoir temperature) or parameters introduced in constitutive relationships (e.g., Young's modulus or Brooks-Corey relationship constants). Uncertainty in the parameters is a major factor that impedes the CS models' predictive capability (Celia & Nordbotton, 2009). CS system parameters are subject to significant uncertainty due to heterogeneity, limited data and imperfect

measurements. As described in Section 1.1.3, uncertainty about $\boldsymbol{m}$ is considered *epistemic* in this thesis.

As an example, reservoir porosity is a parameter with a great influence on the uncertainty of the behaviour of a CS system (Sarkarfarshi *et al.*, 2014). Figure 1-5 illustrates a histogram of the porosity in over 1200 oil reservoirs in the U.S. according to the U.S. National Petroleum Council (NPC, 1984). Kopp *et al*. used this histogram in a quantitative risk analysis (Kopp *et al*., 2010). According to Figure 1-5, uncertainty about the porosity is noticeable. This uncertainty results in an uncertain forecast of system behaviour and risk measures. It can also be seen in Figure 1-5 that this histogram does not follow any standard Probability Distribution Function (PDF), such as Normal or Log-Normal PDFs, which makes efficient and accurate uncertainty analysis and calibration efforts challenging (Kopp *et al*., 2010).

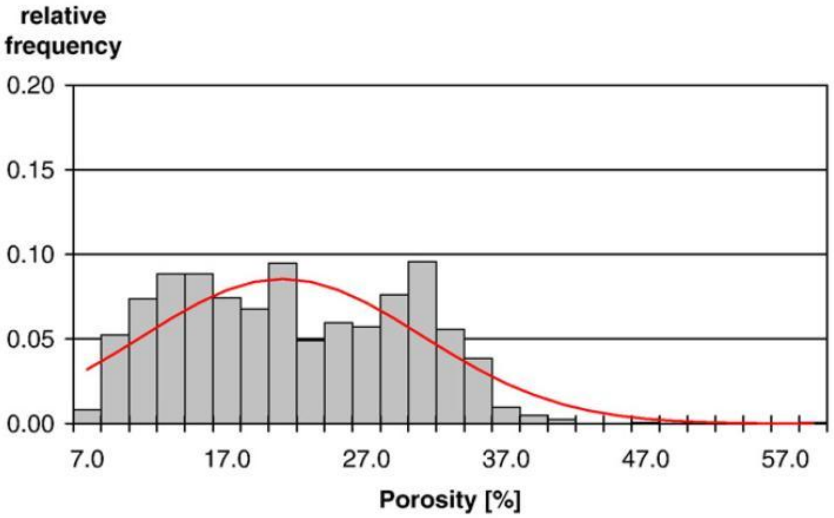


Figure 1-5: Histogram of relative frequency of porosity in over 1200 crude oil reservoirs in U.S. according to the U.S. National Petroleum Council (NPC) public database (NPC, 1984). Solid curve shows a Normal distribution with a mean and a standard deviation similar to the histogram data. From: (Kopp *et al*., 2010).

1.2. Problem Statement

Among the major sources of *epistemic* uncertainty described above (model, parameter, and measurement uncertainty), parameter uncertainty is believed to play the dominant role in the uncertainty of modelling the behaviour of a CS system and quantitative risk assessment (Celia & Nordbotton, 2009; Polson *et al*., 2012). Thus, effective model calibration methods are required to reduce parameter uncertainty and to obtain accurate quantitative risk measures.

Combining equations (1.1) and (1.2), it is possible to relate monitoring data, model output and parameters, and monitoring and modelling errors, as

$$\boldsymbol{d}_{obs}(\boldsymbol{x},t) = \boldsymbol{g}(\boldsymbol{x},t,\boldsymbol{m}) + \boldsymbol{\varepsilon}_{obs} + \boldsymbol{\varepsilon}_{mod} \tag{1.3}$$

where the system model, $\boldsymbol{g}(\boldsymbol{x},t,\boldsymbol{m})$, is a deterministic function of model parameters $\boldsymbol{m}$, and observation and model errors, $\boldsymbol{\varepsilon}_{obs}$ and $\boldsymbol{\varepsilon}_{mod}$, are random errors subject to *epistemic* uncertainty. Typically, calibration of a CS system involves finding either the optimal value of the uncertain parameters deterministically, or estimating the probability distribution of the uncertain parameters based on the monitoring data while accounting for monitoring and modelling errors. In other words, the objective of calibration is either deterministically finding the optimal value (or values) of model parameters $\boldsymbol{m}$ which minimizes the discrepancy between $\boldsymbol{d}_{obs}(\boldsymbol{x},t)$ and $\boldsymbol{g}(\boldsymbol{x},t,\boldsymbol{m})$ in equation (1.3), or finding the probability distribution of model parameters $\boldsymbol{m}$ from equation (1.3) using formal statistical methods.

The optimal value or the probability distribution of $\boldsymbol{m}$ at time t can be used for forecasting the system behaviour beyond time t. Unlike seeking the optimal value of $\boldsymbol{m}$ in a deterministic calibration framework, finding the probability distribution of $\boldsymbol{m}$ in a formal statistical framework allows probabilistically forecasting the system behaviour which can be used for a probabilistic risk analysis (Wilson & Monea, 2004). In this thesis, the term “calibration” is reserved for the latter approach, i.e., the process of finding or estimating the probability distribution of parameters of a deterministic CS model, using a formal statistical framework, and utilizing the site monitoring data. Where required to refer to the former calibration approach, the term “deterministic calibration” is used instead. The concept and process of calibration in this thesis are explained mathematically and in more detail in Section 2.4.

CS system are normally characterized by nonlinear and computationally expensive models, and a large number of uncertain parameters described by non-Gaussian distributions. For the purpose of this thesis, a "linear/nonlinear model" refers to mathematical models with linear/nonlinear responses with respect to model parameters. Typically, the probability distribution of model parameters for such nonlinear and non-Gaussian systems cannot be found exactly from equation (1.3) and must be obtained by approximate means. Accurately finding the probability distribution of $\boldsymbol{m}$ in a formal statistical framework normally requires computationally demanding sampling-based solutions (Tavakoli *et al*., 2013). On the other hand, finding the probability distribution of $\boldsymbol{m}$ using computationally more efficient methods is normally inaccurate due to the major simplifying assumptions (Chen, 2003). Thus, a notable trade-off exists between accuracy and computational efficiency when calibrating complex CS models.

In addition, due to limited monitoring data and limited initial knowledge about the model parameters, calibrating the model only once, when in the early stages of CO_2 injection in a CS project, most likely results in inaccurate and uncertain forecast of future system behaviour. Thus, a single calibration effort is likely insufficient and a continuous calibration, i.e., re-calibrating the model when new monitoring data is available, is likely required to gradually reduce parameter uncertainty using the continuous stream of monitoring data. A continuous calibration process, however, demands several times more computational time and resources compared to calibrating the model only once.

There has been a limited number of works to date focusing on how monitoring data can be incorporated into the calibration of CS models. Johnson & White (Johnson & White, 2012) used Markov Chain Monte Carlo (MCMC) to inverse seismic signals to refine the permeability fields. Bhowmik *et al*. (Bhowmik *et al*., 2011) stochastically calibrated the permeability field of a synthetic CS system using pressure data from injection and monitoring wells. Finally, Espinet and Shoemaker (Espinet & Shoemaker, 2013) and Tavakoli *et al*. (Tavakoli *et al*., 2013) compared a number deterministic and Bayesian algorithms for calibration of CS systems, respectively, using synthetic case studies. Deterministic approaches and/or the trade-off between accuracy and computational efficiency are the major limitations of these works. For instance, the Ensemble Kalman

Filter (EnKF) based solution used in (Tavakoli *et al*., 2013) is computationally more efficient than the MCMC solution used in (Johnson & White, 2012) for calibrating the permeability field in a CS reservoir. However, the basic assumptions of a Kalman Filter (KF) solution, i.e., linear model responses and Gaussian probability distributions, makes the EnKF less accurate compared to a well-designed MCMC algorithm with a sufficiently large number of samples. Similar works in analogous fields (such as reservoir history matching) are also subject to the same limitation, i.e., the trade-off between an approximate and computationally efficient (normally KF-based) solution or an accurate and computationally intensive (normally MCMC-based) solution (Wen & Chen, 2005; Liu & Oliver, 2005; Ma *et al*., 2006; Maucec *et al*., 2007; Liu & McVay, 2009; Emerick & Reynolds, 2012; Emerick & Reynolds, 2013; Rosi *et al*., 2014). Thus, reducing the trade-off between accuracy and computational efficiency is an ongoing area of research in these fields.

The existing trade-off between the accuracy of the calibration results and computational efficiency of the calibration process is one of the primary knowledge gaps in the field of CS and it is a major hindrance to forecasting system behaviour and quantifying risk accurately. Thus, further research is required to adapt and develop flexible and computationally efficient methods to accurately calibrate extremely nonlinear and computationally costly CS models.

1.3. Research Goal and Objectives

The goal of this thesis is to reduce the trade-off between accuracy and computational efficiency of calibrating a CS system. In other words, this thesis aims to develop efficient methods that both reduce the computational cost and improve the accuracy of CS model calibration. This goal is addressed by proposing three complementary strategies, each addressing one of the primary causes of the existing trade-off between the accuracy and computational efficiency (large number of uncertain parameters, computationally demanding calibration methods, and computationally demanding mathematical models). The objectives of this thesis are:

1. To reduce the dimensionality of the parameter space by a parametric sensitivity analysis and by identifying the parameters which contribute the most to the uncertainty of a CS model output. The purpose of this reduction is to ensure only those parameters with most contribution to the uncertainty of a CS model output are considered in a calibration process. A smaller number of uncertain parameters demands a smaller number of model simulations in the calibration process and makes the calibration computationally more efficient.
2. To develop accurate and computationally efficient calibration methods that are flexible when applied to numerical, nonlinear, and computationally expensive CS models, and do not limit the type of the probability distribution used to describe the parameters. To accomplish this objective, an efficient sampling-based method will be developed, that is able to accurately calibrate the complex CS models with smaller number of model simulations compared to conventional approaches.
3. To reduce the computational cost of repeatedly executing the resource demanding system models during the calibration period by introducing an efficient and accurate adaptive meta-modelling technique. This will be accomplished in order to approximate the complex CS model by a simple meta-model that will replace the CS model during most stages of the calibration, reducing the computational demand of the calibration.

Overall, these objectives are complementary, and each can be utilized individually or more than one can be utilized collectively, to reduce the computational cost of the calibration process. Equivalently, they allow available computational resources to be spent more efficiently to increase the calibration accuracy without demanding additional computational budget.

1.4. Notation

The statistical methods and algorithms used in this thesis are widely used in various fields of science and engineering (such as reservoir engineering, hydrology, computer science, communications, applied mathematics and finance). However, the notation used vary between these fields. For instance, it is common to denote the mean of a vector of parameters $\boldsymbol{m}$ by $\boldsymbol{\mu}_m$, $\boldsymbol{\mu}_M$, $\langle \boldsymbol{m} \rangle$, $\bar{\boldsymbol{m}}$, and $E(\boldsymbol{m})$, and to denote the covariance matrix of a vector of parameters $\boldsymbol{m}$ by $\boldsymbol{\Sigma}^m$, $\boldsymbol{\Sigma}_m$, $\mathbf{C}_m$, $\mathbf{C}_{mm}$, $\mathbf{C}_M$, $\boldsymbol{P}_m$, and $\mathrm{Cov}(\boldsymbol{m})$, in different fields.

The notation adopted in this thesis is commonly used in the literature on hydrocarbon reservoir model calibration (i.e., reservoir history matching) field and calibration of CS systems (Abacioglu *et al*., 2001; Liu & Oliver, 2005; Gu & Oliver, 2005; Wen & Chen, 2005; Ma *et al*., 2006; Emerick & Reynolds, 2013; Tavakoli *et al*., 2013). For instance, the mean and the covariance matrix of a vector of parameters $\boldsymbol{m}$ are denoted by $\boldsymbol{\mu}_m$ and $\mathbf{C}_m$, respectively. Where required by methods introduced in this thesis, additional superscripts/subscripts are added to the typical notation. Bold font is used to denote vectors and matrices in this thesis. For example, $\boldsymbol{\mu}_m$ is used to denote the mean of a vector of parameters $\boldsymbol{m}$, and μ_m is used to denote the mean of a scalar parameter m.

1.5. Thesis Structure

After an introduction to CCS and problem statement in Chapter 1, Chapter 2 provides a background on uncertainty analysis and Bayesian techniques to establish the necessary mathematical foundation for the following chapters. Chapter 2 also presents the general formulations, notations, and assumptions used in the thesis. Each of the next chapters, Chapters 3 to 5, address one of the three complementary objectives of the thesis. Even though each chapter addresses the trade-off between accuracy and computational efficiency during the calibration of CS models, each chapter does so with a distinct and unique strategy: reducing the number of uncertain parameters, introducing accurate and efficient calibration methods, and reducing the computational cost of repetitive model simulations. More specifically, Chapter 3 presents a quantitative parametric sensitivity analysis for CS projects. In Chapter 4, a novel and efficient continuous Bayesian calibration method is introduced. Chapter 5 introduces a new computationally efficient and accurate method for reducing the computational cost of continuous Bayesian calibration of CS projects. Chapter 6 summarizes the thesis and identifies potential paths for the future works. Finally, Appendix A and Appendix B provide additional supporting information for topics in Chapters 4 and 5, respectively.

Chapter 2.

Background and Mathematical Preliminaries

Summary

Carbon Sequestration (CS) system responses are normally nonlinear, non-Gaussian and complex (Kopp *et al*., 2010; Oladyshkin *et al*., 2011-b; Tavakoli *et al*., 2013). Quantifying, propagating, and reducing the uncertainties in such systems requires a comprehensive understanding of the existing uncertainty analysis and model calibration methods. Thus, a background on uncertainty propagation and Bayesian calibration is presented in this chapter. The subsequent three chapters will address complementary strategies to decrease parameter uncertainty efficiently once a comprehensive knowledge base has been established. In this chapter, the general formulation, the general notation, and a set of basic assumptions used in the thesis are presented in Section 2.1. In Section 2.2, the concept of uncertainty propagation is introduced. In Section 2.3, several analytical and numerical solutions for uncertainty propagation are discussed and the benefits and limitations of these methods are summarized. These methods provide a mathematical background for a number of Bayesian calibration methods which are described in Section 2.4 and Section 2.5, as well as the sensitivity analysis method used in Chapter 3. In Section 2.4 model calibration using Bayesian statistics is introduced. In Section 2.5, several solutions to Bayesian Calibration are described, and the benefits and limitations of these solutions are summarized. The concepts introduced in Section 2.4 and Section 2.5 provide a mathematical basis for Chapters 4 and 5.

As reported in Chapter 1, the literature in calibration of CS systems is very limited. Thus, where applicable, additional literature is cited and discussed in this chapter from other analogous fields which deal with computationally demanding, nonlinear, high-resolution and complex models such as reservoir engineering and weather forecasting.

2.1. Problem Formulation and Basic Assumptions

Equations (1.1) to (1.3) in Chapter 1 described the relationships between the true behaviour of a CS system, the system model output, and the observations/monitoring data. These equations are elaborated upon in this section to be able to properly and efficiently explain uncertainty propagating and calibrating in CS systems, and will be used in the rest of the thesis.

Let $\boldsymbol{d}_{tru}(\boldsymbol{x}, t_n)$ denote the true behaviour of a CS system at a time t_n and at a location $\boldsymbol{x}$. Due to the cyclic nature of the calibration algorithms introduced later in Chapters 4 and 5, n is used as the time or cycle index in a cyclic process and $\boldsymbol{t}$ is used as a vector of time values in the cyclic process, i.e., $\boldsymbol{t} = \left[t_1, t_2, \dots, t_{N_t}\right]$ where N_t denotes the final cycle of the cyclic process. To make calculations easier, it is desirable to aggregate the system behaviour $\boldsymbol{d}_{tru}(\boldsymbol{x}, t_n)$ at multiple locations into a single vector. Thus, let $\boldsymbol{d}_{tru}^{n} = \left[\boldsymbol{d}_{tru}(\boldsymbol{x}_1, t_n), \boldsymbol{d}_{tru}(\boldsymbol{x}_2, t_n), \dots, \boldsymbol{d}_{tru}\left(\boldsymbol{x}_{N_x}, t_n\right)\right]$ denote the true behaviour of the CS system at time t_n and at locations $\boldsymbol{x}_1$ to $\boldsymbol{x}_{N_x}$ where N_x is the number of locations of interest.

Let $\boldsymbol{d}_{obs}(\boldsymbol{x}, t_n)$ denote the observation (measurement) of $\boldsymbol{d}_{tru}(\boldsymbol{x}, t_n)$ at time t_n and at location $\boldsymbol{x}$. Also let $\boldsymbol{d}_{obs}^{n} = \left[\boldsymbol{d}_{obs}(\boldsymbol{x}_1, t_n), \boldsymbol{d}_{obs}(\boldsymbol{x}_2, t_n), \dots, \boldsymbol{d}_{obs}\left(\boldsymbol{x}_{N_x}, t_n\right)\right]$ denote a vector of measurement of $\boldsymbol{d}_{tru}^{n}$ at time t_n. Measurement of $\boldsymbol{d}_{tru}^{n}$ is assumed to be subject to a random noise (monitoring error). We denote the vector of measurement noises at locations $\boldsymbol{x}_1$ to $\boldsymbol{x}_{N_x}$ by $\boldsymbol{\varepsilon}_{obs}$. Thus, equation (1.2) is restated in a more accurate way as

$$\boldsymbol{d}_{obs}^{n} = \boldsymbol{d}_{tru}^{n} + \boldsymbol{\varepsilon}_{obs} \qquad \boldsymbol{\varepsilon}_{obs} \sim \pi_{obs}(\boldsymbol{\varepsilon}_{obs}) \tag{2.1}$$

where $\boldsymbol{\varepsilon}_{obs} \sim \pi_{obs}(\boldsymbol{\varepsilon}_{obs})$ denotes that $\pi_{obs}(\boldsymbol{\varepsilon}_{obs})$ is the probability distribution of $\boldsymbol{\varepsilon}_{obs}$. Typically, it is assumed monitoring/observation error, $\boldsymbol{\varepsilon}_{obs}$, is not correlated with true system behaviour, $\boldsymbol{d}_{tru}^{n}$, nor time, and monitoring/observation errors at different locations are not correlated (Rougier, 2009). Thus, $\pi_{obs}(\boldsymbol{\varepsilon}_{obs})$ is in fact a multivariate probability distribution of uncorrelated random measurement errors.

Let $\boldsymbol{g}(\boldsymbol{x}, t_n, \boldsymbol{m})$ denote a deterministic model of the behaviour of a CS system at time t_n and location $\boldsymbol{x}$ in the reservoir. The system model is assumed to be parameterized by a set of N_m uncertain parameters, $\boldsymbol{m} = \left[m_1, m_2, \dots, m_{N_m}\right]$, where m_i denotes uncertain model parameter i. Initial conditions, physical system properties and constitutive constants are examples of m_i. Also let $\pi(\boldsymbol{m})$ denote an N_m-dimensional non-cumulative probability distribution that reflects the uncertainty in $\boldsymbol{m}$. For simplicity of notation, let $\boldsymbol{y}^n(\boldsymbol{m})$ denote a vector of model outputs at time t_n and locations of interest, $\boldsymbol{x}_1$ to $\boldsymbol{x}_{N_x}$, i.e., $\boldsymbol{y}^n(\boldsymbol{m}) = \left[\boldsymbol{g}(\boldsymbol{x}_1, t_n, \boldsymbol{m}), \boldsymbol{g}(\boldsymbol{x}_2, t_n, \boldsymbol{m}), \dots, \boldsymbol{g}\left(\boldsymbol{x}_{N_x}, t_n, \boldsymbol{m}\right)\right]$.

The difference between the model and the true system behaviour, $\boldsymbol{d}_{tru}^n - \boldsymbol{y}^n(\boldsymbol{m})$, is called the model error or the *model discrepancy* (Rougier, 2009), which is denoted by $\boldsymbol{\varepsilon}_{mod}^n$, where the superscript n indicates time dependency of $\boldsymbol{\varepsilon}_{mod}^n$. Thus,

$$\boldsymbol{d}_{tru}^n = \boldsymbol{y}^n(\boldsymbol{m}) + \boldsymbol{\varepsilon}_{mod}^n \tag{2.2}$$

which is a generalization of (1.1). Combining (2.1) and (2.2), we can write

$$\boldsymbol{d}_{obs}^n = \boldsymbol{y}^n(\boldsymbol{m}) + \boldsymbol{\varepsilon}_{mod}^n + \boldsymbol{\varepsilon}_{obs} \tag{2.3}$$

which is generalization of (1.3). The basic assumptions for developing the equations in this section are:

A. The system is described by a model, $\boldsymbol{g}(\boldsymbol{x}, t, \boldsymbol{m})$, and a vector of parameters, $\boldsymbol{m}$.
B. Model parameters, m_i, are assumed to be time-invariant random variables.
C. The uncertainty in the model parameters is *epistemic*.
D. Measurement/observation error, $\boldsymbol{\varepsilon}_{obs}$, is a vector of Independent and Identically Distributed (IID) random variables, i.e., mutually independent random variables that have the same probability distribution. It is also assumed that measurement error is not correlated with location nor time.
E. Model error is described with a vector, $\boldsymbol{\varepsilon}_{mod}^n$, with the same dimensions as the mode output. Each element of the model error vector can be a random variable or a variable which depends on time, location, system model, and model parameters.

F. Evolution of the system with time is deterministic in the absence of measurement noise and model error and for a specific value of model parameters.

G. The Bayesian approach is used to define probability. More specifically, unlike the Frequentist inference which defines probability as the frequency of occurrence of a certain event in a large number of trials, here the probability is a measure of the degree of belief in the state of a system or value of a parameter.

Unless stated otherwise, the equations and assumptions presented in this section will be used in the rest of the thesis for the purpose of consistency.

The next four sections present a background on uncertainty propagation and Bayesian calibration, based on the assumptions and formulations presented in this section and to provide the required mathematical foundation for Chapters 3 to 5.

2.2. Propagation of Uncertainty

The purpose of uncertainty propagation is to investigate how the uncertainty of the parameters, $\boldsymbol{m}$, affects the uncertainty of the model output, $\boldsymbol{y}^n(\boldsymbol{m})$. In other words, we are interested in studying the propagation of parameter uncertainty in the CS system model by finding the probability distribution of $\boldsymbol{y}^n(\boldsymbol{m})$, denoted by $\pi\big(\boldsymbol{y}^n(\boldsymbol{m})\big)$, assuming known $\pi(\boldsymbol{m})$. Several methods exist to estimate $\pi\big(\boldsymbol{y}^n(\boldsymbol{m})\big)$ given $\pi(\boldsymbol{m})$. These methods can be used depending on the type of $\pi(\boldsymbol{m})$, the nature of $\boldsymbol{g}(\boldsymbol{x}, t, \boldsymbol{m})$, the available computational resources, and the required accuracy. Some of the most common uncertainty propagation methods are presented and compared in the following sub-section.

2.3. Mainstream Solutions to Propagation of Uncertainty

As described in the previous section, this section describes some the most common approaches for propagation of uncertainty of model parameters into the model outputs. A number of these methods are utilized in the Bayesian update techniques, which are described later in this chapter, and also in the sensitivity analysis method described in Chapter 3.

2.3.1. Analytical Solutions

In very simple cases, the probability distribution of the model output, $\pi\big(\boldsymbol{y}^n(\boldsymbol{m})\big)$, can be obtained analytically. For instance in a simple case where m is a scalar, $N_x = 1$, $g(x,t,m)$ is a scalar and a linear function of m (i.e., $g(x,t_n,m) = a(t_n)m + b(t_n)$) and $\pi(m)$ is Gaussian with a mean of μ_m and a variance of C_m, the probability distribution of $y^n(m)$ is also Gaussian with a mean of $a(t_n)\mu_m + b(t_n)$ and a variance of $a(t_n)^2\mathrm{C}_m$ (Ang & Tang, 2006). As described earlier in Chapter 1, the bold face is removed from model parameters and model outputs since they are assumed scalars in this sub-section.

In CS applications, the mathematical models are nonlinear and the parameter distributions are not typically Gaussian. Thus, analytical propagation of uncertainty is not feasible. In the scalar case, i.e., when $N_m = 1$ and $y^n(m)$ is a scalar, moments of $\pi\big(y^n(m)\big)$ can be approximated using the moments of $\pi(m)$ and the Taylor series expansion of $g(x,t_n,m)$ around the mean of $\pi(m)$, μ_m. The i^{th} moment of a probability distribution $\pi(m)$ around a fixed parameter value m_0 is

$$E[m^i] = \int\limits_{\Omega_m} (m - m_0)^i\, \pi(m)dm \tag{2.4}$$

where Ω_m denotes the parameter space. For instance, the mean of $\pi(m)$ can be obtained when $i = 1$ and $m_0 = 0$,

$$\mu_m = \int\limits_{\Omega_m} m\,\pi(m)dm \tag{2.5}$$

and variance of $\pi(m)$ can be obtained when $i = 2$ and $m_0 = \mu_m$,

$$\mathrm{C}_m = \int_{\Omega_m} (m - \mu_m)^2 \, \pi(m) dm \tag{2.6}$$

Similar to (2.4), the i^{th} moment of $y^n(m)$ around an arbitrary point y_0 is obtained by

$$E[(y^n(m))^i] = \int_{\Omega_m} (y^n(m) - y_0)^i \, \pi(m) dm \tag{2.7}$$

Using $i = 1$ and $y_0 = 0$, the mean of $\pi(y^n(m))$ can be obtained by

$$\mu_{y^n} = \int_{\Omega_m} y^n(m) \, \pi(m) dm \tag{2.8}$$

Approximating $y^n(m)$ by a first order Taylor series expansion about μ_m, μ_{y^n} can be found as

$$\mu_{y^n} = \int_{\Omega_m} \left[y^n(m) + \frac{\partial y^n(m)}{\partial m} (m - \mu_m) + \cdots \right] \pi(m) dm \tag{2.9}$$

Keeping only the first term gives

$$\mu_{y^n} \cong \int_{\Omega_m} y^n(\mu_m) \, \pi(m) dm \tag{2.10}$$

Since $y^n(m)$ is not a function of model parameters, m, it can be removed from the integral and using the fact that the integral of $\pi(m)$ over Ω_m is unity, μ_{y^n} can be approximated by

$$\mu_{y^n} \cong y^n(\mu_m) \tag{2.11}$$

Similarly, the variance of $\pi(y^n(m))$ is obtained from (2.7) as

$$\mathrm{C}_{y^n} = \int_{\Omega_m} \left[y^n(m) - \mu_{y^n} \right]^2 \pi(m) dm \tag{2.12}$$

where $y_0 = \mu_m$ and $i = 2$. Approximating $y^n(m)$ by a first order Taylor series expansion about μ_m and approximating μ_m by (2.10), C_{y^n} can be approximated by

$$\mathrm{C}_{y^n} \cong \int_{\Omega_m} \left[y^n(\mu_m) + \frac{\partial y^n(\mu_m)}{\partial m}(m - \mu_m) + \cdots - y^n(\mu_m) \right]^2 \pi(m) dm \tag{2.13}$$

The first term in the Taylor series is cancelled by $-y^n(\mu_m)$. Thus, keeping the first two term of the Taylor series, the variance of $\pi\big(y^n(m)\big)$ can be estimated as

$$\mathrm{C}_{y^n} \cong \int_{\Omega_m} \left[\frac{\partial y^n(\mu_m)}{\partial m}(m - \mu_m) \right]^2 \pi(m) dm \tag{2.14}$$

Since $\partial y^n(\mu_m)/\partial m$ is not a function of m, it can be removed from the integral. In addition, the rest of the integral is equal to C_m as given by (2.6). Thus, C_{y^n} can be approximated by

$$\mathrm{C}_{y^n} \cong \left(\frac{\partial y^n(\mu_m)}{\partial m} \right)^2 \mathrm{C}_m \tag{2.15}$$

Such linearization of a nonlinear function for approximating the moments of the probability distribution of a function of random variables provides a computationally efficient tool for propagation of uncertainty. However, this assumes that higher order terms in the Taylor series expansion of $y^n(m)$ are negligible. When $y^n(m)$ is a highly non-linear model, more sophisticated estimates of the moments are required (Kuczera, 1988).

2.3.2. Point Estimate Method

As described in the previous sub-section, simple analytical solutions are suboptimal for propagation of uncertainties in highly non-linear systems. Thus, other uncertainty propagations should be explored for this purpose. The Point Estimate method for the Probability Moments (PEPM) is another method for propagation of uncertainty by estimating moments of $\pi\big(y^n(\boldsymbol{m})\big)$ when $y^n(\boldsymbol{m})$ is a scalar, given the few first few moments of $\pi(\boldsymbol{m})$ (Rosenblueth, 1975) and is essentially a form of Gauss quadrature (Christian & Baecher, 1999). In the most commonly used form of PEPM and when only one uncertain parameter exists ($N_m = 1$), PEPM incorporates two quadrature points of the form

$$\hat{m}_{1,2} = \mu_m + \sigma_m \left[\frac{\nu_m}{2} \pm \sqrt{1 + \left(\frac{\nu_m}{2}\right)^2} \right] \tag{2.16}$$

where σ_m denotes the standard deviation of $\pi(m)$ and ν_m denotes the skewness coefficient of $\pi(m)$ which is a function of the third moment of $\pi(m)$,

$$\nu_m = \frac{1}{\sigma_m^3} \int_{\Omega_m} (m - \mu_m)^3 \, \pi(m) dm \tag{2.17}$$

The i^{th} moment of $y^n(m)$ in PEPM is then approximated by

$$E\left[\left(y^n(m)\right)^i\right] \cong W_1\left(y^n(\hat{m}_1)\right)^i + W_2\left(y^n(\hat{m}_2)\right)^i \tag{2.18}$$

where $E\left[\left(y^n(m)\right)^i\right]$ denotes the i^{th} moment of the probability distribution of $y^n(m)$ as defined in (2.7), and the weights, $W_{1,2}$, are obtained by

$$W_{1,2} = \frac{1}{2}\left[1 \pm \frac{\nu_m}{2} \frac{1}{\sqrt{1 + \left(\frac{\nu_m}{2}\right)^2}} \right] \tag{2.19}$$

The equations above are equivalent to approximating $\pi(m)$ with $W_1\delta(m - \hat{m}_1) + W_2\delta(m - \hat{m}_2)$ where $\delta(m)$ is the Dirac delta function (Tsai & Franceschini, 2005). It can be verified that the mean, variance, and skewness coefficient of such approximation of $\pi(m)$ is equal to the mean, variance, and skewness coefficient of $\pi(m)$ (Rosenblueth, 1975).

When the number of uncertain parameters, N_m, is greater than one, a total of 2^{N_m} points are required by PEPM (Christian & Baecher, 1999). Thus, the computational cost of the PEPM increases exponentially with N_m. This shortcoming is addressed in several studies, e.g., (Harrr, 1989; Hong, 1996; Hong, 1998; Christian & Baecher, 1999), a description of which is beyond the scope of this thesis. A review of point estimate methods used in environmental engineering applications, including the presented method (also known as the

Rosenblueth point estimate method) and its variations can be found in (Tsai & Franceschini, 2005).

2.3.3. Unscented Transformation

As described above, the main drawback of PEPM is that the number of quadrature points increases exponentially with the number of dimensions of the parameter space (Julier & Uhlmann, 2004). Unscented Transformation (UT) is a more recently developed method for obtaining estimations of $\pi(\boldsymbol{y}^n(\boldsymbol{m}))$ using the statistics of a finite number of points in the parameter space of $\boldsymbol{m}$ which are chosen deterministically based on $\pi(\boldsymbol{m})$ (Wan & Van Der Merwe, 2001). The main application of UT has been in the Unscented Kalman Filter (UKF) (Julier & Uhlmann, 1996).

In UT, $2N_m + 1$ weighted parameter sets are deterministically chosen, so that their mean is equal to $\boldsymbol{\mu}_m$, i.e., the mean of $\pi(\boldsymbol{m})$, and their covariance matrix is equal to $\mathbf{C}_m$, i.e., the covariance matrix of $\pi(\boldsymbol{m})$. These parameter sets are known as *Sigma points*. Let $\boldsymbol{\chi}_i$ and ϑ_i denote i^{th} Sigma point and weight, respectively. Note that though $\boldsymbol{\chi}_i$ is called a Sigma point, it is in fact a vector of model parameters, $\boldsymbol{m}$. $\boldsymbol{\chi}_i$ and ϑ_i are obtained by

$$
\begin{gathered}
\boldsymbol{\chi}_0 = \boldsymbol{\mu}_m \\
-1 < \vartheta_0 < 1 \\
\boldsymbol{\chi}_i = \boldsymbol{\mu}_m + \left(\sqrt{\frac{N_m}{1-\vartheta_0}\mathbf{C}_m}\right)_i \qquad i = 1,2,\dots,N_m \\
\boldsymbol{\chi}_i = \boldsymbol{\mu}_m - \left(\sqrt{\frac{N_m}{1-\vartheta_0}\mathbf{C}_m}\right)_i \qquad i = N_m+1, N_m+2,\dots,2N_m \\
\vartheta_i = \frac{1-\vartheta_0}{2N_m} \qquad i = 1,2,\dots,2N_m
\end{gathered}
\tag{2.20}
$$

where $\left(\sqrt{\frac{N_m}{1-\vartheta_0}\mathbf{C}_m}\right)_i$ is the i^{th} row of Cholesky decomposition of $\frac{N_m}{1-\vartheta_0}\mathbf{C}_m$. ϑ_0 is a tuning parameter which determines how scattered the Sigma points are about μ_m. It can be verified that these weighted Sigma point and $\pi(\boldsymbol{m})$ have the same mean and covariance matrix.

The Sigma points are then said to be propagated through $\boldsymbol{g}(\boldsymbol{x}, t_n, \boldsymbol{m})$. Let the aggregated system model output at locations $\boldsymbol{x}_1$ to $\boldsymbol{x}_{N_x}$, time t_n and for parameters $\boldsymbol{\chi}_i$ be denoted as

$$\boldsymbol{\Upsilon}_i = \left[\boldsymbol{g}(\boldsymbol{x}_1, t_n, \boldsymbol{\chi}_i), \boldsymbol{g}(\boldsymbol{x}_2, t_n, \boldsymbol{\chi}_i), \ldots, \boldsymbol{g}\left(\boldsymbol{x}_{N_x}, t_n, \boldsymbol{\chi}_i\right)\right] \qquad i = 0,1,\ldots,2N_m \tag{2.21}$$

where $\boldsymbol{\Upsilon}_i$ are the propagated Sigma points. Statistics of $\pi\left(\boldsymbol{y}^n(\boldsymbol{m})\right)$ are approximated from statistics of the weighted $\boldsymbol{\Upsilon}_i$. For instance, the mean and the covariance matrix of $\pi\left(\boldsymbol{y}^n(\boldsymbol{m})\right)$ are estimated by

$$\boldsymbol{\mu}_{\Upsilon} = \sum_{i=0}^{2N_m} \vartheta_i \boldsymbol{\Upsilon}_i \tag{2.22}$$

and

$$\mathbf{C}_{\Upsilon} = \sum_{i=0}^{2N_m} \vartheta_i (\boldsymbol{\Upsilon}_i - \boldsymbol{\mu}_{\Upsilon})^T (\boldsymbol{\Upsilon}_i - \boldsymbol{\mu}_{\Upsilon}) \tag{2.23}$$

respectively.

Similar to the PEPM, UT is equivalent to approximating $\pi(\boldsymbol{m})$ with $\sum_{i=0}^{2N_m} \vartheta_i \delta(\boldsymbol{m} - \boldsymbol{\chi}_i)$ and it can be verified that the mean and covariance matrix of such approximation of $\pi(\boldsymbol{m})$ is equal to the mean and covariance matrix of $\pi(\boldsymbol{m})$.

Unlike linearization which is only a first order accurate estimate of the mean and covariance of $\pi\left(\boldsymbol{y}^n(\boldsymbol{m})\right)$, the UT is up to third order accurate when $\pi(\boldsymbol{m})$ is Gaussian and up to second order accurate for an arbitrary $\pi(\boldsymbol{m})$ (Julier & Uhlmann, 1996). Other variants of UT have been developed which are up to third order accurate for any arbitrary distribution $\pi(\boldsymbol{m})$ and at least fourth order accurate for symmetric $\pi(\boldsymbol{m})$ (Julier & Uhlmann, 1997; Julier, 1998). Thus, UT would outperform Taylor series linearization in the propagation of uncertainty through non-linear CS models. In addition, considering the fact that the number of samples increases linearly with the number of parameters N_m in UT but increases exponentially with the number of parameters N_m in PEPM, UT is computationally more efficient than PEPM.

2.3.4. Monte-Carlo Simulation

Because of the ubiquity by which the Monte-Carlo Simulation (MCS) is used for propagation of uncertainty in CS and similar applications, it is necessary to explore the potential possibilities and shortcomings MCS. MCS is a commonly used approach for propagating the uncertainty of input parameters through nonlinear and complex models (Mark & Mordechai, 2011). MCS was initially developed during World War II and gained significant attention in various scientific and engineering applications (Mark & Mordechai, 2011). MCS relies on repetitive model simulation using samples from $\pi(\boldsymbol{m})$ in order to estimate $\pi\left(\boldsymbol{y}^n(\boldsymbol{m})\right)$.

When m and $y^n(m)$ are scalars, the concept behind MCS can be proven using law of large numbers which states that sample mean almost surely converges to the population mean when the number of samples tends to infinity (Chen, 2003), i.e.,

$$\mu_m = \lim_{N_s \to \infty} \frac{1}{N_s} \sum_{i=1}^{N_s} \widehat{m}_i \qquad \widehat{m}_i \sim \pi(m) \tag{2.24}$$

where $\widehat{m}_i$ are samples from the probability distribution of m. According to definition of μ_m using (2.4),

$$\int_{\Omega_m} m\pi(m) = \lim_{N_s \to \infty} \frac{1}{N_s} \sum_{i=1}^{N_s} \widehat{m}_i \qquad \widehat{m}_i \sim \pi(m) \tag{2.25}$$

Equation (2.25) is in fact a numerical integration where $\widehat{m}_i$ are quadrature points and all quadrature points has equal weights of $1/N_s$.

Similar to (2.24) and (2.25), mean of samples from the probability distribution of a function of a random variable converges to the mean of the probability distribution of that function when the number of samples tends to infinity. In other words,

$$\int_{\Omega_m} f(m)\pi(m)dm = \lim_{N_s\to\infty} \frac{1}{N_s}\sum_{i=1}^{N_s} f(\widehat{m}_i) \qquad \widehat{m}_i \sim \pi(m) \tag{2.26}$$

where $f(m)$ is an arbitrary function of m, $\widehat{m}_i$ are samples from $\pi(m)$ and $f(\widehat{m}_i)$ are samples from the probability distribution of $f(m)$. By replacing $f(m)$ with $(y^n(m) - y_0)^i$ and according to (2.7), any moment of $\pi\big(y^n(m)\big)$ around an arbitrary point y_0 can be approximated using moments of samples from $\pi(y^n)$ as

$$E[(y^n(m))^i] \cong \frac{1}{N_s}\sum_{j=1}^{N_s}\left[y^n(\widehat{m}_j) - y_0\right]^i \qquad \widehat{m}_j \sim \pi(m) \tag{2.27}$$

Thus, the moments of $\pi\big(y^n(m)\big)$ can be approximated by samples from $\pi(m)$ which are propagated through $y^n(m)$.

The traditional sampling method for MCS is the inverse CDF method. CDF stands for Cumulative Distribution Function and is equal to the integral of $\pi(m)$ over Ω_m if $\pi(m)$ is continuous, or the summation of $\pi(m)$ over Ω_m if $\pi(m)$ is discrete. In the inverse CDF method, N_s random numbers (denoted by u_i) are drawn from a uniform distribution between zero and one. Then, N_s samples (denoted by $\widehat{m}_i$) are obtained by inverting the CDF of $\pi(m)$ on the previously drawn random numbers (u_i), such that $u_i = \Pi(\widehat{m}_i)$ where $\Pi(m)$ is the CDF of m. Finally, the model output for those samples ($\widehat{m}_i$) is calculated. Schematic of the inverse CDF method for MCS is illustrated in Figure 2-1.

MCS for a multivariate case, i.e., when $N_m > 1$, follows a similar logic, as explained in (Bohdalová & Šlahor, 2007). Regardless of dimension of the parameter vector, MCS process is equivalent to approximating $\pi(\boldsymbol{m})$ by discrete distribution $\sum_{i=1}^{N_s} \delta(\boldsymbol{m} - \widehat{\boldsymbol{m}}_i)$ and propagating each discrete point $\widehat{\boldsymbol{m}}_j$ through the system model to obtain a discrete approximation of $\pi\big(\boldsymbol{y}^n(\boldsymbol{m})\big)$.

The main benefit of MCS over other uncertainty propagation methods is that it is applicable on all models and probability distributions. The cost of this flexibility is that MCS is computationally demanding. However, MCS is often used in many applications, including CS, due to recent advances in parallel computing (Walton *et al*., 2004; Kopp *et al*., 2007; Celia & Nordbotton, 2009; Dobossy *et al*., 2011).

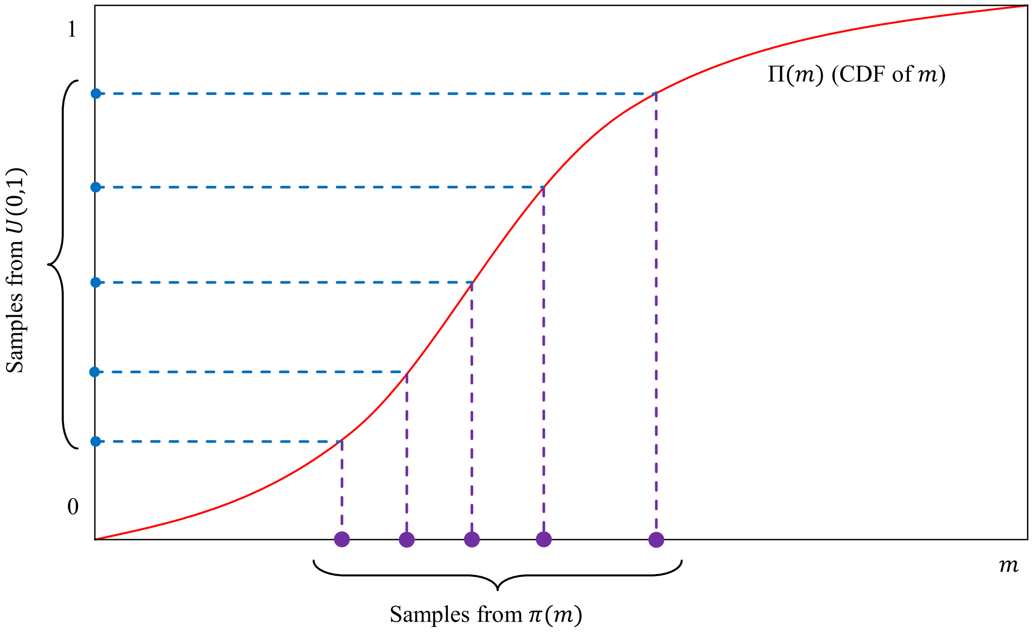


Figure 2-1: Schematic of the inverse CDF method for generating samples from a probability distribution $\pi(m)$ in the MCS method. $U(0,1)$ is a Uniform distribution between zero and one.

2.3.5. Latin Hyper-Cube Sampling

Latin Hyper-Cube Sampling (LHS) is a more recently developed method for sampling from a probability distribution using a "stratified sample without replacement" approach (McKay *et al*., 1979) and can replace MCS in many applications. In LHS, when $N_m = 1$, the CDF of m is divided into N_s equally probable intervals or *stratifications*. Then, one sample is drawn from each stratification. Thus, the number of samples is equal to the number of stratifications. Then, similar to MCS, samples can be propagated through $y^n(m)$ for approximating $\pi\left(y^n(m)\right)$. Figure 2-2 illustrates LHS in a one-dimensional parameter space.

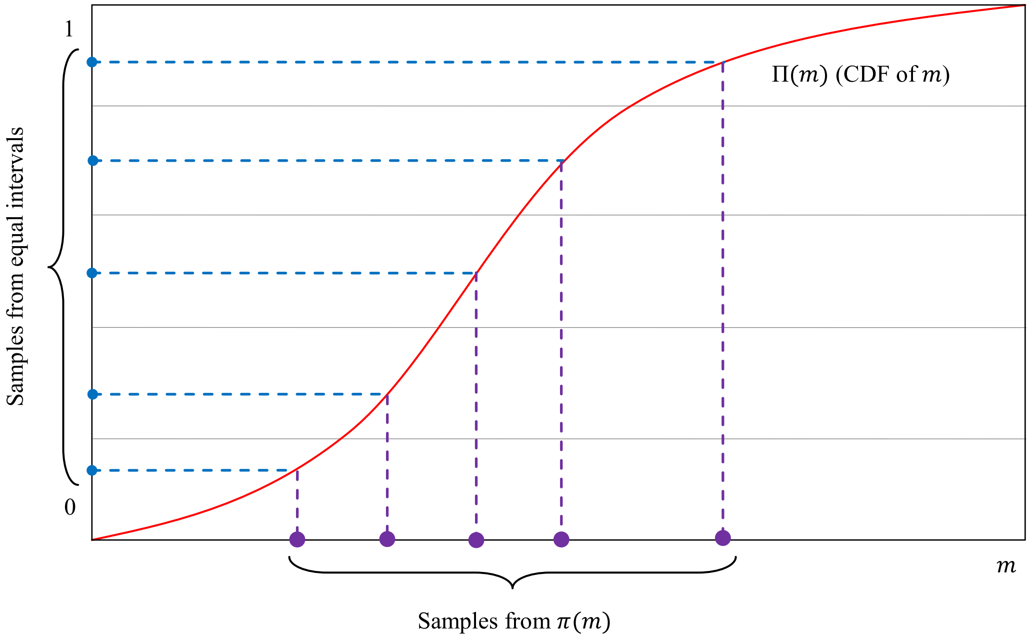


Figure 2-2: Schematic of the inverse CDF method for generating samples from a probability distribution in the LHS method.

When the dimension of the parameter space is greater than one, i.e., $N_m > 1$, and model parameters m_i are independent, each parameter of each sample $\hat{\boldsymbol{m}}_j$ is drawn independently from other parameters of that sample and no stratification is used more than once. For instance, assume $N_m = 2$ and $\boldsymbol{m} = [a, b]$ where a and b are two scalar parameters. Also assume the number of stratifications is five, i.e., $N_s = 5$. First, five samples $\hat{a}_1$ to $\hat{a}_5$ and $\hat{b}_1$ to $\hat{b}_5$ are drawn for each parameter as shown in Figure 2-2 and using the marginal CDFs of a and b. Then, $\hat{\boldsymbol{m}}_1$ to $\hat{\boldsymbol{m}}_5$ are generated as $\hat{\boldsymbol{m}}_j = \left[\hat{A}_j, \hat{B}_j\right]$. It can be seen that the number of samples is only dependent on the number of stratifications and does not change with dimensionality of the parameter space.

The LHS can be more precise than the MCS (Helton & Davis, 2003; Iman, 2008). However, all stratifications must be used in LHS and the results obtained in the middle of sampling (i.e., when the number of samples is smaller than the number of stratifications) could be significantly inaccurate. Furthermore, it is not possible to increase the number of samples, and therefore, accuracy, without restarting the sampling and discarding all previous samples in LHS (Helton & Davis, 2003).

2.3.6. Other Uncertainty Propagation Methods and Comparison

Table 2-1 briefly compares the benefits and drawbacks of the described uncertainty propagation methods. Several other sampling-based uncertainty propagation methods also exist such as Rejection Sampling (RS) (Handschin & Mayne, 1969), Stratified Sampling (Keramat & Kielbasa, 1998) and their variations. A more in depth description of uncertainty propagation methods can be found in the literature (Chen, 2003) and (De Rocquigny, 2012). A number of uncertainty propagation tools presented above have applications in the Bayesian calibration tools which are introduced in the next section.

Table 2-1: A comparison between common uncertainty propagation methods.

Method	**Benefits**	**Drawbacks**
Taylor series linearization	– Computationally very fast	– Requires calculation of the Jacobian matrices (hard for numerical models) – Low accuracy for nonlinear models
Point Estimate Methods	– Computationally fast (for lower dimensional problems) – Does not require calculation of the Jacobian matrices – More flexible than Taylor series linearization	– Computational cost increases exponentially with the parameter space dimensions – Low accuracy for nonlinear models
Unscented Transformation	– Computationally fast (for lower dimensional problems) – Higher accuracy and more flexible than Taylor series linearization – Does not require calculation of the Jacobian matrices	– Not as accurate as sampling for extremely nonlinear models and non-Gaussian systems
Monte-Carlo Simulation	– Flexible (no assumptions on model type or probability distributions) – Does not require calculation of the Jacobian matrices – Accurate	– Computationally demanding
Latin Hyper-Cube Sampling	– Can be more efficient than MCS	– All samples should be drawn simultaneously – When increasing the sample size, previous samples should be discarded

2.4. Bayesian Calibration

Bayesian frameworks allow updating the prior parameter distribution using new evidence and are extremely useful when data about state or parameters of a system is initially sparse or unavailable, e.g., in a CS system. Thus, the calibration methods introduced in Chapters 4 and 5 are based on Bayesian methods. To provide a background on the Bayesian methodologies for model calibration, i.e., Bayesian update or Bayesian calibration, Bayes' formula is described and common solutions to Bayes' formula are presented and compared in the following sub-sections.

2.4.1. Bayes' Theorem

Assume that X_1 and X_2 are two subsets of events in a sample space S. According to Bayes' theorem (a.k.a., Bayes' rule and Bayes' formula), the conditional probability of X_1 given X_2 is

$$P(X_1|X_2) = \frac{P(X_2|X_1)P(X_1)}{P(X_2)} \qquad P(X_2) \neq 0 \tag{2.28}$$

where $P(X_i)$ denotes the probability of X_i and $P\left(X_i \middle| X_j\right)$ denotes the conditional probability of X_i given X_j. According to the total probability law, $P(X_2) = P(X_2|X_1) + P(X_2|\overline{X_1})$, where X_1 and $\overline{X_1}$ are Mutually Exclusive and Collectively Exhaustive (MECE) events, i.e., $X_1 \cap \overline{X_1} = \emptyset$ and $X_1 \cup \overline{X_1} = S$. Thus, equation (2.28) can be re-written as

$$P(X_1|X_2) = \frac{P(X_2|X_1)P(X_1)}{P(X_2|X_1) + P(X_2|\overline{X_1})} \qquad P(X_2) \neq 0 \tag{2.29}$$

Bayes rule can be used for calibrating a deterministic model by finding/estimating the posterior distribution of parameters of the model given noisy measurements, as follows. Assume that the vector of model parameters, $\boldsymbol{m}$, is a vector of continuous random variables[1]. Let $\pi(\boldsymbol{m}|\boldsymbol{d})$ denote the probability distribution of model parameters, $\boldsymbol{m}$, conditioned by the observations (monitoring data) $\boldsymbol{d}$. According to Bayes' formula,

---

[1] This assumption is for simplicity of notation. For a discrete random variables, the following integrals will be changed to summations.

$\pi(\boldsymbol{m}|\boldsymbol{d})$ is proportional to the product of the unconditioned probability distribution of $\boldsymbol{m}$, $\pi(\boldsymbol{m})$ and the probability of $\boldsymbol{d}$ given $\boldsymbol{m}$, which is called the *likelihood* and is denoted by $L(\boldsymbol{d}|\boldsymbol{m})$, i.e.,

$$\pi(\boldsymbol{m}|\boldsymbol{d}) = c^{-1}.\, L(\boldsymbol{d}|\boldsymbol{m})\pi(\boldsymbol{m}) \tag{2.30}$$

where c is a normalizing constant and is given by

$$c = \int_{\Omega_m} L(\boldsymbol{d}|\boldsymbol{m})\pi(\boldsymbol{m})d\boldsymbol{m} \tag{2.31}$$

Finding/estimating the posterior distribution in equations (2.30) is the mathematical definition of calibration in this thesis. In equations (2.30) and (2.31), $\pi(\boldsymbol{m})$ is called the *prior* distribution or simply the *prior* and represents the prior knowledge of $\boldsymbol{m}$ while the evidence $\boldsymbol{d}$ is not accounted for (Rougier, 2009). The prior distribution is subjective and can be assigned based on initial measurements, data from similar sites, expert opinion, etc. The prior distribution is an important part of Bayesian updating and an arbitrary choice of the prior can negatively impact the credibility of the results (Bedi & Harrison, 2013). The *posterior* distribution, $\pi(\boldsymbol{m}|\boldsymbol{d})$, represents the probability distribution of $\boldsymbol{m}$ accounting for the evidence, $\boldsymbol{d}$, and can be interpreted as the calibrated version of the prior given $\boldsymbol{d}$. The likelihood allows us to predict unknown parameters using known evidence (Hall *et al.*, 2012) and can be defined based on the structure of the model error, $\boldsymbol{\varepsilon}_{mod}$, as described below.

2.4.2. Calibration with Unstructured Model Error

It is often assumed that the model error $\boldsymbol{\varepsilon}_{mod}$, given optimal model parameters, is a vector of random variables with a probability distribution $\pi_{mod}(\boldsymbol{\varepsilon}_{mod})$ and is independent from $\boldsymbol{m}$ and time (Kaipio & Somersalo, 2007). Thus, (2.3) can be rewritten as

$$\boldsymbol{d}_{obs}^{n} = \boldsymbol{y}^{n}(\boldsymbol{m}) + \boldsymbol{\varepsilon}_{mod} + \boldsymbol{\varepsilon}_{obs} \qquad \boldsymbol{\varepsilon}_{mod} \sim \pi_{mod}(\boldsymbol{\varepsilon}_{mod}) \qquad \boldsymbol{\varepsilon}_{obs} \sim \pi_{obs}(\boldsymbol{\varepsilon}_{obs}) \tag{2.32}$$

Let $\boldsymbol{\varepsilon}_{tot} = \boldsymbol{\varepsilon}_{mod} + \boldsymbol{\varepsilon}_{obs}$ and let $\pi_{tot}(\boldsymbol{\varepsilon}_{tot})$ denote the probability distribution of $\boldsymbol{\varepsilon}_{tot}$. Thus, $\boldsymbol{d}_{obs}^{n} = \boldsymbol{y}^{n}(\boldsymbol{m}) + \boldsymbol{\varepsilon}_{tot}$ and the likelihood of the measurements $\boldsymbol{d}_{obs}^{n}$ given $\boldsymbol{m}$ is obtained by (Rougier, 2009)

$$L(\boldsymbol{d}_{obs}^{n}|\boldsymbol{m}) = \pi_{tot}(\boldsymbol{\varepsilon}_{tot}) = \pi_{tot}\left(\boldsymbol{d}_{obs}^{n} - \boldsymbol{y}^{n}(\boldsymbol{m})\right) \tag{2.33}$$

Unlike the probability distribution of the monitoring/observation error, $\pi_{obs}(\boldsymbol{\varepsilon}_{obs})$, $\pi_{mod}(\boldsymbol{\varepsilon}_{mod})$ is less likely to be known *a priori*. The modelling discrepancy, $\boldsymbol{\varepsilon}_{mod}$, is commonly ignored, which is realistic if it is negligible compared to measurement noise; otherwise, it should be prescribed. When $\boldsymbol{\varepsilon}_{mod}$ is assumed negligible, $\boldsymbol{\varepsilon}_{tot} \cong \boldsymbol{\varepsilon}_{obs}$ and $\pi_{tot}(\boldsymbol{\varepsilon}_{tot}) \cong \pi_{obs}(\boldsymbol{\varepsilon}_{obs})$. When $\boldsymbol{\varepsilon}_{mod}$ is not negligible, $\pi_{mod}(\boldsymbol{\varepsilon}_{mod})$ can be estimated by minimizing the sum of squared errors using a least squares method and fitting a probability distribution on the resulting residuals (Stedinger *et al.*, 2008) or alternatively, the hyper-parameters of $\pi_{mod}(\boldsymbol{\varepsilon}_{mod})$ or $\pi_{tot}(\boldsymbol{\varepsilon}_{tot})$ can be included in the uncertain parameter vector $\boldsymbol{m}$.

Combining (2.30) and (2.33), the posterior distribution of $\pi(\boldsymbol{m})$ at time t_n and accounting for measurements/observation, $\boldsymbol{d}_{obs}^{n}$, is obtained by

$$\pi(\boldsymbol{m}|\boldsymbol{d}_{obs}^{n}) = \frac{\pi_{tot}\left(\boldsymbol{d}_{obs}^{n} - \boldsymbol{y}^{n}(\boldsymbol{m})\right)\pi(\boldsymbol{m})}{\int_{\Omega_m} \pi_{tot}\left(\boldsymbol{d}_{obs}^{n} - \boldsymbol{y}^{n}(\boldsymbol{m})\right)\pi(\boldsymbol{m})d\boldsymbol{m}} \tag{2.34}$$

It is worth noting that (2.34) is the posterior distribution of $\boldsymbol{m}$ given only the measurements at time t_n. To account for all measurements up to time t_n, it should be considered that elements of $\boldsymbol{\varepsilon}_{tot} = \boldsymbol{\varepsilon}_{mod} + \boldsymbol{\varepsilon}_{obs}$ are time-independent random variables and the probability of a set of independent events is equal to the product of the probabilities of those event. Thus, the likelihood of all observations/measurements is equal to the product of the likelihoods of each of the observations/measurements. Thus, we can re-write equation (2.34) as

$$\pi\left(\boldsymbol{m}|\boldsymbol{d}_{obs}^{1\ldots n}\right) = \frac{\pi^{0}(\boldsymbol{m})\prod_{l=1}^{n}\pi_{tot}\left(\boldsymbol{d}_{obs}^{l} - \boldsymbol{y}^{l}(\boldsymbol{m})\right)}{\int_{\Omega_m}\pi^{0}(\boldsymbol{m})\prod_{l=1}^{n}\pi_{tot}\left(\boldsymbol{d}_{obs}^{l} - \boldsymbol{y}^{l}(\boldsymbol{m})\right)} \tag{2.35}$$

where $\boldsymbol{d}_{obs}^{1 \dots n}$ denotes all monitoring data from time t_1 to t_n and $\pi^0(\boldsymbol{m})$ denotes initial prior distribution of $\boldsymbol{m}$ before obtaining any monitoring data. For simplicity of notation, we denote $\pi(\boldsymbol{m}|\boldsymbol{d}_{obs}^{1 \dots n})$ by $\pi^n(\boldsymbol{m})$. It should also be noted that the posterior distribution at time t_n becomes the prior distribution for time t_{n+1}, i.e., $\pi^{n+1}(\boldsymbol{m}) = c^{-1}L(\boldsymbol{d}_{obs}^{n+1}|\boldsymbol{m})\pi^n(\boldsymbol{m})$.

2.4.3. Calibration with Structured (Autoregressive) Model Error

The model error, $\boldsymbol{\varepsilon}_{mod}$, may not be a vector of time independent random variables in many applications. For instance, the nature of the numerical models used in applications such as CS requires discretization of the governing equations in time (and space), solving the equations for each time step and proceeding to the next time step. Thus, the model discrepancy accumulates during such numerical solutions and $\boldsymbol{\varepsilon}_{mod}$ becomes time-dependent. Consequently, it is more realistic to look at the model discrepancy as a first order autoregressive time series (Vrugt *et al.*, 2009), where the error vector at each time step is a function of the error vector at the previous time step plus a random noise, $\boldsymbol{\varepsilon}_{mod}^0$,

$$\boldsymbol{\varepsilon}_{mod}^n = u\left(\boldsymbol{\varepsilon}_{mod}^{n-1}\right) + \boldsymbol{\varepsilon}_{mod}^0 \qquad \boldsymbol{\varepsilon}_{mod}^0 \sim \pi_{mod}^0\left(\boldsymbol{\varepsilon}_{mod}^0\right) \tag{2.36}$$

where $\boldsymbol{\varepsilon}_{mod}^n = \boldsymbol{d}_{tru}^n - \boldsymbol{y}^n(\boldsymbol{m})$. For simplicity of calculations, assume $N_m = 1$, and system behaviour and model output are scalars. According to (Sorooshian & Dracup, 1980), if the error has a linear first order autoregressive structure, i.e., $u(\varepsilon_{mod}) = u^* \varepsilon_{mod}$ and $\pi_{tot}^0(\varepsilon_{tot}^0) = N\left(0, \mathrm{C}_{\varepsilon_{tot}^0}\right)$, i.e., a Normal distribution with a mean of zero and a variance of $\mathrm{C}_{\varepsilon_{tot}^0}$, where $\varepsilon_{tot}^0 = \varepsilon_{mod}^0 + \varepsilon_{obs}$, the likelihood of all observations up to time t_n given m is obtained by

$$L\left(d_{obs}^{1 \dots n}|m\right) = \frac{\left(1 - u^{*2}\right)^{\frac{1}{2}}}{\left(2\pi \mathrm{C}_{\varepsilon_{tot}^0}\right)^{\frac{n}{2}}} e^{\left(-\frac{1}{2\mathrm{C}_{\varepsilon_{tot}^0}}\left[(1-u^*)^2\left(d_{tru}^1 - y^1(m)\right)^2 - \sum_{l=2}^{n}\left(\varepsilon_{mod}^l - u^* \varepsilon_{mod}^{l-1}\right)^2\right]\right)} \tag{2.37}$$

Then, the posterior distribution of m at time t_n follows from (2.30) and (2.31),

$$\pi\left(m|d_{obs}^{1\dots n}\right) = \frac{L\left(d_{obs}^{1\dots n}|m\right)\pi^0(m)}{\int_{\Omega_m} L\left(d_{obs}^{1\dots n}|m\right)\pi^0(m)dm} \tag{2.38}$$

where $L(d_{obs}^{1\dots n}|m)$ is given by (2.37). If u^* is not known in *a priori*, it can be treated as one of the uncertain parameters. Similar to the unstructured model error and for simplicity of notation, we denote $\pi(m|d_{obs}^{1\dots n})$ by $\pi^n(m)$. Since the errors addressed in this thesis are primarily unstructured, extension of equation (2.37) for $N_m > 1$ is not discussed.

2.5. Mainstream Solutions to Bayesian Calibration

Several solutions exist to solve the Bayesian calibration equations presented in the last two previous sub-sections. To provide a technical background on these solutions and their potentials and shortcomings, the mainstream Bayesian calibration solutions are presented and compared in the subsequent sub-sections.

2.5.1. Conjugate Distributions

In cases where the prior and the likelihood are specific pairs of distributions, the posterior has an analytical closed form solution which is similar to the prior. These pairs of distributions are known as *conjugate* distributions, and were first introduced in (Raiffa & Schlaifer, 1961). Conjugate distributions are widely listed in statistical literature, e.g., (Ang & Tang, 2006). Conjugate distributions provide convenient solutions to many statistical problems. However, they limit the choice of the prior and the likelihood to a limited list of distributions which are not realistic in many real-life cases.

A widely used example of conjugate distributions are Gaussian distributions; if the prior and likelihood are both Gaussian, then the posterior is also Gaussian (Ang & Tang, 2006). Thus, Gaussian distributions are *self-conjugate*. This property of the Gaussian distributions is the basis for the Kalman Filter (KF) which is described in the next sub-section.

2.5.2. Kalman Filter

The Kalman Filter family are recursive Bayesian algorithms for updating the mean and covariance matrix of the posterior distribution of the state or the parameters of a system based on uncertain forecast and noisy data and were introduced by Kalman (Kalman, 1960). The original KF assumes Gaussian conjugate distributions.

In this sub-section, the standard notation of KF is used for consistency with the KF literature. At time t_n, assume that $\boldsymbol{d}^n$ denotes the state of a system, e.g., saturation or pressure in a CS system at certain locations and at time t_n, as a column vector. Assume $\boldsymbol{d}^n$ is a linear function of the previous state, $\boldsymbol{d}^{n-1}$, and the model control inputs, $\boldsymbol{u}^n$, as

$$\boldsymbol{d}^n = \boldsymbol{F}^n \boldsymbol{d}^{n-1} + \boldsymbol{B}^n \boldsymbol{u}^n + \boldsymbol{w}^n \qquad \boldsymbol{w}^n \sim N(0, \mathbf{C}_w^n) \tag{2.39}$$

where $\boldsymbol{F}^n$ is the linear state transition model, $\boldsymbol{B}^n$ is the linear control input model and $\boldsymbol{w}^n$ is the Gaussian modelling noise with a mean of zero and a covariance matrix of $\mathbf{C}_w^n$, at time t_n. Assume that the measurements are also linearly related to the system states plus a Gaussian noise, $\boldsymbol{v}^n$,

$$\boldsymbol{d}_{obs}^n = \boldsymbol{H}^n \boldsymbol{d}^n + \boldsymbol{v}^n \qquad \boldsymbol{v}^n \sim N(0, \mathbf{C}_v^n) \tag{2.40}$$

where $\mathbf{C}_v^n$ is the covariance matrix of the measurement noise and $\boldsymbol{H}^n$ maps the system state to the measurements less measurement noise. For instance, if $\boldsymbol{d}^n$ is the saturation of CO_2 in a CS system in 200 grid points and $\boldsymbol{d}_{obs}^n$ is the saturation at the 200th grid point, then $\boldsymbol{H}^n$ is a 1 by 200 vector with zero elements, except for the 200th element which is equal to one.

At each time step, a KF has two stages: the *prediction (forecast)* stage and the *measurement update* stage. The forecast stage at time t_{n-1} estimates the mean and covariance matrix of the system state or parameters at time t_n as

$$\begin{aligned} \boldsymbol{d}^{n|n-1} &= \boldsymbol{F}^n \boldsymbol{d}^{n-1|n-1} + \boldsymbol{B}^n \boldsymbol{u}^n \\ \mathbf{C}_d^{n|n-1} &= \boldsymbol{F}^n \mathbf{C}_d^{n-1|n-1} (\boldsymbol{F}^n)^T + \mathbf{C}_w^n \end{aligned} \tag{2.41}$$

where $\boldsymbol{d}^{n|n-1}$ and $\mathbf{C}_d^{n|n-1}$ are the prior mean and covariance matrix of system state at time t_n and $\boldsymbol{d}^{n-1|n-1}$ and $\mathbf{C}_d^{n-1|n-1}$ are posterior mean and covariance matrix of system state at time t_{n-1}.

The measurement update (or simply, update) stage estimates the current mean and covariance matrix of the system state or parameters at time t_n as

$$\begin{aligned} \boldsymbol{d}^{n|n} &= \boldsymbol{d}^{n|n-1} + \boldsymbol{K}^n\left(\boldsymbol{d}_{obs}^n - \boldsymbol{H}^n\boldsymbol{d}^{n|n-1}\right) \\ \mathbf{C}_d^{n|n} &= (\boldsymbol{I} - \boldsymbol{K}^n\boldsymbol{H}^n)\mathbf{C}_d^{n|n-1} \end{aligned} \tag{2.42}$$

where $\boldsymbol{K}^n$ is the optimal *Kalman gain* at time t_n. Kalman gain indicates how much weight should be given to the measurements, $\boldsymbol{d}_{obs}^n$, relative to forecast, $\boldsymbol{d}^{n|n-1}$. Using the normal conjugate formulation and considering the fact that a linear function of a normally-distributed random variable has a normal distribution itself, it can be verified that $\boldsymbol{K}^n$ is given by

$$\boldsymbol{K}^n = \mathbf{C}_d^{n|n-1}(\boldsymbol{H}^n)^T\left(\boldsymbol{H}^n\mathbf{C}_d^{n|n-1}(\boldsymbol{H}^n)^T + \mathbf{C}_v^n\right)^{-1} \tag{2.43}$$

In the KF literature, the prior is referred to as *background* and the posterior is referred to as *analysis*. For instance, the prior and posterior covariance matrices are referred to as *background covariance matrix* and *analysis covariance matrix*, respectively.

KF is essentially equivalent to using normal conjugate distributions in the measurement update stage, as described in Section 2.5, and in the forecast stage KF is essentially equivalent to using analytical uncertainty propagation for a normal distribution, as described in Section 2.3. Thus, if the system model and the measurement function are nonlinear, the KF cannot be used directly. In such cases, the nonlinear models can be linearized about $\boldsymbol{d}^{n|n-1}$ using a first order Taylor series expansion. This version of the KF is known as the Extended Kalman Filtering (EKF) (Chen, 2003). The main drawback of the EKF is that it is not accurate when the models are considerably nonlinear. Moreover, calculating the Jacobian of the model in the Taylor series expansion is nontrivial if a numerical model is being used, as would be expected of CS.

An alternative solution to EKF is linearizing the models using UT instead of the Taylor series expansion. This version of KF is the Unscented Kalman Filter (UKF) (Julier & Uhlmann, 1996). In UKF, approximated mean and covariance matrix of non-linear functions of random variables using UT replace the true mean and covariance matrix of that function in KF formulation, i.e., in (2.41) and (2.42). The UKF formulation is explained in details in Section 4.2.2 and is not repeated here. Unlike the EKF which is first order accurate, UKF is at least second order accurate for non-Gaussian Probability Distribution Functions (PDFs) and third order accurate for Gaussian PDFs (Chen, 2003).

Another solution for applying KF to nonlinear CS models is known as Ensemble Kalman Filter (EnKF) (Evensen, 1994), where statistics of samples from prior and posterior distributions are used instead of the statistics of those distributions. EnKF has found profound application when the number of variables are very large. For example, EnKF has been used in weather forecast (Houtekamer *et al*., 2005), reservoir engineering (Naevdal *et al*., 2005; Aanonsen *et al*., 2009) and hydrology (Reichle *et al*., 2002). It has been also used for the calibration of CS systems (Tavakoli *et al*., 2013).

In EnKF, statistics of an *ensemble* (i.e., a set of samples) are assumed to represent the mean and covariance matrix of the system state and/or parameters. Let $\boldsymbol{D}^{n|n-1} = \left[\left(\boldsymbol{d}_1^{n|n-1}\right)^T, \left(\boldsymbol{d}_2^{n|n-1}\right)^T, \ldots, \left(\boldsymbol{d}_{N_s}^{n|n-1}\right)^T\right]$ be a matrix with each column being one of the N_s ensemble members at time t_n. Each of these ensemble members is a sample from the prior distribution of model parameters/system states at time t_n. $\boldsymbol{D}^{n|n-1}$ is called the background or prior ensemble. Also, let $\boldsymbol{D}_{obs}^n = \left[\left(\boldsymbol{d}_{obs,1}^n\right)^T, \left(\boldsymbol{d}_{obs,2}^n\right)^T, \ldots, \left(\boldsymbol{d}_{obs,N_s}^n\right)^T\right]$ be a matrix with each column being the measurements plus a random Gaussian measurement noise. In EnKF, the analysis or the posterior ensemble is obtained by

$$\boldsymbol{D}^{n|n} = \boldsymbol{D}^{n|n-1} + \boldsymbol{K}^n\left(\boldsymbol{D}_{obs}^n - \boldsymbol{H}^n\boldsymbol{D}^{n|n-1}\right) \tag{2.44}$$

The mean and covariance matrix of the members of $\boldsymbol{D}^{n|n} = \left[\left(\boldsymbol{d}_1^{n|n}\right)^T, \left(\boldsymbol{d}_2^{n|n}\right)^T, \ldots, \left(\boldsymbol{d}_{N_s}^{n|n}\right)^T\right]$ are then used as the analysis state and analysis covariance matrix of the system. In other words,

$$\boldsymbol{d}^{n|n} \cong \frac{1}{N_s}\sum_{i=1}^{N_s} \boldsymbol{d}_i^{n|n} \tag{2.45}$$

and

$$\mathbf{C}_d^{n|n} \cong \frac{1}{N_s}\sum_{i=1}^{N_s} \left(\boldsymbol{d}_i^{n|n} - \boldsymbol{d}^{n|n}\right)\left(\boldsymbol{d}_i^{n|n} - \boldsymbol{d}^{n|n}\right)^T \tag{2.46}$$

Similar to UKF, EnKF does not require calculating the Jacobian of the models. For very high dimensional state vectors, EnKF is computationally more efficient than UKF (Kim, 2011). However, EnKF could be less accurate than UKF (Ambadan & Tang, 2009).

Replacing the system state with the system parameters $\boldsymbol{m}$, measurement update stage of KF-based solutions can be used for estimating the parameters of a system, as described in (Wan & Van Der Merwe, 2001). In this case, $\boldsymbol{d}^{n|n-1}$, $\boldsymbol{d}^{n|n}$, $\mathbf{C}_d^{n|n-1}$, and $\mathbf{C}_d^{n|n}$ represent the mean of $\pi^{n-1}(\boldsymbol{m})$, the mean of $\pi^{n}(\boldsymbol{m})$, the covariance matrix of $\pi^{n-1}(\boldsymbol{m})$, and the covariance matrix of $\pi^{n}(\boldsymbol{m})$, respectively. This process is described in detail in Section 4.2.2.

2.5.3. Multi-Grid Method

When the parameter space is discrete or can be discretized, a group of solutions known as Multi-Grid (MG) methods (Fedorenko, 1964) can provide the exact solution for Bayes' formula. MG methods is presented in a simple form in the following.

In a discrete parameter space, the integral in the denominator of Bayes' formula changes to a summation. Assume Ω_m is discrete, so that $\pi(\boldsymbol{m})$ can be expressed as

$$\pi(\boldsymbol{m}) = \sum_{i=1}^{N_s} W_i^0 \delta(\boldsymbol{m} - \widehat{\boldsymbol{m}}_i) \tag{2.47}$$

where $\widehat{\boldsymbol{m}}_i$ are the discrete points, W_i^0 are probability masses so that $\sum_{i=1}^{N_s} W_i^0 = 1$, and $\delta(\boldsymbol{m})$ is the Dirac delta function. Note that in this thesis $\widehat{\boldsymbol{m}}_i$ is used for denoting a sample of the parameter $\boldsymbol{m}$ and is a vector in general which should not be confused with the earlier definition of m_i as a component of $\boldsymbol{m}$. In this case, the normalizing constant in Bayes' formula is given by

$$c = \sum_{i=1}^{N_s} L(\boldsymbol{d}_{obs}^n|\widehat{\boldsymbol{m}}_i)\pi(\widehat{\boldsymbol{m}}_i) \tag{2.48}$$

Thus, the posterior distribution of $\boldsymbol{m}$ in (2.30) is given by

$$\pi(\boldsymbol{m}|\boldsymbol{d}_{obs}^n) = \sum_{i=1}^{N_s} W_i \delta(\boldsymbol{m} - \widehat{\boldsymbol{m}}_i) \tag{2.49}$$

where $\delta(\boldsymbol{m})$ is the Dirac delta function and W_i are the weights obtained by

$$W_i = \frac{L(\boldsymbol{d}_{obs}^n|\widehat{\boldsymbol{m}}_i)W_i^0}{\sum_{i=1}^{N_s} L(\boldsymbol{d}_{obs}^n|\widehat{\boldsymbol{m}}_i)W_i^0} \tag{2.50}$$

While MG methods provide the exact solution to Bayes' in a discrete parameter space, they are normally approximate for continuous parameter spaces and require a dense discretization for high accuracy. This requirement makes MG computationally expensive, especially if a uniform grid is used for discretization. Efforts were made to adaptively choose or modify the grid points to reduce the computational cost of MG (Bucy & Youssef, 1974; Kramer & Sorenson, 1988; Cai *et al.*, 1995). However, MG is not used in more recent works as frequently as more efficient sampling-based methods such as Markov-Chain Monte-Carlo or Importance Sampling.

2.5.4. Markov-Chain Monte-Carlo

Markov-Chain Monte-Carlo (MCMC) refers to a class of methods that are used for sampling from a probability distribution by constructing a Markov chain (Chen, 2003) and are widely applied for sampling from the posterior distribution in Bayes' formula. A Markov chain is a memory-less sequence, as the current state of the sequence depends only on the previous state and not on the states before that. In other words, a sequence $[\widehat{\boldsymbol{m}}_1, \widehat{\boldsymbol{m}}_2, \dots, \widehat{\boldsymbol{m}}_{N_s}]$ construct a Markov chain only if

$$P(\hat{\boldsymbol{m}}_{N_s}|\hat{\boldsymbol{m}}_{N_s-1}, \hat{\boldsymbol{m}}_{N_s-2}, \dots, \hat{\boldsymbol{m}}_1) = P(\hat{\boldsymbol{m}}_{N_s}|\hat{\boldsymbol{m}}_{N_s-1}) \tag{2.51}$$

where $P(\hat{\boldsymbol{m}}_i)$ stands for probability of the sequence being at state $\hat{\boldsymbol{m}}_i$ on step i.

Before explaining the application of MCMC in sampling from the posterior distribution in Bayes' formula, MCMC is presented in its general form in the following.

One of the main properties of a Markov chain is the *transition probability*, i.e., the probability that the chain moves from point $\hat{\boldsymbol{m}}_i$ to $\hat{\boldsymbol{m}}_j$ and is denoted by a probability distribution $q(\hat{\boldsymbol{m}}_j|\hat{\boldsymbol{m}}_i)$. It can be shown that if $q(\hat{\boldsymbol{m}}_j|\hat{\boldsymbol{m}}_i)$ is symmetric, i.e., $q(\hat{\boldsymbol{m}}_j|\hat{\boldsymbol{m}}_i) = q(\hat{\boldsymbol{m}}_i|\hat{\boldsymbol{m}}_j)$, the following algorithm which is known as the Metropolis algorithm (Metropolis & Ulam, 1949; Metropolis *et al.*, 1953), generates a Markov chain, where its members, after discarding a sufficiently large number of initial samples (N_b) known as the *burn-in* period, are samples from a probability distribution $\pi(\boldsymbol{m})$ known up to a normalizing constant. In other words, the Metropolis algorithm can be used to generate a Markov chain $[\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2, \dots, \hat{\boldsymbol{m}}_{N_b-1}, \hat{\boldsymbol{m}}_{N_b}, \hat{\boldsymbol{m}}_{N_b+1}, \dots, \hat{\boldsymbol{m}}_{N_b+N_s}]$ where $[\hat{\boldsymbol{m}}_{N_b+1}, \dots, \hat{\boldsymbol{m}}_{N_b+N_s}]$ are N_s random samples from $\pi(\boldsymbol{m})$.

The steps of the Metropolis algorithm are:

A. At the first step ($i = 1$), choose an initial sample $\hat{\boldsymbol{m}}_1$

B. At each step $i > 1$, draw a candidate sample $\boldsymbol{m}^*$ from a *transition kernel* or *proposal distribution* $q(\boldsymbol{m}^*|\hat{\boldsymbol{m}}_{i-1})$. Drawing the candidate sample from the proposal distribution can be done using methods such as inverse CDF method, as discussed in Section 2.3.4.

C. Calculate the acceptance probability, α_{MCMC}, as

$$\alpha_{MCMC} = \min\left(\frac{\pi(\boldsymbol{m}^*)}{\pi(\hat{\boldsymbol{m}}_{i-1})}, 1\right) \tag{2.52}$$

D. Accept $\boldsymbol{m}^*$ as $\hat{\boldsymbol{m}}_i$ with a probability of α_{MCMC}, i.e., generate a random number between zero and one and accept $\boldsymbol{m}^*$ as $\hat{\boldsymbol{m}}_i$ if that number is not larger than α_{MCMC}; otherwise, use $\hat{\boldsymbol{m}}_{i-1}$ as $\hat{\boldsymbol{m}}_i$.

E. Repeat steps 2 to 4 until the desired sample count plus a burn-in count N_b is achieved.

The original Metropolis algorithm was modified by Hasting (Hastings, 1970) to relax the symmetry condition on the transition kernel. The modified algorithm, known as Metropolis-Hasting, is similar to the original Metropolis algorithm except for the acceptance probability which is obtained by

$$\alpha_{MCMC} = \min\left(\frac{\pi(\boldsymbol{m}^*)q(\boldsymbol{m}^*|\widehat{\boldsymbol{m}}_{i-1})}{\pi(\widehat{\boldsymbol{m}}_{i-1})q(\widehat{\boldsymbol{m}}_{i-1}|\boldsymbol{m}^*)}, 1\right) \tag{2.53}$$

Convergence of the chain can be checked using various tests, a comprehensive review of which can be found in (Cowlesa & Carlinb, 1996). The choice of the proposal distribution is an important factor to the convergence rate of the Metropolis-Hastings algorithm (Brooks *et al.*, 2011). $q(\widehat{\boldsymbol{m}}_{i-1}|\boldsymbol{m}^*) = \pi(\boldsymbol{m}^*)$ seems a straightforward choice which results in $\alpha_{MCMC} = 1$, i.e., a 100% acceptance rate. However, MCMC is usually used when $\pi(\boldsymbol{m})$ is available only up to a normalizing constant. Finding a good proposal distribution for the Metropolis-Hastings MCMC algorithm has been an ongoing field of research in recent years (Brooks *et al.*, 2011), an in-depth discussion of which is beyond the scope of this chapter.

Now, we explaining the application of MCMC in sampling from the posterior distribution in Bayes' formula. Assume that $\pi(\boldsymbol{m}|\boldsymbol{d})$ is a posterior distribution, given by (2.30), known up to the normalizing constant. It can be verified that the normalizing constant will cancel out in the numerator and denominator when obtaining α_{MCMC} from (2.52) and (2.53). Thus, the Metropolis algorithm can be used for sampling from a posterior distribution even if the normalizing constant (which is the primary hard to calculate part of Bayes' formula) is unknown. Thus, MCMC methods, and specifically, Metropolis-Hasting algorithm and its variants, have become one of the standard methods to sample from the posterior distributions in applications such as groundwater and environmental modelling (Clark, 2005; Hassan *et al.*, 2009; Wu & Zeng, 2013) and reservoir history matching (Liu & McVay, 2009; Oliver & Chen, 2011). The literature is rich in other variations of MCMC, e.g., Reversible Jump MCMC (Green, 1995), Gibbs sampler (Geman & Geman, 1984), Hybrid Monte Carlo (Durbin & Koopman, 1997) and adaptive variations of MCMC (Andrieu & Moulines, 2006; Brooks *et al.*, 2011), a description of which is beyond the scope of this chapter.

The burn-in size of MCMC depends on the problem, choice of the initial chain member and the transition kernel. The burn-in size can be significant based on the problem and how the algorithm is designed (Liu & McVay, 2009). Thus, using MCMC for sampling from the posterior distribution of uncertain parameters of a computationally demanding model (such as CS models) is subject to the challenge of demanding a large number of model simulation in the burn-in period which negatively impacts the computational efficiency of the sampling process (Ma *et al.*, 2006; Liu & McVay, 2009).

2.5.5. Importance Sampling

Importance Sampling (IS), which was initially introduced by (Marshall, 1956), is a solution for sampling from a probability distribution, $\pi(\boldsymbol{m})$, which is available at least up to the normalizing constant. Similar to MCMC, IS can be used for sampling from the posterior distribution, given by (2.30), in Bayesian updating. In IS, samples are drawn from a proposal distribution, $q(\boldsymbol{m})$, which is close to $\pi(\boldsymbol{m})$ and is easy to sample from. Each sample is then weighted by the ratio of $\pi(\boldsymbol{m})$ over $q(\boldsymbol{m})$ and the weights are normalized to sum up to one. The proposal distribution encourages samples from more important areas of the parameter space, i.e., areas with a larger $\pi(\boldsymbol{m})$, to reduce the variance of the estimator and increase efficiency of sampling.

Before explaining the application of the IS in Bayesian updating, first the IS itself is explained. In the IS method, $\pi(\boldsymbol{m})$ is approximated as

$$\pi(\boldsymbol{m}) \cong \sum_{i=1}^{N_s} W_i \delta(\boldsymbol{m} - \widehat{\boldsymbol{m}}_i) \qquad \widehat{\boldsymbol{m}}_i \sim q(\boldsymbol{m}) \tag{2.54}$$

where $\widehat{\boldsymbol{m}}_i$ are the samples randomly drawn from $q(\boldsymbol{m})$ by using methods such as the inverse CDF method as explained in Section 2.3.4, and $\delta(\boldsymbol{m})$ is the Dirac delta function. W_i are the normalized *importance weights*, obtained by

$$W_i = \frac{w_i}{\sum_{i=1}^{N_s} w_i} \qquad i = 1 \dots N_s \tag{2.55}$$

where

$$w_i = \pi(\widehat{\boldsymbol{m}}_i)/q(\widehat{\boldsymbol{m}}_i) \tag{2.56}$$

The statistics of $\pi(\boldsymbol{m})$ can then be estimated using the statistics of the samples $\{\widehat{\boldsymbol{m}}_1, \widehat{\boldsymbol{m}}_2, \dots, \widehat{\boldsymbol{m}}_{N_s}\}$ and the importance weights $\{W_1, W_2, \dots, W_{N_s}\}$. For instance, the IS estimation of the mean and covariance matrix of $\pi(\boldsymbol{m})$ are

$$\boldsymbol{\mu}_m \cong \sum_{i=1}^{N_s} W_i \widehat{\boldsymbol{m}}_i \tag{2.57}$$

and

$$\mathbf{C}_m \cong \sum_{i=1}^{N_s} W_i (\widehat{\boldsymbol{m}}_i - \boldsymbol{\mu}_m)\,(\widehat{\boldsymbol{m}}_i - \boldsymbol{\mu}_m)^T \tag{2.58}$$

respectively.

It can be shown that $\pi(\boldsymbol{m})/q(\boldsymbol{m})$ must be finite (Rougier, 2009), which means that the proposal needs to have heavier tails than the sampling distribution (Owen & Zhou, 2000). In order to satisfy this condition, a heavy-tailed proposal distribution can be used (Rougier, 2009), or the proposal distribution can be combined with a heavy-tailed distributions such as the uniform distribution (Hesterberg, 1995; Owen & Zhou, 2000) as

$$q^*(\boldsymbol{m}) = (1 - \eta) q(\boldsymbol{m}) + \eta \hat{q}(\boldsymbol{m}) \tag{2.59}$$

where η is the mixture ratio and varies between zero and one, $\hat{q}(\boldsymbol{m})$ is the heavy-tailed mixture distribution, and $q^*(\boldsymbol{m})$ is the combined proposal distribution.

IS is especially useful when a distribution is hard to sample from but is available up to the normalizing constant, such as the posterior distribution in Bayes' formula. Thus, the application of IS in the Bayesian updating is described in the following.

Replacing $\pi(\boldsymbol{m})$ with $\pi(\boldsymbol{m}|\boldsymbol{d}_{obs}^{n})$ in (2.56), the un-normalized weights of samples from the posterior distribution are calculated as

$$w_i = c^{-1}\frac{\pi_{tot}\big(\boldsymbol{d}_{obs}^{n} - \boldsymbol{y}^{n}(\widehat{\boldsymbol{m}}_i)\big)\pi^{n-1}(\widehat{\boldsymbol{m}}_i)}{q(\widehat{\boldsymbol{m}}_i)} \qquad i = 1 \dots N_s \tag{2.60}$$

and the normalized weights are calculated according to equation (2.55) as

$$W_i = \frac{c^{-1}\dfrac{\pi_{tot}\big(\boldsymbol{d}_{obs}^{n} - \boldsymbol{y}^{n}(\widehat{\boldsymbol{m}}_i)\big)\pi^{n-1}(\widehat{\boldsymbol{m}}_i)}{q(\widehat{\boldsymbol{m}}_i)}}{c^{-1}\sum_{i=1}^{N_s}\dfrac{\pi_{tot}\left(\boldsymbol{d}_{obs}^{n} - \boldsymbol{y}^{n}(\widehat{\boldsymbol{m}}_i)\right)\pi^{n-1}(\widehat{\boldsymbol{m}}_i)}{q(\widehat{\boldsymbol{m}}_i)}} \qquad i = 1 \dots N_s \tag{2.61}$$

It can be seen that the normalizing constant (c) will be cancelled from numerator and denominator when calculating the importance weights. Thus, similar to MCMC, the posterior distribution can be approximated using IS without calculating the normalizing constant.

Equation (2.61) presents the IS weights when samples are drawn from $\pi(\boldsymbol{m}|\boldsymbol{d}_{obs}^{n})$. To sample from $\pi^{n}(\boldsymbol{m})$, i.e., the posterior distribution of $\boldsymbol{m}$ accounting for all monitoring data up to time t_n, equation (2.61) can be modified to

$$W_i = \frac{\dfrac{\pi^{0}(\widehat{\boldsymbol{m}}_i)\prod_{l=1}^{n}\pi_{tot}\left(\boldsymbol{d}_{obs}^{l} - \boldsymbol{y}^{l}(\widehat{\boldsymbol{m}}_i)\right)}{q(\widehat{\boldsymbol{m}}_i)}}{\sum_{i=1}^{N_s}\dfrac{\pi^{0}(\widehat{\boldsymbol{m}}_i)\prod_{l=1}^{n}\pi_{tot}\left(\boldsymbol{d}_{obs}^{l} - \boldsymbol{y}^{l}(\widehat{\boldsymbol{m}}_i)\right)}{q(\widehat{\boldsymbol{m}}_i)}} \qquad i = 1 \dots N_s \tag{2.62}$$

It has been shown that the optimal proposal distribution in IS should be close to the posterior distribution (Chen, 2003). However, the posterior distribution is usually unknown. Researchers have tried to iteratively adjust the proposal distribution to make it closer to the posterior distribution and so improve the efficiency of the IS. These efforts resulted in the Adaptive Importance Sampling (AIS) algorithms for iteratively optimizing the proposal distribution. The VEGAS algorithm (Lepage, 1978; Lepage, 1980) is an iterative AIS which utilizes the posterior distribution of each iteration to improve the proposal distribution of the next iteration. Pennanen & Koivu used a stochastic gradient decent method to adjust the proposal density and minimize the estimator variance

(Pennanen & Koivu, 2006). More recently developed AIS algorithms, such as Mixture Population Monte-Carlo (M-PMC) (Cappé *et al*., 2008) and Adaptive Multiple Importance Sampling (AMIS) (Cournet *et al*., 2012) utilized a mixture of the posterior distributions obtained at each iteration to adaptively adjust the proposal distribution. The iterative nature of these algorithms requires the sampling to be carried out several times until the proposal distribution gets close to the posterior distribution. This procedure increases the computational cost of model simulation several times compared to non-iterative algorithms. Thus, the abovementioned AIS algorithms are not optimal, in terms of computational efficiency, to be used with computationally costly models such as those used in CS.

2.5.6. Other Bayesian Calibration Methods and Comparison

While other methods exist to estimate Bayes' formula, most of them are based on one of the methods described above. Table 2-2 briefly compares the benefits and the drawbacks of the described methods for Bayesian update.

Table 2-2: A comparison between common Bayesian update solutions.

Method	**Benefits**	**Drawbacks**
Conjugate families	– Computationally very fast – Analytically tractable	– Applicable only for a limited number of distributions and specific types of models
Kalman Filter (measurement update stage)	– Computationally very fast – Analytically tractable	– Applicable only on linear models and Gaussian distributions
Extended Kalman Filter (measurement update stage)	– Computationally very fast	– Requires calculation of Jacobian matrixes (non-trivial for numerical models) – Low accuracy for extremely nonlinear models and non-Gaussian distributions

Method	Benefits	Drawbacks
Unscented Kalman Filter (measurement update stage)	– Computationally faster than sampling-based methods (except for systems with very large dimensions) – At least second degree accurate for nonlinear models and non-Gaussian distributions – Does not require calculation of Jacobian matrixes	– Computationally costly for very high dimensional systems such as weather forecast models
Ensemble Kalman Filter (measurement update stage)	– Computationally efficient for very high dimensional systems such as weather forecast	– Low accuracy for extremely nonlinear models and non-Gaussian distributions
Multi-Grid method	– No assumptions on the model type – Accurate in a dense grid	– Computationally costly in a high resolution grid
Markov-Chain Monte-Carlo	– No assumptions on the model type or on the distributions – Can be computationally more efficient than Multi-Grid method	– Burn-in period discards a large number of samples
Bayesian Importance Sampling	– No assumptions on the model type or on the distributions – Computationally more efficient than Multi-Grid method for high dimensional systems – A well-designed IS is Computationally more efficient than MCMC since there is no burn-in period	– Requires a proposal distribution close to the posterior distribution to be computationally efficient, specifically in high dimensional systems

2.6. Discussion and Concluding Remarks

This chapter provided the mathematical basis for Chapters 3 to 5 of the thesis by introducing a general formulation and general set of assumptions for mathematical representation of a CS system and by presenting the background for the statistical methods used for the propagation and mitigation of uncertainty.

The uncertainty propagation methods presented in Section 2.2 and Section 2.3 are applicable on various applications namely probabilistic forecast, quantitative risk assessment and several sensitivity analysis techniques. As reported in Section 2.3.4, MCS is one of the state-of-the-art uncertainty analysis methods in many applications and has been the primary tool for the uncertainty propagation in CS systems to date (Walton *et al.*, 2004; Kopp *et al.*, 2007; Celia & Nordbotton, 2009; Dobossy *et al.*, 2011). Thus, the parametric sensitivity analysis conducted in the next chapter (Chapter 3) adapts a sensitivity measure based on MCS technique. In addition, several uncertainty propagation methods described in Section 2.2 and Section 2.3 provided a background for a number of Bayesian update techniques, e.g., UT provided a background for UKF and Taylor series linearization provided a background for EKF.

On the other hand, the Bayesian update methods presented in Section 2.3.4 and Section 2.5 provided a background for the Bayesian calibration techniques introduced in Chapters 4 and 5. A shown in Table 2-2, each of the presented Bayesian update techniques has advantages and disadvantages in terms of accuracy and computational efficiency. However, these methods all tend to be suboptimal for a CS system when considering accuracy and computational efficiency simultaneously: computationally efficient methods are normally inaccurate and simulation-based methods are normally computationally demanding. As formerly discussed in Chapter 1, the trade-off between accuracy and computational efficiency is the common shortcoming of the efforts to date for the calibration of CS and comparable systems. Thus, Chapter 4 is an attempt to introduce a hybrid solution for Bayesian calibration of CS systems, a solution which benefits from the computational efficiency of an approximate solution and the accuracy of a sampling-based solution. Then, Chapter 5 will benefit from efficient sampling objective introduced in Chapter 4 to reduce the number of CS model simulations during the calibration solution.

Chapter 3.

Parametric Sensitivity Analysis

This chapter is the mirror of the following article. Minor changes are made, mostly in the Summary and Section 3.1, for greater consistency with the body and notation of the thesis. Section 3.4.4 does not exist in the article. In this chapter, Sections 3.3.1 and 3.4.3 are co-authored contributions and I am the main contributor to the rest of the sections.

Sarkarfarshi, M., Malekzadeh, F.A., Gracie, R. and Dusseault, M.B. 2014. Parametric Sensitivity Analysis for CO_2 Sequestration. *International Journal of Greenhouse Gas Control*, 23: 61-71.

Summary

As described in Chapter 1, the overall goal of this thesis is to reduce the trade-off between computational cost and accuracy when calibrating CS systems. The first strategy to achieve this goal is to reduce the dimensionality of the CS system's parameter space, which will be achieved by identifying the parameters that contribute the most to the uncertainty of a CS model output, and excluding parameters with little or no effect on the uncertainty of a CS model output from the calibration. Thus, the purpose of this chapter is to present a sensitivity analysis to quantify the impact of different uncertain parameters on the overall uncertainty of CO_2 plume evolution and to provide a basis for reducing the dimensionality and computational cost of future calibration and forecast efforts.

The sensitivity analysis method used in this chapter is based on the Monte-Carlo Simulation (MCS), which was formerly described in Chapter 2. The measure of sensitivity used in this chapter includes both the role of the parameter in the model, and the parameter uncertainty, in order to distinguish between parameters with equal influence in the model output, yet that have different degrees of intrinsic uncertainty. In a case study, the sensitivity of the plume interface location, the maximum breakthrough distance of CO_2, and the moment of inertia of the CO_2 plume with respect to intrinsic physical system parameters and parameters introduced in constitutive relationships is investigated. The probability distributions of parameters of the Nisku aquifer which is targeted for CO_2 injection in Alberta, Canada are used in the case study. It is also shown that the contribution of constitutive relationship parameters in plume evolution uncertainty can be as high as that of the physical characteristics of the system.

3.1. Introduction

As reported in Chapter 1, any prediction of the long-term fate of the injected CO_2 or assessment of risk in CS systems is subject to several sources of uncertainty and parameter uncertainty is believed to be the dominant source of uncertainty in a CS system (Celia & Nordbotton, 2009; Polson *et al.*, 2012). Uncertainty in the parameters of geological systems such as CS comes from limited knowledge of the aquifer properties and conditions, heterogeneity of the aquifer, sparse and inaccurate measurements of the aquifer, data interpolation and model calibration strategies (Alshuhail *et al.*, 2009; Lavoie & Keith, 2010; Zhao *et al.*, 2010; Sarkarfarshi & Gracie, 2013).

In this chapter, a quantitative sensitivity analysis is presented to quantify the impact of different uncertain parameters on the uncertainty of CO_2 plume evolution in deep saline aquifers. This quantitative comparison which is the first objective for accurate and computationally efficient calibration of CS models in this thesis provides a basis for reducing the dimensionality and computational cost of future calibration and forecast efforts. In other words, this comparison allows to focus our attention on the parameters that have the greatest impact on the uncertainty and reduce the computational cost of probabilistic model forecast and model calibration techniques. In addition, methodology and results of this chapter can contribute to prioritizing data collection activities.

We hypothesize that the influence of constitutive relationship parameters in CO_2 plume evolution uncertainty can be as high as the influence of the intrinsic physical characteristics of the system. Two parameter groups considered in this study are:

A. *Physical Parameters:* Intrinsic physical characteristics of the system such as porosity, reservoir depth, reservoir geometry and reservoir temperature.
B. *Constitutive Relationship Parameters*: Constitutive relationship parameters introduced in constitutive equations and in idealized parameter correlations, such as parameters in relative permeability-saturation and capillary pressure-saturation relationships.

To date, sensitivity analysis, in a CS context, directly or indirectly, has been conducted only in a limited number of studies. For instance, (Kopp *et al*., 2007) conducted a sensitivity analysis investigation where the sensitivity of CO_2 arrival time, free-phase CO_2 to total injected CO_2 fractions, dissolved CO_2 fractions, and pressure build-up in a 2D dipping aquifer model, defined by 15 parameters, was examined using the extended Morris Method (Morris, 1991; Campolongo *et al*., 2005). Kopp *et al*. concluded that horizontal permeability, injection interval and reservoir dimensions were the most influential parameters, whereas porosity, sorting factor (i.e., the exponent in the Brooks and Corey relationship), injection temperature, and dip angle were the least influential parameters. This approach provides a qualitative basis for global ranking of parameters, but the site-specific degree of uncertainty of parameters was not reflected in the measure of sensitivity.

(Sifuentes *et al*., 2009) and (Han *et al*., 2011) studied the contribution of different parameters to the residual (capillary) trapping and solubility trapping mechanisms in CS projects with a sensitivity analysis. However, the trapping mechanisms and time scales of interest in these studies were different from those investigated here.

(Golder Associates, 2010) performed a sensitivity analysis on leakage through abandoned wells, using Wabamun Area Sequestration Project data. They used tornado charts and XY function charts to graphically present the sensitivity of leakage from an abandoned well, and also to present the sensitivity of the maximum radius of the CO_2 plume to different parameters. Parameter bounds were selected based on the mean and quantiles of the probability distribution of the parameters. Such representation provides a quick and easy guide for the decision makers for identifying the degree of correlation between each parameter and the model output. However, it does not account for the shape of the probability distribution of the parameters since parameter values in the XY plot are chosen from an equally-spaced grid and tornado charts just incorporated the mean value of the parameter and upper and lower bounds of the parameter ranges.

Sensitivity of CO_2 saturation distribution, plume arrival time, site injectivity and storage capacity to the heterogeneity in porosity and permeability fields is investigated by a small number of authors. For instance, (Lengler *et al*., 2010) investigated the impact of heterogeneity of petrophysical properties of the Ketzin CS site in Germany on the saturation distribution of the CO_2 and arrival time of the plume to a vertical column 50 m from the injection well. They concluded that permeability heterogeneity reduced the CO_2 arrival time at observation wells, decreased the sharpness of the plume interface, and reduced injectivity. (Deng *et al*., 2012) showed that storage capacity can also be affected by heterogeneity. Using a hypothetical CS site in the Rock Springs Uplift, they showed that heterogeneity in porosity and permeability fields can reduce the storage capacity by 64% and reduce the maximum injection time by a factor of three, compared to previous homogeneous estimates by (Surdam *et al*., 2007).

These studies provide valuable insight into the correlations between characteristics of a CS system and various measures of plume extension, trapping mechanisms and leakage. However, none of these studies incorporates parameter roles in the model and parameter distributions at the same time in the measure of sensitivity, or compares the effect of physical and constitutive relationship parameters in the uncertainty of CS model outputs.

The sensitivity analysis presented in this chapter distinguishes between parameters with the same significance in the model structure, but different degrees of uncertainty. For example, consider reservoir effective porosity and the CO_2 injection rate where the former is usually much more uncertain than the latter. Moreover, the approach taken in this chapter requires a realistic characterization of parameter uncertainty; therefore the properties of the Nisku aquifer in Alberta are used to characterize parameter distributions in a hypothetical CO_2 geosequestration site. Though the sensitivity analysis and a parameter ranking is conducted based on site-specific data; the approach is general and applicable to other CS sites.

To account for both the role of the parameters in the model and parameter uncertainty, sensitivity measures such as global sensitivity indices (Homma & Saltelli, 1996; Sobol, 2001) or Relative Deviation (RD) (Hamby, 1994) can be used. Due to simplicity and easiness of application, the RD method is selected for the sensitivity analysis in this chapter. RD is defined to be equal to the Coefficient Of Variation (COV) in the model output, subjected to parameter variations, and provides a reliable sensitivity measure (Hamby, 1994; Hamby, 1995). Parameters are ranked based on three criteria:

I. the vertically averaged RD of the CO_2 plume boundaries, measured at different times,
II. the RD of the maximum radial extension of the plume, measured at different times; and
III. the RD of the moment of inertia of the plume, measured at different times.

RD is a sampling based method requiring repetitive model runs. Such a method requires a computationally efficient model to conduct repetitive simulations in a reasonable amount of time. A two-phase flow, two-dimensional Finite Element (FE) model is used to simulate CO_2 plume evolution in a saline aquifer. To make the sensitivity analysis computationally tractable, the system model assumes an axisymmetric geometry, homogeneous rock properties, a single injection well, and an incompressible rock matrix. The system model is discussed in detail in Section 3.3.

In Section 3.2, sensitivity analysis using RD method is described and formulated. Next, a synthetic CS case study is described in Section 3.3 and criteria for ranking the parameters in the sensitivity analysis are described in Section 3.3.4. Sensitivity analysis results are described in Section 3.3.5 and discussed in Section 3.4. Finally, conclusions are summarized in Section 3.5.

3.2. Methodology

The objective of a sensitivity analysis is to identify which mathematical model input parameters most significantly impact the uncertainty in the model outputs. The sources of uncertainty of input parameters cannot be related in a straightforward way to the uncertainty in the model output because each input parameter may be associated with a different degree of uncertainty. Dissimilarities in uncertainty arise due to different temporal and spatial variability of the parameters, different measurement/monitoring data used to calibrate the parameters and so on. To compare the influence of different parameters on model output uncertainty, it is necessary to take into account both how much uncertainty is attributed to each parameter, as well as the role of the parameter in the mathematical structure of the model.

In the RD algorithm, model output uncertainty is measured through the coefficient of variation, while sampling input parameters One-At-a-Time (OAT) from parameter probability distribution functions.

The RD method and the OAT sampling approach are now described in a more mathematical form.

As described previously in Chapter 2, let $g(\boldsymbol{x}, t_n, \boldsymbol{m})$ denote a deterministic system model at time t_n and at location $\boldsymbol{x}$ and with an scalar output, which is parameterized by a set of N_m time-invariant and location-invariant uncertain parameters, $\boldsymbol{m} = \left[m_1, m_2, \dots, m_{N_m}\right]$. Assume each m_i has a marginal probability distribution $\pi_i(m_i)$.

Let $\boldsymbol{D}_i^{x,n} = \left\{g\left(\boldsymbol{x}, t_n, \left[\mu_{m_1}, \mu_{m_2}, \dots, \mu_{m_{i-1}}, \widehat{m}_{i,j}, \mu_{m_{i+1}}, \dots, \mu_{m_{N_m}}\right]\right) : j = 1 \dots N_s\right\}$ denote a set of N_s model evaluations where for each sample j, all inputs $m_{k \neq i}$ are chosen equal to the mean value of their probability distribution, μ_{m_k}, and m_i is randomly sampled from $\pi_i(m_i)$, i.e., $\widehat{m}_{i,j} \sim \pi_i(m_i)$. Thus, $\boldsymbol{D}_i^{x,n}$ is essentially the output of a Monte Carlo Simulation (MCS).

The relative deviation of m_i at time t_n and location $\boldsymbol{x}$ is donated by $RD_i^{x,n}$ and is equal to the standard deviation of $\boldsymbol{D}_i^{x,n}$, normalized by the mean of $\boldsymbol{D}_i^{x,n}$ and it is considered a measure of the sensitivity of $g(\boldsymbol{x}, t_n, \boldsymbol{m})$ to m_i. In other words,

$$RD_i^{x,n} = \frac{\sigma_{D_i^{x,n}}}{\mu_{D_i^{x,n}}} \tag{3.1}$$

where $\sigma_{D_i^{x,n}}$ and $\mu_{D_i^{x,n}}$ are standard deviation and mean of $\boldsymbol{D}_i^{x,n}$, respectively. Consequently, both the degree of uncertainty in parameter m_i and the role of m_i in the mathematical model are reflected in the sensitivity measure $RD_i^{x,n}$. The correlation between parameters is not considered here. The accuracy of the RD method depends upon N_s and a larger N_s produces more accurate results.

Algorithm 1 illustrates a pseudo-code of the RD method used in this chapter. The first step of the RD method is to obtain N_s random samples from each of the N_m probability distribution, $\pi_i(m_i)$. Next, for each time interval of interest (t_n), for each coordinate in the domain of interest ($\boldsymbol{x}_k$) and for each parameter (m_i), the model is evaluated N_s times using the N_s random samples of the parameter ($\widehat{m}_{i,j}$) while rest of the parameters are held constant at their mean values. Once the N_s model evaluations have been recorded, the RD for parameter m_i can be evaluated for each coordinate and time of interest.

Algorithm1: The pseudo-code for the RD method used in this chapter. Modified from: (Sarkarfarshi *et al.*, 2014).

```
Set $N_s$ = the number of desired simulations for each parameter and set $N_m$ = number of parameters
Set $\boldsymbol{t} = [t_1, t_2, \dots, t_{N_t}]$ as the time grid and set $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{N_x}]$ as the space grid
For $i = 1$ to $N_m$
    For $j = 1$ to $N_s$
        Draw $\widehat{m}_{i,j}$ from $\pi_i(m_i)$
    End
End
For $n = 1$ to $N_t$
    For $k = 1$ to $N_x$
        For $i = 1$ to $N_m$
            For $j = 1$ to $N_s$
                Evaluate $g\left(\boldsymbol{x}, t_n, \left[\mu_{m_1}, \mu_{m_2}, \dots, \mu_{m_{i-1}}, \widehat{m}_{i,j}, \mu_{m_{i+1}}, \dots, \mu_{m_{N_m}}\right]\right)$
            End
            Set $\boldsymbol{D}_i^{x,n} = \{g(\boldsymbol{x}, t_n, [\mu_{m_1}, \mu_{m_2}, \dots, \mu_{m_{i-1}}, \widehat{m}_{i,j}, \mu_{m_{i+1}}, \dots, \mu_{N_m}]) : j = 1 \dots N_s\}$
            Evaluate $RD_i^{x_k,n} = \sigma_{D_i^{x,n}} / \mu_{D_i^{x,n}}$ for $m_i$, $t_n$ and $\boldsymbol{x}_k$
        End
    End
End
```

3.3. Case Study

3.3.1. System Model

Model Description: The injection and migration of CO_2 are assumed to take place in an infinite, horizontal and homogeneous saline aquifer (Figure 3-1). The aquifer is assumed to be rigid, bounded above and below by impermeable formations, of constant thickness (B) and with a constant porosity (ϕ). Both the CO_2 (non-wetting phase) and the host brine (wetting phase) are assumed to be incompressible Newtonian fluids of constant density (ρ_n and ρ_w, respectively) and constant viscosity (μ_n and μ_w, respectively), which is reasonable if pressure and temperature do not vary greatly during the injection process. The CO_2 and the brine phases are taken to be immiscible. The total CO_2 injection rate, Q, is constant in time. Injection is assumed to occur from a vertical well that fully penetrates the whole thickness of the target formation; thus, injection is modelled as a line source with an injection rate of Q/B per unit length.

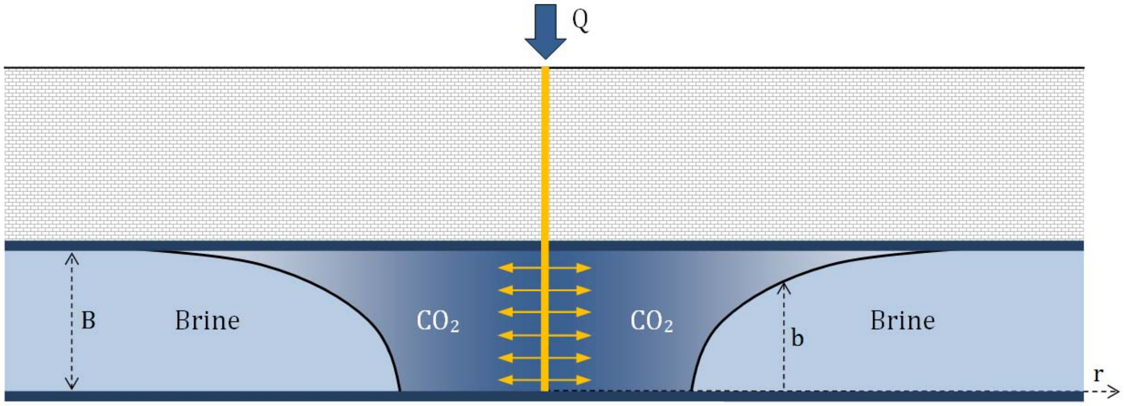


Figure 3-1: Migration of CO_2 into an infinite, horizontal and homogeneous saline aquifer. Color gradient in CO_2 zone shows the saturation transition. From: (Sarkarfarshi *et al.*, 2014).

We wish to model the impact of gravity drainage on the evolution of a CO_2 plume in the early stages of injection. Because the density of the injected CO_2 is less than that of the host brine ($\rho_n < \rho_w$), the injected CO_2 phase will rise towards the top of the aquifer. Thus, the CO_2 will spread at a faster rate at the top of the aquifer compared to the bottom, resulting in a CO_2/brine front similar to that illustrated in Figure 3-1. In the presence of capillary pressure effects, the CO_2/brine front is not a sharp interface where the brine saturation, S_w, abruptly jumps from 1 to S_r, but consists of a transition zone where S_w varies from 1 to S_r where S_r denotes the residual saturation for the brine in the porous medium.

The volumetric flux, $\boldsymbol{q}_i$, of phase i within the aquifer is given by a generalized Darcy's Law

$$\boldsymbol{q}_i = \frac{kk_{ri}}{\mu_i}(\boldsymbol{\nabla} p_i - \rho_i \boldsymbol{g}), \qquad i = wp \ or \ np \tag{3.2}$$

where $i = wp$ corresponds to the wetting phase (brine), $i = np$ corresponds to the non-wetting phase (CO_2), k is the intrinsic permeability, $\boldsymbol{g}$ is the acceleration due to gravity, k_{ri} is the relative permeability of phase i and p_i is the pressure of phase i.

Relative permeability is often a function of saturation. Let S_w denote the saturation of the brine; the effective saturation is defined as

$$S_e = \begin{cases} \dfrac{S_w - S_r}{1 - S_r} & S_w > S_r \\ 0 & S_w \le S_r \end{cases} \tag{3.3}$$

Laboratory measurements in the Nisku aquifer (Bennion & Bachu, 2005; Bennion & Bachu, 2006) demonstrate that k_{ri} is a nonlinear function of saturation. The value of k_{ri} can be reasonably represented by the Corey-Brooks correlation:

$$k_{rw} = \mathrm{S_e}^{\alpha_w} \tag{3.4}$$

and

$$k_{rn} = \beta(1 - S_e)^{\alpha_n} \tag{3.5}$$

where β is the endpoint relative permeability of the non-wetting phase (CO_2) and α_i are empirical parameters. In this study, it is assumed that $\alpha_w = \alpha_n = \alpha$ for simplicity.

The capillary pressure is defined as

$$p_c = p_{np} - p_{wp} \tag{3.6}$$

The capillary pressure is generally non-zero and is known to be a function of saturation. We will assume that capillary pressure is governed by a Brooks-Corey function of the form

$$p_c = AS_w^{-\sigma} \tag{3.7}$$

where A is the entry capillary pressure. Constitutive parameters α, β, A and σ must be suitably chosen for a given target formation. These four parameters plus the viscosity of CO_2 are classified as constitutive relationship parameters.

Flow of the CO_2 and brine is governed by the conservation of mass for each phase, which can be written as

$$\phi\rho_i \frac{\partial S_i}{\partial t} + \rho_i \boldsymbol{\nabla}.\boldsymbol{q}_i = Q_i, \qquad i = w \; or \; n \tag{3.8}$$

where Q_i is the source/sink mass flux of phase i and S_i is the saturation of phase i.

Finite Element Solution: Approximate solutions of (3.8) incorporating three-dimensional domains, arbitrary boundary condition, heterogeneities and multiple sources and sinks can be obtained using numerical simulators such as TOUGH2 (Pruess & Spycher, 2007); however, such a model is computationally too expensive for the sensitivity analysis considered here. Instead, we solve (3.8) for an axisymmetric domain where injection occurs at the center of the domain.

A numerical solution of equation (3.8) is obtained using the Finite Element Method (FEM), with appropriate stabilization (Morton, 2010). A structured mesh of linear finite elements discretizes the domain. The FEM code has been validated against the recently developed semi-analytical solution of Malekzadeh and Dusseault (Malekzadeh & Dusseault, 2013). The MD2013 model assumed that the injection of CO_2 occurs at the center of an axisymmetric domain, which is initially completely saturated with brine. The semi-analytical solution MD2013 is orders of magnitude more computationally efficient than the FEM model and is thus generally more useful for sensitivity and risk analyses. However, the semi-analytical solution exhibits some instability over the range of values of β and α

encountered in the distribution of these parameters for the Nisku aquifer. Therefore, the slower but more stable FEM solution is used in this study.

A schematic of typical saturation profile of the MD2013 and FEM solutions is illustrated in Figure 3-2. It can be seen that there is a saturation transition zone between regions of fully saturated brine ($S_w = 1$) and fully saturated CO$_2$ ($S_w = S_r$). $b(r,t)$ is the boundary of the plume, separating the areas of $S_w = 1$ and $S_w < 1$. Not obvious from Figure 3-2, is that the solution of (3.8) involves a small shock front where the brine saturation jumps abruptly from $S_w = 1$ to S_{front} at the CO$_2$ front. The location and magnitude of the shock front are obtained naturally from the semi-analytical solution MD2013; however, in the FEM solution, the shock is smoothed out. It does not appear that the smoothing of the shock front significantly impacts the accuracy of the FEM model.

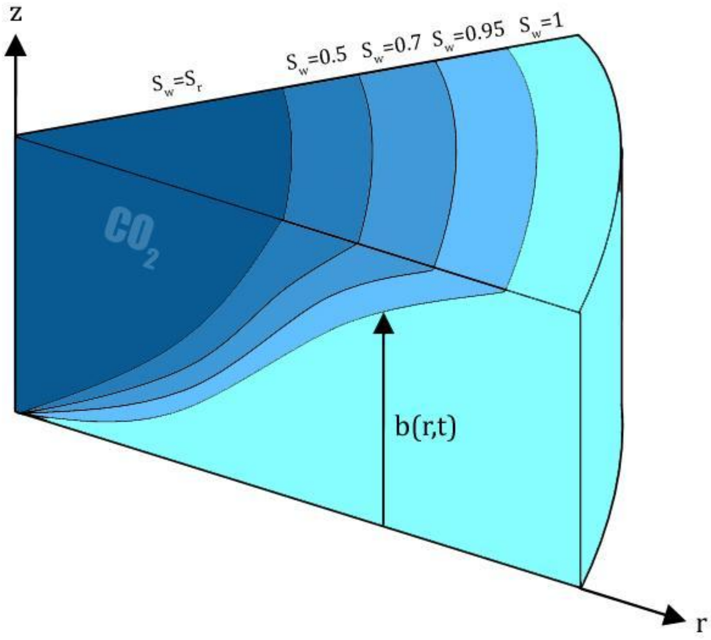


Figure 3-2: Axisymmetric CO$_2$ transition zone and plume boundary. From: (Sarkarfarshi *et al*., 2014).

In this study, the aquifer is modelled as an axisymmetric and homogeneous domain with a radius of 1 km. No flow boundary conditions are assumed on the upper and lower boundaries of the domain. A uniform mass flux is applied along the left most boundary (r=0) and hydrostatic pressure is prescribed along the right most boundary (r=1 km). The domain is chosen sufficiently large in the radial direction so that the saturation front is far from the boundary and there is negligible effect on the motion of the saturation front. A uniform mesh of triangular elements, with 100 equally-spaced nodes in the radial direction and 5 equally-spaced nodes in the vertical direction, is used.

3.3.2. The Nisku Aquifer

The Nisku aquifer is a carbonate unit in the Western Canada Sedimentary Basin; due to oil exploration and production, it has been subjected to detailed analysis and data collection. The study site is located near Edmonton, Alberta and was selected as the CO_2 injection site for the Project Pioneer[2] (MIT, 2012; Project Pioneer, 2012). Nisku aquifer characteristics were gathered in a study led by the University of Calgary called the Wabamun Area Sequestration Project (WASP) (Lavoie & Keith, 2010) based on wireline geophysical logs, core analysis, drill stem tests and petrographic studies collected since 1950 (Eisinger & Jensen, 2009), which provides an appropriate statistical basis for this study. These characteristics are partially available from (Golder Associates, 2010). In addition, some other characteristics (including the relative permeability and capillary pressure curves of two core samples) are available from (Bennion & Bachu, 2005) and (Bennion & Bachu, 2006). In the WASP study area, the saline aquifer is 1730 m deep (Eisinger & Jensen, 2009) and is 64 m thick (Golder Associates, 2010) on average, consisting of both a dolomite with a porosity of 3%~5% and a permeability of 5~15 mD, and a hypersaline carbonate mudstone with a porosity of < 2% and a permeability of < 5 mD (Eisinger & Jensen, 2009).

3.3.3. Parameters

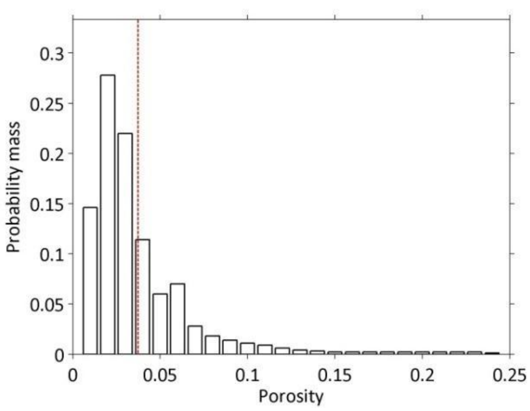
The parameters considered in the sensitivity analysis are tabulated in Table 3-1 and the probability distribution of each parameter is illustrated in Figure 3-3.

---

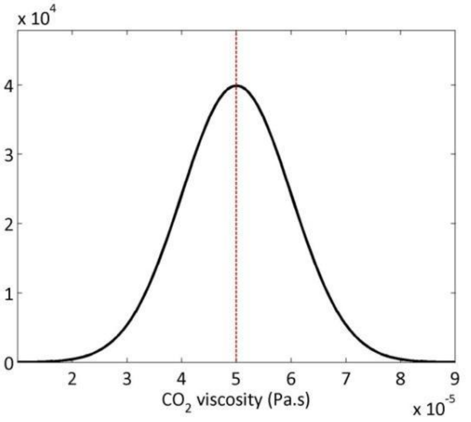
[2] The project was cancelled in 2012 due to economic reasons (MIT, 2012). However the availability of site statistics makes it an appropriate choice for this study.

Table 3-1: Parameters and probability distributions/values used in the sensitivity analysis. D, U, 1 and 2 stand for Deterministic parameter, Uncertain parameter, 1st group (physical) parameter and 2nd group (constitutive relationship) parameter, respectively. From: (Sarkarfarshi *et al*., 2014).

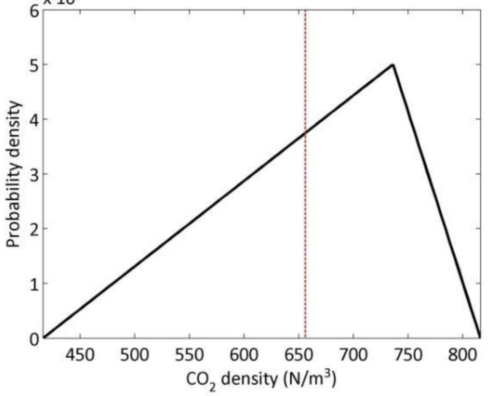
Parameter	**Type**	**Probability distribution/Value**	**Mean**	**Unit**	**Reference**
Q	D1	0.1	-	Mt/year	-
B	U1	Lognormal (4.147, 0.14)	63.88	m	(Golder Associates, 2010)
k	D1	440	-	mD	(Golder Associates, 2010)
ϕ	U1	Discrete PMF	0.038	-	(Golder Associates, 2010)
S_r	U1	Lognormal (-0.924, 0.200)	0.41	-	(Bennion & Bachu, 2005; Alberta Geological Survey, 2012)
μ_{wp}	D2	7.49×10^{-4}	-	Pa.s	(Bennion & Bachu, 2006; CREWES, 2012; Golder Associates, 2010)
μ_{np}	U2	Normal(5×10^{-5}, 1×10^{-5})	5×10^{-5}	Pa.s	(Golder Associates, 2010)
ρ_{wp}	D1	1090	-	kg/m^3	(Bennion & Bachu, 2006; CREWES, 2012; Golder Associates, 2010)
ρ_{np}	U1	Triangular (416, 736, 816)	656	kg/m^3	(Golder Associates, 2010)
α	U2	Lognormal (0.992, 0.010)	2.71	-	(Bennion & Bachu, 2005; Alberta Geological Survey, 2012)
β	U2	Lognormal (-1.938, 0.385)	0.155	-	(Bennion & Bachu, 2005; Alberta Geological Survey, 2012)
A	U2	Lognormal (13.249, 0.632)	0.693	MPa	(Bennion & Bachu, 2006; Alberta Geological Survey, 2012)
σ	U2	Lognormal (0.346, 0.01)	1.42	-	(Bennion & Bachu, 2006; Alberta Geological Survey, 2012)


Probability mass
Porosity
0
0.05
0.1
0.15
0.2
0.25
0.3

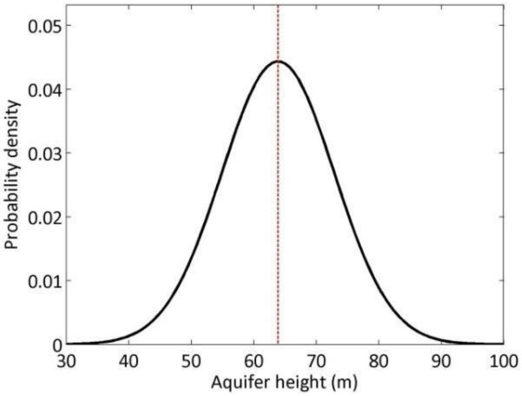
(a)


x 10^4
Probability density
CO₂ viscosity (Pa.s)
x 10^-5
0
1
2
3
4
5
6
7
8
9

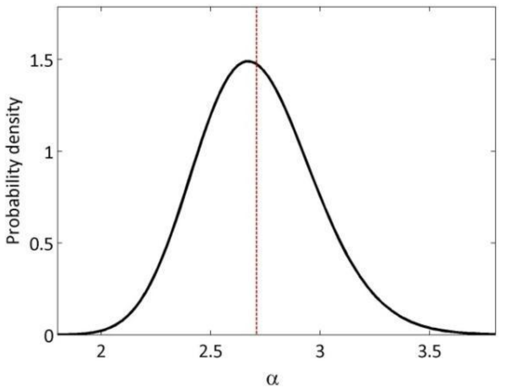
(b)


x 10^-3
Probability density
CO₂ density (N/m³)
0
1
2
3
4
5
6
450
500
550
600
650
700
750
800

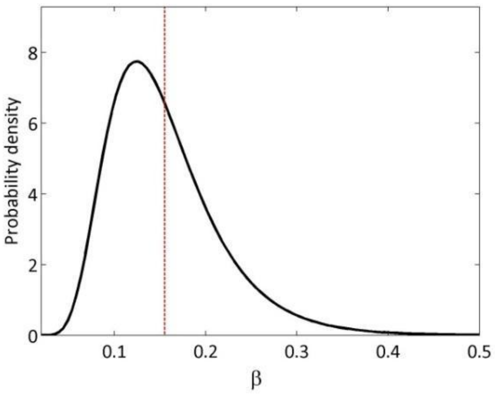
(c)


Probability density
Aquifer height (m)
0
0.01
0.02
0.03
0.04
0.05
30
40
50
60
70
80
90
100

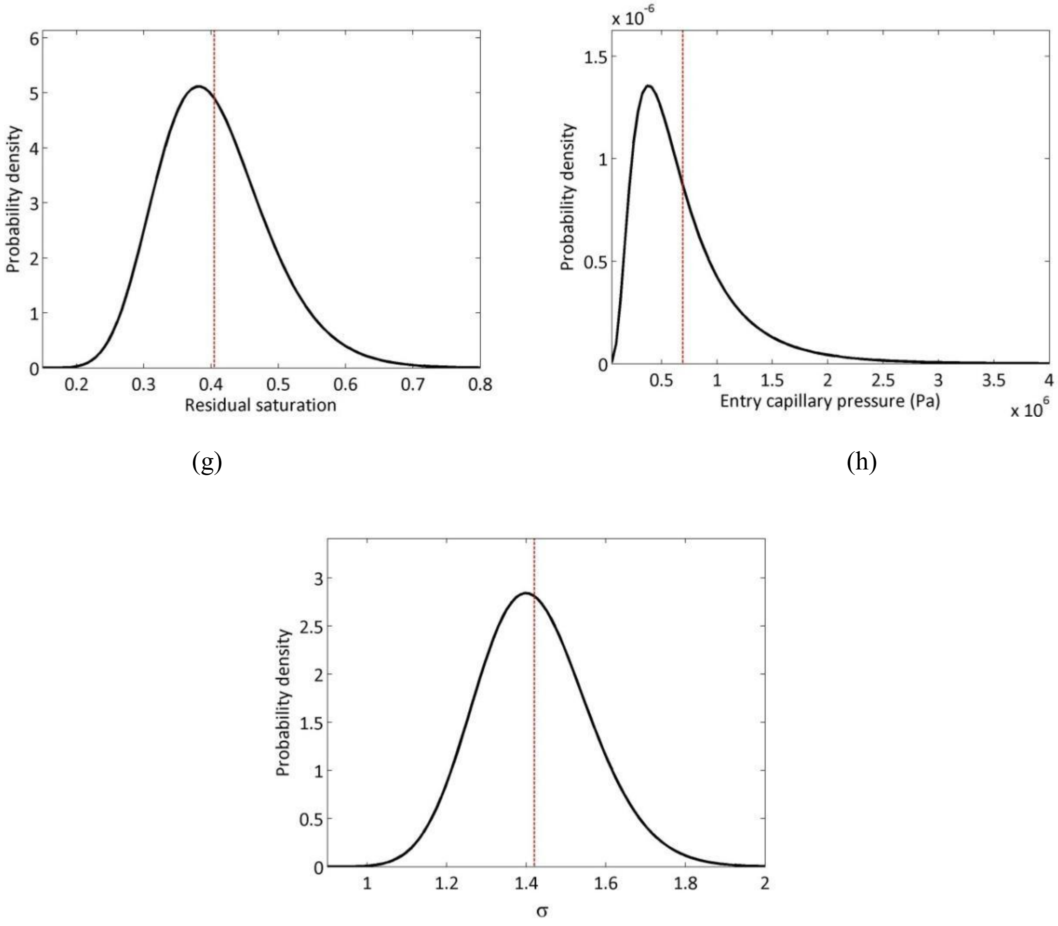
(d)


Probability density
α
0
0.5
1
1.5
2
2.5
3
3.5

(e)


Probability density
β
0
2
4
6
8
0.1
0.2
0.3
0.4
0.5

(f)



(g)

(h)

(i)

Figure 3-3: Probability distribution of Nisku aquifer characteristics used in this study: (a) porosity, (b) CO_2 viscosity, (c) CO_2 density, (d) aquifer thickness, (e) α, (f) β, (g) residual saturation, (h) entry capillary pressure, and (i) σ. Red dashed lines show the mean value of the distributions. From: (Sarkarfarshi *et al.*, 2014).

Mean values of the intrinsic permeability (k), the probability mass function for the porosity (ϕ) and probability density functions for CO_2 viscosity (μ_{np}), CO_2 density (ρ_{np}), and formation thickness (B) of the Nisku aquifer are reported in (Golder Associates, 2010). Changes in the formation thickness results in changes in element height in simulations related to formation thickness. Such variation in element size is small and did not significantly impact the accuracy in the FE computations and so does not significantly impact the RD results. The formation porosity is represented by a PMF (Figure 3-3a) and the CO_2 viscosity, the CO_2 density and the formation thickness are represented by three PDFs (Figure 3-3b to Figure 3-3d).

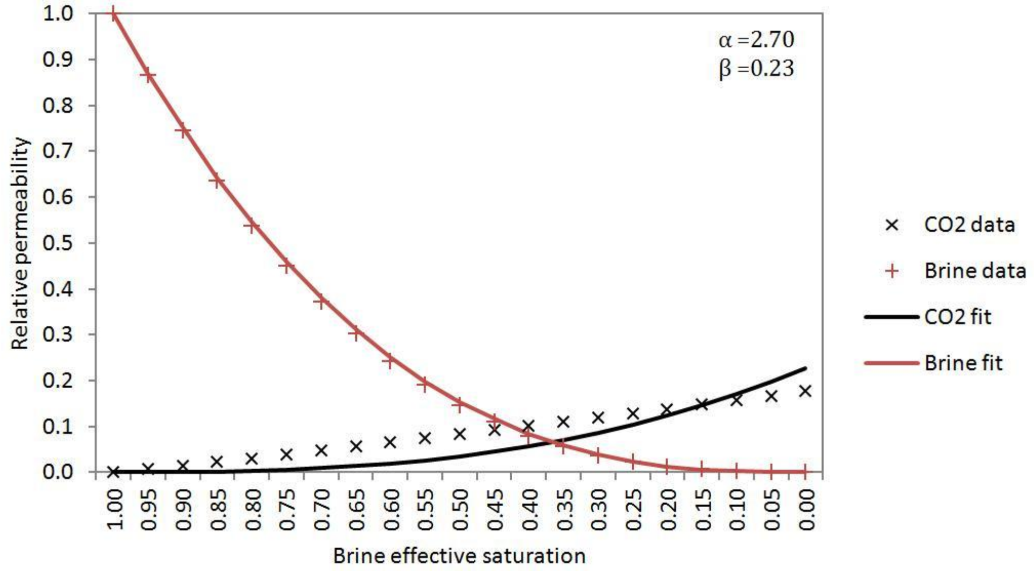
Brine viscosity (μ_{wp}) and density (ρ_{wp}) are assumed to be constant with values calculated from the CREWES fluid properties calculator (CREWES, 2012). In-situ Nisku temperature and brine salinity are assumed to be 56 °C and 136,817 ppm, respectively (Bennion & Bachu, 2006). In-situ Nisku pressure is assumed to be 18.2 MPa (Golder Associates, 2010). The resulting brine viscosity and density are 1090 kg/m^3 and 7.49×10^{-4} Pa.s, respectively.

No PDFs/PMFs are reported in the (Golder Associates, 2010) for constitutive law parameters α, β, A, σ nor residual brine saturation (S_r). However, α, β and S_r can be estimated from relative permeability data collected from two core samples, (Bennion & Bachu, 2008) and (Michael *et al.*, 2009). Entry capillary pressure (A) can also be calculated from capillary pressure-saturation data of the same cores (Bennion & Bachu, 2006). Due to the limited data, we have subjectively assumed lognormal PDFs for each of these parameters. Lognormal PDFs are selected to avoid negative values.

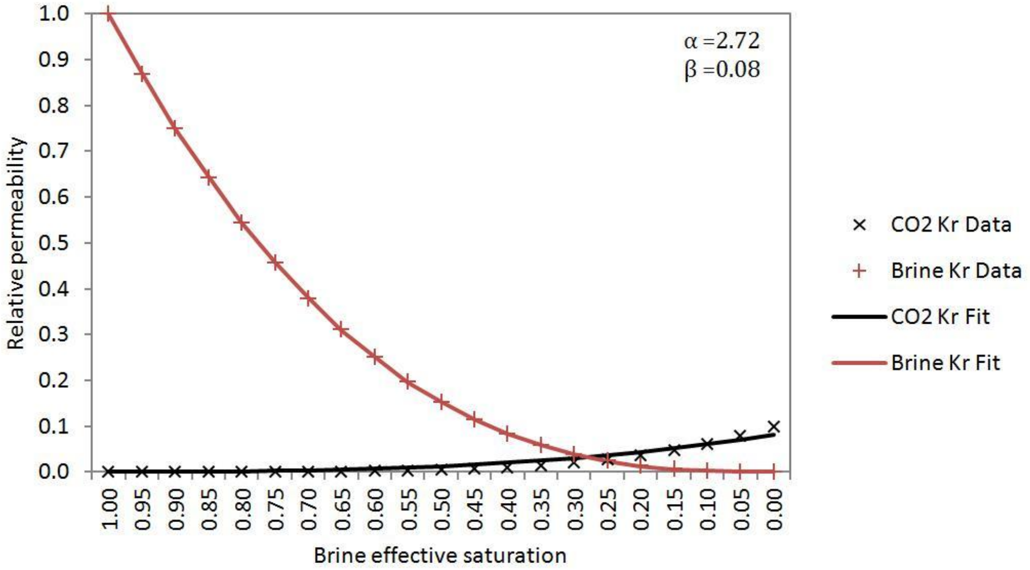
Using a least squares method, the parameters α and β are estimated by fitting equations (3.4) and (3.5) to the relative permeability-saturation data for each core (Figure 3-4). Values of α = 2.70 and 2.72 were obtained for the first and the second core, respectively. These values are taken as indicators of a small uncertainty in the effective value of α parameter for the Nisku aquifer. A lognormal distribution is assigned to α with a mean equal to the mean of the two cores and with a COV of 0.1—indicating little uncertainty (Figure 3-3e).

Similarly, values for β of 0.23 and 0.08 were estimated for the first and the second core, respectively, which indicate much more uncertainty compared to α. As illustrated in Figure 3-3f, a lognormal distribution is assigned to β with a mean equal to the mean of β values of each core and with a COV of 0.4 to capture the large uncertainty in the effective value of β.

Values for residual saturation (S_r) of 0.32 and 0.49 were also estimated, showing less uncertainty compared to β, but more uncertainty compared to α. A lognormal distribution is assigned to the S_r with a mean equal to the mean of obtained values and a COV of 0.2 (Figure 3-3g).



(a)

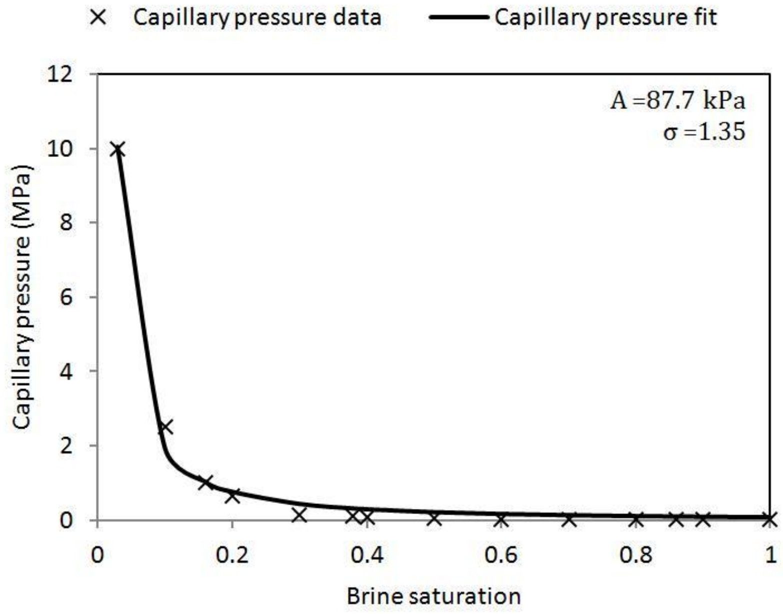


(b)

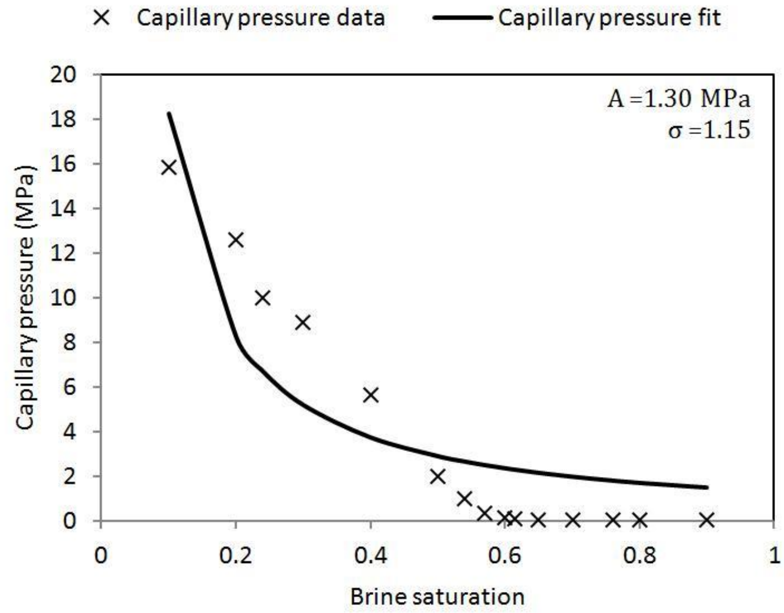
Figure 3-4: Relative permeability data and fitted curves for the (a) first and (b) second core samples of the Nisku aquifer according to (Alberta Geological Survey, 2012) and (Bennion & Bachu, 2005) data. From: (Sarkarfarshi *et al*., 2014).

Using a least squares method, parameters A and σ are estimated for each of the cores by fitting equation (3.7) to the capillary pressure-saturation data (Figure 3-5). Values of 87.7 kPa and 1.3 MPa for A and 1.35 and 1.48 for σ were obtained for each sample, respectively. Two lognormal PDFs are assigned to A and σ (Figure 3-3h and Figure 3-3i);

the means of these PDFs are equal to the mean of the obtained values and the COV of A and σ selected, according to their uncertainty, as 0.7 and 0.1, respectively.



(a)



(b)

Figure 3-5: Capillary pressure data and fitted curves for the (a) first and (b) second core sample of the Nisku aquifer, according to (Bennion & Bachu, 2006) data. From: (Sarkarfarshi *et al*., 2014).

3.3.4. Ranking Criteria

Three criteria are chosen to rank the sensitivity of the system to various parameters:

I. The first criterion is the *average plume radius sensitivity* and is equal to vertically averaged sensitivity of the CO_2 plume radius, evaluated N_t times (once at each t_n) for each uncertain parameter m_i. This criterion indicates the vertically averaged sensitivity of the plume radius.

II. The second criterion is the *plume tip sensitivity* (sensitivity of the maximum radial extension of the plume) and is equal to sensitivity of the radius of the CO_2 plume at the top of the aquifer, evaluated N_t times (once at each t_n) for each uncertain parameter m_i. This criterion is selected because leakage of CO_2 from different pathways (e.g., abandoned wells, faults, fractures, etc.) initiates from the top of the aquifer. Thus, the location of the CO_2 plume tip (the maximum breakthrough distance of CO_2 at the top of the aquifer) and its uncertainty is of paramount importance in determining leakage initiation time and for risk assessment (Kopp *et al.*, 2010).

III. The third criterion is the *moment of inertia sensitivity* and is equal to the sensitivity of the moment of inertia of the plume, evaluated N_t times (once at each t_n) for each uncertain parameter m_i. The moment of inertia of the plume is defined in equation (3.9) and can be easily approximated from the finite element solution as

$$I(t_n) = \int_V \rho_n r^2 dV = \int_{z=0}^{B} \int_{r=0}^{1\,km} 2\pi\phi\rho_n r^3 S_{np}(r,z,t_n) dr\, dz \tag{3.9}$$

where $I(t_n)$ is the moment of inertia of the plume mass at time t_n, V is the volume of CO_2, and $S_{np}(r,z,t)$ indicates the saturation of the CO_2 phase at radius r from the injection well, height z from the aquifer base, and time t. This criterion represents the overall sensitivity of the CO_2 mass spread in the aquifer at different times.

3.3.5. Results

A sensitivity analysis by the RD method is conducted by running the model 50 times for each uncertain parameter, i.e., $N_s = 50$. The supercritical CO_2 injection rate is assumed to be 0.1 Mt/year. The time grid ($\boldsymbol{t}$) is chosen equal to $[1, 2, \ldots, 12]$ months and the vertical node coordinates are used as the space grid ($\boldsymbol{X}$). Figure 3-6 shows the effective saturation distribution for the base case in which all parameters are kept at their mean value.

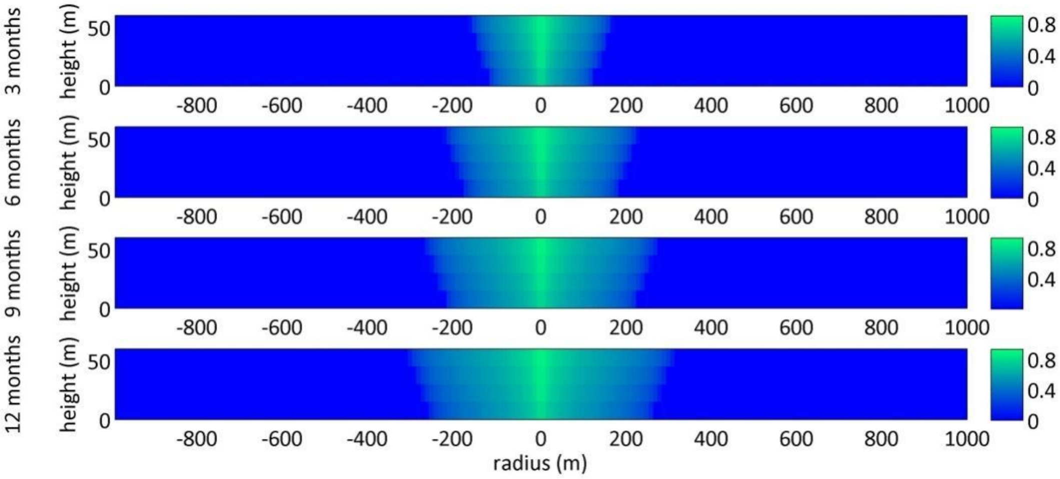


Figure 3-6: Effective saturation distribution of CO_2 for the base case, in which all parameters are kept at their mean value, after 3, 6, 9 and 12 months of injection. From: (Sarkarfarshi *et al.*, 2014).

Figure 3-7, Figure 3-8 and Figure 3-9 illustrate the sensitivity to the uncertain parameters, according to criteria I, II and III, respectively. It can be seen that the ranking of the parameters is almost time-invariant. Small fluctuations in Figure 3-7 and Figure 3-8 are most likely a result of the definition of the plume boundary, which is subject to an error of <10 m (equal to length of the elements in the radial direction).

In order to compare the relative role of the parameters more readily based on the criteria, the sensitivity measures can be normalized by their summation. Table 3-2 lists the un-normalized and the normalized sensitivity measures for each of the three criteria, at the end of year 1.

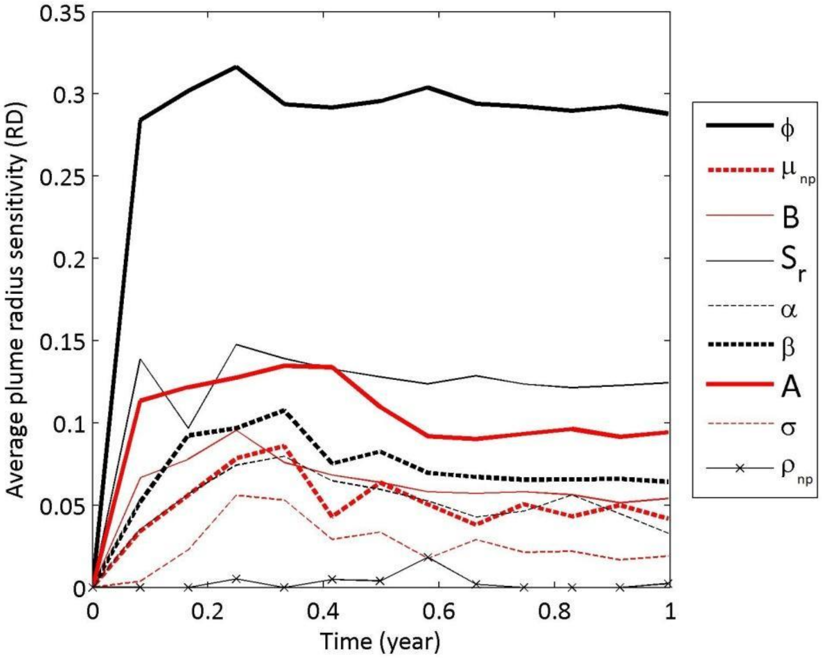


Figure 3-7: Criterion I: average plume radius sensitivity. From: (Sarkarfarshi *et al*., 2014).

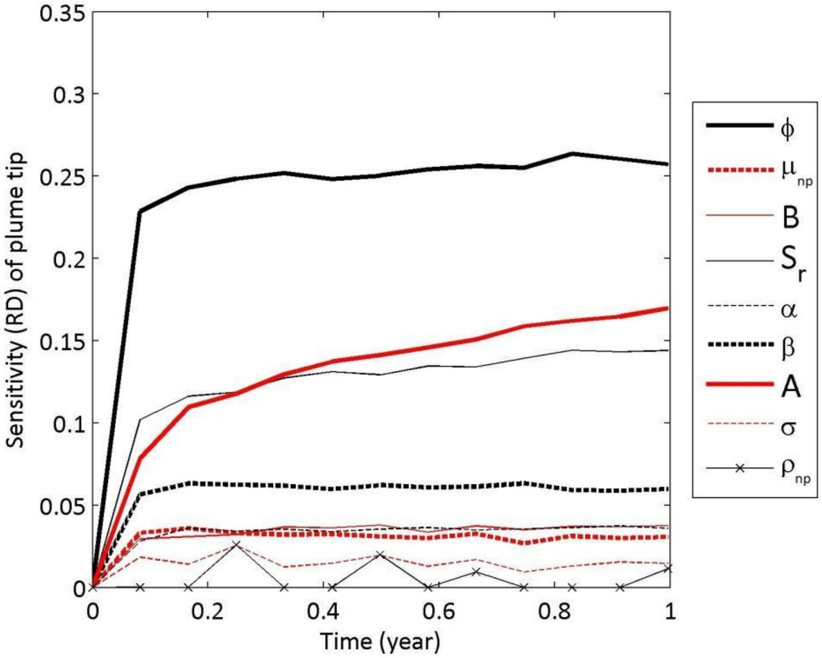


Figure 3-8: Criterion II: plume tip sensitivity. From: (Sarkarfarshi *et al*., 2014).

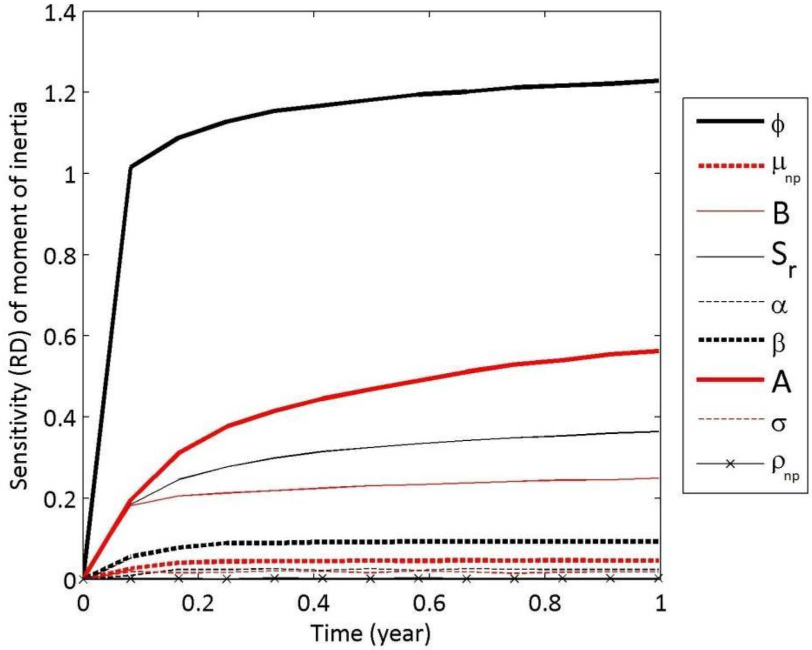


Figure 3-9: Criterion III: moment of inertia sensitivity. From: (Sarkarfarshi *et al.*, 2014).

Table 3-2: Original and normalized values of measure of sensitivity and the rank of uncertain parameters according to criteria I, II and III. Modified from: (Sarkarfarshi *et al.*, 2014).

Criterion		**I**			**II**			**III**		
Parameter		**Rank**	**Measure of sensitivity**		**Rank**	**Measure of sensitivity**		**Rank**	**Measure of sensitivity**	
			Original	**Normalized**		**Original**	**Normalized**		**Original**	**Normalized**
Physical	B	5	0.054	0.075	5	0.037	0.049	4	0.249	0.096
	ϕ	1	0.288	0.400	1	0.257	0.338	1	1.23	0.475
	S_r	2	0.124	0.172	3	0.144	0.189	3	0.364	0.140
	ρ_{np}	9	0.002	0.003	9	0.011	0.015	9	0.002	0.001
Constitutive	α	7	0.033	0.046	6	0.036	0.047	7	0.024	0.009
	β	4	0.064	0.089	4	0.060	0.079	5	0.093	0.036
	σ	8	0.019	0.026	8	0.015	0.019	8	0.018	0.007
	A	3	0.094	0.131	2	0.170	0.223	2	0.563	0.218
	μ_{np}	6	0.419	0.058	7	0.031	0.040	6	0.046	0.018

3.4. Discussion

3.4.1. Sensitivity Analysis Results

According to all three criteria, the formation porosity, ϕ, is the parameter that most influences the uncertainty in the migration of CO_2. The difference between the measures of sensitivity for porosity and the rest of the parameters is significant. This difference is caused by the role of porosity in the model structure, as well as its degree of uncertainty. As a marine carbonate reservoir, the Nisku aquifer is subject to extreme heterogeneity in porosity (Dominguez *et al.*, 1992), and consequently, extreme uncertainty in the effective value of porosity in the homogeneous system model.

Brine residual saturation, S_r, seems to be the second most influential parameter in the physical parameter group, as it is ranked second in criterion I and third in criteria II and III, respectively. Aquifer thickness, B, is ranked third among physical parameters and is in the middle of the overall ranking. Density of CO_2, ρ_{np}, is ranked as the least important physical parameter and the least important parameter in the overall ranking.

Among the five constitutive relationship parameters investigated in this study, entry capillary pressure, A and endpoint relative permeability of the CO_2, β, received the highest ranks, as a result of their considerable degree of uncertainty, in addition to their role in the model formulation. A is ranked as the second most influential parameter in the uncertainty of the plume tip and plume moment of inertia and the third most important parameter in the average location of plume boundaries whereas β is in the middle of the overall ranking for all three criteria. It can also be seen that the sensitivity of the plume tip and plume moment of inertia to A increases with time. Thus, if the same trend continues after year 1, A could become more significant. Constitutive parameters α and σ are far less significant, compared to A and β, according to all three criteria. This shows the *factors* in the constitutive equations of power law form, e.g., equations (3.5) and (3.7), are more influential on the uncertainty of the plume fate than the *exponents* in the same equations, according to the approach and data used here. The ranking of viscosity of CO_2, μ_{np}, was in the middle between constitutive parameters and was not as significant as A and β. The role of μ_{np} is expected to be more significant in the uncertainty of the pressure plume.

One of the goals of this case study was to test if the influence of constitutive relationship parameters in plume spread uncertainty is as important as that of the physical parameters. Apparently, for all three criteria, A is the second or the third most influential parameter and holds more than 10% of the total normalized measures of sensitivity at year 1. In addition, at the end of year 1, β is more influential than half of the physical parameters in criteria I and II. Thus, the role of some constitutive relationship factors such as A and β can be as important as the role of physical parameters in the uncertainty of the plume fate. This implies that spending time and effort on a proper choice of constitutive relationship parameters can play a key role in mitigating the uncertainty in forecasting the CO_2 plume fate. The role of constitutive relationship parameters in the risk assessment of CS projects can be also highlighted through taking into account the role of these parameters in the uncertainty assessment of the plume tip location (criterion II).

The overall ranking presented in Figure 3-7, Figure 3-8, Figure 3-9 and Table 3-2 can help prioritize future data collection and calibration efforts. It provides a basis for reducing the dimensionality of the parameter space in order to reduce the computational cost of further uncertainty analysis and model calibration by identifying major sources of uncertainty.

3.4.2. Sensitivity Analysis Approach

The main purpose of this study is to present a methodology for comparing the influence of uncertain parameters in the uncertainty of the CO_2 plume fate in CS projects. Results of the sensitivity analysis determine which parameters should be focused on more in future uncertainty analyses, model calibration and risk assessments. This process becomes more important when using computationally expensive, full-scale reservoir simulators. A relatively simple hypothetical aquifer model with simple geometry is used to demonstrate the methodology; the parameter probability distributions are taken from data collected from the Nisku aquifer, which is targeted for CS. The results presented are specific to the aquifer data, boundary conditions, geometry and modelling objective we have used in this study and may not be readily extrapolated to other cases. For instance, if the injection pressure is fixed at a a constant value instead of fixing the the injection rate, the parameter rankings could be different. However, the approach taken is quite general and can be extended to other CS projects.

Results obtained from this approach are dependent on the degree of uncertainty in the input parameters. Thus, they are valid for the current state of knowledge about the parameters and they should be updated each time probability distributions are revised. As mentioned previously in Chapter 2, the Bayesian approach to probability is used in this thesis; hence, the probability distributions are indicators of degree of belief (and not frequency) in the effective value of the parameters.

As mentioned in Section 3.3.3, the probability distributions of several parameters were selected subjectively based on limited data. This is likely to be a situation encountered in future CS projects. For example, data for parameters estimated from experiments on core samples (such as relative permeability and capillary pressure functions) may be limited because acquiring core samples requires penetrating the caprock of the formation and creates a potential leakage pathway for the CO_2, so there is a large motivation to limit the collection of core samples. In such cases, the PDFs of parameters will likely be assigned by experts based on limited available data or data from other similar rocks. Recent papers (Oladyshkin *et al*., 2011-b; Oladyshkin *et al*., 2012) present efficient methods for reducing the subjectivity of the PDFs in data sparse situations, with an application to CO_2 geosequestration context. Their methodology is applicable to the present sensitivity analysis approach, specifically in the site selection stage when most PDFs come from expert judgment and global databases.

3.4.3. The System Model

Some assumptions of the mathematical model are discussed here for completeness. Note that the assumptions introduced in the mathematical model do not affect the generality of the sensitivity analysis methodology.

Pressure effects on the density and viscosity of the fluids were neglected. The maximum pressure build-up in the reservoir after 1 year of injection in the base case of this sensitivity analysis was 2.6% of the initial formation pressure. Thus, neglecting the pressure effect on the density and viscosity of the fluids should not significantly affect the results. However, this assumption will not be robust when the permeability of the formation is considerably smaller or the injection rate is considerably larger. Moreover, maximum pressure build-up is a constraint on the injection rate in order to make sure it does not exceed the fracture pressure of the caprock and the injection system tolerance. Thus, in more realistic injection cases where the pressure build-up is significant, the ranking could be different.

The temperature dependence of the density and viscosity of the fluids has also been neglected. It is assumed for simplicity that CO_2 will be injected at the temperature of the host formation. In reality, CO_2 will likely be injected at different temperatures from the in-situ temperature of the host formation. Studies such as (Gor *et al.*, 2013) have highlighted the importance of temperature in the geological CO_2 storage process. Thus, the role of temperature differences between injected supercritical CO_2 and the host formation should be revisited in future studies.

The rock matrix was assumed to be rigid and consequently the stress dependence of permeability was neglected. This assumption is adequate for the small pressure changes encountered in this study; however, if the rock mass undergoes a significant temperature change then the role of mechanical deformations will become important. Mechanical stresses are of paramount significance in the integrity of the caprock and injectivity of the aquifer in the presence of thermal effects, or when injecting supercritical CO_2 into less permeable aquifers or at higher rates.

The aquifer was assumed to be homogeneous and isotropic, in order to limit the number of uncertain parameters. However, heterogeneity and anisotropy affect the degree of uncertainty in the effective value of the parameters. For instance, while porosity was assumed constant in this study, the significant amount of uncertainty in the PMF of the porosity was a result of extreme heterogeneity in the Nisku formation.

From a physical point of view, solubility of gas in saline water results in a pressure drop and changes to the composition of the media. Moreover, vaporization of saline in a dry gas leads to precipitation of salt and that can block the pores and decrease the permeability of the media (Mathias *et al*., 2009; Peysson *et al*., 2011). CO_2 and brine phases were considered immiscible in the model. Due to the limited miscibility of CO_2 and saline and the relatively short simulated time of one year, compositional changes have negligible effect on saturation distribution and so immiscibility of the phases is a valid assumption. This assumption may not be valid when compositional changes of the media occur or when longer simulation time are investigated; in these cases, this assumption should be revisited. Neglecting compositional effects results in slightly higher overall saturation and lower pressure near the injection well.

An axisymmetric domain with radius of 1 km was used in this study. Considering the injection rate, injection time and porosity values used in our simulations, the plume front was always far from the boundaries. At larger injection rates and longer injection times however, a domain larger than 1 km is required to ensure that the plume boundary does not reach the domain boundaries.

3.4.4. Contribution to Model Calibration

As described earlier in Chapter 1, the numerical CS models are normally characterized by extreme nonlinearities and parameter distributions in CS systems are not normally well-behaved PDFs. Thus, accurate calibration of CS models normally demands sampling-based solutions which require repetitive simulation of the system model. In general, a larger number of parameters require a larger number of model simulation in a sampling-based calibration algorithm. Taking into account the computational cost of model simulation, reducing dimensionality of the parameter space based on a quantitative sensitivity analysis, as presented in this chapter, can reduce the number of required model simulations to achieve a certain degree of accuracy and precision.

It should be noted that while a reduced parameter space obtained based on the results of the sensitivity analysis can reduce the computational cost of further calibration efforts, it does not eliminate the issue of computational cost of CS model calibration. Thus, the next chapters (Chapters 4 and 5) will focus on developing efficient and accurate calibration methods and reducing the computational cost of model simulation during the calibration process of a CS system for furthermore reduction in the computational cost of accurate model calibration.

3.5. Concluding Remarks

A quantitative methodology is presented in this chapter to compare the contribution of different parameters to the uncertainty of CO_2 plume evolution in CS projects. The methodology accounts for both the degree of uncertainty in each parameter and role of the parameter in the model structure. Results of such analyses can be used to guide and reduce time and effort spent on model calibration to obtain trustworthy model forecasts and risk assessments in CS systems.

The method presented in this chapter is applied to a hypothetical CS site with parameter distributions inspired by Nisku aquifer, a reservoir targeted for CO_2 injection in Alberta, Canada. Two parameter groups are considered: one, physical characteristics of the system; and, two, constitutive relationship parameters introduced in the constitutive laws and empirical relationships. Three ranking criteria are used to compare the role of each parameter in the uncertainty of CO_2 plume evolution: I. the average plume radius sensitivity, II. the plume tip sensitivity and III. the moment of inertia sensitivity. In addition, the overall roles of physical and constitutive relationship parameter groups are compared and it is found that the second group is as influential as the first group in the CO_2 plume evolution uncertainty.

In the presented case study, porosity and residual saturation were the most influential physical parameters whereas entry capillary pressure and the endpoint relative permeability of CO_2 were the most influential constitutive relationship parameters. Results also showed that the roles of entry capillary pressure and the endpoint relative permeability of CO_2 (from the constitutive relationship parameter group) are as important the roles of the physical parameters.

The advantages of the presented sensitivity analysis can be summarized as:

- Ranking parameters of a CS system according to the sensitivity of the model output to each parameter.
- Accounting for both parameter role in the model and degree of uncertainty about each parameter.
- Reducing the dimensionality of the parameter space and avoiding unnecessary calculations in further calibration and probabilistic forecast efforts.
- Demonstration of the importance of constitutive relationship parameters in the uncertainty of a CS system.

Current work is subject to limitations that can be addressed in future research, namely:

- Parameter correlations are not considered in the presented case study.
- The computationally fast system model in the case study is based on simplifying assumptions that such as rigidity of the aquifer, incompressibility of the fluids and isothermality of the injection process which can be relaxed in future works using a more realistic system model.
- Using a site-specific geometry and including the heterogeneity in the system model can increase the validity of the results in the case study.

Results of a quantitative sensitivity analysis similar to the case study presented in this chapter can be used as a basis for reducing the dimensionality of the parameter space for the purpose of reducing the computational cost of model calibration efforts. This potential reduction is complementary to developing computationally efficient calibration algorithms and computationally efficient system models. With a reduced number of parameters, but an inefficient calibration algorithm and a computationally demanding model, the computational demand and accuracy trade-off of calibrating CS systems can be still an issue. Thus, Chapter 4 and Chapter 5 will focus on developing computationally efficient calibration algorithms and computationally efficient replacements for the system models, respectively.

Chapter 4.

Unscented transformation Importance Sampling (UIS)

This chapter is the mirror of the following article. Minor changes are made, mostly in the Summary and Sections 4.1, 4.2 and 4.6.4, for more consistency with the body and notation of the thesis and eliminating redundant information previously presented in Chapter 2. Section 4.3 does not exist in the article. I am the main contributor to all sections of this chapter.

Sarkarfarshi, M. and Gracie, R. 2015. Unscented Importance Sampling for Parameter Calibration of Carbon Sequestration Systems. *Stochastic Environmental Research and Risk Assessment*, 29(3): 975-93.

Summary

As described in Chapter 1, a trade-off is commonly required between the computational cost and accuracy of model calibration efforts for CS and similar engineered systems. Chapter 3 focused on the first objective of this thesis, i.e., reducing this trade-off by presenting a methodology for identifying the most significant parameters effecting the uncertainty of a CS model output and providing a basis for reducing the dimensionality of the parameter space. However, even in a reduced parameter space, accurate calibration of a CS model using conventional calibration approaches is computationally demanding. Thus, in this chapter, we focus on the second objective of this thesis by developing an accurate and computationally efficient calibration method which is suitable for nonlinear and computationally expensive CS models and does not assume a specific type of probability distribution (e.g., Gaussian) for the parameters.

Utilizing the methods and formulations presented earlier in Chapter 2, a computationally efficient Bayesian Importance Sampling (IS) method is developed and applied to CS, which continuously mitigates parameter uncertainty by incorporating time-lapsed noisy monitoring data. Three case studies are generated to demonstrate the application of this method, to compare it with other conventional calibration algorithms, and to study the effect of tuning parameters. It is shown that the calibration method developed in this chapter outperforms traditional IS in terms of accuracy and computational efficiency, and outperforms an accurate member of Kalman Filter (KF) family in terms of accuracy.

4.1. Introduction

Chapter 1 described that the parameters of a CS system are uncertain at the start of injection. Methods for calibration, i.e., using monitoring data to reduce the uncertainty of the parameters, are therefore desirable. The primary contribution of this chapter is the development of a new Bayesian calibration method: the Unscented transformation Importance Sampling (UIS) method. The UIS is the second objective among the three complementary objectives in this thesis that aid in the accurate and computationally efficient calibration of CS systems. The UIS continuously, accurately, and efficiently calibrates the uncertain and non-Gaussian parameter distributions of complex and nonlinear system models by using continuously streaming monitoring data. Accuracy, computational efficiency, and compatibility with nonlinear system models and non-Gaussian probability distributions are the key features of the UIS. The UIS is described in terms of a general system model and is applied to a linear analytical system model, a nonlinear analytical system model, and a multiphase flow system model of CO_2 injection into a saline aquifer.

As mentioned in Section 1.2 of the thesis, there has been only a limited number of works to date focusing on how monitoring data can be incorporated into the calibration of CS models (Bhowmik *et al*., 2011; Johnson & White, 2012; Espinet & Shoemaker, 2013; Tavakoli *et al*., 2013). Due to the limitations of the abovementioned works, (e.g., being deterministic, being computationally demanding, or using assumptions such as linearity of the models or Gaussian structure of the probability distributions), more research is required to study, adapt and develop computationally efficient methods that are accurate and able to calibrate nonlinear and numerical CS models.

UIS is a Bayesian calibration algorithm, which incorporates both expert opinion and noisy time-lapse monitoring data. In UIS, the proposal distribution of the Bayesian Importance Sampling (IS) method is continuously updated utilizing the latest monitoring data to improve sample quality and to mitigate the demand for larger sample size. The proposal distribution is constructed by mixing a computationally efficient version of Kalman Filters (KF) and a heavy-tailed *defensive distribution* (Hesterberg, 1995). The defensive distribution has two roles. The first role is to ensure the robustness of the importance sampling step and the second role is to allow the algorithm to seek parameter values that are not assigned a high probability initially but might gain a large likelihood and improve the sample quality. The mixture ratio between KF output and the defensive distribution is a tuning parameter of UIS. While a version of the KF is used for constructing the proposal density, actual probability distribution of the parameters is not required to be Gaussian. As a result, UIS is more versatile for parameter estimation than most KF-based methods.

In Section 4.2, Bayesian calibration and UIS are described mathematically. In Sections 4.3 and 4.4, the application of UIS is presented in two analytical case studies. In Section 4.5, UIS is applied on a more sophisticated CS case study and the sensitivity of UIS to tuning parameters is investigated. Results of the case study in Section 4.5 are discussed in Section 4.6. Finally, conclusions are presented in Section 4.7.

4.2. Methodology

4.2.1. Problem Formulation

As described previously in Chapter 2, let $\boldsymbol{d}_{tru}^{n}$ denote the true behaviour of CS system (e.g., the CO_2 saturation distribution in a saline aquifer) at time t_n and at locations $\boldsymbol{x}_1$ to $\boldsymbol{x}_{N_x}$ so that $\boldsymbol{d}_{tru}^{n} = \left[\boldsymbol{d}_{tru}(\boldsymbol{x}_1, t_n), \boldsymbol{d}_{tru}(\boldsymbol{x}_2, t_n), \dots, \boldsymbol{d}_{tru}\left(\boldsymbol{x}_{N_x}, t_n\right)\right]$, where N_x is the number of measurement locations. Let $\boldsymbol{d}_{obs}^{n} = \left[\boldsymbol{d}_{obs}(\boldsymbol{x}_1, t_n), \boldsymbol{d}_{obs}(\boldsymbol{x}_2, t_n), \ldots, \boldsymbol{d}_{obs}(\boldsymbol{x}_{N_x}, t_n)\right]$ denote measurements of $\boldsymbol{d}_{tru}^{n}$ at time t_n and at locations $\boldsymbol{x}_1$ to $\boldsymbol{x}_{N_x}$ with a random measurement noise, $\boldsymbol{\varepsilon}_{obs}$ and let $\pi_{obs}(\boldsymbol{\varepsilon}_{obs})$ denote the probability distribution of $\boldsymbol{\varepsilon}_{obs}$. It is assumed that $\boldsymbol{\varepsilon}_{obs}$ is not correlated with $\boldsymbol{d}_{tru}^{n}$ nor time (Rougier, 2009). In addition, let $\boldsymbol{y}^{n}(\boldsymbol{m})$ denote the output of a deterministic mathematical model of the system, $\boldsymbol{g}(\boldsymbol{x}, t, \boldsymbol{m})$, in a vector form, at time t_n and at locations $\boldsymbol{x}_1$ to $\boldsymbol{x}_{N_x}$, i.e., $\boldsymbol{y}^{n}(\boldsymbol{m}) = \left[\boldsymbol{g}(\boldsymbol{x}_1, t_n, \boldsymbol{m}), \boldsymbol{g}(\boldsymbol{x}_2, t_n, \boldsymbol{m}), \ldots, \boldsymbol{g}(\boldsymbol{x}_{N_x}, t_n, \boldsymbol{m})\right]$, where $\boldsymbol{m} = [m_1, m_2, \ldots, m_{N_m}]$ is the model parameter vector and $\pi(\boldsymbol{m})$ denotes a N_m-dimensional probability distribution that reflects the *epistemic* uncertainty in $\boldsymbol{m}$.

As formerly described in Section 2.4.2, it is assumed that the discrepancy between $\boldsymbol{d}_{tru}^{n}$ and $\boldsymbol{y}^{n}(\boldsymbol{m})$, which is denoted by $\boldsymbol{\varepsilon}_{mod}$, is also a vector of random variables that are independent from $\boldsymbol{m}$ and time given optimal model parameters (Kaipio & Somersalo, 2007). Let $\pi_{mod}(\boldsymbol{\varepsilon}_{mod})$ denote the probability distribution of $\boldsymbol{\varepsilon}_{mod}$. Thus,

$$\boldsymbol{d}_{obs}^{n} = \boldsymbol{y}^{n}(\boldsymbol{m}) + \boldsymbol{\varepsilon}_{mod} + \boldsymbol{\varepsilon}_{obs} \qquad \boldsymbol{\varepsilon}_{mod} \sim \pi_{mod}(\boldsymbol{\varepsilon}_{mod}) \qquad \boldsymbol{\varepsilon}_{obs} \sim \pi_{obs}(\boldsymbol{\varepsilon}_{obs}) \tag{4.1}$$

or

$$\boldsymbol{d}_{obs}^{t} = \boldsymbol{y}^{n}(\boldsymbol{m}) + \boldsymbol{\varepsilon}_{tot} \qquad \boldsymbol{\varepsilon}_{tot} \sim \pi_{tot}(\boldsymbol{\varepsilon}_{tot}) \tag{4.2}$$

where $\boldsymbol{\varepsilon}_{tot} = \boldsymbol{\varepsilon}_{mod} + \boldsymbol{\varepsilon}_{obs}$ is the combined model and observation error and reflects the uncertainty of model and measurement when the optimal $\boldsymbol{m}$ is used (Liu *et al.*, 2004). Recalling (2.34) and (2.35), the posterior distribution of $\boldsymbol{m}$ given the monitoring data at time t_n can be obtained by

$$\pi(\boldsymbol{m}|\boldsymbol{d}_{obs}^{n}) = \frac{\pi(\boldsymbol{m})\pi_{tot}\left(\boldsymbol{d}_{obs}^{n} - \boldsymbol{y}^{n}(\boldsymbol{m})\right)}{\int_{\Omega_m} \pi(\boldsymbol{m})\pi_{tot}\left(\boldsymbol{d}_{obs}^{n} - \boldsymbol{y}^{n}(\boldsymbol{m})\right)} \tag{4.3}$$

and the posterior distribution of $\boldsymbol{m}$ given all monitoring data up to time t_n can be obtained by[3]

$$\pi\left(\boldsymbol{m}|\boldsymbol{d}_{obs}^{1\dots n}\right) = \frac{\pi^0(\boldsymbol{m})\prod_{l=1}^{n}\pi_{tot}\left(\boldsymbol{d}_{obs}^{l} - \boldsymbol{y}^{l}(\boldsymbol{m})\right)}{\int_{\Omega_m}\pi^0(\boldsymbol{m})\prod_{l=1}^{n}\pi_{tot}\left(\boldsymbol{d}_{obs}^{l} - \boldsymbol{y}^{l}(\boldsymbol{m})\right)} \tag{4.4}$$

The objective in this section is to develop a hybrid method to approximate (4.4) by combining a computationally efficient method and a flexible and accurate method, i.e., KF and IS, respectively.

4.2.2. Unscented Kalman Filter (Measurement Update Stage)

Recall from Section 2.5.2, Kalman Filter (KF) is applicable only to linear and Gaussian systems. While KFs are usually used for estimation of states of dynamic systems, the measurement update stage of a KF can also be used for parameter estimation, as described in (Wan & Van Der Merwe, 2001). To formulate the measurement stage of the KF for solving (4.3), assume:

- $\pi(\boldsymbol{m})$ is Gaussian, i.e., $\pi(\boldsymbol{m}) = N\left(\boldsymbol{\mu}_m^b, \mathbf{C}_m^{\mathrm{b}}\right)$.
- $\pi_{tot}(\boldsymbol{\varepsilon}_{tot})$ is Gaussian and has a mean of zero, i.e., $\pi_{tot}(\boldsymbol{\varepsilon}_{tot}) = N\left(\mathbf{0}, \mathbf{C}_{\varepsilon_{tot}}\right)$
- The system model is linear, i.e., $\boldsymbol{y}^n(\boldsymbol{m}) = \left[\boldsymbol{G}(\boldsymbol{x}_1, t_n)\boldsymbol{m}, \boldsymbol{G}(\boldsymbol{x}_2, t_n)\boldsymbol{m}, \dots, \boldsymbol{G}\left(\boldsymbol{x}_{N_x}, t_n\right)\boldsymbol{m}\right]$

Then, the posterior distribution in (4.4) is also Gaussian and the mean ($\boldsymbol{\mu}_m^a$) and the covariance matrix ($\mathbf{C}_m^a$) of $\pi\left(\boldsymbol{m}|\boldsymbol{d}_{obs}^n\right)$ are

$$\begin{aligned} \boldsymbol{\mu}_m^a &= \boldsymbol{\mu}_m^b + \boldsymbol{K}\left(\boldsymbol{d}_{obs}^n - \boldsymbol{y}^{n,b}(\boldsymbol{m})\right) \\ \mathbf{C}_m^a &= \mathbf{C}_m^b - \boldsymbol{K}\left(\mathbf{C}_{y^n}^b + \mathbf{C}_{\varepsilon_{tot}}\right)\boldsymbol{K}^T \end{aligned} \tag{4.5}$$

---

[3] Equations (4.3) and (4.4) are obtained based on the assumption that errors are unstructured. This assumption will be used in the rest of this chapter, since the artificial errors used in the case studies are unstructured. Equations (4.3) and (4.4) can be modified as described in section 2.4.3 if a structured error exists.

where $\boldsymbol{y}^{n,b}(\boldsymbol{m})$ is the prior mean of $\boldsymbol{y}^{n}(\boldsymbol{m})$, calculated as

$$\boldsymbol{y}^{n,b}(\boldsymbol{m}) = \left[\boldsymbol{G}(\boldsymbol{x}_1, t_n)\boldsymbol{\mu}_m^b, \boldsymbol{G}(\boldsymbol{x}_2, t_n)\boldsymbol{\mu}_m^b, \dots, \boldsymbol{G}(\boldsymbol{x}_{N_x}, t_n)\boldsymbol{\mu}_m^b\right] \tag{4.6}$$

$\mathbf{C}_{y^n}^b$ is the prior covariance matrix of $\boldsymbol{y}^{n}(\boldsymbol{m})$, calculated as

$$\mathbf{C}_{y^n}^b = \left[\boldsymbol{G}(\boldsymbol{x}_1, t_n), \boldsymbol{G}(\boldsymbol{x}_2, t_n), \dots, \boldsymbol{G}(\boldsymbol{x}_{N_x}, t_n)\right]\mathbf{C}_m^b\left[\boldsymbol{G}(\boldsymbol{x}_1, t_n), \boldsymbol{G}(\boldsymbol{x}_2, t_n), \dots, \boldsymbol{G}(\boldsymbol{x}_{N_x}, t_n)\right]^T \tag{4.7}$$

and $\boldsymbol{K}$ is the optimal Kalman gain, calculated as

$$\boldsymbol{K} = \mathbf{C}_{m,y^n}\left(\mathbf{C}_{y^n}^b + \mathbf{C}_{\varepsilon_{tot}}\right)^{-1} \tag{4.8}$$

where $\mathbf{C}_{m,y^n}$ is the prior covariance matrix between $\boldsymbol{m}$ and $\boldsymbol{y}^{n}(\boldsymbol{m})$ and is calculated as

$$\mathbf{C}_{m,y^n} = \mathbf{C}_m^b\left[\boldsymbol{G}(\boldsymbol{x}_1, t_n), \boldsymbol{G}(\boldsymbol{x}_2, t_n), \dots, \boldsymbol{G}(\boldsymbol{x}_{N_x}, t_n)\right]^T \tag{4.9}$$

A nonlinear CS system can be linearized using a first order Taylor series expansion, as done in the Extended Kalman Filtering (EKF) formulation. An alternative solution is linearizing the system model using Unscented Transformation (UT), as described in Section 2.3.3. Unlike the EKF, which is first order accurate, UKF is at least second order accurate for non-Gaussian probability distributions and third order accurate for Gaussian probability distributions (Chen, 2003). Another clear advantage of the UKF over the EKF is that it can be used with any nonlinear function (system model) and does not require the calculation of the Jacobian or Hessian of the function (Chen, 2003), which is non-trivial when the system model is numerical.

Recalling the Unscented Transformation (UT) formulation in Section 2.3.3, let $2N_m + 1$ weighted Sigma points be deterministically chosen such that their mean is equal to $\boldsymbol{\mu}_m^b$ and their covariance matrix is equal to $\mathbf{C}_m^b$ as

$$\begin{gathered}
\boldsymbol{\chi}_0 = \boldsymbol{\mu}_m^b \\
-1 < \vartheta_0 < 1 \\
\boldsymbol{\chi}_i = \boldsymbol{\mu}_m^b + \left(\sqrt{\frac{N_m}{1-\vartheta_0}\mathbf{C}_m^b}\right)_i \qquad i = 1,2,\dots,N_m \\
\boldsymbol{\chi}_i = \boldsymbol{\mu}_m^b - \left(\sqrt{\frac{N_m}{1-\vartheta_0}\mathbf{C}_m^b}\right)_i \qquad i = N_m+1, N_m+2,\dots,2N_m \\
\vartheta_i = \frac{1-\vartheta_0}{2N_m} \qquad i = 1,2,\dots,2N_m
\end{gathered} \tag{4.10}$$

where $\left(\sqrt{\frac{N_m}{1-\vartheta_0}\mathbf{C}_m^b}\right)_i$ is the i^{th} row of the square root of $\frac{N_m}{1-\vartheta_0}\mathbf{C}_m^b$, obtained by Cholesky decomposition, ϑ_i are weights of the Sigma points and ϑ_0 determines how scattered the Sigma points are about $\boldsymbol{\chi}_0$. The Sigma points are then propagated through the system model,

$$\boldsymbol{\Upsilon}_i = \left[\boldsymbol{g}(\boldsymbol{x}_1, t_n, \boldsymbol{\chi}_i), \boldsymbol{g}(\boldsymbol{x}_2, t_n, \boldsymbol{\chi}_i), \dots, \boldsymbol{g}\left(\boldsymbol{x}_{N_x}, t_n, \boldsymbol{\chi}_i\right)\right] \qquad i = 0,1,\dots,2N_m \tag{4.11}$$

The mean, covariance matrix, and cross covariance matrix of the Sigma points, $\boldsymbol{\chi}_i$, and the propagated Sigma points, $\boldsymbol{\Upsilon}_i$, are calculated as

$$\begin{gathered}
\boldsymbol{\mu}_\chi = \sum_{i=0}^{2N_m} \vartheta_i \boldsymbol{\chi}_i \\
\boldsymbol{\mu}_\Upsilon = \sum_{i=0}^{2N_m} \vartheta_i \boldsymbol{\Upsilon}_i \\
\mathbf{C}_\Upsilon = \sum_{i=0}^{2N_m} \vartheta_i [\boldsymbol{\Upsilon}_i - \boldsymbol{\mu}_\Upsilon]^T [\boldsymbol{\Upsilon}_i - \boldsymbol{\mu}_\Upsilon] \\
\mathbf{C}_{\chi,\Upsilon} = \sum_{i=0}^{2N_m} \vartheta_i \left[\boldsymbol{\chi}_i - \boldsymbol{\mu}_\chi\right]^T [\boldsymbol{\Upsilon}_i - \boldsymbol{\mu}_\Upsilon]
\end{gathered} \tag{4.12}$$

where $\boldsymbol{\mu}_\chi$ is the mean of the weighted Sigma points, $\boldsymbol{\mu}_\Upsilon$ is the mean of the weighted propagated Sigma points, $\mathbf{C}_\Upsilon$ is the covariance matrix of the weighted propagated Sigma points, and $\mathbf{C}_{\chi,\Upsilon}$ is the cross covariance matrix of the weighted Sigma points and the weighted propagated Sigma points.

Substituting (4.12) into (4.5) and (4.8), the posterior mean ($\boldsymbol{\mu}_m^a$) and covariance matrix ($\mathbf{C}_m^a$) of the posterior distribution of $\boldsymbol{m}$ in (4.3) are approximated as

$$\begin{aligned} \boldsymbol{\mu}_m^a &= \boldsymbol{\mu}_m^b + \boldsymbol{K}(\boldsymbol{d}_{obs}^n - \boldsymbol{\mu}_\Upsilon) \\ \mathbf{C}_m^a &= \mathbf{C}_m^b - \boldsymbol{K}(\mathbf{C}_\Upsilon + \mathbf{C}_{\varepsilon_{tot}})\boldsymbol{K}^T \end{aligned} \tag{4.13}$$

where the Kalman gain is obtained by

$$\boldsymbol{K} = \mathbf{C}_{\chi,\Upsilon}(\mathbf{C}_\Upsilon + \mathbf{C}_{\varepsilon_{tot}})^{-1} \tag{4.14}$$

Equations (4.13) and (4.14) are the measurement update stage of an UKF. For simplicity of notation, we denote the posterior Gaussian distribution $N(\boldsymbol{\mu}_m^a, \mathbf{C}_m^a)$ obtained from (4.13) and (4.14) by $UT(\pi(\boldsymbol{m}), \boldsymbol{d}_{obs}^n)$.

As described in Section 2.5.2, an alternative KF for nonlinear systems is the Monte-Carlo approach to KF, known as Ensemble Kalman Filter (EnKF) (Evensen, 1994). EnKF has found application when the number of variables is very large, e.g., in weather forecast (Houtekamer *et al*., 2005), reservoir engineering (Naevdal *et al*., 2005; Aanonsen *et al*., 2009) and hydrology (Reichle *et al*., 2002; Shu *et al*., 2005). In a recent study, (Tavakoli *et al*., 2013) demonstrated how EnKF and similar ensemble-based algorithms such as Ensemble Smoothers (ES) (Van Leeuwen & Evensen, 1996) can be applied to CS. Similar to UKF, EnKF does not require the calculation of the Jacobian of the models. For very high dimensional state vectors, EnKF is computationally more efficient than UKF (Kim, 2011). However, EnKF can be less accurate than UKF (Ambadan & Tang, 2009; Mesbah *et al*., 2011).

4.2.3. Importance Sampling

The KF-based approximations of the mean and covariance matrix of the posterior distribution in (4.3) and (4.4) can be inadequate for highly nonlinear models and non-

Gaussian probability distributions (Leisenring & Moradkhani, 2011). A common solution to this issue is drawing random samples from the posterior distribution of the parameters and describing its statistics with the statistics of the samples (Rougier, 2009). Since the posterior distribution in (4.4) is usually unavailable, indirect sampling approaches such as IS or Markov Chain Monte Carlo (MCMC) can be used. As described in Section 2.5.5, in IS we draw weighted samples from a proposal distribution which is close to the posterior and is easy to sample from. The proposal distribution encourages samples from areas of more importance and the weighting ensures that the importance sampling estimator is unbiased.

Recalling from Section 2.5.5, the general formulation for IS for sampling from $\pi^n(\boldsymbol{m})$ can be written as

$$\pi^n(\boldsymbol{m}) \cong \sum_{i=1}^{N_s} W_i \delta(\boldsymbol{m} - \widehat{\boldsymbol{m}}_i) \qquad \widehat{\boldsymbol{m}}_i \sim q(\boldsymbol{m}) \tag{4.15}$$

where $q(\boldsymbol{m})$ is the proposal distribution, $\widehat{\boldsymbol{m}}_i$ are the samples, and W_i are normalized weights which are calculated as

$$W_i = \frac{\dfrac{\pi^0(\widehat{\boldsymbol{m}}_i) \prod_{l=1}^{n} \pi_{tot}\left(\boldsymbol{d}_{obs}^l - \boldsymbol{y}^l(\widehat{\boldsymbol{m}}_i)\right)}{q(\widehat{\boldsymbol{m}}_i)}}{\sum_{i=1}^{N_s} \dfrac{\pi^0(\widehat{\boldsymbol{m}}_i) \prod_{l=1}^{n} \pi_{tot}\left(\boldsymbol{d}_{obs}^l - \boldsymbol{y}^l(\widehat{\boldsymbol{m}}_i)\right)}{q(\widehat{\boldsymbol{m}}_i)}} \qquad i = 1 \dots N_s \tag{4.16}$$

As described earlier in Section 2.5.5, the posterior distribution is the optimal proposal distribution in IS (Chen, 2003). However, the posterior distribution is normally unavailable in IS. Thus, we aim to approximate the posterior distribution using a computationally efficient method and use it as the proposal distribution of IS.

4.2.4. Unscented transformation Importance Sampling

The proper choice of the proposal distribution is the key to an efficient importance sampler. If the proposal distribution is not close to the true posterior, most samples will be drawn from unimportant areas, which results in decreased accuracy and computational efficiency. The significance of a proper choice of the proposal distribution is illustrated in Figure 4-1 and Figure 4-2. Figure 4-1 illustrates sampling from a proposal distribution which is close to the actual sampling distribution and Figure 4-2 shows a poor choice for a proposal distribution. It can be seen that samples in Figure 4-1 are better distributed and the weights are closer in magnitude compared to those illustrated in Figure 4-2. In Figure 4-2C, however, all samples are located on the left side of the mean of the actual sampling distribution, making the approximation of the mean using samples biased. Consequently, more samples are required in the case illustrated in Figure 4-2 compared to that shown in Figure 4-1 in order to ensure that samples are well distributed throughout the important areas of the actual sampling distribution. The optimal choice of the proposal distribution for sampling from $\pi^n(\boldsymbol{m})$ is $q(\boldsymbol{m}) = \pi^n(\boldsymbol{m})$ (Smith *et al.*, 1997) which is not practical because if $\pi^n(\boldsymbol{m})$ is computable then there is no need for IS (Owen & Zhou, 2000; Van Der Merwe *et al.*, 2000). Our effort is therefore focused on finding a proposal distribution close to $\pi^n(\boldsymbol{m})$, to enhance the accuracy and computational efficiency of the sampler.

An intuitive and common choice of the proposal distribution is the prior distribution itself (Van Der Merwe *et al.*, 2000), which reduces (4.16) to the likelihood function. This "sample from the prior, weight by the likelihood" (Rougier, 2009) approach, however, does not contain the latest information from the measurements.

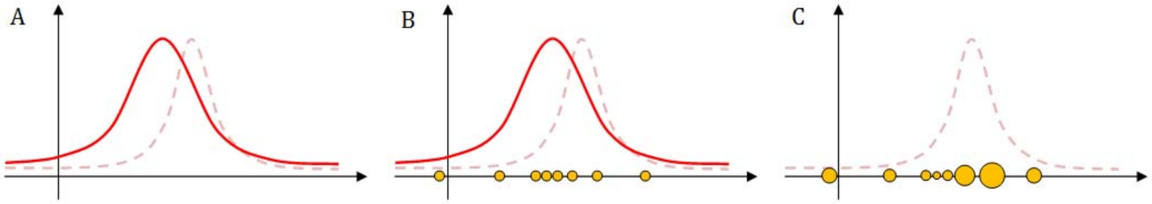


Figure 4-1: Importance Sampling using a proposal distribution (solid curve) close to the actual sampling distribution (dashed curve). Samples in (B) are drawn from the proposal distribution and are weighted accordingly in (C). Size of the samples indicates sample weights. From: (Sarkarfarshi & Gracie, 2015).

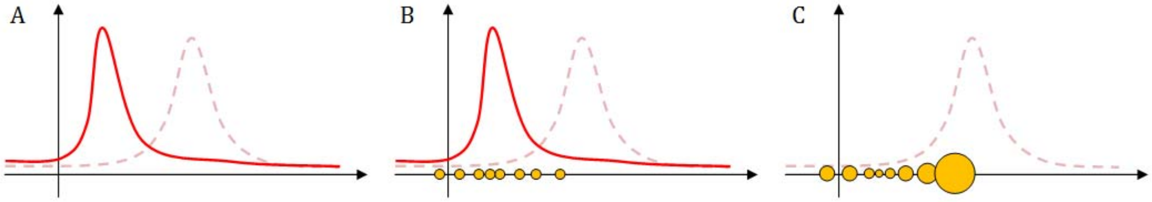


Figure 4-2: Importance Sampling using a poor proposal distribution (solid curve) which is not close to the actual sampling distribution (dashed curve). Samples in (B) are drawn from the proposal distribution and are weighted accordingly in (C). Size of the samples indicates sample weights. Since the sample is clearly biased, more samples are required which negatively affects the computational efficiency of the sampler. From: (Sarkarfarshi & Gracie, 2015).

Revising the proposal distribution periodically using the information from measurements can improve the quality of the samples. For instance, $\pi^{n-1}(\boldsymbol{m})$ can be used as the proposal distribution at time t_n, since it is likely to be closer to $\pi^n(\boldsymbol{m})$ than the initial proposal. However, $\pi^{n-1}(\boldsymbol{m})$ does not contain the latest measurement information, i.e., $\boldsymbol{d}_{obs}^n$. Another solution is iteratively improving the proposal distribution as in Adaptive Importance Sampling (AIS) and Population Monte-Carlo (PMC) methods (Lepage, 1980; Cappé *et al.*, 2004; Pennanen & Koivu, 2006; Douc *et al.*, 2007; Cappé *et al.*, 2008), as described earlier in Section 2.5.5. However, the iterative nature of these algorithms requires repeating the IS several times that makes them computationally more demanding than the original IS and consequently, they are less favorable for CS applications.

Alternatively, an approximation of $\pi^n(\boldsymbol{m})$ which is obtained from a computationally faster solution can be used as the proposal distribution. We propose using the measurement update stage of the UKF for this purpose and call this approach *Unscented transformation Importance Sampling* (UIS). As shown in the flowchart in Figure 4-3, each update cycle in UIS includes two major steps:

A. UKF step: the posterior distribution of the IS step from the previous cycle, $\pi^{n-1}(\boldsymbol{m})$, is utilized as the prior of the UKF measurement update at time t_n. For the initial cycle, the initial prior, $\pi^0(\boldsymbol{m})$, is used instead. Next, the Gaussian approximation of the posterior, $UT(\pi^{n-1}(\boldsymbol{m}), \boldsymbol{d}_{obs}^n)$, is obtained using the measurement update stage of the UKF, as described by (4.13) and (4.14). Thus, the

UKF step uses $\boldsymbol{d}_{obs}^{n}$ to update the prior in which previous monitoring data are accounted for.

B. IS step: $UT(\pi^{n-1}(\boldsymbol{m}), \boldsymbol{d}_{obs}^{n})$ from the UKF step is used as the proposal distribution for the Bayesian IS as described by (4.4). In this step, the initial prior is used as the prior in Bayes' theorem and all monitoring data ($\boldsymbol{d}_{obs}^{1}$ to $\boldsymbol{d}_{obs}^{n}$) are accounted for in the weights of the samples.

When new monitoring data becomes available, the algorithm will be repeated. The stopping criteria are flexible and application-specific. For instance, criteria such as reaching a certain number of cycles or reaching certain parameter variances can be used.

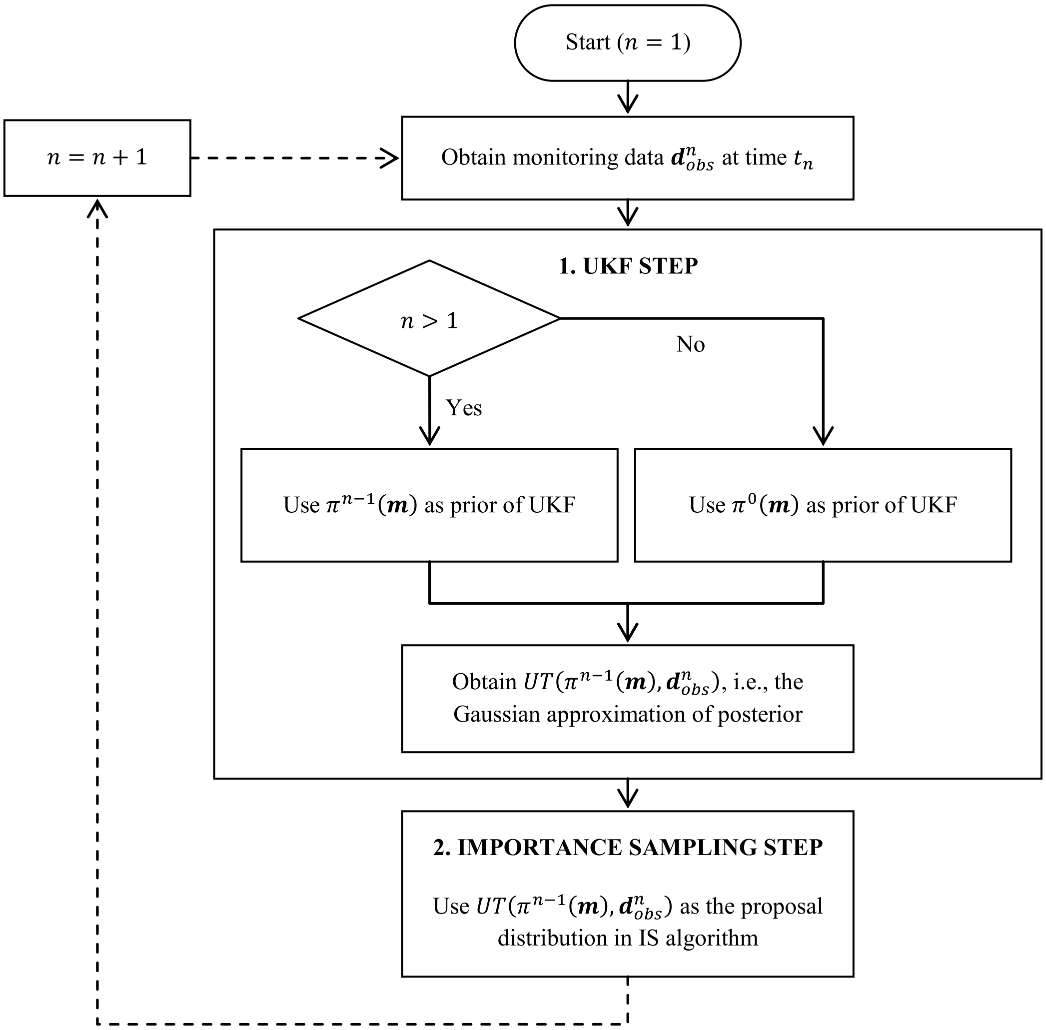


Figure 4-3: Flowchart of Unscented transformation Importance Sampling (UIS). Modified from: (Sarkarfarshi & Gracie, 2015).

By substituting proper equations into each block of Figure 4-3, the complete algorithm of UIS is obtained and is summarized in Algorithm 4-1.

Algorithm 4-1: Algorithm of the UIS. Modified from: (Sarkarfarshi & Gracie, 2015).

```
1. Set $n = n + 1$ where $n$ is monitoring time step (start with $n = 1$)
2. Obtain monitoring data $\boldsymbol{d}_{obs}^{n}$.
3. UKF step
    3.1. Set the prior of UKF step $\pi_b^n(\boldsymbol{m}) = N(\boldsymbol{\mu}_m^{n-1}, \mathbf{C}_m^{n-1})$ where $\boldsymbol{\mu}_m^{n-1} = \sum_{i=1}^{N_s} W_i^{n-1}\widehat{\boldsymbol{m}}_i^{n-1}$ and $\mathbf{C}_m^{n-1} = \sum_{i=1}^{N_s} W_i^{n-1}\left(\widehat{\boldsymbol{m}}_i^{n-1} - \boldsymbol{\mu}_b^n\right)\left(\widehat{\boldsymbol{m}}_i^{n-1} - \boldsymbol{\mu}_b^n\right)^T$ when $n > 1$. $\boldsymbol{\mu}_m^{n-1}$ and $\mathbf{C}_m^{n-1}$ are mean and covariance matrix of $\pi^0(\boldsymbol{m})$ when $n = 1$.
    3.2. Select Sigma points and their weights as follows:
         $\boldsymbol{\chi}_0 = \boldsymbol{\mu}_m^{n-1} \qquad -1 < \vartheta_0 < 1 \qquad \vartheta_i = \frac{1 - \vartheta_0}{2N_m} \qquad i = 1,2,\dots,2N_m$
         $\boldsymbol{\chi}_i = \boldsymbol{\mu}_m^{n-1} \pm \left(\sqrt{\frac{N_m}{1 - \vartheta_0}\mathbf{C}_m^{n-1}}\right)_i \quad +(\,i = 1,,\dots,N_m) \; -(\,i = N_m + 1,\dots,2N_m)$
    3.3. Propagate Sigma points through nonlinear system model and obtain $\mathbf{Y}_i = \left[\boldsymbol{g}(\boldsymbol{x}_1, t_n, \boldsymbol{\chi}_i), \boldsymbol{g}(\boldsymbol{x}_2, t_n, \boldsymbol{\chi}_i), \dots, \boldsymbol{g}\left(\boldsymbol{x}_{N_\chi}, t_n, \boldsymbol{\chi}_i\right)\right]$, $i = 0,1,\dots,2N_m$
    3.4. Calculate statistics of Sigma points and propagated Sigma points as follows.
         $\boldsymbol{\mu}_\chi = \sum_{i=0}^{2N_m} \vartheta_i \boldsymbol{\chi}_i \qquad \boldsymbol{\mu}_{\mathrm{Y}} = \sum_{i=0}^{2N_m} \vartheta_i \mathbf{Y}_i$
         $\mathbf{C}_{\mathrm{Y}} = \sum_{i=0}^{2N_m} \vartheta_i [\mathbf{Y}_i - \boldsymbol{\mu}_{\mathrm{Y}}]^T[\boldsymbol{Y}_i - \boldsymbol{\mu}_{\mathrm{Y}}] \qquad \mathbf{C}_{\chi,\mathrm{Y}} = \sum_{i=0}^{2N_m} \vartheta_i \left[\boldsymbol{\chi}_i - \boldsymbol{\mu}_\chi\right]^T[\boldsymbol{Y}_i - \boldsymbol{\mu}_{\mathrm{Y}}]$
    3.5. Calculate mean and covariance matrix of the Gaussian posterior of UKF step as $\boldsymbol{\mu}_m^n = \boldsymbol{\chi}_0 + \boldsymbol{K}(\boldsymbol{d}_{obs}^n - \boldsymbol{\mu}_{\mathrm{Y}})$ and $\mathbf{C}_m^n = \mathbf{C}_m^{n-1} - \boldsymbol{K}\left(\mathbf{C}_{\mathrm{Y}} + \mathbf{C}_{\varepsilon_{\mathrm{obs}}}\right)\boldsymbol{K}^T$ where $\boldsymbol{K} = \mathbf{C}_{\chi,\mathrm{Y}}\left(\mathbf{C}_{\mathrm{Y}} + \mathbf{C}_{\varepsilon_{\mathrm{obs}}}\right)^{-1}$
    3.6. Set $UT(\pi_b^n(\boldsymbol{m}), \boldsymbol{d}_{obs}^n) = N(\boldsymbol{\mu}_m^n, \mathbf{C}_m^n)$
4. IS step
    4.1. Set the proposal distribution $q^n(\boldsymbol{m}) = UT(\pi_b^n(\boldsymbol{m}), \boldsymbol{d}_{obs}^n)$
    4.2. Sample $\left[\widehat{\boldsymbol{m}}_1^n, \widehat{\boldsymbol{m}}_2^n, \dots, \widehat{\boldsymbol{m}}_{N_s}^n\right]$ from $q^n(\boldsymbol{m})$
    4.3. Simulate the model for each sample from time zero to $t_n$.
    4.4. Calculate un-normalized importance weights by
         $w_i^n = \pi^0(\widehat{\boldsymbol{m}}_i^n)/q^n(\widehat{\boldsymbol{m}}_i^n)\prod_{l=1}^{n} \pi_{\varepsilon_{tot}}\left(\boldsymbol{d}_{obs}^l - \boldsymbol{y}^n(\widehat{\boldsymbol{m}}_i^n)\right)$
    4.5. Normalize weights to obtain $\left[W_1^n, W_2^n, \dots, W_{N_s}^n\right]$
5. Approximate $\pi^n(\boldsymbol{m})$ with $\sum_{i=1}^{N_s} \mathrm{W}_i^n \delta(\boldsymbol{m} - \widehat{\boldsymbol{m}}_i^n)$
Repeat steps 1 to 4 for each calibration cycle until stopping criteria are met.
```

This algorithm is inspired by the Unscented Particle Filtering (UPF) (Julier & Uhlmann, 1996) and ensures that the proposal distribution improves with time and contains information from all the monitoring data up to t_n. The final posterior distribution in the UIS is generated in the IS step. Thus, unlike the KF family, no assumption of the model structure or distribution types is necessary. Moreover, the sampling based approach in both steps (UKF and IS) lets different numerical models to be incorporated in a straightforward way. The abovementioned characteristics make the UIS flexible for parameter calibration in various applications such as CS.

In the case that the true posterior is multimodal, the presented version of UIS is still applicable. However, the performance of the algorithm diminishes because the Gaussian proposal cannot accurately represent multi-modal distributions. For such cases, the Gaussian approximation of the posterior in the UKF step can be replaced by a Gaussian mixture model.

4.2.5. Defensive Importance Sampling in UIS

As described in Section 2.5.5, the proposal distribution in IS needs to have a heavier tail than the actual sampling distribution for the solution to be robust (Owen & Zhou, 2000). To ensure the tails of the proposal distribution $q(\boldsymbol{m})$ are heavier than the sampling distribution, we can mix $q(\boldsymbol{m})$ with a heavy-tailed distribution $\hat{q}(\boldsymbol{m})$, such as a uniform distribution, in a process called “defensive mixture sampling” (Hesterberg, 1995) or “defensive importance sampling” (Owen & Zhou, 2000),

$$q^*(\boldsymbol{m}) = (1-\eta)q(\boldsymbol{m}) + \eta\hat{q}(\boldsymbol{m}) \tag{4.17}$$

where η is the defensive mixture ratio and $0 < \eta < 1$. Moreover, the defensive mixture distribution searches the areas of the parameter space with very small probability density for potentially better parameter sets. A defensive mixture distribution can be added to UIS by modifying the sub-step 4.1 of Algorithm 4-1 as

$$q^*(\boldsymbol{m}) = (1-\eta)UT(\pi_b^n(\boldsymbol{m}), \boldsymbol{d}_{obs}^n) + \eta\hat{q}(\boldsymbol{m}) \tag{4.18}$$

where η is the mixture ratio and is one of the tuning parameters. A very small η does not affect the sample population and a very large η is likely to waste numerous model runs by encouraging samples with negligible posterior weight. In the literature, $0.1 < \eta < 0.5$ is recommended (Hesterberg, 1995; Owen & Zhou, 2000).

4.3. Analytical Case Study I

Before moving to CS case study, a simplified linear problem with a known solution is presented to demonstrate the application and accuracy of UIS.

4.3.1. Case Description

Assume a simple deterministic, linear, and scalar system described by $d_{tru}(x,t) = 8x + 10t$ where x is a location varying between 1 and 20 and t is time. $d_{tru}(x,t)$ is modelled by a scalar function denoted by $g(x,t,[m_1,m_2]) = m_1x + m_2t$ and m_1 and m_2 are uncertain, time-independent and location-independent model parameters. All parameters are dimensionless.

In three time steps, $t_1 = 1$, $t_2 = 2$ and $t_3 = 3$, $d_{tru}(x,t)$ is measured at two locations, $x = 7$ and $x = 10$, as shown in Figure 4-4. Thus, there are a total of 6 measurements, two at each time step. These measurements are denoted by $d_{obs}(7,t_n)$ and $d_{obs}(10,t_n)$ where $n = 1,2,3$. Model discrepancy is assumed zero and measurements are assumed to be subject to an IID Gaussian noise, i.e., $\boldsymbol{d}_{obs}^n = \boldsymbol{y}^n([m_1,m_2]) + \boldsymbol{\varepsilon}_{obs}$ where $\boldsymbol{d}_{obs}^n = \left[d_{obs}(7,t_n), d_{obs}(10,t_n)\right]$, $\boldsymbol{y}^n([m_1,m_2]) = [g(7,t_n,[m_1,m_2]), g(10,t_n,[m_1,m_2])]$, $\boldsymbol{\varepsilon}_{obs} \sim \pi_{obs}(\boldsymbol{\varepsilon}_{obs})$ and $\pi_{obs}(\boldsymbol{\varepsilon}_{obs})$ is a bivariate Normal distribution with a mean of [0,0] and a covariance matrix with diagonal elements of $\sqrt{10}$ and off-diagonal elements of zero.

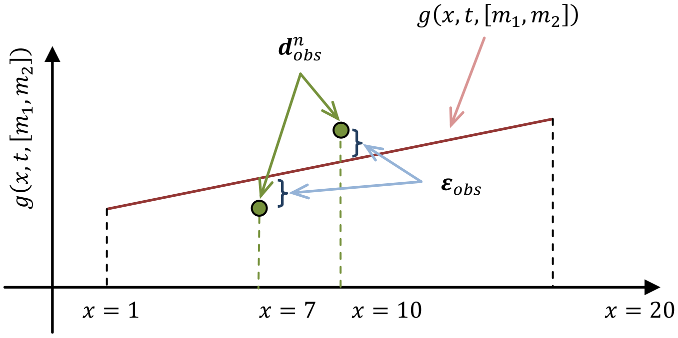


Figure 4-4: Schematic of the system behaviour and noisy measurements at an arbitrary time t.

Values of m_1 and m_2 are assumed uncertain with a joint Gaussian prior distribution denoted by $\pi^0([m_1,m_2])$ with a mean of $\boldsymbol{\mu}^0_{[m_1,m_2]}=[12{,}12]$ and a covariance matrix $\mathbf{C}^0_{[m_1,m_2]}$ with diagonal elements of 1 and off-diagonal elements of zero. This prior is equivalent to using independent marginal Gaussian priors with a mean of 12 and a standard deviation of 1 for m_1 and m_2.

At each of the three time steps, new monitoring data is obtained and the posterior distribution of m_1 and m_2 can be obtained. In other words, our objective is to find the posterior distribution of m_1 and m_2 at each time step, denoted by $\pi^n([m_1,m_2])$, where $n=1{,}2{,}3$. We call this process a *calibration cycle* for each of the time steps.

In each time step, Bayes' formula in equation (4.4) can be re-written as

$$\pi^n([m_1,m_2])=\frac{\pi^0([m_1,m_2])\prod_{l=1}^{n}\prod_{x=7,10}\pi_{\varepsilon_{obs}}\left(d_{obs}(x,t_l)-g(x,t_l,[m_1,m_2])\right)}{c} \qquad (4.19)$$

where c is the normalizing constant.

An alternative way to look at this problem is to assume that the monitoring data at each point in time and space is a realization of a Gaussian random variable with a mean of $g(x,t,[m_1,m_2])$ and standard deviation of 10 and m_1 and m_2 are uncertain parameters of the mean of this Gaussian distribution. Thus, according to Bayes' theorem, we can obtain the joint posterior distribution of m_1 and m_2 given the noisy monitoring data.

Since $\pi^0([m_1, m_2])$ and $\pi_{obs}(\boldsymbol{\varepsilon}_{obs})$ are both Gaussian and $g(x, t, [m_1, m_2])$ is a linear function of m_1 and m_2, the posterior distribution of m_1 and m_2 in equation (4.19) is a bivariate Gaussian distribution and can be evaluated exactly. This exact solution is equivalent to the measurement update stage of the Kalman Filter and can be obtained by

$$\pi^n([m_1, m_2]) = \frac{\pi^0([m_1, m_2]) \prod_{l=1}^{n} \prod_{x=7,10} N\big(d_{obs}(x, t_l) - g(x, t_l, [m_1, m_2]), 0,10\big)}{c} \tag{4.20}$$

By expanding and simplifying equation (4.20), it can be shown that $\pi^n([m_1, m_2])$ is Gaussian with a covariance matrix of

$$\mathbf{C}^n_{[m_1,m_2]} = \left[\left(\mathbf{C}^0_{[m_1,m_2]}\right)^{-1} + \frac{\sum_{l=1}^{n} \sum_{x=7,10} \begin{bmatrix} x^2 & xt_l \\ xt_l & t_l^2 \end{bmatrix}}{10^2}\right]^{-1} \tag{4.21}$$

and a mean of

$$\boldsymbol{\mu}^n_{[m_1,m_2]} = \mathbf{C}^n_{[m_1,m_2]} \left(\mathbf{C}^0_{[m_1,m_2]}\right)^{-1} \boldsymbol{\mu}^0_{[m_1,m_2]} \tag{4.22}$$

The probability distribution of the system model forecast at a specific time $t_{n^*} \geq t_n$ might be of interest. Since $g(x, t, [m_1, m_2])$ is a linear function of m_1 and m_2, the posterior distribution of model forecast is also Gaussian with a mean of

$$\mu_y^{n,n^*} = [x \quad t_{n^*}] \boldsymbol{\mu}^n_{[m_1,m_2]} \tag{4.23}$$

and a covariance matrix of

$$\mathbf{C}_y^{n,n^*} = [x \quad t_{n^*}] \mathbf{C}^n_{[m_1,m_2]} \begin{bmatrix} x \\ t_{n^*} \end{bmatrix} \tag{4.24}$$

Now, we want to demonstrate how UIS performs in comparison with this exact solution (which is equivalent to measurement update stage of UKF in this case study) and ordinary IS with $\pi^0([m_1, m_2])$ as a static proposal distribution. Thus, we solved this problem three times using the measurement update stage of UKF, ordinary IS with 100 samples, and UIS with 100 samples and no defensive mixture distribution.

4.3.2. Results and Discussion

Figure 4-5 illustrates the joint prior and the joint posterior distributions of m_1 and m_2 for each calibration cycle in each of these three solutions. Since the posterior distributions obtained by the measurement update stage of UKF are exact in this problem, we use them as a reference. It can be seen in Figure 4-5 that in UIS, posterior samples resemble the true posterior distribution very closely while IS samples are unable to do so. It can also be seen that the location of the prior samples drawn from the proposal distribution in UIS (regardless of the weights) are close to the true posterior distribution, while this is not the case in IS.

To compare the accuracy of IS and UIS against the exact solution, the Root Mean Square Error (RMSE) of the expected value of the posterior predictive distribution of $g(x, t_n, [m_1, m_2])$ at $n = 3$, i.e., the forecast of $g(x, t_n, [m_1, m_2])$ at time $t_n = 3$, is calculated at the end of each calibration cycle, as

$$RMSE^n = \sqrt{\frac{\sum_{x=1}^{20}\left[d_{tru}(x,3) - \mu_g^{n,3}\right]^2}{20}} \tag{4.25}$$

where $\mu_g^{n,3}$ is obtained from equation (4.23).

As shown in Figure 4-6, the RMSE of measurement update stage of UKF (exact solution) and UIS are very close and notably smaller than IS. Thus, we conclude that UIS results were very close to exact solution and notably more accurate than ordinary IS.

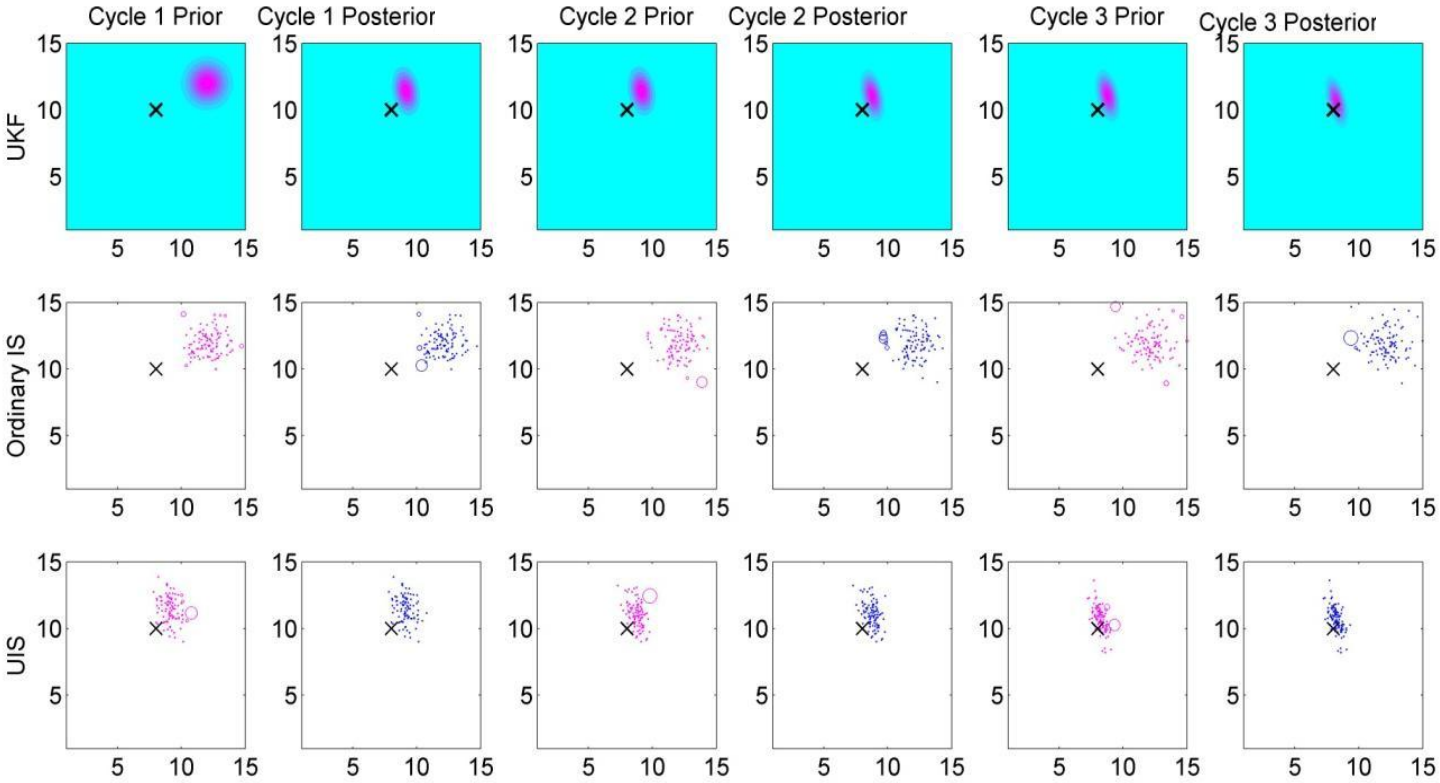


Figure 4-5: Prior and posterior distributions of the presented linear problem using measurement update stage of UKF, IS and UIS. Horizontal and vertical axis indicate m_1 and m_2, respectively. The black Xs indicates the location of true parameter values. In IS and UIS, the size of the markers indicates the magnitudes of the importance weight. Weights of the samples in prior distributions of IS and UIS include $1/q^n(\hat{\boldsymbol{m}}_i^n)$ term.

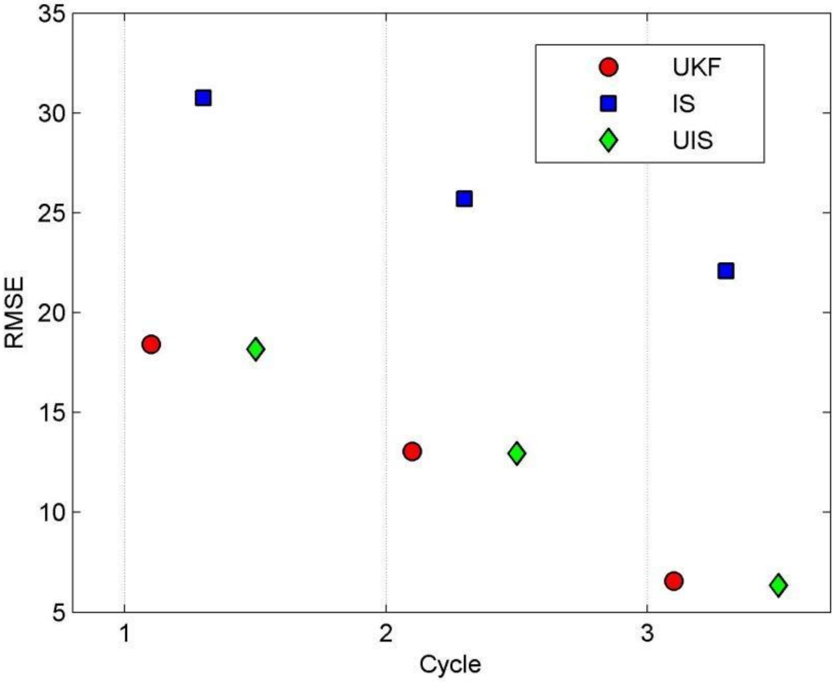


Figure 4-6: The RMSE of model output at the end of each calibration cycle for the presented linear problem. From: (Sarkarfarshi & Gracie, 2015).

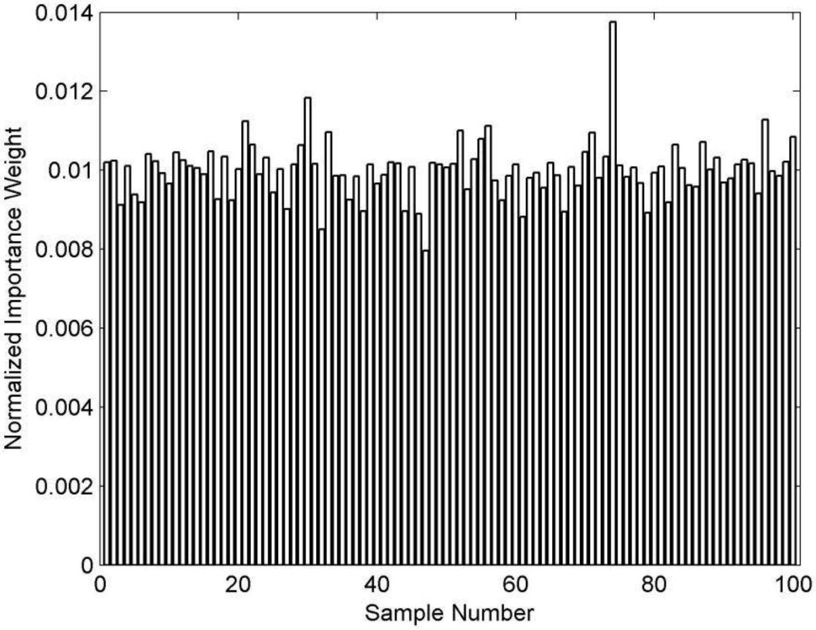
To further investigate the performance of UIS and IS, we plotted the importance weights of the samples at the last calibration cycle for UIS and IS in Figure 4-7. It can be seen in Figure 4-7a that the weights are very close together in UIS. In other words, the variance of the weights is very small. The reason for this almost even weight distribution is that the proposal distribution in UIS were close to the true posterior distribution and thus, un-normalized importance weights for all samples are close to one. In ordinary IS, however, one of the samples is carrying more than 95% of the importance weights as shown in Figure 4-7b due to a poor proposal distribution. In practice, 97 out of 100 samples in IS had a near zero weight and thus, only three out of 100 samples were effective in the calculations. To avoid this problem in ordinary IS, the number of samples should be increased significantly the generate more effective samples. Thus, for the ordinary IS to obtain the same accuracy as UIS, a significantly larger sample size is required (about 33 times more in this problem).

According to the discussion above, the effective samples size in IS and UIS can be quantified by a dimensionless number, defined as

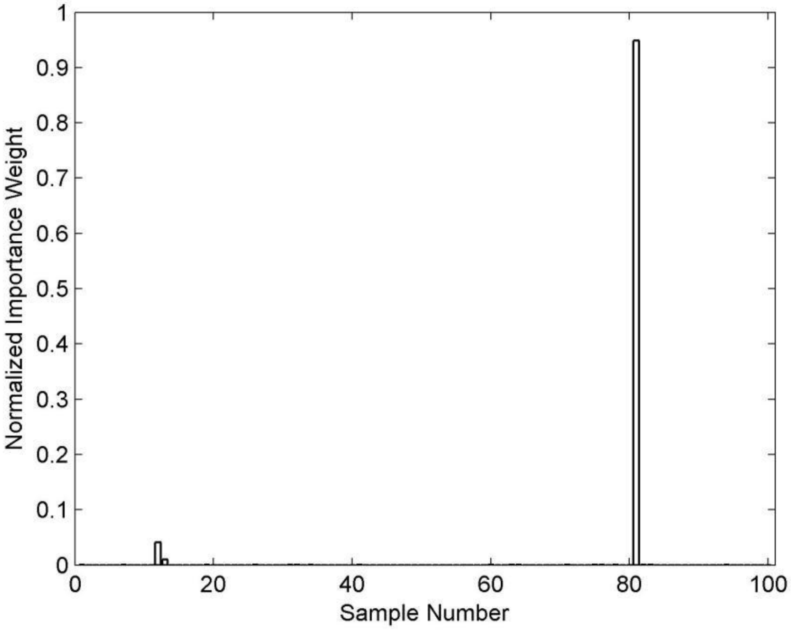
$$N_{eff} = \frac{1}{\sum_{i=1}^{N_s} W_i^2} \tag{4.26}$$

N_{eff} is in fact a measure for sample quality of a particle filter (Arulampalam *et al*., 2002). When all importance weights are equal, N_{eff} is equal to N_s meaning that the samples are drawn from the true posterior. When most samples are drawn from non-important areas of the posterior (e.g., in ordinary IS in this problem), most W_i become very close to zero while a few W_i carry most of the weight. In such cases, N_{eff} becomes closer to one indicating that only a few samples are effective and sample quality is poor. In the extreme case that one sample carries all the weight (i.e., one W_i equals to one and all other W_i equal zero), N_{eff} is equal to one. This situation is known as *filter degeneracy* in particle filters (Arulampalam *et al*., 2002) and should be avoided. The adaptive nature of UIS is expected to increase N_{eff} since the proposal distribution is adaptively adjusted to more closely resemble the true posterior.

We define the effective sample ratio, denoted by R_{eff}, as the N_{eff} normalized by the total number of samples, i.e., $R_{eff} = N_{eff}/N_s$. Figure 4-8 plots R_{eff} for IS and UIS in calibration cycle 3. It can be seen that R_{eff} is close to one for UIS, meaning that all samples were effective in approximation of the posterior distribution, while R_{eff} is close to zero for IS, meaning that very few samples were effective in approximation of the posterior distribution.



(a)



(b)

Figure 4-7: Importance weights in calibration cycle 3 for the described linear problem for (a) UIS and (b) IS.

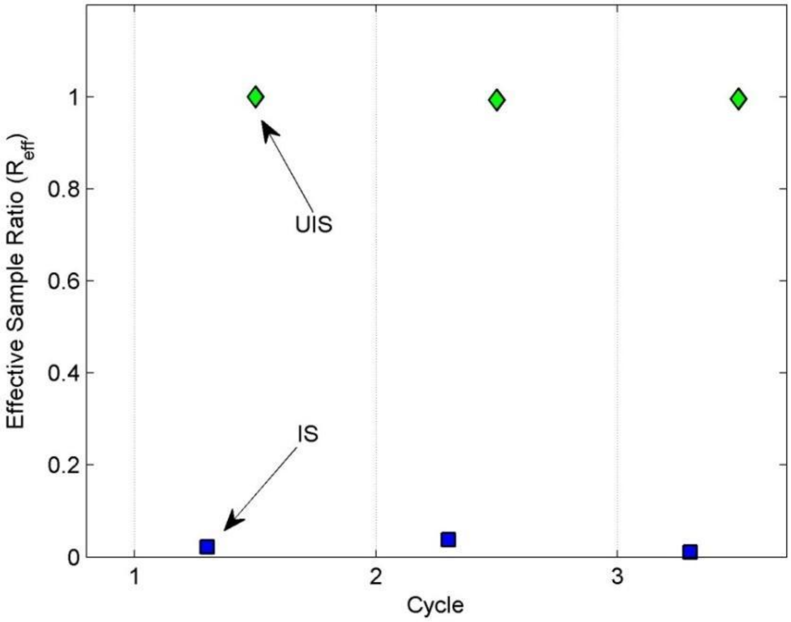


Figure 4-8: The effective sample ratio (R_{eff}) for IS and UIS in calibration cycle 3 for the described linear problem.

4.4. Analytical Case Study II

Now, we present another analytical case study with a significantly nonlinear analytical function as the system model. This case study enables us to investigate the performance of UIS in nonlinear situations for various numbers of uncertain parameters with a reasonable computational effort.

4.4.1. Case Description

Assume a deterministic system described by

$$d_{tru}(x,t) = \sum_{i=1}^{N_m-1} a_i^2 \cos(a_{i+1})^2 \sin(x) + a_{i+1}^2 \cos(a_i)^2 \sin(t) \tag{4.27}$$

where N_m is the total number of a_i, x is a dimensionless location varying between 1 and 20, t is dimensionless time and

$$a_i = \begin{cases} 8 & i = 1,3,5,\ldots \\ 10 & i = 2,4,6,\ldots \end{cases} \tag{4.28}$$

$d_{tru}(x,t)$ is modelled by a scalar function denoted by $g(x,t,\boldsymbol{m})$ with a similar structure as

$$g(x,t,\boldsymbol{m}) = \sum_{i=1}^{N_m-1} m_i^2 \cos(m_{i+1})^2 \sin(x) + m_{i+1}^2 \cos(m_i)^2 \sin(t) \tag{4.29}$$

where m_i are the uncertain model parameters and $\boldsymbol{m} = \left[m_1, m_2, \ldots, m_{N_m}\right]$ is the uncertain parameter vector. In four time steps, $n = 1, 2, 3,4$, $d_{tru}(x,t)$ is measured at two locations, $x = 7$ and $x = 13$. These measurements are denoted by $d_{obs}(7,t_n)$ and $d_{obs}(13,t_n)$ for time step n. Model discrepancy is assumed zero and measurements are assumed to be subject to an IID Gaussian noise denoted by $\boldsymbol{\varepsilon}_{obs}$, i.e., $\boldsymbol{d}_{obs}^n = \boldsymbol{y}^n(\boldsymbol{m}) + \boldsymbol{\varepsilon}_{obs}$ where $\boldsymbol{d}_{obs}^n = \left[d_{obs}(7,t_n), d_{obs}(13,t_n)\right]$, $\boldsymbol{y}^n(\boldsymbol{m}) = [g(7,t_n,\boldsymbol{m}), g(10,t_n,\boldsymbol{m})]$, $\boldsymbol{\varepsilon}_{obs} \sim \pi_{obs}(\boldsymbol{\varepsilon}_{obs})$ and $\pi_{obs}(\boldsymbol{\varepsilon}_{obs})$ is a bivariate Normal distribution with a mean of [0,0] and a covariance matrix with diagonal elements of 10 and off-diagonal elements of zero. The prior distribution of $\boldsymbol{m}$ is assumed to be Gaussian denoted by $\pi^0(\boldsymbol{m})$ with a mean of $[12, \ldots, 12]$ and a covariance matrix with diagonal elements of 1 and off-diagonal elements of zero.

At each of the four time steps, new monitoring data is obtained. Our objective is to find the posterior distribution of $\boldsymbol{m}$ at each time step, denoted by $\pi^n(\boldsymbol{m})$ by using Algorithm 1 and for $n = 1, 2, 3,4$.

4.4.2. Calibration Scenarios and Comparison Metrics

We solved this problem using the measurement update stage of UKF, ordinary IS with 200 samples and UIS with 200 samples and no defensive mixture distribution. Three sizes of the uncertain parameter vector are considered $N_m = 10$, $N_m = 20$ and $N_m = 50$. Table 4-1 summarizes all calibration scenarios considered in this case study.

Table 4-1: Calibration scenarios in the simplified case study. Modified from: (Sarkarfarshi & Gracie, 2015).

Scenario Name	**Method**	**Number of parameters (N_m)**	**Sample count (N_s)**	**Simulations per cycle**
UKF10	UKF	10	-	21
IS10	IS	10	200	200
UIS10	UIS	10	200	221
UKF20	UKF	20	-	41
IS20	IS	20	200	200
UIS20	UIS	20	200	241
UKF50	UKF	50	-	101
IS50	IS	50	200	200
UIS50	UIS	50	200	301

To compare the accuracy of the methods, the Root Mean Square Error (RMSE) of the model output at time $n = 4$ is calculated at the end of each calibration cycle, using the expected value (mean) of the posterior distribution of $g(x, t_n, \boldsymbol{m})$ by

$$RMSE^n = \sqrt{\frac{\sum_{x=1}^{20}\left[d_{tru}(x, 4) - \mu_g^{n,4}\right]^2}{20}} \tag{4.30}$$

where $\mu_g^{n,4}$ is the mean of the posterior distribution of $g(x, t, \boldsymbol{m})$, obtained using Monte Carlo simulation, when $\pi^n(\boldsymbol{m})$ is used as the distribution of model parameters.

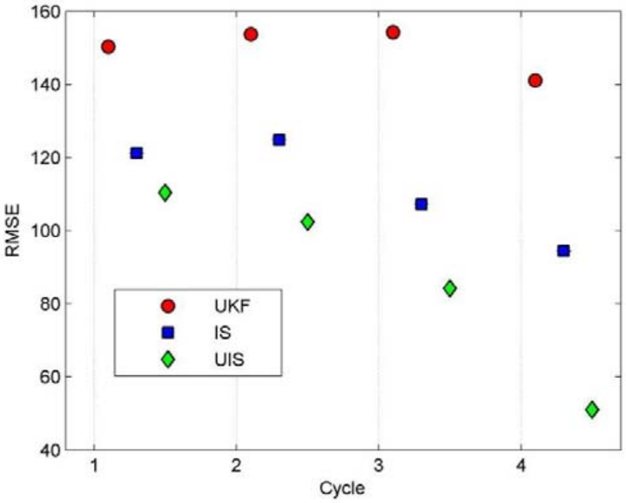
4.4.3. Results and Discussion

Figure 4-9 plots the RMSE of all scenarios listed in Table 4-1. It can be seen that the RMSE of UIS is considerably smaller than the RMSE of the UKF measurement stage and IS in all scenarios. In scenarios with 10 and 20 parameters, IS was more accurate than UKF but worse than UIS in terms of accuracy. In the scenarios with 50 parameters, IS performed worse than both UIS and UKF.

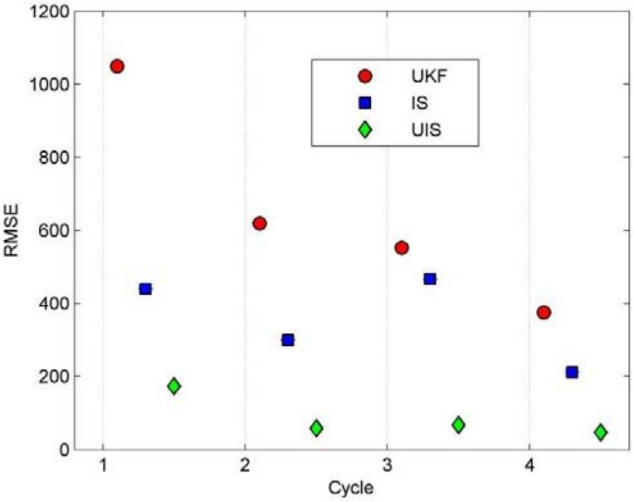
From Table 4-1, it can be seen that the measurement update stage of UKF used the smallest number of model simulation. However, the UIS was always more accurate than the UKF. For instance in the case of 20 uncertain parameters, it can be seen from Figure 4-9(b) that the error measured by the RMSE of UIS20 was more than 10 times smaller than the error obtained in UKF20. The reason why UIS is more accurate than either IS or UKF is that UIS samples from a proposal distribution that is closer to the posterior, without assuming linearity of the model and Gaussian distributions. Thus, UIS is able to capture the true posterior more accurately than the measurement update stage of UKF and IS without a dramatic increase in the number of simulations.

It can also be seen from Table 4-1 that UIS used more simulations (10.5%, 20.5% and 50.5% in the scenarios with 10, 20 and 50 parameters, respectively) than ordinary IS, but is notably more accurate in all scenarios. It can also be seen that the relative accuracy of UIS over the ordinary IS method increases with the number of parameters. For instance, the RMSE of UIS10 was less than two times smaller than the RMSE of IS10, while the RMSE of UIS50 was more than 30 times smaller than the RMSE of IS50. The reason for this increasing relative accuracy is that UIS adaptively revised the proposal distribution while ordinary IS sampled from the initial prior in all cycles. Thus, when the number of uncertain parameters increased, IS was unable to estimate the larger-dimensional posterior distributions with 200 samples while UIS sampled from important areas of the posterior distributions and yielded a much smaller RMSE with the same number of samples.

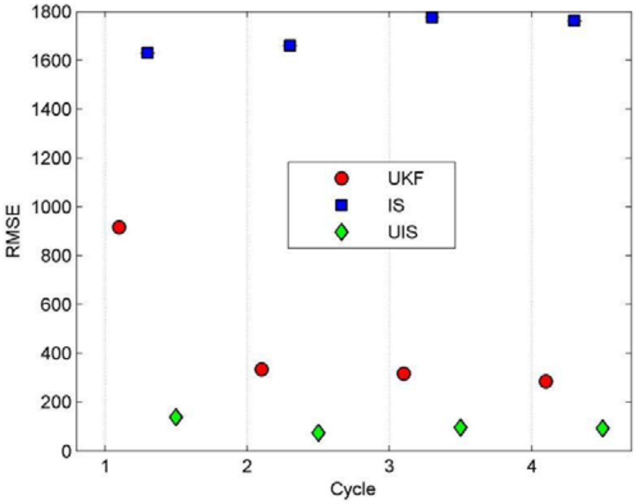
For a more in-depth comparison of UIS with UKF and IS, including an investigation of effect of tuning parameters on the performance of UIS, a more sophisticated CS case study is presented in the next section.



(a) 10 parameters



(b) 20 parameters



(c) 50 parameters

Figure 4-9: The RMSE of model output at the end of each calibration cycle for the scenarios with (a) 10, (b) 20 and (c) 50 parameters. The RMSE of UIS was well below RMSE of both UKF measurement stage and IS in all cycles of all scenarios. From: (Sarkarfarshi & Gracie, 2015).

4.5. CS Case Study

A synthetic case study of CO_2 injection in a deep saline aquifer is generated and used for two purposes; first, comparing the effectiveness of UIS with the measurement update of UKF and IS with a static proposal; and second, studying the effect of the number of samples (N_s) and the defensive mixture ratio (η) on the performance of UIS. Nine calibration scenarios are designed for this task.

4.5.1. Case Description

A horizontal and rigid saline aquifer is assumed, initially saturated with brine. The aquifer is 20 m thick and its base is located at a depth of 2 km. The aquifer domain is assumed to be 1020 m × 1020 m. Four facies are assumed within the aquifer, distributed as shown in Figure 4-10. Each facies is assumed homogeneous with a constant porosity and permeability. The porosities of the facies are assumed to be $\phi_1 = 0.18$, $\phi_2 = 0.13$, $\phi_3 = 0.15$ and $\phi_4 = 0.21$, respectively. For all four facies, the porosity-permeability correlation is assumed to be $\log_{10}(k_i) = 7 + 7\log_{10}(\phi_i)$ where k_i is the intrinsic permeability of facies i in mD and ϕ_i is the porosity of facies i as a bulk volume fraction (SPE International, 2013).

Initial salinity of the brine is assumed to be 50,000 ppm. The CO_2 injection rate (Q) is assumed to be 0.035 Mt/year and the injection is assumed to take place continuously for 5 years through a vertical well located at the center of the domain and penetrating the whole thickness of the aquifer.

All parameters are assumed to be known except the porosity and the permeability of the four facies. The porosities are considered to be the uncertain parameters and the permeability of each facies is obtained using the porosity-permeability correlation given above. The prior distribution of the uncertain parameter vector ($\boldsymbol{m} = [\phi_1, \phi_2, \phi_3, \phi_4]$) is assumed to be a multivariate Gaussian distribution with a mean of $[0.15, 0.15, 0.15, 0.15]$, diagonal covariance matrix elements of 0.02^2 and off-diagonal covariance matrix elements of zero.

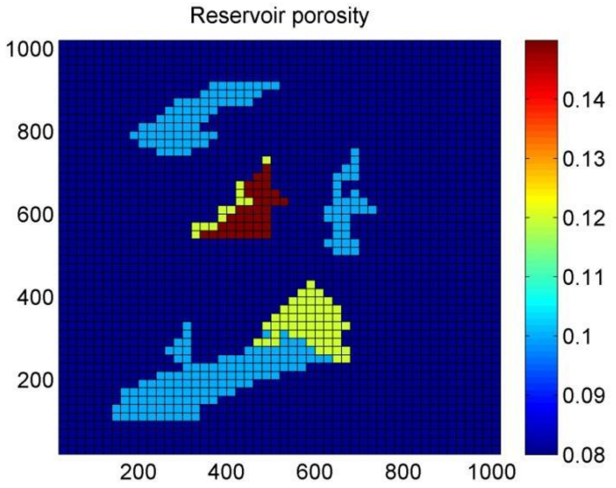


Figure 4-10: True porosity distribution in the reservoir. From: (Sarkarfarshi & Gracie, 2015).

Model error is assumed zero since the same model is used for calibration and generating the monitoring data. Pressure and saturation at the end of each year are measured at the injection well and two other monitoring wells, located 110 m away from the injection well in X and Y directions, respectively. Figure 4-11 illustrates a schematic of the aquifer dimensions, the injection well and the monitoring wells. The saturation measurement error is assumed to be Gaussian with a mean of zero and a standard deviation of 0.02. The pressure measurement error is also assumed to be Gaussian with a mean of zero and a standard deviation of 500 kPa.

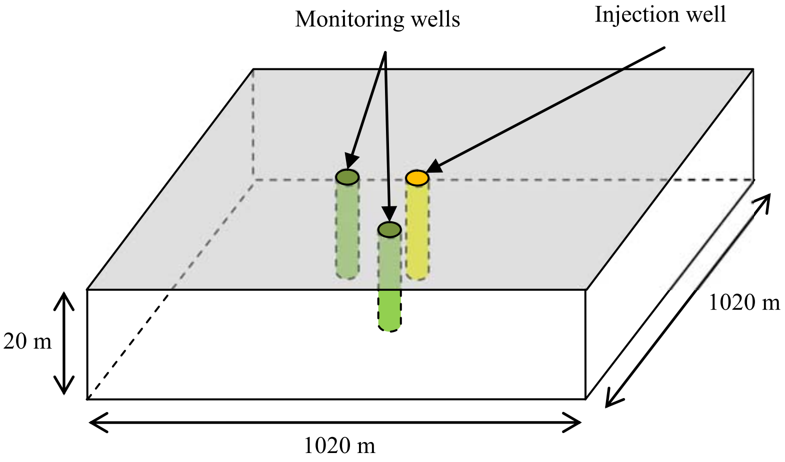


Figure 4-11: Schematic of the aquifer dimensions, injection well and monitoring wells. From: (Sarkarfarshi & Gracie, 2015).

4.5.2. System Model

The system model is governed by the mass conservation equations for the multiphase and multicomponent flow in a porous medium of porosity ϕ, which are given by

$$\frac{\partial\left(\phi \sum_{i=1}^{N_{ph}} X_{\beta}^{i}\rho_i S_i\right)}{\partial t}+\sum_{i=1}^{N_{ph}} \nabla.\left(X_{\beta}^{i}\rho_i \boldsymbol{q}_i\right)=Q_i \tag{4.31}$$

where N_{ph} is the number of phases, X_{β}^{i} is the mass fraction of component β in phase i, ρ_i is the density of phase i, S_i is the saturation of phase i and Q_i is the source/sink mass flux of phase i. The volume flux of phase i, $\boldsymbol{q}_i$, is given by Darcy's law,

$$\boldsymbol{q}_i=\frac{kk_{ri}}{\mu_i}(\nabla p_i-\rho_i \boldsymbol{g}) \tag{4.32}$$

where i is the phase index, k is the intrinsic permeability, $\boldsymbol{g}$ is the gravity acceleration and k_{ri} is the relative permeability of phase i. p_i is the pressure of phase i and is obtained by

$$p_i=p_{ref}-p_c \tag{4.33}$$

where p_{ref} is the reference pressure and p_c is the capillary pressure. The relative permeability and the capillary pressure of each phase are assumed functions of saturation.

Schlumberger's ECLIPSE reservoir simulator is used to solve the above flow equations to obtain the pressure and the saturation of each phase (CO_2 and water) and the molar fraction of each component (salt). The GASWAT keyword in ECLIPSE allows simulating multiphase equilibrium of gas and aqueous phases using a modified Peng-Robinson equation of state (Soreide & Whitson, 1992; Schlumberger, 2013). No flow boundary conditions are specified on the top and the bottom of the reservoir, open boundary conditions are specified on the reservoir boundaries and an adaptive implicit solution procedure is used (Tavakoli *et al.*, 2013). A more detailed description of the flow equations and the solution options in ECLIPSE is beyond the scope of this chapter and can be found in the ECLIPSE technical manual (Schlumberger, 2013).

4.5.3. Calibration Scenarios and Comparison Metrics

Nine calibration scenarios are described in Table 4-2. All scenarios use UIS as the calibration method, except scenario 1, which uses the measurement update step of UKF and scenario 2, which uses IS with $\pi^0(\boldsymbol{m})$ as the proposal. The 5th scenario is chosen as the base case for UIS where 200 samples and no defensive mixture distribution are used. Scenarios 3, 4 and 6 are similar to the base case, except for the number of samples. Scenarios 7, 8 and 9 have different defensive mixture ratios compared to the base case and use a uniform mixture distribution bounded between 0.01 and 0.5 for the porosities of all facies. For all scenarios, ϑ_0 in the UKF step is assumed to be zero.

Table 4-2: Calibration scenarios. From: (Sarkarfarshi & Gracie, 2015).

Scenario # and name	**Method**	**Sample count (N_s)**	**Defensive mixture ratio (η)**	**Model simulations per cycle**
1: UKF	UKF	-	-	9
2: IS	IS	200	0	200
3: N50	UIS	50	0	59
4: N100	UIS	100	0	109
5: BASE	UIS	200	0	209
6: N500	UIS	500	0	509
7: η0.1	UIS	200	0.1	209
8: η0.25	UIS	200	0.25	209
9: η0.5	UIS	200	0.5	209

Scenarios above are constructed so that three sets of comparisons can be conducted:

1. Comparing measurement stage of UKF, IS (with the prior as proposal) and UIS (scenarios 1,2 and 3)
2. Comparing the effect of the number of samples in UIS (scenarios 3, 4, 5 and 6)
3. Comparing the effect of the defensive mixture ratios (scenarios 5, 7 and 8)

Three metrics are chosen for these comparisons:

A. The first metric is the normalized Root Mean Square Error (RMSE) of the saturation and the pressure fields after 5 years of injection, evaluated after each calibration cycle, similar to equation (4.30) in the previous case study. The error expectation of each field is first obtained and is normalized by the variance of monitoring error corresponding to that field and the sum of the normalized RMSE of both fields in time and space is used as the comparison metric.

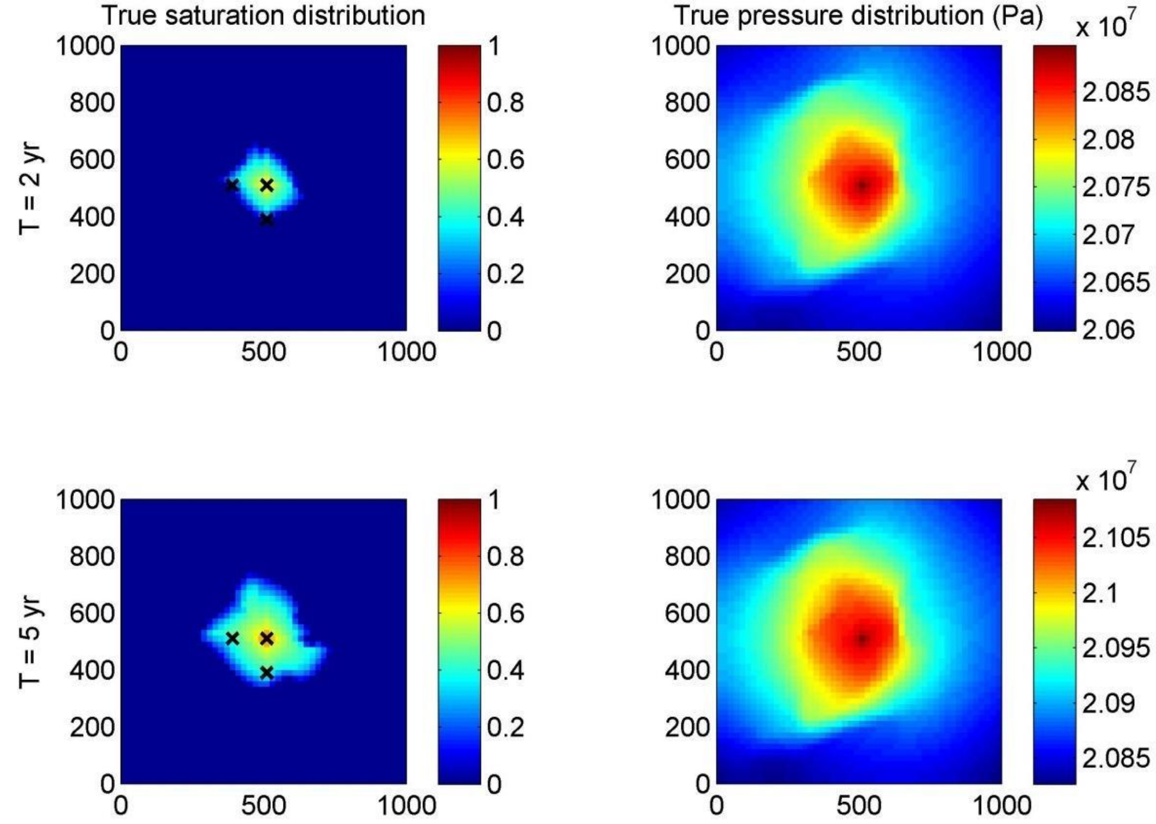
B. The second metric is the effective sample ratio (R_{eff}).

C. The third metric is the expectation of the absolute error and standard deviation of the error for both the saturation and pressure fields. These fields are plotted in Section 4.6 and discussed qualitatively.

4.5.4. Results

The synthetic "True" system behaviour is generated using true system characteristics and is perturbed by the monitoring noise, as described in Section 4.5.1. Figure 4-12 is a snapshot of the true saturation and pressure distribution of CO_2 within the reservoir after 2 and 5 years of injection. Black "X"s in the saturation plots indicate monitoring locations. The CO_2 plume reached the monitoring wells by the second year of injection. Thus, the first calibration cycle is mostly based on the pressure measurements because the pressure plume spreads considerably faster than the saturation plume.

Calibration is conducted for all scenarios listed in Table 4-2. After each calibration and cycle for each scenario, the model is stochastically simulated (i.e., Monte Carlo simulation) up to year 5 using the weighted samples from the posterior distribution of that cycle. The expectation and standard deviation of the error between the true and the forecasted saturation and pressure fields are then generated for each scenario. In all scenarios, the saturation error expectation and the saturation uncertainty (standard deviation) were more significant near the plume boundaries. The pressure error expectation and uncertainty were more complex and did not follow such a standard pattern.



(a) (b)

Figure 4-12: Snapshots of true (a) effective saturation and (b) pressure distribution (Pa) of the reservoir after 2 and 5 years of injection. Black "X"s on saturation plots indicate monitoring locations. From: (Sarkarfarshi & Gracie, 2015).

The RMSE of all scenarios and R_{eff} for scenarios 2 to 9 were calculated after each cycle and are listed in Table 4-3. This table will be used as the reference for further comparisons in Section 4.6. The overall trend of the table suggests that UIS performed better than both UKF and IS, increasing the number of samples increased the performance of UIS and a moderate defensive mixture ratio increased the performance of UIS at the cost of a slight decrease in R_{eff}.

Table 4-3: RMSE and ratio of effective samples for all scenarios. From: (Sarkarfarshi & Gracie, 2015).

Scenario	**Cycle 1**		**Cycle 2**		**Cycle 3**		**Cycle 4**		**Cycle 5**	
	RMSE	R_{eff}	RMSE	R_{eff}	RMSE	R_{eff}	RMSE	R_{eff}	RMSE	R_{eff}
1: UKF	77.12	-	42.49	-	37.86	-	37.57	-	31.12	-
2: IS	78.48	0.04	73.68	0.005	50.30	0.005	57.18	0.005	27.14	0.005
3: N50	77.74	0.03	65.27	0.09	46.68	0.13	40.84	0.042	32.81	0.05
4: N100	76.59	0.50	37.16	0.25	35.11	0.25	28.44	0.13	23.94	0.18
5: BASE	80.72	0.30	38.55	0.09	30.21	0.17	29.60	0.65	22.05	0.34
6: N500	75.42	0.26	34.91	0.15	20.64	0.34	28.45	0.79	21.88	0.44
7: η0.1	83.61	0.19	33.89	0.17	20.88	0.50	20.00	0.57	23.71	0.27
8: η0.25	75.50	0.28	22.83	0.061	31.42	0.16	29.01	0.46	21.17	0.29
9: η0.5	78.05	0.21	39.62	0.13	29.74	0.13	28.05	0.16	22.04	0.15

4.6. Discussion

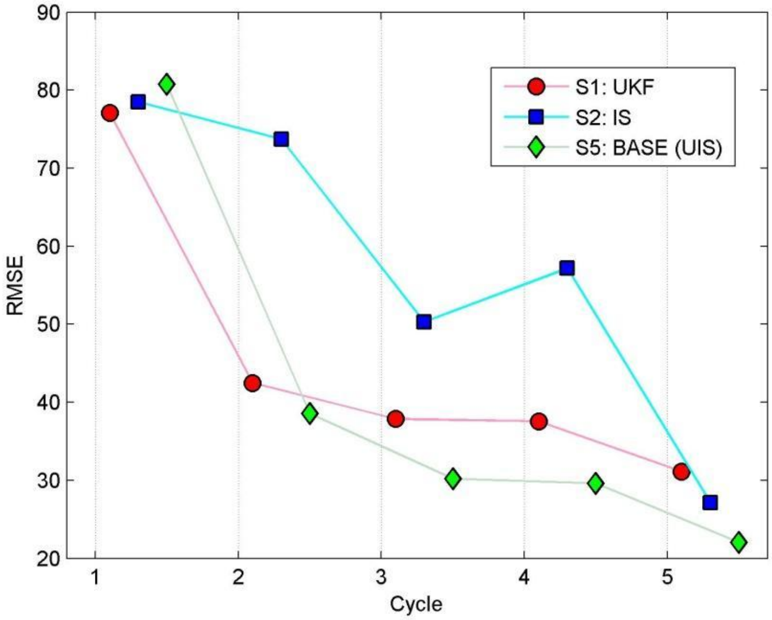
Now we compare the calibration scenarios to assess the performance of UIS against UKF measurement update step and IS, and also investigate the effect of the tuning parameters (the number of samples and the defensive mixture ratio) on the performance of UIS. We also discuss the computational cost of these algorithms.

4.6.1. Comparing Calibration Algorithms

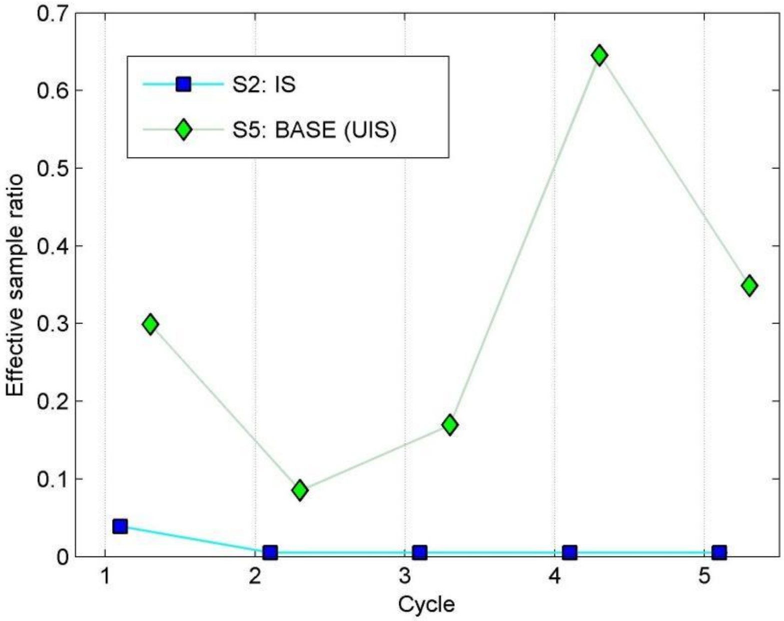
The RMSE for UKF, IS and the base case of UIS (scenarios 1, 2 and 5) and the effective ratio of the samples for IS and the base case of UIS are extracted from Table 4-3 and plotted in Figure 4-13. It can be seen that UIS had a smaller RMSE compared to both UKF and IS. With the exception of the last year, UKF measurement update stage performed better than IS. This is because the static proposal distribution in IS was much wider than the real posterior. Thus, most samples were drawn from non-important areas of the posterior distribution, resulting in a large ratio of samples having weights of zero or close to zero. Thus, the posterior distribution was represented by only a few samples. This shortcoming is apparent in Table 4-3 and Figure 4-13b. It can be seen that after the second year, the number of effective samples for IS was very close to one. The number of samples should be increased to increase N_{eff} for this case, which will also increase the computational burden. UIS on the other hand used the same number of samples but adaptively adjusted the proposal with the Gaussian approximation of the real posterior. This approach resulted in significantly larger R_{eff} compared to IS, which is more desirable.

Figure 4-14 shows the absolute error expectation and the error standard deviation for the saturation and pressure fields at the end of year five and after five calibration cycles for UKF, IS and UIS. In the IS scenario, the uncertainty in both fields were almost zero and the error expectation for both fields were noticeably higher than UKF and UIS, because of having one sample carrying most of the importance weight. The uncertainty in both fields were higher in UKF than in UIS. It is worth noting that while sampling-based solutions tend to underestimate the uncertainty, UKF results do not necessarily represent the true uncertainty either, because of the model linearization and the Gaussian assumptions.

The accuracy of UKF is expected to deteriorate more compared to that of UIS when the distributions become non-Gaussian. The performance of UIS in the case of non-Gaussian densities will not be affected as much as UKF, since no assumptions is made on the type of the distributions and UKF stage of UIS generates the proposal not the posterior.



(a)



(b)

Figure 4-13: (a) RMSE and (b) R_{eff} for scenarios 1 (UKF), 2 (IS with the prior used as proposal) and 5 (Base case for UIS). Data point connections and the small offset between the data points of the same cycle are for clarity. UIS demonstrate smaller RMSE compared to UKF and IS and has significantly larger ratio of effective samples compared to IS. From: (Sarkarfarshi & Gracie, 2015).

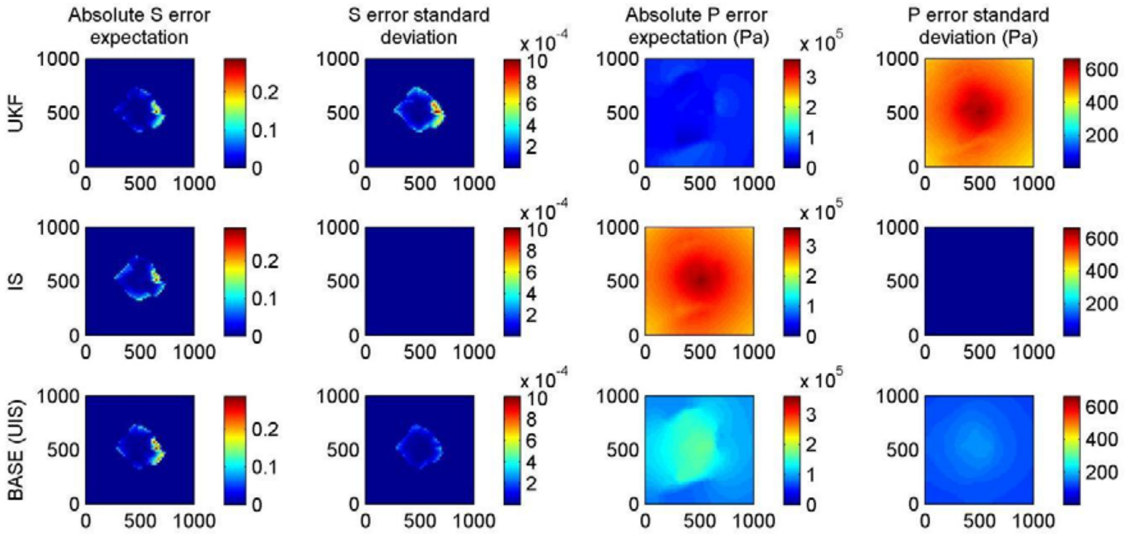
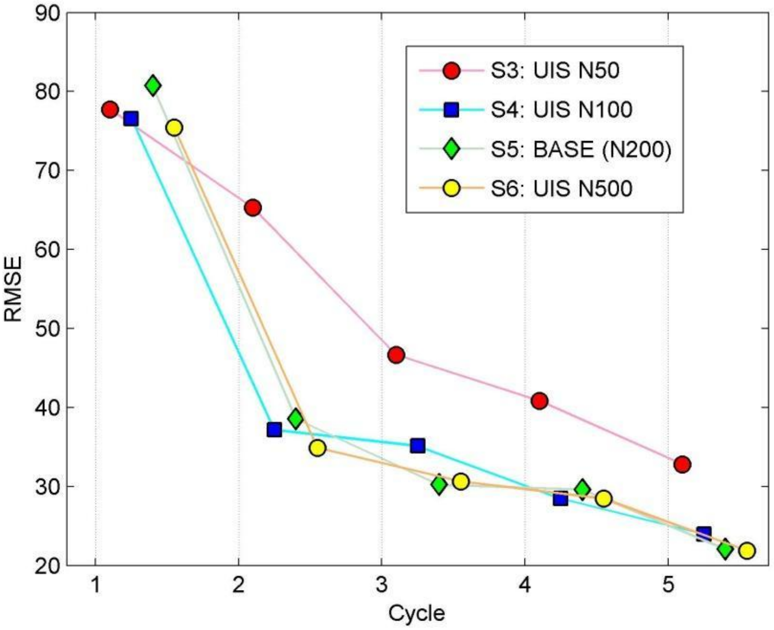


Figure 4-14: Absolute error expectation and standard deviation of saturation (S) and pressure (P) fields at the end of year 5 and after 5 calibration cycles for UKF, IS and BASE (UIS) scenarios. From: (Sarkarfarshi & Gracie, 2015).

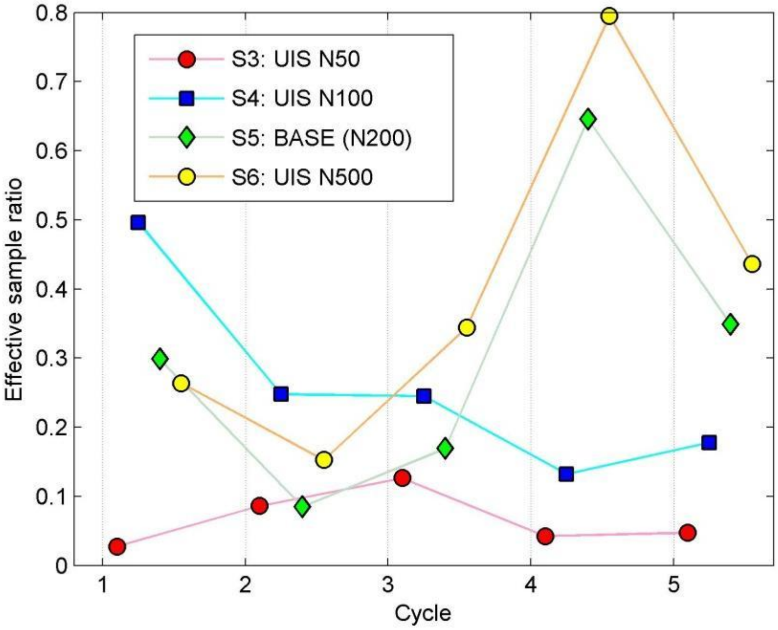
4.6.2. Effect of the Number of Samples on UIS

Figure 4-15 illustrates the RMSE and R_{eff} for UIS with 50, 100, 200 and 500 samples (scenarios 3, 4, 5 and 6), extracted from Table 4-3. The general trend of the RMSE in Figure 4-15a shows that increasing the number of samples increases the accuracy of UIS, which was not unexpected in a Monte-Carlo based method. However, the difference between scenarios deteriorates when the number of samples increases, e.g., the difference between RMSE of N50 and N100 scenarios is more significant compared to the difference between RMSE of N100, N200 and N500 scenarios. Taking into account both accuracy and computational efficiency, N100 scenario maintained the best balance among rest of the scenarios.

According to Figure 4-15b, the effective sample ratio increased with the number of samples. This is because the weights are more evenly distributed among the samples when more samples are drawn from the posterior distribution. As a result, the variance of the weights decreased and R_{eff} increased with N_s.



(a)



(b)

Figure 4-15: (a) RMSE and (b) R_{eff} for UIS with 50, 100, 200 and 500 samples. Data point connections and the small offset between the data points of the same cycle are for clarity. Increasing the number of samples decreased RMSE in general. The difference between the case with 200 samples and 500 samples was not significant. Ratio of effective samples increased with the number of samples in general. From: (Sarkarfarshi & Gracie, 2015).

Figure 4-16 shows the absolute error expectation and the error standard deviation of both fields at the end of year five and after five calibration cycles for UIS with 50 to 500 samples. The uncertainty of both the pressure and the saturation fields are noticeably underestimated and the error expectations are noticeably higher in the N50 scenario compared to other scenarios, due to a very small sample size. Scenarios with more samples are expected to more accurately represent the true uncertainty in the pressure and saturation fields. Moreover, it can be seen that the error expectation is generally decreasing in both fields with increasing the sample size.

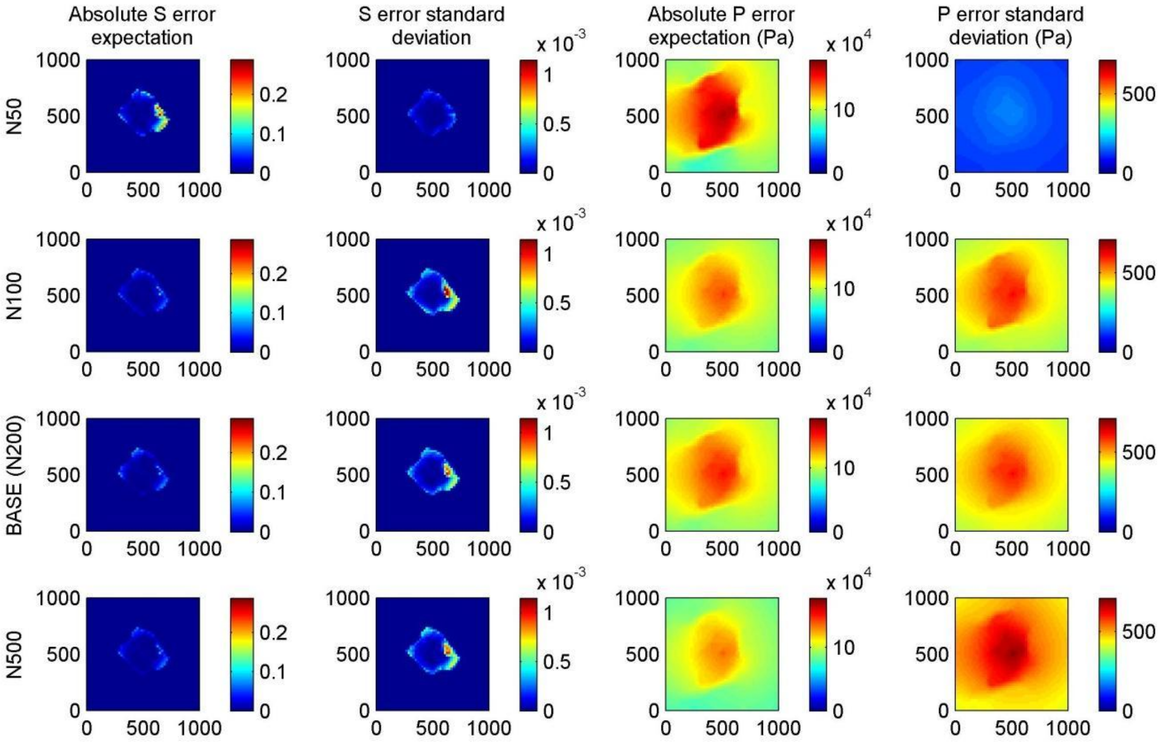


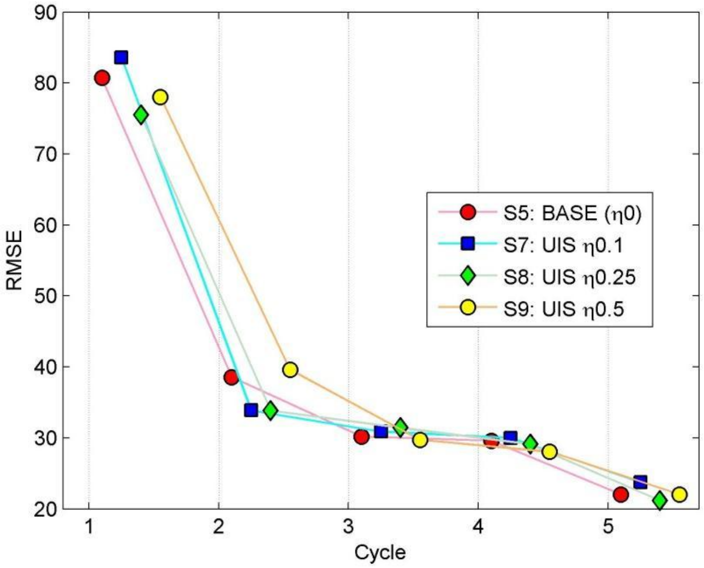
Figure 4-16: Absolute error expectation and standard deviation of saturation (S) and pressure (P) fields at the end of year 5 and after 5 calibration cycles for UIS scenarios with 50, 100, 200 and 500 samples. From: (Sarkarfarshi & Gracie, 2015).

4.6.3. Effect of the Defensive Mixture Ratio on UIS

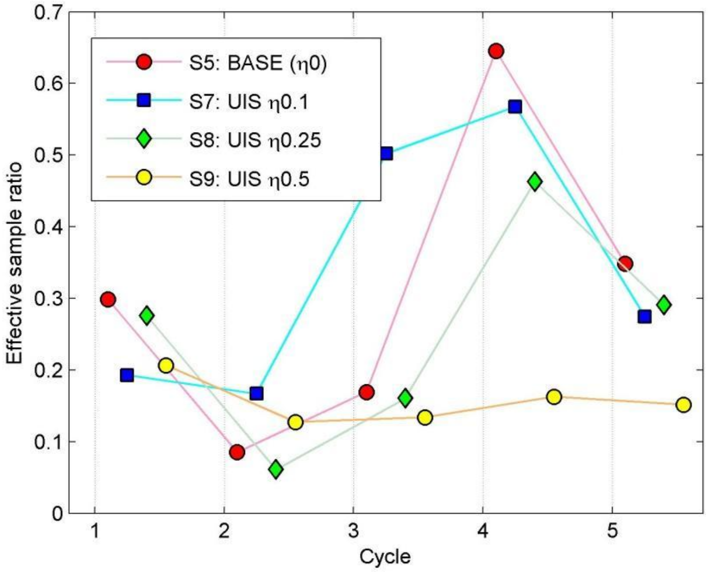
The effect of the defensive mixture ratios of zero, 0.1, 0.25 and 0.5 (scenarios 5, 7, 8 and 9) on the RMSE and R_{eff} of UIS are plotted in Figure 4-16, extracted from Table 4-3. It can be seen that the $\eta 0.1$ and $\eta 0.25$ scenarios resulted in a smaller RMSE compared to the other two scenarios. R_{eff} for $\eta 0.1$ and $\eta 0.25$ scenarios were slightly smaller than the base case; however, increasing η to 0.5 decreased R_{eff} significantly. Increasing the defensive mixture ratio is generally expected to increase the robustness of UIS at the price of a slight decrease in R_{eff}, provided that the proposal is close to the real posterior. When this is not the case, e.g., when the system model is poor or the initial prior deviates from the posterior significantly, the defensive distribution helps search a wider range of the parameter space and decrease the chance of degeneracy of the UIS.

Figure 4-18 shows the absolute error expectation and the error standard deviation of the saturation and the pressure fields at the end of year five and after five calibration cycles for UIS with zero to 0.5 defensive mixture ratios. In general, the error expectations for both the saturation and the pressure fields decreased with increasing the defensive mixture ratio.

According to current results, we recommend using a defensive mixture ratio between $\eta = 0.1$ and $\eta = 0.25$ with a uniform defensive distribution. While larger defensive mixture ratios might result in slightly smaller errors in some cases (e.g., when the prior is very poorly chosen), they tend to decrease R_{eff} noticeably.



(a)



(b)

Figure 4-17: (a) RMSE and (b) R_{eff} for UIS with defensive mixture ratios of zero, 0.1, 0.25 and 0.50, respectively. Data point connections and the small offset between the data points of the same cycle are for clarity. Increasing the defensive ratio up to 0.25 slightly decreased the ratio of the effective samples, but the decline for $\eta = 0.5$ is more significant. The RMSE for $\eta = 0.25$ presented the best overall RMSE. From: (Sarkarfarshi & Gracie, 2015).

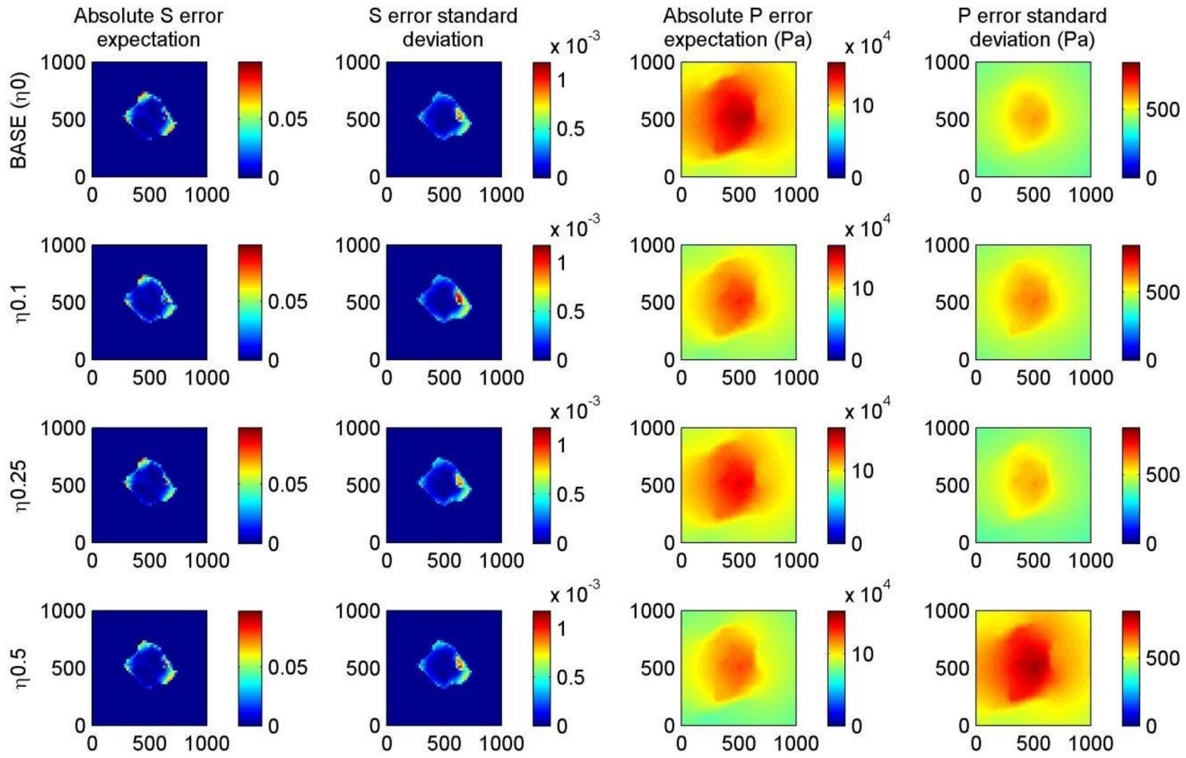


Figure 4-18: Absolute error expectation and standard deviation of saturation (S) and pressure (P) fields at the end of year 5 and after 5 calibration cycles for UIS scenarios with defensive mixture ratios of zero, 0.1, 0.25 and 0.5, respectively. All defensive mixture densities were uniform bounded between 0.01 and 0.5. From: (Sarkarfarshi & Gracie, 2015).

4.6.4. Computational Cost

A factor that should be considered when comparing the algorithm accuracies is the computational cost. The most computationally costly stage of all the algorithms discussed above was the model simulation stage. The computational cost of optimal size UIS (N100) was almost half of the computational cost of IS (109 model simulation per cycle versus 200 model simulation per cycle), while the differences in accuracies were significant. The difference between the computational cost of UKF and the computational cost of optimal size UIS (N100), however, was significant (109 model simulations per cycle versus 9 model simulations per cycle). UKF, and in general, KF-based algorithms, are computationally more efficient than sampling-based algorithms such as IS and UIS. The reason for this computational efficiency is the linearity and Gaussian assumptions in the KF-based algorithms, which makes them computationally less costly. Sampling-based algorithms on the other hand require more samples in order to capture the true posterior distribution more accurately without these assumptions. Thus, the difference between the computational costs of N100 UIS and UKF was not un-expected. The accuracy of the UIS, however, was notably better than UKF, even though the prior distribution and likelihood were Gaussian in this case study.

The computational demand of calibration of CS models can still be considerable for higher resolution models and high-dimensional parameter spaces. Thus, using only a computationally efficient calibration method such as UIS and calibrating smaller number of parameters based on a sensitivity analysis might be insufficient when computational resources are limited. As discussed above, model simulation was the most computationally demanding step during the calibration of the CS case study presented in this chapter. The adaptive Response Surface Method (RSM) introduced in the next chapter (Chapter 5) addresses this problem by replacing the computationally costly *full model* with a computationally efficient, yet accurate, *meta-model*.

4.7. Concluding Remarks

The Bayesian Unscented transformation Importance Sampling (UIS) method for mitigating parameter uncertainty was presented. The methodology revises the joint prior probability distribution of uncertain parameters utilizing noisy monitoring data.

UIS was applied to two analytical case studies and a synthetic CO_2 injection case study in order to be benchmarked against Importance Sampling (IS) with a static proposal and the measurement update stage of an Unscented Kalman Filter (UKF). The CO_2 injection case study was also used to investigate the impact of the number of samples and the defensive mixture ratio on the performance of UIS. Nine scenarios were designed for these comparisons. In the linear analytical case study, UIS performed very close to the exact solution. In the nonlinear analytical case study and CO_2 injection case study, UIS outperformed IS and UKF. It was demonstrated that increasing the number of samples enhances UIS performance, and using a moderate defensive mixture ratio increases UIS performance slightly with the price of slight decrease in the efficient number of the samples. It was suggested that the defensive mixture density becomes more significant when the prior distribution is chosen more poorly. In the case study, UIS performed best when a defensive mixture ratio between $\eta = 0.1$ and $\eta = 0.25$ was used, coupled with a uniform defensive mixture distribution.

The benefits of UIS can be summarized as:

- Bayesian calibration: UIS utilizes monitoring data to periodically revise the parameter probability distributions and reduces the parameter uncertainty. Moreover, the revised parameter distributions demonstrate which areas of a geosequestration site are poorly understood and require heavier monitoring. Thus, it can also contribute to an adaptive monitoring program.
- Model flexibility: no assumptions were made on the system model. Thus, UIS can be used with linear/nonlinear/analytical/semi-analytical/numerical system models.
- Distribution flexibility: no assumptions were made on the type of probability distributions (prior, posterior, and likelihood). Thus, various continuous, discrete, formal, and empirical probability distributions can be used as the prior or likelihood.
- Accuracy and computational efficiency: It was shown that UIS can be more accurate than IS with a static proposal distribution. As a result, to obtain a certain level of accuracy, UIS is likely to require a considerably smaller number of samples compared to IS with a static proposal. It was also shown that UIS is more accurate than the measurement update stage of UKF, even with a Gaussian prior and likelihood.

Current work is subject to limitations that can be addressed in future research, namely:

- Current formulation does not account for systematic error autocorrelation and it is possible to revisit the formulation in future work to include error autocorrelation using the equations describe earlier in Section 2.4.3.
- As described earlier, the current algorithm could underperform for multi-modal posterior distributions. In those cases, a bank of UKFs can replace the single UKF

for generating the proposal distribution. However, the added computational demand for such cases should be also taken into account.

The UIS provides an efficient tool to calibrate CS models by sampling from the important areas of the parameter space, i.e., areas with a large posterior density. This process reduces the demand for a larger sample count compared to the requires sample count of conventional Bayesian sampling-based algorithms. Depending on the problem characteristics, the important areas of the parameter space can have substantial overlaps. Such overlaps result in redundant model simulations in the same areas of the parameter space. These redundant simulations can be avoided, by using the computationally demanding system model only in a subset of the cycles and interpolating the model output in rest of the cycles to further reduce the computational cost of the calibration process. Thus, an adaptive Response Surface Method (RSM) will be introduced in the next chapter (Chapter 5) to reduce the computational cost of sampling-based calibration algorithms such as UIS by adaptively designing, updating and using an accurate and computationally efficient meta-model during the calibration process.

Chapter 5.

An Adaptive Response Surface Method for Continuous Bayesian Model Calibration

This chapter is prepared in the format of an article to be submitted to *Stochastic Environmental Research and Risk Assessment* journal, with minor changes, for publication. I am the main contributor to all sections of this chapter.

Summary

Chapter 4 presented an accurate and computationally efficient tool for calibrating CS models. The Unscented transformation Importance Sampling (UIS) method presented in Chapter 4 promotes sampling from more important areas of the parameter space in a sampling-based calibration process to reduce the demand for a larger number of model simulations during the calibration process. Combining an efficient calibration method such as UIS and a reduced parameter space based on the sensitivity analysis presented in Chapter 3, the number of required model runs during the calibration process of a CS system can be considerably reduced compared to using conventional calibration approaches. Reducing the computational cost of each model run, on the other hand, can further reduce the computational cost of calibrating CS models. Thus, this chapter focuses on the third objective of the thesis which is the reduction of the computational cost of running the system model thereby reducing the trade-off between accuracy and computational efficiency of the calibration.

In this chapter, the Bayesian Adaptive Response Surface Method (BARSM) which is an adaptive Response Surface Method (RSM) is developed to mitigate the computational cost of sampling-based and continuous calibration of CS models. It is demonstrated that the BARSM has a negligible effect on the accuracy of the calibration while providing a significant increase in efficiency. In the BARSM, a meta-model replaces the

computationally costly full model during the majority of the calibration cycles. In the rest of the cycles, the full model is used and samples of these cycles are utilized for adaptively updating the meta-model. We integrated the BARSM with the UIS calibration algorithm which was presented in Chapter 4. Then, a synthesized case of supercritical CO_2 injection in a saline aquifer is used to assess the performance of the BARSM and to compare it with a classical RSM approach and the UIS without RSM integration. It is demonstrated that the error of the meta-model fitted using the BARSM, when samples are drawn from the posterior parameter distribution, is negligible and considerably smaller than the monitoring error. We also showed that UIS, the BARSM-integrated UIS and the classical RSM-integrated UIS generate a Pareto front with computational efficiency and accuracy as the objective functions.

5.1. Introduction

As discussed in the previous chapters, model calibration for the purpose of uncertainty mitigation is a critical part of CS projects. The nonlinear nature of numerical CS system models is problematic when utilizing fast analytical calibration methods and so, sampling based algorithms are attractive solutions for calibration of nonlinear models. The computational burden of numerical model simulation is a common obstacle among the efforts for the calibration of CS models using sampling-based algorithms (Johnson & White, 2012; Espinet & Shoemaker, 2013; Tavakoli *et al.*, 2013; Sarkarfarshi & Gracie, 2013; Sarkarfarshi & Gracie, 2015). The computational cost of repetitive model simulation in the sampling-based algorithms can be reduced by using a smaller number of parameters, by using computationally efficient calibration algorithms that require a small number of model simulation, and by using a computationally efficient system model.

As described in Chapter 4, computationally efficient algorithms such as Unscented transformation Importance Sampling (UIS) are capable of accurately obtaining the posterior distribution of uncertain model parameters with a smaller number of model simulations than less accurate and computationally more costly algorithms such as ordinary Importance Sampling (IS). Nevertheless, the computational demand of calibration of CS models is still considerable for high resolution models and high-dimensional parameter spaces.

Few simplified analytical and semi-analytical models for supercritical CO_2 injection into saline aquifers are developed with the purpose of reducing the computational burden for model simulation and model calibration. For example, see: (Nordbotton *et al*., 2005) and (Malekzadeh & Dusseault, 2013). These models provide fast and effective tools for sensitivity analysis, uncertainty analysis and risk assessment in the early stages of a CS project, e.g., during site selection (Sarkarfarshi *et al*., 2014). However, the simplifying assumptions such as incompressibility of the fluids, axisymmetric geometry, homogeneity of the aquifer, the presence of a single injection well and the absence of potential leakage sources make them suboptimal for long-term calibration and forecast of a CS site behaviour.

Another common solution to reduce the computational cost of model simulation is to replace the computationally demanding “full model” with a faster “meta-model”, which approximates the true response of the full model. The meta-models are known by various names, such as surrogate models, Response Surface Method (RSM) models, models of the model, and emulators (Kleijnen, 1987; Jin *et al*., 2001; Razavi *et al*., 2012-a). In the RSM, the full model response is simulated at a finite number of points in the parameter space that are chosen by a Design of Experiment (DOE) (Barton, 1994). Then the meta-model is fitted or “trained” based on the model responses at these data points, and using an “objective” or “cost” function and a “training algorithm”. Meta-model fitting is essentially finding the unknown parameters of a function (i.e., the meta-model) to maximize an objective function, or equivalently, minimize a cost function. This approach aims for a meta-model that is computationally more efficient compared to the full model and is differentiable/integrable in some cases so that the problem becomes analytically tractable (Oladyshkin *et al*., 2011-a; Razavi *et al*., 2012-a). Some well-known approaches to meta-model fitting are Polynomial Regression (PR) (Jin *et al*., 2003), Artificial Neural Networks (ANN) (Smith, 1993), Adaptive Neuro Fuzzy Inference System (ANFIS) (Abraham, 2005), Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991), Radial basis functions (Buhmann, 2003) and Kriging (Stein, 1999). A comprehensive review of metamodelling approaches in computationally intensive environmental problems can be found in (Razavi *et al*., 2012-b).

Simple meta-models can be efficiently trained and used for uncertainty analysis (Isukapalli *et al*., 2000), sensitivity analysis (Oladyshkin *et al*., 2012), risk analysis (Oladyshkin *et al*., 2011-a) and uncertainty mitigation (Balakrishnan *et al*., 2003). The simple meta-models, however, are not optimal for calibrating complex, highly nonlinear and high-dimensional numerical CS models since they are not able to capture sever nonlinearities and complicated correlations in the whole parameter space. More complex meta-models on the other hand, are computationally intensive to train and do not necessarily improve the calibration results when the computational budget is limited (Razavi *et al*., 2012-a). Consequently, the choice between a simple and complex meta-model is usually a trade-off between accuracy and computational efficiency.

Building and continuously revising a simple meta-model with higher accuracy in a subset of the parameter space, which is used the most during the calibration process, can result in an accurate and computationally efficient solution for Bayesian calibration of complicated systems such CS. Based on this idea, we introduce the Bayesian Adaptive Response Surface Method (BARSM) in this chapter as an adaptive approach for efficiently utilizing the RSM to reduce the simulation time of ensemble-based and continuous Bayesian calibration algorithms with non-stationary ensemble members (samples). By continuous, we mean calibration that is repeated when new observations (i.e., monitoring data) are obtained. Similar to Chapter 4, every repetition of calibration is called a “cycle” here. The BARSM is the third objective among the three complementary objectives of this thesis that seek to achieve efficient and computationally efficient calibration of CS systems.

To avoid confusion, it should be noted that the BARSM is different from the Adaptive Response Surface Method (ARSM) (Wang *et al*., 2002; Wang, 2003). The ARSM is aimed for iterative global optimization problems and revises the meta-model during optimization iterations, while BARSM is aimed for continuous (but not necessarily iterative) Bayesian calibration and does not revise the meta-model during a calibration cycle.

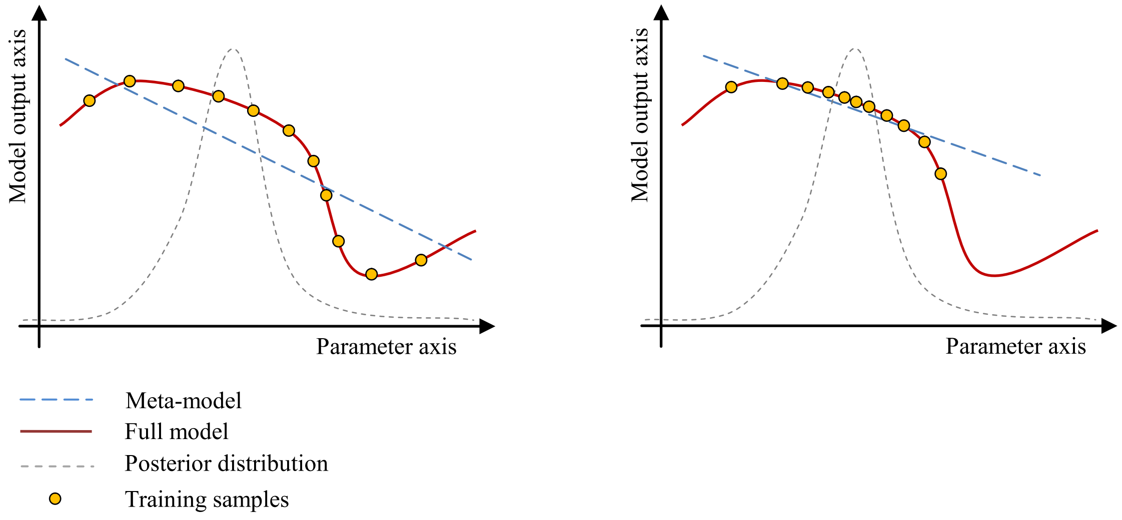
In sampling-based calibration algorithms, each sample represents a combination of model parameters, and the model is simulated for each sample. The accuracy of the meta-model is more important in the areas of the parameter space where the samples are most drawn from. In the BARSM, we take advantage of the fact that in sampling-based calibration algorithms, the samples are likely to be drawn from areas with larger posterior density, allowing the training algorithm to adaptively adjust the training samples during calibration and to increase the accuracy of the meta-model where it is most simulated. Figure 5-1 illustrates this idea by comparing the linear meta-model fit to (a) uniformly chosen training samples (which we call a "classical" or "traditional" approach) and (b) samples drawn from the more important areas of the parameter space. It can be seen that while the meta-model in Figure 5-1 (a) approximates the full model better in the whole parameter space, the meta-model in Figure 5-1 (b) is more accurate in areas with higher posterior density.

For this purpose, we divide the calibration cycles into two groups:

A. *Full cycles*, i.e., cycles which use the full model during the calibration. Full cycles are no different than full calibration algorithm, except for simulating the full model for a longer time frame (up to the final calibration time). This longer simulation is used to calibrate the meta-model.

B. *Meta-cycles*, i.e., cycles utilizing a meta-model during the calibration, to enhance the computational efficiency. The meta-cycles are computationally faster than the full cycles, assuming that the process of training and simulating the meta-model is faster than using the full model.

In Section 5.2, the BARSM is described and formulated. Next, in Section 5.3, we integrate the BARSM with the UIS. A synthetic case study in a CS context is described in Section 5.4 and is used to study the performance of the BARSM against a traditional RSM integration method and the full UIS in terms of accuracy and computational efficiency. Finally, conclusions are presented in Section 5.5.



(a) (b)

Figure 5-1: Schematic of fitting a meta-model using (a) linear regression on uniformly chosen samples from the whole parameter space and (b) linear regression on random samples from areas of the parameter space with higher probability density. It can be seen that the meta-model in (b) is more accurate than (a) in areas of the parameter space with higher probability density. These areas are more likely to be sampled from during the calibration process. Thus, the meta-model in (b) is more accurate than (a) for the purpose of model calibration.

5.2. Methodology

Assume that measurements of the true behaviour of a system taking place at discrete times t_1, t_2, ... , t_{N_t} where t_n denotes the time elapsed at the n^{th} cycle. Similar to the previous chapters, we denote the true system behaviour at time t_n by $\boldsymbol{d}_{tru}^{n}$, measurements at time t_n by $\boldsymbol{d}_{obs}^{n}$ and system model output at time t_n by $\boldsymbol{y}^{n}(\boldsymbol{m})$ where $\boldsymbol{y}^{n}(\boldsymbol{m}) = [\boldsymbol{g}(\boldsymbol{x}_1, t_n, \boldsymbol{m}), \boldsymbol{g}(\boldsymbol{x}_2, t_n, \boldsymbol{m}), \dots, \boldsymbol{g}(\boldsymbol{x}_{N_x}, t_n, \boldsymbol{m})]$ and $\boldsymbol{g}(\boldsymbol{x}, t_n, \boldsymbol{m})$ is the system model with parameters $\boldsymbol{m} = [m_1, m_2, \dots, m_{N_m}]$. We also assume the discrepancy between $\boldsymbol{d}_{obs}^{n}$ and $\boldsymbol{y}^{n}(\boldsymbol{m})$ is a vector of random variables, independent from time, location and $\boldsymbol{m}$, similar to Chapter 4. Thus,

$$\boldsymbol{d}_{obs}^{n} = \boldsymbol{y}^{n}(\boldsymbol{m}) + \boldsymbol{\varepsilon}_{tot} \qquad \boldsymbol{\varepsilon}_{tot} \sim \pi_{tot}(\boldsymbol{\varepsilon}_{tot}) \tag{5.1}$$

where $\boldsymbol{\varepsilon}_{tot}$ is the combined model and monitoring error and $\pi_{tot}(\boldsymbol{\varepsilon}_{tot})$ is the probability distribution of $\boldsymbol{\varepsilon}_{tot}$. Thus, as described in Chapter 2 and according to Bayes' formula, if the prior distribution of model parameters is denoted by $\pi^0(\boldsymbol{m})$, the posterior probability distribution of model parameters given all monitoring data up to time t_n can be obtained by

$$\pi^n(\boldsymbol{m}) = \frac{\pi^0(\boldsymbol{m}) \prod_{l=1}^{n} \pi_{tot}\left(\boldsymbol{d}_{obs}^l - \boldsymbol{y}^l(\boldsymbol{m})\right)}{\int_{\Omega_m} \pi^0(\boldsymbol{m}) \prod_{l=1}^{n} \pi_{tot}\left(\boldsymbol{d}_{obs}^l - \boldsymbol{y}^l(\boldsymbol{m})\right)} \tag{5.2}$$

where $\pi^n(\boldsymbol{m}) = \pi(\boldsymbol{m}|\boldsymbol{d}_{obs}^{1\dots n})$. At time t_{n+1}, new monitoring data is obtained and Bayes' formula can be used to obtain $\pi^{n+1}(\boldsymbol{m})$. Similar to the previous chapter, we call each of these cycles a "calibration cycle" here.

Assume a sampling-based Bayesian calibration algorithm, such as UIS or MCMC, which we call the "full calibration algorithm" here, is used for solving equation (5.2) in each calibration cycle. Assume that at each cycle n, the full calibration algorithm draws N_s samples from the parameter space and simulates the full-model for each sample from time zero to time t_n.

As shown in the flowchart of BARSM in Figure 5-2, calibration cycles are divided into full cycles and meta-cycles. The first cycle is always a full cycle and each of the following cycles could be either a full cycle or a meta-cycle. In an arbitrary full cycle n, the calibration algorithm is run using the full model. However, instead of running the full model up to t_n, we run the model up to the final calibration time, t_{N_t} and add the samples and corresponding full model outputs to a "sample pool". Next, in another arbitrary meta-cycle n, the sample pool data is used to train a meta-model and the meta-model replaces the full model in that cycle. Thus, full model run is avoided in the meta-cycles. If the computational burden of fitting and running the meta-model is significantly lower than the full model, the computational cost of the meta-cycles would be negligible compared to the full cycles.

After a few meta-cycles, the posterior distribution might change noticeably or get more concentrated compared to the posterior distributions in the previous cycles. Consequently, adding more samples from the latest posterior distribution to the sample pool makes the meta-model more accurate in the most recent important areas of the parameter space. So, BARSM uses another full cycle after a certain number of meta-cycles to adaptively improve the accuracy of the meta-model.

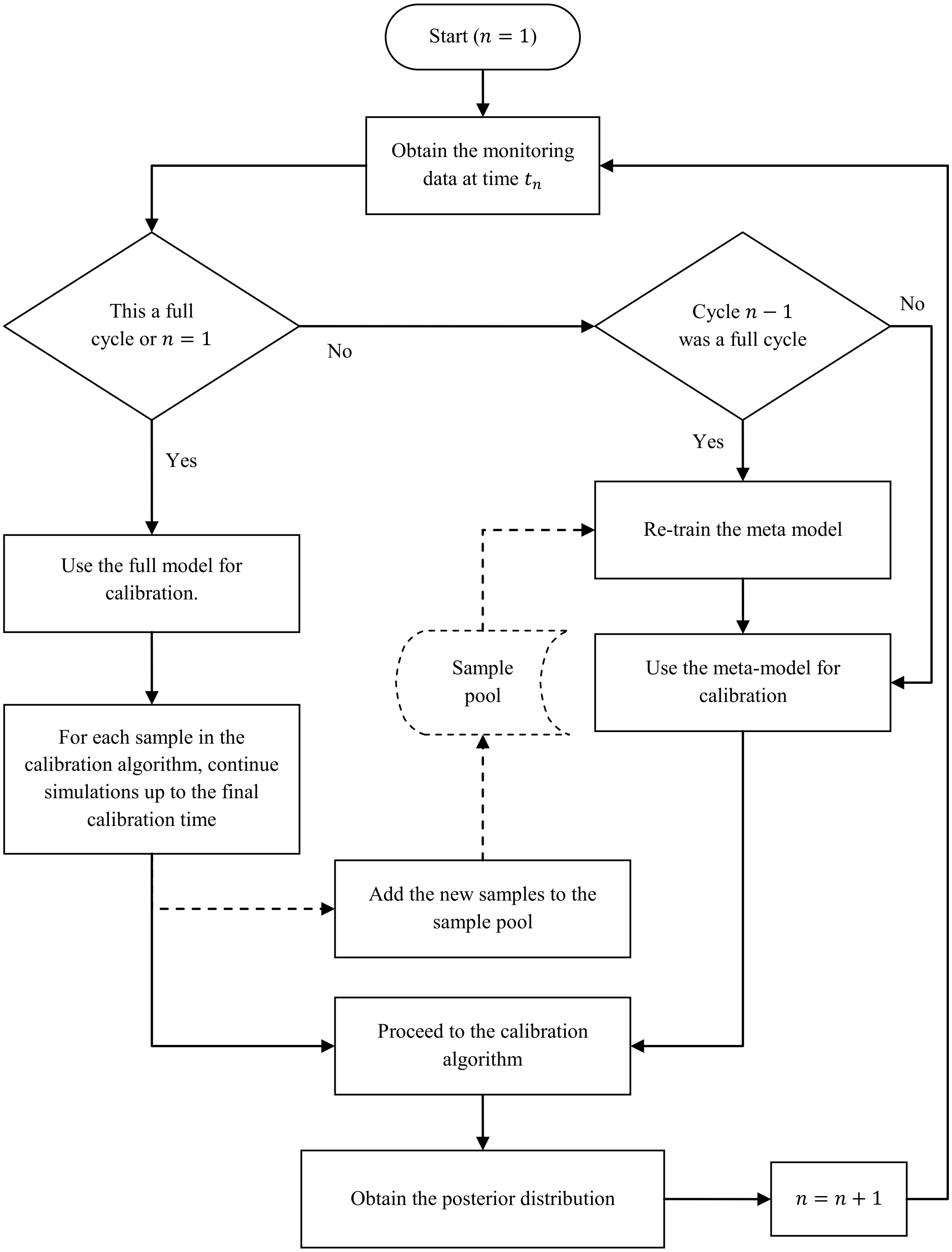


Figure 5-2: Flowchart of the BARSM for mitigating the computational burden of Bayesian calibration algorithms for computationally expensive models.

For simplicity of the calculations, assume that $t_n = n$ years. Also assume that one full cycle is used after every n_m meta-cycles and the total number of calibration cycles, N_t, is a multiple of $n_m + 1$. Also assume that the computational cost of a full model run increases proportionally with the time frame the model is being simulated (e.g., in transient numerical models) and each calibration cycle requires N_s model simulations. After N_t cycles, the total years of full model run is equal to $N_s N_t\,(N_t + 1)/2$ when the full calibration algorithm is used and is equal to $N_s(n_m + 1)$ when the BARSM is used. Thus, the computational cost of calibration of the model in BARSM is reduced compared to the full calibration algorithm, if $N_t(N_t + 1)/(2n_m + 2) > 1$, or equivalently, if

$$n_m > \frac{N_t - 1}{N_t + 1} \tag{5.3}$$

Since $(N_t - 1)/(N_t + 1)$ is between zero and one, n_m should be equal to or greater than one. Thus, the BARSM mitigates the computational cost of full model run in sampling-based Bayesian calibration algorithms if at least one meta-cycle exists after each full cycle. The ratio of reduction in computational cost of full model run is given by

$$R^{BARSM} = \frac{(n_m + 1)(N_t + 1)}{2N_t} \tag{5.4}$$

Figure 5-3 plots R^{BARSM} for $N_t = 1$ to 30 and $n_m = 1$ to 29. It can be seen that the computational cost of full model run can be reduced significantly using the BARSM. n_m is a parameter of the BARSM and should be chosen on a problem-specific basis.

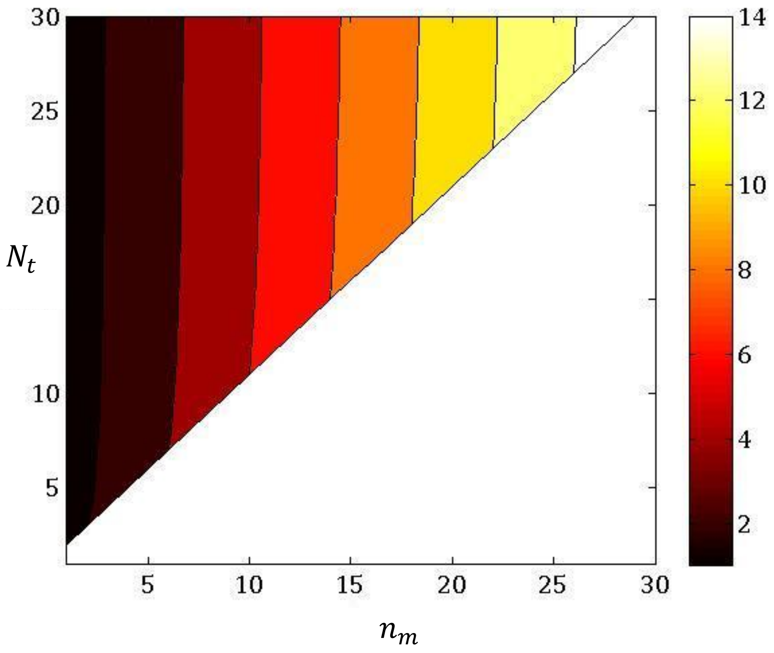


Figure 5-3: The ratio of reduction in computational cost of full model run in the BARSM, R^{BARSM}, for N_t cycles when n_m meta-cycle are used after each full cycle.

5.3. Integration of BARSM and UIS

The BARSM will be inefficient if the posterior sample collapses, i.e., one of the samples gains almost all of the posterior importance weight while the posterior importance weight of all other samples are close to zero. The BARSM will be also inefficient if the samples are most drawn from areas of the parameter space with smaller posterior distribution. Since in the UIS the samples are drawn from a continuous proposal distribution close to the posterior distribution, UIS is an effective method for avoiding collapse and ensuring that samples are most drawn from areas of the parameter space which have a greater posterior distribution. Thus, we integrate BARSM with the UIS in this section to demonstrate how it can be integrated with a sampling-based Bayesian calibration algorithm.

Recalling from the previous chapter, the propagated Sigma points in the UKF step of the UIS at time t_n are obtained by

$$\boldsymbol{\Upsilon}_i = \left[\boldsymbol{g}(\boldsymbol{x}_1, t_n, \boldsymbol{\chi}_i), \boldsymbol{g}(\boldsymbol{x}_2, t_n, \boldsymbol{\chi}_i), \dots, \boldsymbol{g}(\boldsymbol{x}_{N_x}, t_n, \boldsymbol{\chi}_i)\right] \qquad i = 0,1, \dots, 2N_m \tag{5.5}$$

where $\boldsymbol{\chi}_i$ is the i^{th} Sigma point and $\boldsymbol{\Upsilon}_i$ is the propagated Sigma. In addition, the importance weight of a sample $\widehat{\boldsymbol{m}}_i$ drawn in the IS step of the UIS and from an approximate Gaussian posterior distribution at time t_n is obtained by

$$W_i = {c_W}^{-1}\frac{\pi^0(\widehat{\boldsymbol{m}}_i)}{q^n(\widehat{\boldsymbol{m}}_i)}\prod_{l=1}^{n}\pi_{tot}\left(\boldsymbol{d}_{obs}^{l} - \boldsymbol{y}^{l}(\widehat{\boldsymbol{m}}_i)\right) \qquad i = 1 \dots N_s \tag{5.6}$$

where W_i is the normalized importance weight, $c_W = \sum_{i=1}^{N_s} W_i$ is a normalizing constant and $q^n(\boldsymbol{m})$ is the proposal distribution of the IS step at time t_n. Equations (5.5) and (5.6) are the only UIS equations which require running the system model.

At an arbitrary full cycle of the BARSM, typical UIS formulation is used with the full system model. However, the simulation of the model for each of the samples is continued until t_{N_t} and the samples and corresponding full model output are added to the sample pool. The sample pool is basically a set of data vectors $\left\{[\widehat{\boldsymbol{m}}_i, \boldsymbol{y}^{N_t}(\widehat{\boldsymbol{m}}_i)]: i = 1 \dots N_{sp}\right\}$ where each $\widehat{\boldsymbol{m}}_i$ is the i^{th} sample (i.e., a parameter set) and $\boldsymbol{y}^{N_t}(\widehat{\boldsymbol{m}}_i)$ is the corresponding full model output simulated up to t_{N_t}.

At arbitrary meta-cycles n, we fit a series of meta-models denoted by $\widehat{\boldsymbol{g}}(\boldsymbol{x}, t_l, \boldsymbol{m})$ to the sample pool data where l is the cycle number between one and N_t and $\boldsymbol{x}$ is one of the N_x spatial grid points in the domain where the model output is required[4]. Thus, the total number of meta-models are $N_t N_x$. For instance, if the problem domain is a two dimensional 50×50 grid, $N_t = 4$, model output is the temperature at each grid point, measurements are taken in 25 grid points and meta-model is chosen to be linear, we fit a linear function to the sample pool data for each of the 25 measurement grid points and for each of the times t_1 to t_4. Thus, a total of 4×25=100 linear meta-models are fitted to the sample pool data.

Next, we use the meta-model $\widehat{\boldsymbol{y}}^l(\boldsymbol{m})$ instead of the full model in the UIS algorithm during the cycle n, where $\widehat{\boldsymbol{y}}^l(\boldsymbol{m}) = \left[\,\widehat{\boldsymbol{g}}(\boldsymbol{x}_1, t_l, \boldsymbol{m}), \widehat{\boldsymbol{g}}(\boldsymbol{x}_2, t_l, \boldsymbol{m}), \dots, \widehat{\boldsymbol{g}}(\boldsymbol{x}_{N_x}, t_l, \boldsymbol{m})\right]$. For instance, equation (5.6) for obtaining the importance weights in the IS step of the UIS changes to

---

[4] To use the meta-model for forecasting the system behaviour beyond t_n, the $l \leq n$ condition is released and the meta-model is trained up to the end of forecast time.

$$W_i = c^{-1}\frac{\pi^0(\widehat{\boldsymbol{m}}_i)}{q^n(\widehat{\boldsymbol{m}}_i)}\prod_{l=1}^{n}\hat{\pi}_{tot}^{l}\left(\boldsymbol{d}_{obs}^{l} - \widehat{\boldsymbol{y}}^{l}(\widehat{\boldsymbol{m}}_i)\right) \qquad i = 1 \ldots N_s \tag{5.7}$$

where $\hat{\boldsymbol{\varepsilon}}_{tot}^{l} = \boldsymbol{\varepsilon}_{tot} + \hat{\boldsymbol{\varepsilon}}_{mod}^{l}$ is the discrepancy between the measurements $\boldsymbol{d}_{obs}^{l}$ and $\widehat{\boldsymbol{y}}^{l}(\boldsymbol{m})$, $\hat{\boldsymbol{\varepsilon}}_{mod}^{l}$ is the discrepancy between the full model and the meta-model and $\hat{\pi}_{tot}^{l}(\hat{\boldsymbol{\varepsilon}}_{tot}^{l})$ is the probability distribution of $\hat{\boldsymbol{\varepsilon}}_{tot}^{l}$. The probability distribution of $\boldsymbol{\varepsilon}_{tot}$ is assumed to be known in UIS. If $\hat{\boldsymbol{\varepsilon}}_{mod}^{l}$ is negligible compared to $\boldsymbol{\varepsilon}_{tot}$, it can be neglected and $\pi_{tot}(\boldsymbol{\varepsilon}_{tot})$ can be used instead of $\hat{\pi}_{tot}^{l}(\hat{\boldsymbol{\varepsilon}}_{tot}^{l})$ in equation (5.7) and in the UKF step of the UIS. Otherwise, to obtain $\hat{\pi}_{tot}^{l}(\hat{\boldsymbol{\varepsilon}}_{tot}^{l})$, we need to find $\hat{\pi}_{mod}^{l}(\hat{\boldsymbol{\varepsilon}}_{mod}^{l})$, i.e., the probability distribution of $\hat{\boldsymbol{\varepsilon}}_{mod}^{l}$.

Since polynomial meta-models are computationally efficient to simulate and can be fitted without carrying out computationally costly iterative training algorithms, we use polynomial meta-model with least square fitting algorithm in BARSM-integrated UIS. Least square fitting for polynomial regression is explained in details in Appendix B. Since the meta-model output is not necessarily equal to the full model output at the training data points when using least square regression, we approximate $\hat{\pi}_{mod}^{l}(\hat{\boldsymbol{\varepsilon}}_{mod}^{l})$ by a probability distribution fitted on the discrepancy of the full model and the meta-model at the sample pool data. In other words, we fit a probability distribution on $\hat{\boldsymbol{\varepsilon}}_{mod}^{l} = \boldsymbol{y}^{l}(\widehat{\boldsymbol{m}}_i) - \widehat{\boldsymbol{y}}^{l}(\widehat{\boldsymbol{m}}_i)$ where $\widehat{\boldsymbol{m}}_i$ is the i^{th} parameter set in the sample pool. Now, by having $\hat{\pi}_{mod}^{l}(\hat{\boldsymbol{\varepsilon}}_{mod}^{l})$ and $\pi_{tot}(\boldsymbol{\varepsilon}_{tot})$, we can find $\hat{\pi}_{tot}^{l}(\hat{\boldsymbol{\varepsilon}}_{tot}^{l})$ using analytical methods or simulation-based methods such as MCS.

5.4. Case Study

A synthetic case study of CO_2 injection into a saline aquifer is used to study the accuracy and computational efficiency of the BARSM-integrated UIS, to compare it with a classical RSM integration and full UIS (no RSM), and investigate the effect of frequency of full cycles on the performance of the BARSM.

5.4.1. Case Description

The geometry and initial conditions of this case study are mostly similar to the case study in Section 4.5 of the previous chapter. Supercritical CO_2 is injected at a constant rate of 0.035 Mt/year for 112 month in an injection well located in the center of the domain. Porosities of the four facies are assumed the uncertain parameters with a Gaussian multivariate prior distribution. Prior mean is assumed to be [0.15,0.15,0.15,0.15] and prior covariance matrix is assumed to be a diagonal matrix with diagonal members of 0.05^2.

Monitoring is conducted every 8 months at two monitoring wells similar to Figure 4-11. One calibration cycle is performed at the end of each 8 months. Saturation monitoring error is assumed Gaussian with a zero mean and a standard deviation of 0.03, and pressure monitoring error is assumed Gaussian with a zero mean and a standard deviation of 0.5 MPa.

Schlumberger's ECLIPSE E300 reservoir simulator is used for simulating the case described above with GASWAT keyword and Peng-Robinson equation of state (Soreide & Whitson, 1992; Schlumberger, 2013), no flow boundary condition on top and bottom of the aquifer, and open boundary condition on the sides of the aquifer.

5.4.2. Calibration Scenarios

Six calibration scenarios are described in Table 5-1. All scenarios use UIS calibration algorithm with 200 samples in each cycle and no defensive mixture distribution. Scenario 1 uses original UIS algorithm without RSM integration. Scenarios 2, 3 and 4 use BARSM-integrated UIS with different numbers of full cycles, as reported in Table 5-1. Scenarios 5 and 6 use UIS with a classical RSM fitting approach, i.e., fitting a response surface at the beginning of the calibration using 200 and 500 samples from the initial prior distribution, respectively, and using this meta-model in all calibration cycles. In scenarios 2 to 6, a quadratic meta-model is used.

Table 5-1: Calibration scenarios.

Scenario Number and Name	Calibration Method	RSM Integration	Full Cycles	Number of Training samples per Full Cycle	Total Years of Full Model Simulation
1: UIS	UIS	-	1 to 14	-	14,630
2: BARSM1	UIS	BARSM	1, 6, 11	209	5,852
3: BARSM2	UIS	BARSM	1, 11	209	3,901
4: BARSM3	UIS	BARSM	1	209	1,951
5: CL1	UIS	Classical	0*	200	1,867
6: CL2	UIS	Classical	0*	500	4,667

* By cycle 0, we mean training the meta-model before initiating the calibration.

5.4.3. Results

The first metric we use for comparing the scenarios described in the previous sub-section is the accuracy of the meta-model in comparison with the full model, in the areas of the posterior distribution where the model is more likely to be simulated during the calibration. To do so, for each cycle, 200 samples are drawn from the posterior distribution of model parameters in scenario 1 which always uses the full model. Meta-models used in each cycle of scenarios 2 to 6 are simulated using samples drawn at that same cycle. The discrepancy between full model output and the meta-model output (separately for saturation and pressure) is then calculated and averaged in the whole field, as

$$E_i^n = \frac{\sum_{j=1}^{N_x}\left(g\left(\boldsymbol{x}_j, t_n, \widehat{\boldsymbol{m}}_i^{n,UIS}\right) - \hat{g}\left(\boldsymbol{x}_j, t_n, \widehat{\boldsymbol{m}}_i^{n,UIS}\right)\right)}{N_x} \tag{5.8}$$

$$\widehat{\boldsymbol{m}}_i^{n,UIS} \sim \pi^{n,UIS}(\boldsymbol{m}) \qquad i = 1\ to\ 200 \qquad n = 1\ to\ 14$$

where $\widehat{\boldsymbol{m}}_i^{n,UIS}$ is the i^{th} sample drawn from $\pi^{n,UIS}(\boldsymbol{m})$, i.e., the posterior distribution of $\boldsymbol{m}$ at cycle n of the UIS scenario, and E_i^n is the average error of the sample i at cycle n. The mean and the 95% central confidence interval[5] of the errors are then calculated for each cycle, as

$$\mu_E^n = \frac{\sum_{i=1}^{200} E_i^n}{200}$$

$$\sigma_E^n = \sqrt{\frac{\sum_{i=1}^{200}\left(E_i^n - \mu_E^n\right)^2}{200}} \tag{5.9}$$

$$CI_{E,0.975}^n \cong \mu_E^n + 1.96\sigma_E^n$$

$$CI_{E,0.025}^n \cong \mu_E^n - 1.96\sigma_E^n$$

where μ_E^n, σ_E^t, $CI_{E,0.975}^n$ and $CI_{E,0.025}^n$ are the mean, standard deviation, and lower and upper bounds of the 95% confidence interval of the meta-model error at cycle n, respectively. Figure 5-4 illustrates the mean and the 95% central confidence interval of the meta-model error, as described above, for (a) saturation and (b) pressure fields, respectively. Three observations can be made from these figures:

I. The errors of the meta-model in BARSM scenarios are clearly smaller than those of classical scenarios. Mean saturation error is close to zero for all scenarios. The 95% confidence interval is stable and below ±2% in BARSM1 and below ±3.5% in BARSM2. The 95% confidence interval is below ±4.1% in BARSM3 but it has a growing trend. 95% confidence interval of CL1 and CL2 scenarios are growing and reached ±4.9%, without a significant difference between two scenarios. Moreover, the mean pressure error is close to zero and the 95% confidence interval is stable and less than 10 kPa in BARSM1 and BARSM2 and ~25 kPa in BARSM3. In CL1 and CL2 scenarios, however, both mean and 95% confidence interval of the pressure error are continuously growing and reached 6.4 MPa and 5.36 MPa, respectively. Thus, accuracy of the meta-model in all BARSM scenarios was higher than the accuracy of the meta-model in the classical scenarios in areas of the parameter space where most samples are drawn from. In addition, increasing the frequency of the full cycles increased the accuracy of BARSM. It should also be

---

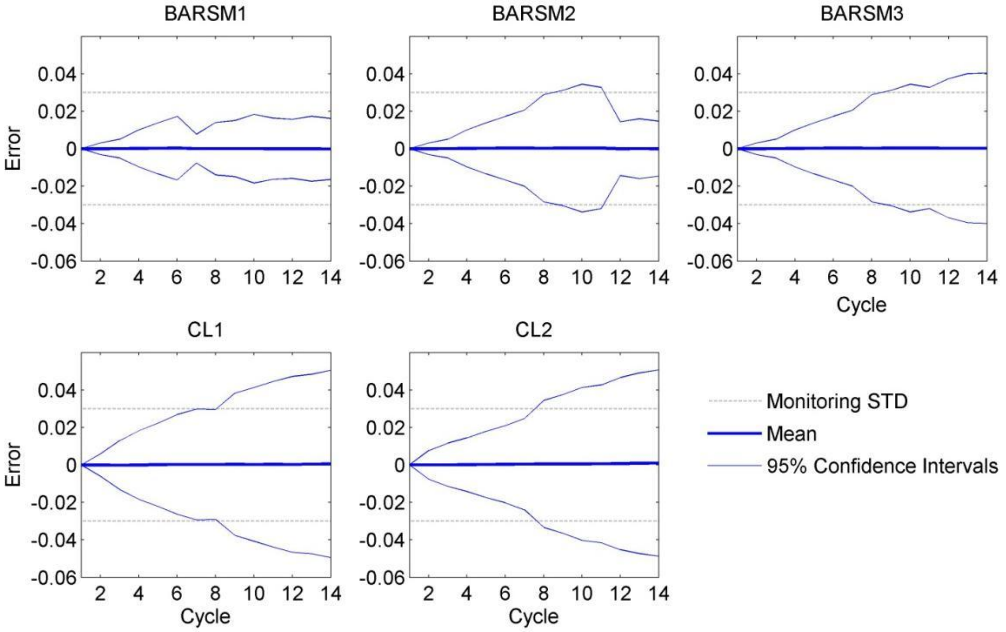
[5] Since the distribution of the errors was close to a Normal distribution, the 2.5% and 97.5% quantiles were obtained by subtracting 1.96 standard deviations from the mean and adding 1.96 standard deviations to the mean, respectively.

noted that using a larger sample pool in CL scenarios was not effective in obtaining a more accurate meta-model.

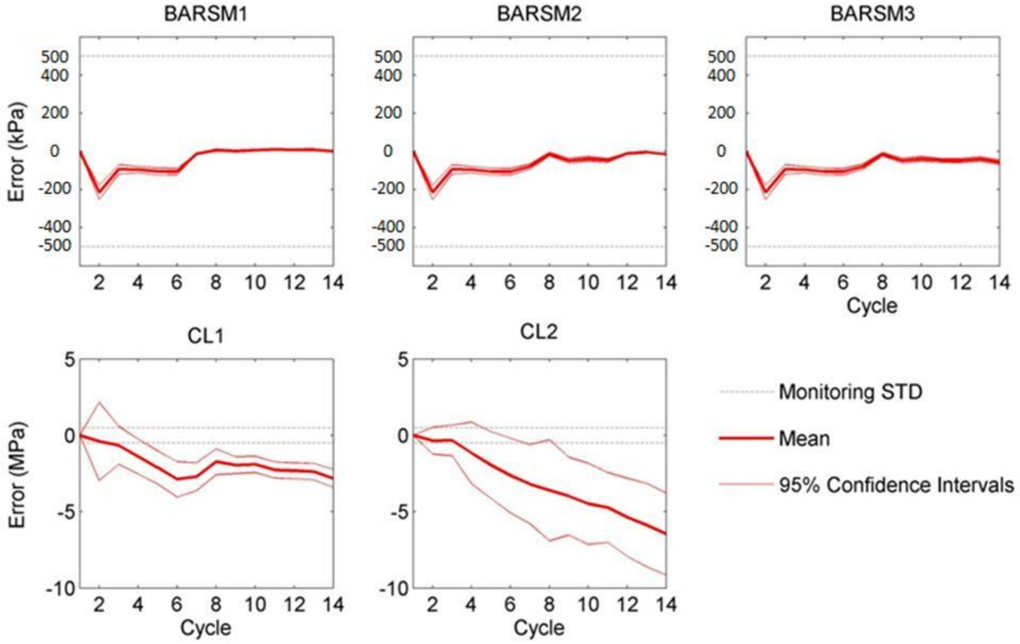
II. Full cycles in BARSM scenarios increased meta-model accuracy noticeably. This is specifically clear in Figure 5-4(a) where the 2.5% and 97.5% quantiles of BARSM1 and BARSM2 scenarios are narrowed after full cycles. Moreover, a comparison between meta-model errors in BARSM3 and classical scenarios reveals that the only full cycle in BARSM3 (which is the first cycle) increased the meta-model accuracy compared to not having any full cycles, i.e., CL1 and CL2 scenarios. Thus, adding the full cycles in BARSM was an effective approach in decreasing the error introduced by using meta-models during the calibration process. In addition, by adjusting the frequency of the full cycles, meta-model error can be confined and prevented from increasing continuously.

III. Third, the errors between the full model and the BARSM meta-models were well below the monitoring error. In CL scenarios on the other hand, the saturation errors were comparable to the monitoring error and the pressure errors were well above the monitoring error. Thus, we can conclude that the additional uncertainty introduced by using the meta-model in BARSM scenarios was negligible compared to the monitoring error, while this was not the case in CL scenarios.

A noteworthy difference between Figure 5-4(a) and Figure 5-4(b) is the magnitude of difference between the meta-model errors in saturation and pressure fields. For instance, the ratio of 95% confidence intervals of CL2 and BARSM1 scenarios is ~3 in Figure 5-4(a) while it is ~2100 in Figure 5-4(b). The reason for this notable difference is the fact that the parameters being calibrated in this case study are porosities of the facies and permeability of each facies was obtained by a logarithmic equation, as described in Section 4.5.1 of the previous chapter. Considering this equation and the fact that pressure gradient increases with the inverse of permeability, a small change in porosity which causes small changes in the saturation field can cause large changes in the pressure field. Meta-models in the BARSM scenarios were notably more accurate where they were more simulated and thus, they captured these steep changes in the pressure field well while it was not the case for CL scenarios. As a result, the difference between accuracy of the meta-models in re-generating the pressure field was more visible compared to the saturation field.



(a)



(b)

Figure 5-4: Mean and central 95% confidence interval of meta-models errors for (a) saturation and (b) pressure field, obtained using equation (5.9). Standard deviation of the monitoring error is plotted by dashed grey lines. It can be seen that meta-models in BARSM scenarios were more accurate than CL scenarios, accuracy of meta-model increased with frequency of full cycles and meta-model errors in BARSM scenarios were smaller than the monitoring error.

The second metric we use to compare the calibration scenarios is the mean and standard deviation of the absolute error in the posterior forecast of the pressure and saturation fields at the end of calibration, as illustrated in Figure 5-5. It can be seen that the posterior error fields of the BARSM scenarios are close to the UIS scenario. Thus, stochastic simulation of pressure and saturation fields after calibration using the BARSM-integrated UIS and original UIS were close. The CL scenarios, on the other hand, had larger errors compared to the BARSM scenarios and had noticeably larger errors compared to the UIS scenario.

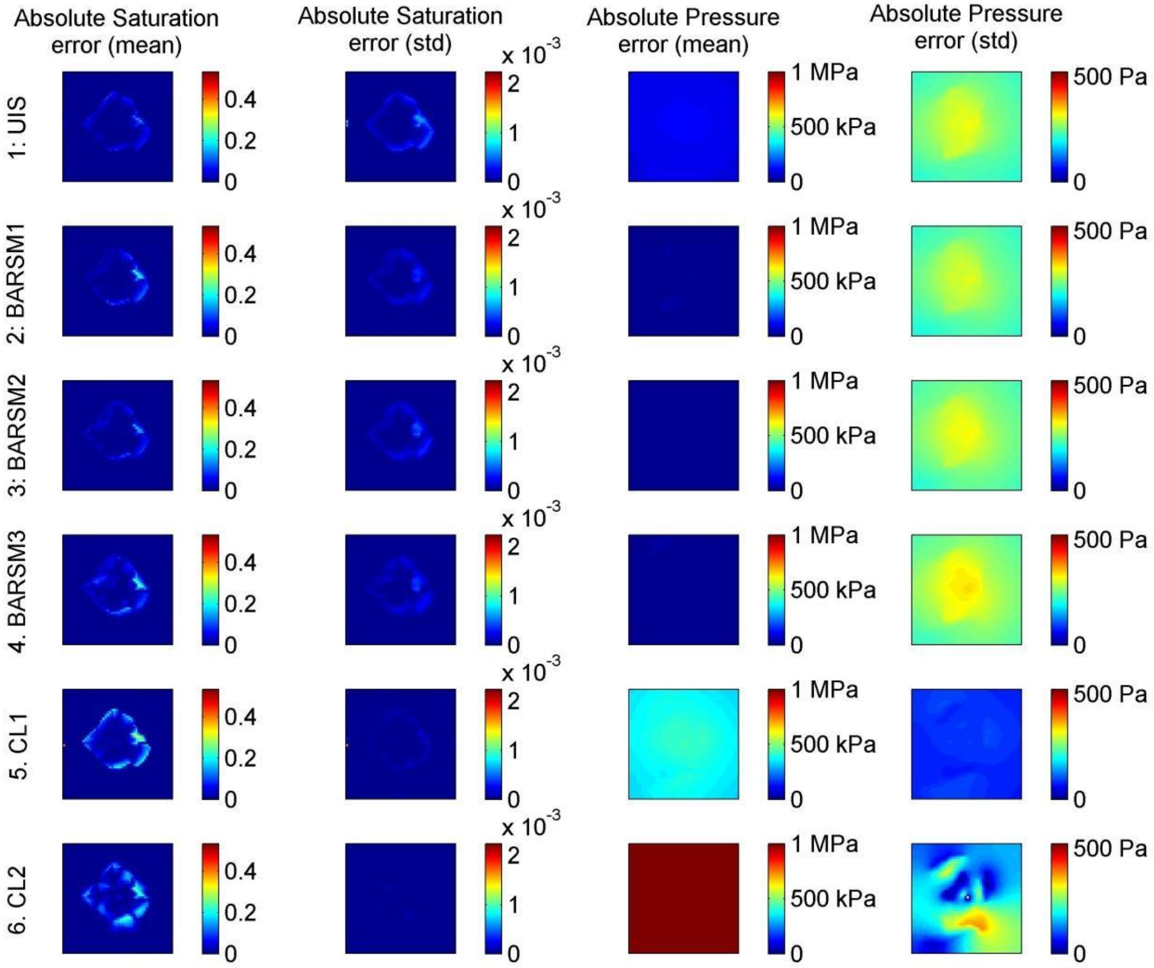


Figure 5-5: Absolute error expectation and error standard deviation in the posterior forecast of pressure (P) and saturation (S) fields after 14 calibration cycles.

The third metric we use to compare the calibration scenarios is the Root Mean Square Error (RMSE) of the posterior forecast of the pressure and saturation fields at the end of injection, normalized by the standard deviation of monitoring error, and calculated after each cycle, similar to equation (4.30) in the second analytical case study of the previous chapter. Figure 5-6 and Figure 5-7 plot the normalized RMSE of the posterior pressure and saturation fields, respectively, for all scenarios. Again, it can be seen that the normalized RMSE obtained in BARSM scenarios are close to the UIS scenario while that the normalized RMSE obtained in CL scenarios are notably higher than the UIS scenario and the BARSM scenarios.

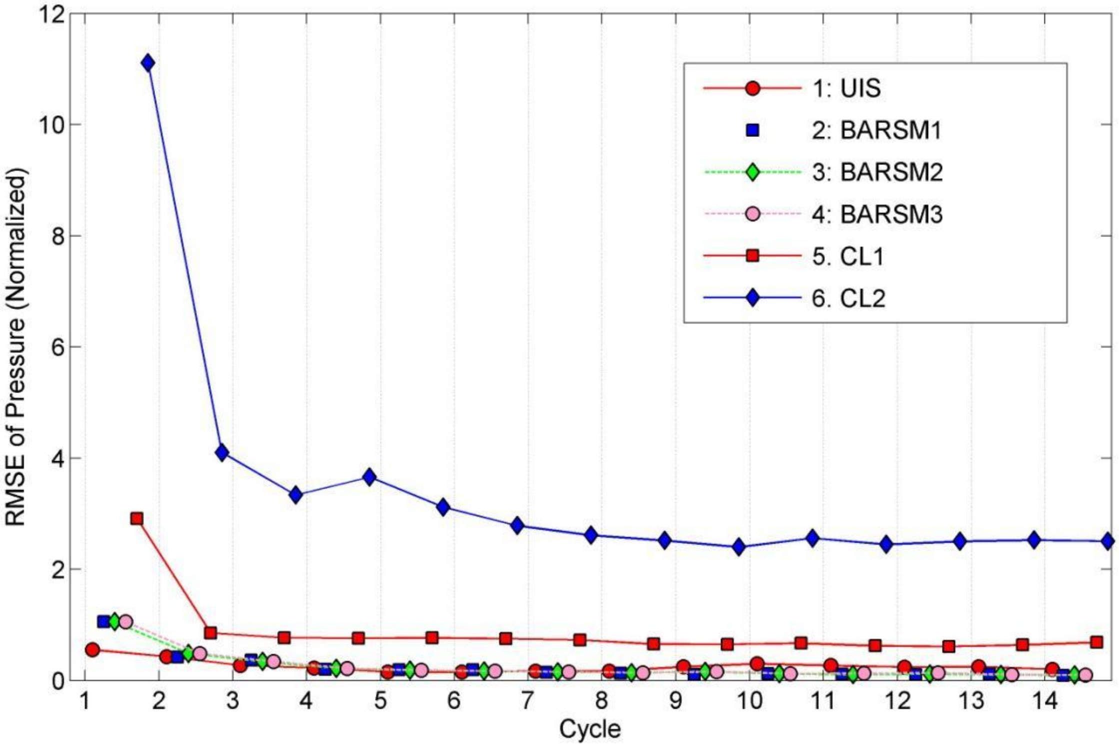


Figure 5-6: Root Mean Square Error (RMSE) of the posterior forecast of pressure field expectation, obtained from samples of the posterior distribution, and normalized by standard deviation of the pressure monitoring error.

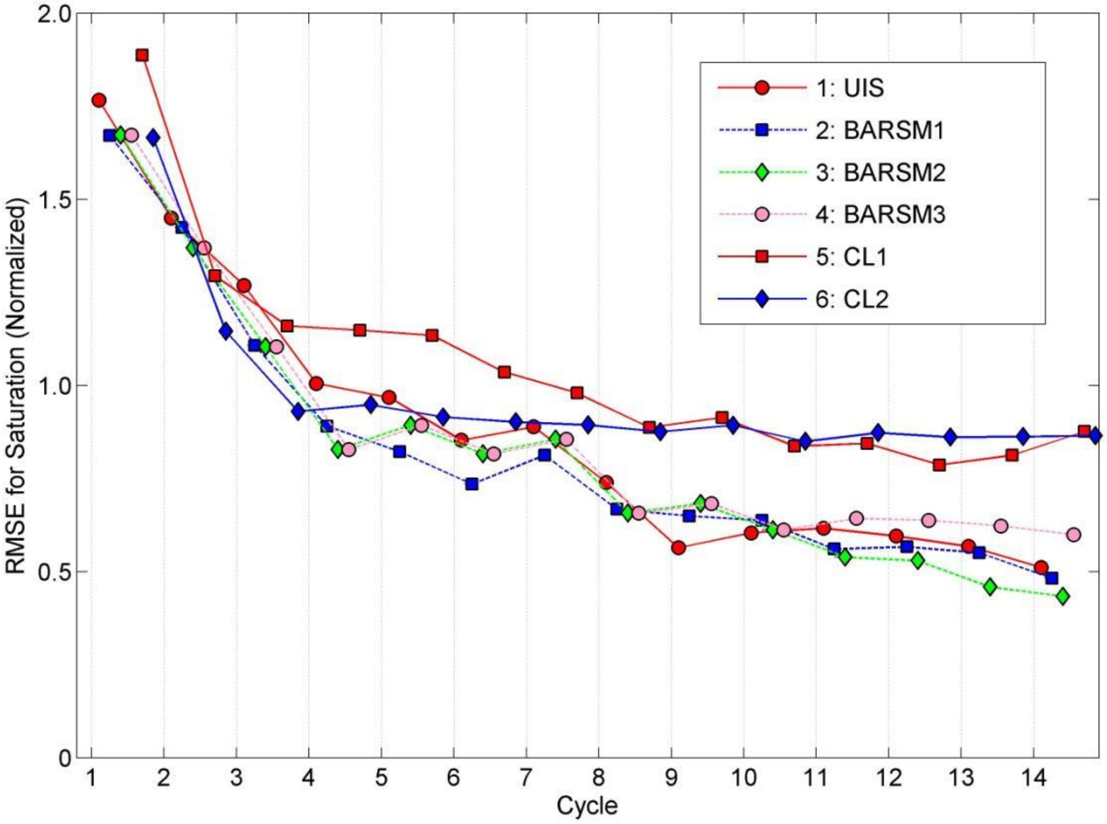


Figure 5-7: Root Mean Square Error (RMSE) of posterior forecast of the saturation field expectation, obtained from samples of the posterior distribution, and normalized by standard deviation of the saturation monitoring error.

5.4.4. Discussion

The question that remains to be answered is how to choose the optimal methodology for CS model calibration among the presented scenarios and based on the presented results. UIS scenario used the full model during all calibration cycles and was the most accurate approach overall, hence, the most computationally costly approach among the six. The computational cost of BARSM and CL scenarios were significantly lower than UIS scenario and CL1 and BARSM3 scenarios had the least computationally cost. The error introduced by using the meta-model and the overall calibration error in all BARSM scenarios were noticeably smaller than the CL scenarios and the BARSM scenarios had meta-model errors smaller than monitoring error.

Choosing the optimal scenario here is essentially a multi-objective optimization problem, with computational efficiency and accuracy as the two objective functions. To discuss this better, each of the scenarios in the case study are simulated two additional times. Additional simulations are to reduce the effect of random sampling on the optimization results. The smallest RMSE of the forecast of saturation and pressure fields of each scenario (with reference to the best of the three UIS scenario runs) at the end of calibration is then considered as a base for comparison. In other words, the absolute differences between the RMSE of the saturation and pressure fields of each scenario, and the RMSE of the saturation and pressure fields of the best UIS scenario, are calculated for each of the three runs for each scenario and at the end of calibration, and the smallest result is used as a cost function for accuracy of that scenario. Then, the total years of full model simulation is used as a cost function for computational efficiency. These two cost functions are shown in a scatter plot in Figure 5-8 for all three runs of all scenarios. For each scenario, the simulation with the smallest "accuracy cost function" is considered as the representative of that scenario.

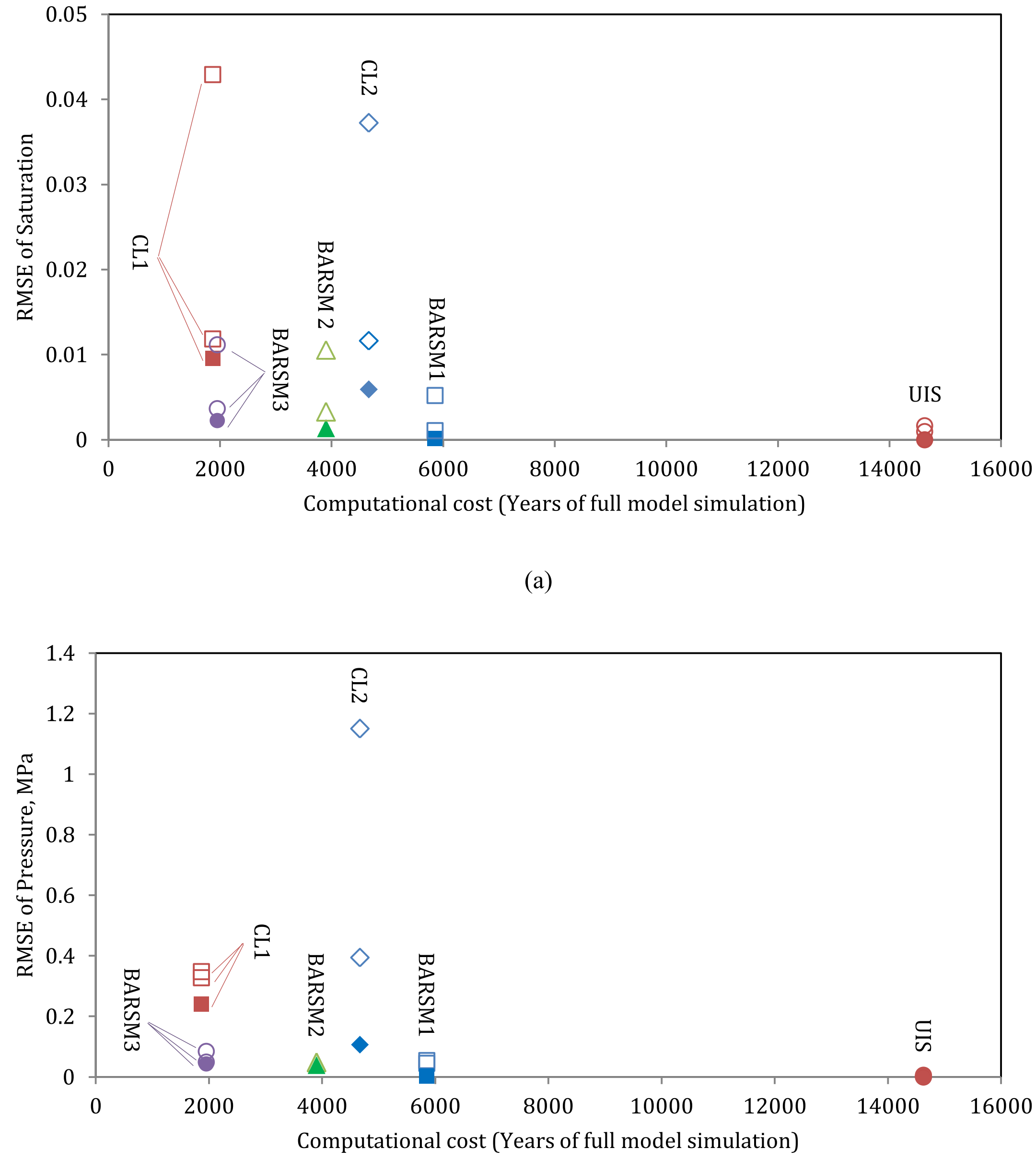
From Figure 5-8, it can be seen that UIS, BARSM1, BARSM2, BARSM3 and CL1 scenarios generate a Pareto frontier (Tamaki *et al*., 1996), i.e., none of these five scenarios "dominates" the other ones in both cost functions. Thus, preference over one of the above-mentioned five scenarios requires either additional criteria (e.g., available computational resources and required accuracy), or a subjective preference.

It is possible to "scalarize" the cost functions, i.e., merge the two cost functions into a single combined cost function. A discussion over methods of scalarization is beyond the scope of this chapter and can be found in multi-objective optimization literature, e.g., (Jahn, 1985) and (Miettinen & Mäkelä, 2002). As an example, we choose a simple and commonly used scalarization method, i.e., weighted sum of the normalized cost functions. Assume O_j^i denotes the i^{th} cost function for the j^{th} scenario. First, we normalize the cost functions as

$$\bar{O}_j^i = \frac{O_j^i - min\left(\{O_k^i: k = 1 \dots 6\}\right)}{max\left(\{O_k^i: k = 1 \dots 6\}\right) - min\left(\{O_k^i: k = 1 \dots 6\}\right)} \tag{5.10}$$

to obtain a dimensionless cost function, $\bar{O}_j^i$, which varies between zero and one (Marler & Arora, 2004). Then, we define the combined cost function as

$$O_j^* = \bar{O}_j^1 + \bar{O}_j^2 \tag{5.11}$$



(a)

(b)

Figure 5-8: Error versus computational cost for all scenarios in (a) saturation and (b) pressure fields. The solid marks show the best run of each scenario and the hollow marks show other runs of each scenario. The absolute difference between the RMSE of each scenario at the end of calibration and the RMSE of the best UIS scenario at the end of calibration is used as a cost function for accuracy of that scenario, and the total years of full model simulations is used as a cost function for computational efficiency of that scenario.

to give equal weights to accuracy and computational efficiency. Table 5-2 lists O_j^* for both pressure and saturation fields and for all scenarios. It can be seen that BARSM 2 and BARSM3 were the best scenarios among the six. BARSM1 scenario had the third ranking and CL2 scenario had the fourth ranking. Finally, UIS and CL2 scenarios had the same ranks. It is clear that the combined cost functions and the weights we used here are not unique and other weights or scalarization functions might result differently.

According to the results discussed in this section, we conclude that BARSM2 and BARSM3 scenarios had the best combined accuracy and computational efficiency. BARSM2 scenario reduced the computational cost of full model simulation by 73% compared to the original UIS scenario without a significant effect on the accuracy of the calibration results. BARSM3 scenario reduced the computational cost of full model simulation by 87% with slightly larger errors compared to BARSM2. BARSM1 had an accuracy close to BARSM2 and BARSM3 scenarios, but reduced the computational cost of full model simulation by 60% which was not as good as BARSM2 and BARSM3 scenarios. CL1 was the most computationally efficient scenario, but had a notably less accuracy in term of calibration results and meta-model output. Thus, it did not rank better than the UIS scenario overall. CL2 performed better than CL1 in terms of accuracy, but had a worse computational efficiency. Overall, CL2 ranked better than CL1. However, CL2 was not a member of the Pareto frontier since it performed worse than BARSM2 and BARSM3 scenarios in terms of both cost functions. In addition, it can be seen from Figure 5-8 that the diversities of the RMSE of different runs in the BARSM scenarios were noticeably lower than those of the CL scenarios. Thus, BARSM scenarios were more consistent than the CL scenarios to forecast the mean of the pressure and the saturation fields after calibration.

Table 5-2: The combined cost function for all scenarios, obtained by sum of the normalized cost functions.

Scenario number and name	**Combined cost function**			
	Saturation		**Pressure**	
	Cost	**Rank**	**Cost**	**Rank**
1: UIS	1.00	5	1.00	5
2: BARSM1	0.33	3	0.33	3
3: BARSM2	0.17	1	0.32	2
4: BARSM3	0.24	2	0.19	1
5: CL1	1.00	5	1.00	5
6: CL2	0.84	4	0.66	4

5.5. Concluding Remarks

In this chapter, we presented the Bayesian Adaptive Response Surface Method (BARSM), which is an efficient adaptive approach for incorporating the Response Surface Method (RSM) to mitigate the computational demand of Bayesian, sampling-based, and continuous calibration algorithms, with an application in CS. In BARSM, Bayesian calibration cycles were divided into full cycles and meta-cycles. Samples drawn in the full cycles were used to train the meta-model, and the meta-model was used instead of the full model in the meta-cycles to reduce the computational cost of the full model simulation. We presented the BARSM in a general formulation and also integrated the Unscented transformation Importance Sampling (UIS) with the BARSM.

A synthetic case study of supercritical CO_2 injection into a deep saline aquifer was conducted to investigate how the BARSM performs compared to the UIS and calibration with a classical and non-adaptive RSM. We also investigated the effect of frequency of full cycles on the accuracy of calibration and accuracy of the meta-model. Results showed that the BARSM scenarios performed very close to the full UIS in terms of calibration results and model forecast, with up to 7.7 times less computational cost. The meta-models trained by the BARSM were able to accurately reproduce the output of the full model on samples of the posterior distribution, and with an error noticeably smaller than the monitoring error.

The results showed that increasing the frequency of the full cycles increase the accuracy of the BARSM, but the BARSM with a small number of full cycles (e.g., 1 or 2 full cycles in 14 cycles) still performs close to the full calibration algorithm. The meta-model accuracy and model forecast accuracy in the classic RSM approach were significantly worse than the BARSM. The classic RSM cases had slightly lower computational cost compared to the BARSM. Increasing the number of training samples in the classic RSM approach slightly improved the outcome.

Considering accuracy and computational efficiency as the objectives for choosing the optimal calibration method, we concluded that the three abovementioned approaches (i.e., full UIS, the BARSM integration UIS and the classical RSM) generate a Pareto frontier. We combined these objectives into a single and subjective cost function and showed that one of the BARSM integrated scenarios performed the best among the rest of the scenarios. We also concluded that the BARSM was a reasonable choice for mitigating the computational cost of CS model calibration because it was able to reduce calibration time notably with negligible effect on the accuracy compared to the full calibration algorithm.

The benefits of the BARSM and the presented results can be summarized as:

- Reducing the computational cost of full model simulation in sampling-based Bayesian calibration of CS models with small effect on the accuracy.
- Reducing the demand for simpler models which need to ignore many aspects of the physics of CS to achieve analytical and computationally fast solutions.
- Avoiding redundant simulation of the full model in similar areas of the parameter space.
- Demonstrating that the choice between calibration algorithms based on accuracy and computational efficiency can be a multi-objective optimization problem with no unique solution.

Current work is subject to limitations that can be addressed in future research, namely:

- With a large number of samples, the BARSM can become more efficient by adding only a subset of the samples of the full cycles to the sample pool.

- Strategies are required to determine which cycles should be a full cycle, either before or during the calibration process.

Chapter 6.

Summary, Conclusions and Future Directions

6.1. Summary and Conclusion

This thesis contributed to reducing the trade-off between accuracy and computational efficiency when calibrating CS models by way of three complementary objectives. In Chapter 3, a methodology for quantitatively assessing the contribution of parameter uncertainty in the model output uncertainty is discussed with application to CO_2 plume evolution. This method accounts for the degree of uncertainty in each of the parameters, as well as the role of each parameter in the model structure. Using a synthetic case study based on Nisku aquifer characteristics, we ranked the parameters based on their contribution to the uncertainty of CO_2 plume evolution. Such ranking can be used to reduce parameter space dimensionality as well as to guide time and effort spent for mitigating the parameter uncertainty and conduct probabilistic model forecast in future work. We also showed that the influence of some constitutive relationship parameters in the uncertainty of CO_2 plume evolution is as high as the contribution of the physical characteristics of the system. Thus, these parameters should be focused on in the uncertainty analysis and model calibration efforts as much as the physical characteristics of the system.

Chapter 4 presents an efficient Bayesian Importance Sampling (IS), the Unscented transformation Importance Sampling (UIS), for mitigating parameter uncertainty with an application in geological CO_2 sequestration. UIS is a method for continuously updating the prior parameter probability distributions obtained by expert opinion and initial measurements at the beginning of the project, utilizing Bayesian statistics and noisy monitoring data. The UIS algorithm is flexible and can be used with any linear/nonlinear/analytical/semi-analytical/numerical system model and any continuous/discrete probability distribution. It also provides an appropriate balance between accuracy and computational efficiency. In a synthetic case study, we benchmarked UIS and showed that it is more accurate than the measurement update stage of Unscented

Kalman Filter (UKF) and ordinary Importance Sampling (IS). We also showed that while increasing the number of samples increases the accuracy of the results, UIS is efficient enough that even with a small number of samples (100 in our case study) performs almost as accurate as the UIS with a larger number of samples (500 in our case study). We adjusted the tuning parameters in our case study and concluded that mixing the proposal distribution in the IS stage of the UIS with a uniform distribution by a ratio of 0.1 to 0.25 enhances the accuracy of the results and helps the algorithm search the parameter space more efficiently.

In Chapter 5, an efficient approach is introduced to incorporate the Response Surface Method (RSM) for mitigating the computational cost of sampling-based continuous Bayesian calibration algorithms, with an application to CS. Our approach is to adaptively modify the meta-model in some of calibration cycles, in order to make it more accurate in areas with higher posterior probability density. This meta-model is then used in the rest of the calibration cycles instead of the original computationally costly model. We showed that this approach can mitigate the computational cost of model calibration more than 90%. In a synthetic case study, we benchmarked this approach against a classical RSM integration approach and against calibration without RSM integration. We showed that our RSM integration approach performed almost as accurately as the original calibration algorithm with up to 7.7 time smaller computational cost, while the classical RSM integration approach was unable to do so. Moreover, the response surface model trained by our approach reproduced the original model output at samples of the posterior distribution accurately with an error noticeably smaller than the monitoring error. We also discussed how to choose the optimal calibration method among the approaches presented in this chapter and we concluded that these approaches generate a Pareto frontier when accuracy and computational efficiency are both accounted for. Thus, picking the optimal approach will be subjective and project specific. Using a subjective combined cost function, a RSM-integrated UIS using our approach was the best among other methods. We concluded that our approach for adaptively training the meta-model is a reasonable choice for mitigating the computational cost of model calibration in CS projects, since it reduces the computational cost noticeably with a negligible error compared to the full model.

6.2. Recommendations for Future Research

In Chapter 3, the sensitivity analysis was conducted using a computationally fast system model that was based on approximations such as rigidity of the aquifer, incompressibility of the fluids and isothermality of the injection process. These assumptions can be relaxed in further works using a more realistic system model. Moreover, using site-specific geometry and including heterogeneity in the system model increases the validity of the results. The combined effect of the uncertainty of the physical model parameters on the uncertainty of the model output can be also compared against that of the constitutive model parameters in future research.

In Chapter 4, UIS might underperform if the posterior distribution is multi-modal. In such cases, a bank of UKFs can replace the single UKF in the first stage of UIS. Thus, new formulations compatible with a bank of UKFs can be developed. The performance of such a modified version should also be investigated against Gaussian-mixture based Kalman filters. Further work on UIS could also address the fact that the UKF will not be computationally efficient when number of uncertain parameters is very large. In such cases, an Ensemble Kalman Filter (EnKF) can replace the UKF in the UIS formulation, even though it might be less accurate. Moreover, escape mechanisms are developed and presented in Appendix A to avoid collapse of the UIS in extreme cases. In the future, these mechanisms should be investigated in greater detail.

The calibration results obtained in Chapters 4 and 5 are obtained using synthetic CS projects, since we did not have access to real site monitoring data and the computational demand of using a high-resolution model of a real site in Chapters 4 and 5 was beyond our available computational resources. The case studies in Chapters 4 and 5 can be repeated using real site data if the required resources are available. In those cases, the performance of UIS and the presented RSM-integration method can be re-assessed, and can be compared with the results of Chapters 4 and 5 case studies.

Finally, in Chapter 5, the meta-model training approach in the full cycles can be evolved to use the full model only in an optimally chosen subset of the samples in order to increase computational efficiency. Solutions to choose that optimal subset of the samples can be investigated further in future research.

Bibliography

Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C. and Vallès, B. 2009. The Ensemble Kalman Filter in Reservoir Engineering - a Review. *SPE Journal*, 14(3): 393-412.

Abacioglu, Y., Oliver, D. and Reynolds, A. 2001. Efficient Reservoir History Matching Using Subspace Vectors. *Computational Geosciences*, 5: 151-72.

Abraham, A. 2005. Adaptation of Fuzzy Inference System Using Neural Learning. In Nedjah, N. and Mourelle, L.D.M. *Fuzzy Systems Engineering*. Springer Berlin Heidelberg, Heidelberg, Germany.

Alberta Geological Survey. Test Case for Comparative Modelling of CO_2 Injection, Migration and Possible Leakage - Wabamun Lake Area, Alberta, Canada. Available at: http://www.ags.gov.ab.ca/co2_h2s/wabamun/Wabamun_base.html [Accessed 2012].

Alshuhail, A., Lawton, D. and Isaac, H. 2009. Seismic Characterizations of the Nisku Formation. University of Calgary Institute for Sustainable Energy, Environment and Economy (ISEEE), Calgary, AB.

Ambadan, J.T. and Tang, Y. 2009. Sigma-Point Kalman Filter Data Assimilation Methods for Strongly Nonlinear Systems. *Journal of the Atmospheric Sciences*, 66(2): 261-85.

Andrieu, C. and Moulines, É. 2006. On the Ergodicity Properties of Some Adaptive MCMC Algorithms. *The Annals of Applied Probability*, 16(3): 1059-732.

Ang, A.H. and Tang, W.H. 2006. Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering. 2nd ed. John Wiley & Sons, New York, NY.

Arulampalam, M.S., Maskell, S., Gordon, N. and Clapp, T. 2002. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing*, 50(2): 174-88.

Balakrishnan, S., Roy, A., Ierapetritou, M.G., Flach, G.P. and Georgopoulos, P.G. 2003. Uncertainty Reduction and Characterization for Complex Environmental Fate and Transport Models: An Empirical Bayesian Framework Incorporating the Stochastic Response Surface Method. *Water Resources Research*, 39(12).

Barton, R.R. 1994. Metamodeling: A State of the Art Review. *WSC '94 Proceedings of the 26th Conference on Winter Simulation*. Orlando, FL.

Bedi, A. and Harrison, J.P. 2012. Dealing with Epistemic Uncertainty in Fractured Rock Masses. *Proceedings Eurorock 2012*. Stockholm, Sweden.

Bedi, A. and Harrison, J.P. 2013. A Comparison of Bayesian Techniques and Non-Probabilistic Models in Rock Engineering Design. *Proceedings of 47th US Rock Mechanics/Geomechanics Symposium (ARMA 2013)*. San Francisco, CA.

Bennion, D.B. and Bachu, S. 2005. Relative Permeability Characteristics for Supercritical CO_2 Displacing Water in a Variety of Potential Sequestration Zones in Western Canada Sedimentary Basin. *Proceedings of Society of Petroleum Engineers (SPE) Annual Conference*. Dallas, TX.

Bennion, D.B.B. and Bachu, S. 2006. Supercritical CO_2 and H_2S-Brine Drainage and Imbibition Relative Permeability Relationships for Intergranular Sandstone and Carbonate Formations. *Proceedings of 68th EAGE Conference and Exhibition incorporating (SPE EUROPEC)*. Vienna, Austria.

Bennion, D.B. and Bachu, S. 2006. The Impact of Interfacial Tension and Pore-Size Distribution/Capillary Pressure Character on CO_2 Relative Permeability at Reservoir Conditions in CO_2-Brine Systems. *Proceedings of SPE/DOE Symposium on Improved Oil Recovery*. Tulsa, OK.

Bennion, D.B. and Bachu, S. 2008. Drainage and Iimbibition Relative Permeability Relationships for Supercritical CO_2 and H_2S/Brine Systems in Intergranular Sandstone, Carbonate, Shale and Anhydrite Formations. *SPE Reservoir Evaluation and Engineering*, 11(3): 487-96.

Bhowmik, S., Mantilla, C.A. and Srinivasan, S. 2011. Tracking CO_2 Plume Migration During Geologic Sequestration Using a Probabilistic History Matching Approach. *Stochastic Environmental Research and Risk Assessment*, 25: 1085-90.

Bohdalová, M. and Šlahor, Ľ. 2007. Monte Carlo Simulations of the Multivariate Distributions with Different Marginals. Comenius University.

Brooks, S., Gelman, A., Jones, G. and Meng, X.L. 2011. Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC, Boca Raton, FL.

Bucy, R.S. and Youssef, H. 1974. Nonlinear Filter Representation via Spline Functions. *Proceedings of 5th Symposium in Nonlinear Estimation*. Los Angeles, CA.

Buhmann, M.D. 2003. Radial Basis Functions: Theory and Implementations. Cambridge University Press, Cambrdige, UK.

Cai, Z., LeGland, F. and Zhang, H. 1995. An Adaptive Local Grid Refinement Method for Nonlinear Filtering.

Campolongo, F., J, C., Saltelli, A. and Schoutens, W. 2005. Enhancing the Morris Method. In Hanson, K.M. and Hemez, F.M. *Sensitivity Analysis of Model Output*. Los Alamos National Laboratory, Los Alamos, NM.

Cappé, O., Douc, R., Guillin, A., Marin, J.M. and Robert, C.P. 2008. Adaptive Importance Sampling in General Mixture Classes. *Statistics and Computing*, 18(4): 447-59.

Cappé, O., Guillin, A., Marin, J.M. and Robert, C.P. 2004. Population Monte Carlo. *Journal of Computational and Graphical Statistics*, 13(4): 907-29.

Celia, M.A. and Nordbotton, J.M. 2009. Practical Modelling Approaches for Geological Storage of Carbon Dioxide. *Ground Water*, 47(5): 627-38.

Chadwick, R.A., Arts, R., Eiken, O., Kirby, G.A., Lindeberg, E. and Zweigel, P. 2004. 4D Seismic Imaging of an Injected CO_2 Plume at the Sleipner Field, Central North Sea. *Geological Society*, 29: 311-20.

Chen, Z. 2003. Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond. *Statistics*, 182(1): 1-69.

Christian, J.T. and Baecher, G.B. 1999. The Point-Estimate Method as Numerical Quadrature. *Journal of Geotechnical and Geoenvironmental Engineering*, 125(9): 779-86.

Clark, J.S. 2005. Why Environmental Scientists Are Becoming Bayesians. *Ecology Letters*, 8: 2-14.

CO_2CRC. CO_2CRC Image Library. Available at: http://www.co2crc.com.au/imagelibrary/ [Accessed 2013].

Cournet, J.M., Marin, J.M., Mira, A. and Robert, C.P. 2012. Adaptive Multiple Importance Sampling. *Scandinavian Journal of Statistics*, 39(4): 798-812.

Cowlesa, M.K. and Carlinb, B.P. 1996. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association*, 91(434): 883-904.

CREWES. *CREWES Fluid Properties Calculator*. Available at: http://www.crewes.org/ResearchLinks/ExplorerPrograms/FlProp/FluidProp.htm [Accessed 2012].

De Rocquigny, E. 2012. Modelling Under Risk and Uncertainty: An Introduction to Statistical, Phenomenological and Computational Methods. John Wiley & Sons, New York, NY.

Deng, H., Stauffer, P.H., Dai, Z., Jiao, Z. and Surdam, R.C. 2012. Simulation of Industrial-Scale CO_2 Storage: Multiscale Heterogeneity and its Impact on Storage Capacity, Injectivity and Leakage. *International Journal of Grrenhouse Gas Control*, 10: 397-418.

Dobossy, M.E., Celia, M.A. and Norbotton, J.M. 2011. An Efficient Software Framework for Performing Industrial Risk Assessment of Leakage for Geological Storage of CO_2. *Energy Procedia*, 4: 4207-14.

Dominguez, G.C., Samaniego, F., Mazzullo, S.J., Rieke, H.H. and Chilingarian, G.V. 1992. Carbonate Reservoir Characterization: A Geological Engineering Analysis. 1st ed. Elsevier, Amsterdam, Netherlands.

Douc, R., Guillin, A., Marin, J.M. and Robert, C.P. 2007. Minimum Variance Importance Sampling via Population Monte Carlo. *ESAIM: Probability and Statistics*, 11: 427-47.

Durbin, J. and Koopman, S.J. 1997. Monte Carlo Maximum Likelihood Estimation for non-Gaussian State Space Models. *Biometrika*, 84(3): 1403-12.

Ebigbo, A., Class, H. and Helmig, R. 2007. CO_2 Leakage Through an Abandoned Well: Problem-Oriented Benchmarks. *Computational Geosciences*, 11(2): 103-15.

Ehlig-Economides, C. and Economides, M.J. 2010. Sequestering Carbon Dioxide in a Closed Underground Volume. *Journal of Petroleum Science and Engineering*, 70(1-2): 123-30.

Eisinger, C.L. and Jensen, J.L. 2009. Data Integration, Petrophysics and Geomodelling-Wabamun Area CO_2 Sequestration Project (WASP). University of Calgary, Calgary, AB.

Emerick, A.A. and Reynolds, A.C. 2012. Combining the Ensemble Kalman Filter with Markov-Chain Monte Carlo for Improved History Matching and Uncertainty Characterization. *SPE Journal*, 17(2): 418-40.

Emerick, A.A. and Reynolds, A.C. 2013. Investigation of the Sampling Performance of Ensemble-Based Methods with a Simple Reservoir Model. *Computational Geosciences*, 17(2): 325-50.

Espinet, A.J. and Shoemaker, C.A. 2013. Comparison of Optimization Algorithms for Parameter Estimation of Multiphase Flow Models with Application to Geological Carbon Sequestration. *Advances in Water Resources*, 54: 133-48.

Evensen, G. 1994. Sequential Data Assimilation with Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics. *Journal of Geophysical Research*, 99(C5): 143-62.

Fedorenko, R.P. 1964. The Rate of Convergence of an Iterative Process. *USSR Computational Math and Mathematical Physics*, 4(3): 227-35.

Freifeld, B.M., Trautz, R.C., Kharaka, Y.K., Phelps, T.J., Myer, L.R., Hovorka, S.D. and Collins, D.J. 2005. The U-Tube: A Novel System for Acquiring Borehole Fluid Samples from a Deep Geologic CO_2 Sequestration Experiment. *Journal of Geophysical Research*, 110(B10).

Friedman, J.H. 1991. Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1): 1-67.

Gasda, S.E. 2008. Numerical Models for Evaluating CO_2 Storage in Deep, Saline Aquifers: Leaky Wells and Large-Scale Geological Features. PhD Thesis. Princeton University, Princeton, NJ.

Geman, S. and Geman, D. 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6): 721-41.

Ghaderi, S. and Leonenko, Y. 2009. Reservoir Modelling, Wabamun Area CO_2 Sequestration Project (WASP). University of Calgary, Calgary, AB.

Glegola, M. 2012. Gravity Observations for Hydrocarbon Reservoir Monitoring. PhD Thesis. TU Delft, Delft, Netherlands.

Golder Associates, 2010. Wabamun Area Sequestration Project: Risk-Based Leakage Model. Project Number: 08-1334-0082.

Gor, G.Y., Elliot, T.R. and Prévost, J.H. 2013. Effects of Thermal Stresses on Caprock Integrity During CO_2 Storage. *International Journal of Greenhouse Gas Control*, 12: 300-09.

Gracie, R. and Craig, J.R. 2011. Using the Extended Finite Element Method for Simulation of Transient Well Leakage. *Advances in Water Resources*, 34(9): 1207-14.

Green, P.J. 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, 82(4): 711-32.

Gu, Y. and Oliver, D.S. 2005. History Matching of the PUNQ-S3 Reservoir Model Using the Ensemble Kalman Filter. *SPE Journal*, 10(2): 217-24.

Hall, B., Statisticat and LLC, 2012. Bayesian Inference. R Package Version 12.11.05.

Hamby, D.M. 1994. A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models. *Enviromental Monitoring and Assessment*, 32(2): 135-54.

Hamby, D.M. 1995. A Comparison of Sensitivity Analysis Techniques. *Health Physics*, 68(2): 195-204.

Handschin, J.E. and Mayne, D.Q. 1969. Monte Carlo Techniques to Estimate Conditional Expectation in Multi-State Nonlinear Filtering. *International Journal of Control*, 9(5): 547-59.

Han, W.S., Kim, K.Y., Esser, R.P., Park, E. and McPherson, B.J. 2011. Sensitivity Study of Simulation Parameters Controlling CO_2 Trapping Mechanisms in Saline Formations. *Transport in Porous Media*, 90(3): 807-29.

Harrr, M.E. 1989. Probabilistic Estimates for Multivariate Analyses. *Applied Mathematical Modelling*, 13(5): 313-18.

Hassan, A.E., Bekhit, H.M. and Chapman, J.B. 2009. Using Markov Chain Monte Carlo to Quantify Parameter Uncertainty and Its Effect on Predictions of a Groundwater Flow Model. *Environmental Modelling & Software*, 24(6): 749-63.

Hastings, W.K. 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1): 97-109.

Helton, J.C. and Davis, F.J. 2003. Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems. *Reliability Engineering and System Safety*, 81(1): 23-69.

Hesterberg, T. 1995. Weighted Average Importance Sampling and Defensive Mixture Distributions. Stanford University, Stanford, CA.

Homma, T. and Saltelli, A. 1996. Importance Measures in Global Sensitivity Analysis of Nonlinear Models. *Reliability Engineering and System Safety*, 52(1): 1-17.

Hong, H.P. 1996. Point-Estimate Moment-Based Reliability Analysis. *Civil Engineering Systems*, 13(4): 81-294.

Hong, H.P. 1998. An Efficient Point Estimate Method for Probabilistic Analysis. *Reliability Engineering and System Safety*, 59(3): 261-67.

Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L. and Hansen, B. 2005. Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations. *Monthly Weather Review*, 133(3): 604-20.

Hoversten, G.M., Milligan, P., Byun, J., Washbourne, J., Knauer, L.C. and Harness, P. 2004. Cross-Well Electromagnetic Seismic Imaging: An Examination of Coincident Surveys at a Steam Flood Project. *Geophysics*, 69: 406-414.

Humez, P., Audigane, P., Lions, J., Chiaberge, C. and Bellenfant, G. 2011. Modelling of CO_2 Leakage up Through an Abandoned Well From Deep Saline Aquifer to Shallow Fresh Ground Water. *Transport in Porous Media*, 90(1): 153-81.

IEA, 2012. CO_2 Emissions From Fuel Combustion Highlights. International Energy Agency.

Iman, R.L. 2008. Latin Hypercube Sampling. In Iman, R.L. *Encyclopedia of Quantitative Risk Analysis and Assessment*. John Wiley & Sons, New York, NY.

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

IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Intergovermental Panel on Climate Change.

Isukapalli, S.S., Roy, A. and Georgopoulos, P.G. 2000. Efficient Sensitivity/Uncertainty Analysis Using the Combined Stochastic Response Surface Method and Automated Differentiation: Application to Environmental and Biological Systems. *Risk Analysis*, 20(5): 591-602.

Jahn, J. 1985. Scalarization in Multi Objective Optimization. *Mathematics of Multi Objective Optimization*, 289: 45-88.

Jin, R., Chen, W. and Simpson, T.W. 2001. Comparative Studies of Metamodeling Techniques Under Multiple Modeling Criteria. *Structural and Multidisciplinary Optimization*, 23(1): 1-13.

Jin, R., Du, X. and Chen, W. 2003. The Use of Metamodeling Techniques for Optimization. *Structural Multidiciplinary Optimization*, 25(2): 99-116.

Johnson, J.W. and White, D. 2012. History Matching and Performence Validation. In Hitchon, B. Best Practices for Validating CO_2 Geological Storage: Observations and Guidance from the IEAGHG Weyburn-Midale CO_2 Monitoring and Storage Project. Geoscience Publishing, Sherwood Park, AB.

Julier, S. 1998. A Skewed Approach to Filtering. *Proceedings of SPIE 3373, Signal and Data Processing of Small Targets*. Orlando, FL.

Julier, S. and Uhlmann, J.K. 1996. A General Method for Approximating Nonlinear Transformations of Probability Distributions. Department of Engineering Science, University of Oxford, Oxford, UK.

Julier, S. and Uhlmann, J.K. 1997. Consistent Debiased Method for Converting Between Between Polar and Cartesian Coordinate Systems. *Proceedings of the SPIE Conference on Acquisition, Tracking, and Pointing*. Orlando, FL.

Julier, S.J. and Uhlmann, J.K. 2004. Unscented Filtering and Nonlinear Estimation. *Proceedings of the IEEE*, 92(3): 401-22.

Kaipio, J.P. and Somersalo, E. 2007. Statistical Inverse Problems: Discretization, Model and Inverse Crimes. *Journal of Computational and Applied Mathematics*, 198(2): 493-504.

Kalman, R.E. 1960. A New Approach to Linear Filtering and Predicition Problems. *Journal of Fluids Engineering*, 82(1): 35-45.

Keramat, M. and Kielbasa, R. 1998. A Study of Stratified Sampling in Variance Reduction Techniques for Parametric Yield Estimation. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 45(5): 575-83.

Kim, I.S. 2011. Large Scale Data Assimilation with Application to the Ionosphere-thermosphere. Ann Arbor, MI.

Kopp, A., Binning, P.J., Johannsen, K., Helmig, R. and Class, H. 2010. A Contribution to Risk Analysis for Leakage Through Abandoned Wells in Geological CO_2 Storage. *Advances in Water Resources*, 33(8): 867-79.

Kopp, A., Class, H. and Helmig, R. 2007. Sensitivity Analysis of CO_2 Injection Processes in Brine. *Proceedings of European Geosciences Union (EGU) General Assembly*. Vienna, Austria.

Kramer, S.C. and Sorenson, H.W. 1988. Recursive Bayesian estimation using piece-wise constant approximations. *Automatica*, 24(6): 789-901.

Kuczera, G. 1988. On the Validity of First-Order Prediction Limits for Conceptual Hydrologic Models. *Journal of Hydrology*, 103(3): 229-47.

Lavoie, R. and Keith, D. 2010. Executive Summary of Wabamun Area CO_2 Sequestration Project (WASP). University of Calgary, Calgary, AB.

Leisenring, M. and Moradkhani, H. 2011. Snow Water Equivalent Prediction Using Bayesian Data Assimilation Methods. *Stochastic Environmental Research and Risk Assessment*, 25(2): 253-70.

LeNeveu, D.M. 2008. CQUESTRA, a Risk and Performance Assessment Code for Geological Sequestration of Carbon Dioxide. *Energy Conversion and Management*, 49(1): 32-46.

Lengler, U., De Lucia, M. and Kühn, M. 2010. The Impact of Heterogeneity on the Distribution of CO_2: Numerical Simulation of CO_2 Storage at Ketzin. *International Journal of Greenhouse Gas Control*, 4(6): 1016-25.

Lepage, G.P. 1978. A New Algorithm for Adaptive Multidimensional Integration. *Journal of Computational Physics*, 27(2): 192-203.

Lepage, G.P. 1980. VEGAS: An Adaptive Multidimensional Integration Program.

Liu, X., Cardiff, M.A. and Kitanidis, P.K. 2004. Parameter Estimation in Nonlinear Environmental Problems. *Stochastic Environmental Research and Risk Assessment*, 24(7): 1003-22.

Liu, C. and McVay, D.A. 2009. Continuous Reservoir Simulation Model Updating and Forecasting Using a Markov Chain Monte Carlo Method. *Proceedings of SPE Reservoir Simulation Symposium*. The Woodlands, TX.

Liu, N. and Oliver, D.S. 2005. Critical Evaluation of the Ensemble Kalman Filter on History Matching of Geological Facies. *Proceedings of SPE Reservoir Simulation Symposium*. Houston, TX.

Ma, X., Al-Harbi, M., Datta-Gupta, A. and Efendiev, Y. 2006. A Multistage Sampling Method for Rapid Quantification of Uncertainty in History Matching Geological Models. *Proceedings of SPE Annual Technical Conference and Exhibition*. San Antonio, TX.

Malekzadeh, F.A. and Dusseault, M.B. 2013. A Solution for the Transition Zone Isosats in Two-Phase Primary Drainage in the Presence of Gravity. *Computational Geosciences*, 17(5): 757-71.

Mark, S. and Mordechai, S. 2011. Application of Monte Carlo Method in Science and Engineering. InTech, Rijeka, Croatia.

Marler, R.T. and Arora, J.S. 2004. Survey of Multi-Objective Optimization Methods for Engineering. *Structural and Multidisciplinary Optimization*, 26(6): 369-95.

Marshall, A. 1956. The Use of Multi-Stage Sampling Schemes in Monte Carlo Computations. *Proceedings of Symposium on Monte Carlo Methods*. New York, NY.

Mathias, S.A., Hardisty, P.E., Trudell, M.R. and Zimmerman, R.W. 2009. Approximate Solutions for Pressure Buildup During CO_2 Injection in Brine Aquifers. *Transport in Porous Media*, 79(2): 265-84.

Maucec, M., Douma, S.G., Hohl, D., Leguijt, J., Jimenez, E. and Datta-Gupta, A. 2007. Streamline-Based History Matching and Uncertainty-Markov-Chain Monte Carlo Study of an Offshore Turbidite Oil Field. *Proceedings of SPE Annual Technical Conference and Exhibition*. Anaheim, CA.

McKay, M.D., Beckman, R.J. and Conover, W.J. 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2): 239-45.

Mesbah, A., Huesmana, A.E.M., Kramerb, H.J.M. and Van den Hofa, P.M.J. 2011. A Comparison of Nonlinear Observers for Output Feedback Model-Based Control of Seeded Batch Crystallization Processes. *Journal of Process Control*, 21(4): 652-66.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. and Teller, A., Teller, H, 1953. Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21(6): 1087-91.

Metropolis, N. and Ulam, S. 1949. The Monte Carlo Method. *Journal of American Statistics Association*, 44(247): 335-41.

Michael, K..B.S., Buschkuehle, B.E., Haug, K. and Talman, S. 2009. Comprehensive Characterization of a Potential Site for CO_2 Geological Storage in Central Alberta, Canada. In AAPG Studies in Geology 59. Carbon Dioxide Sequestration in Geological Media: State of the Science. American Association of Petroleum Geologists (AAPG), Tulsa, OK.

Miettinen, K. and Mäkelä, M.M. 2002. On Scalarizing Functions in Multiobjective Optimization. *OR Spectrum*, 24(2): 193-213.

MIT. Project Pioneer Fact Sheet: Carbon Dioxide Capture and Storage Project. Available at: http://sequestration.mit.edu/tools/projects/transalta.html [Accessed 2012].

Morris, M.D. 1991. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2): 161-74.

Morton, K.W. 2010. The Convection-Diffusion Petrov-Galerkin Story. *IMA Journal of Numerical Analysis*, 30(1): 231-40.

Naevdal, G., Aanonsen, S.I., Vefring and E.H. 2005. Reservoir Monitoring and Continuous Model Updating Using Ensemble Kalman Filter. *SPE Journal*, 10(1): 66-74.

Nooner, S.L., Eiken, O., Hermanrud, C., Sasagawa, G.S., Stenvold, T. and Zumberge, M.A. 2007. Constraints on the in situ Density of CO_2 within the Utsira Formation from Time-Lapse Seafloor Gravity Measurements. *International Journal of Greenhouse Gas Control*, 1: 198-214.

Nordbotton, J.M., Celia, M.A. and Bachu, S. 2005. Injection and Storage of CO_2 in Deep Saline Aquifers: Analytical Solution for CO_2 Plume Evolution Injection. *Transport in Porous Media*, 58(3): 339-60.

Nordbotton, J.M., Flemisch, B., Gasda, S.E., Nilsen, H.M., Fan, Y., Pickup, G.E., Wiese, B., Celia, M.A., Dahle, H.K., Eigestad, G.T. and Pruess, K. 2012. Uncertainties in Practical Simulation of CO_2 Storage. *International Journal of Greenhouse Gas Control*, 9: 234-42.

NPC, 1984. U.S. National Petroleum Council Public Database. Available at: http://www.netl.doe.gov [Accessed 2012].

Oladyshkin, S., Class, H., Helmig, R. and Nowak, W. 2011-a. An Integrative Approach to Robust Design and Probabilistic Risk Assessment for CO_2 Storage in Geological Formations. *Computational Geosciences*, 15(3): 565-77.

Oladyshkin, S., Class, H., Helmig, R. and Nowak, W. 2011-b. A Concept for Data-Driven Uncertainty Quantification and its Application to Carbon Dioxide Storage in Geological Formations. *Advances in Water Resources*, 34(11): 1508-18.

Oladyshkin, S., de Barros, F.P.J. and Nowak, W. 2012. Global Sensitivity Analysis: A Flexible and Efficient Framework with an Example From Stochastic Hydrogeology. *Advances in Water Resources*, 37: 10-22.

Oliver, D.S. and Chen, Y. 2011. Recent Progress on Reservoir History Matching: a Review. *Computational Geosciences*, 15(1): 185-221.

Owen, A. and Zhou, Y. 2000. Safe and Effective Importance Sampling. *Journal of the American Statistical Association*, 95(449): 135-43.

Pennanen, T. and Koivu, M. 2006. An Adaptive Importance Sampling Technique. In Niederreiter, H. and Talay, D. Monte Carlo and Quasi-Monte Carlo Methods. Springer Berlin Heidelberg, Heidelberg, Germany.

Peysson, Y., Bazin, B., Magnier, C., Kohler, E. and Youssef, S. 2011. Permeability Alteration due to Salt Precipitation Driven by Drying in the Context of CO_2 Injection. *Energy Procedia*, 4: 4387-94.

Polson, D., Curtisa, A. and Vivaldac, C. 2012. The Evolving Perception of Risk During Reservoir Evaluation Projects for Geological Storage of CO_2. *International Journal of Greenhouse Gas Control*, 9: 10-23.

Prestona, C., Moneab, M., Jazrawib, W., Brown, K., Whittakerd, S., Whitee, D., Lawf, D., Chalaturnykg, R. and Rostron, B. 2005. IEA GHG Weyburn CO_2 Monitoring and Storage Project. *Fuel Processing Technology*, 86: 1547-68.

Project Pioneer. Available at: http://www.projectpioneer.ca/ [Accessed 2012].

Pruess, K. and Spycher, N. 2007. ECO2N - A Fluid Property Module for the TOUGH2 Code for Studies of CO_2 Storage in Saline Aquifers. *Energy Conversion and Management*, 48(6): 1761-67.

Raiffa, H. and Schlaifer, R. 1961. Applied Statistical Decision Theory. MIT Press, Cambridge, MA.

Raikes, S., Mathieson, A., Roberts, D. and Ringrose, P. 2008. Integration of 3D Seismic with Satellite Imagery at In Salah CO_2 Sequestration Project, Algeria. *SEG Annual Meeting*. Las Vegas, NV.

Razavi, S., Tolson, B.A. and Burn, D.H. 2012-a. Numerical Assessment of Metamodelling Strategies in Computationally Intensive Optimization. *Environmental Modelling & Software*, 34: 67-86.

Razavi, S., Tolson, B.A. and Burn, D.H. 2012-b. Review of Surrogate Modeling in Water Resources. *Water Resources Research*, 48(7).

Reichle, R.H., McLaughlin, D.B. and Entekhabi, D. 2002. Hydrologic Data Assimilation with the Ensemble Kalman Filter. *Monthly Weather Review*, 130(1): 103-16.

Rosenblueth, E. 1975. Point Estimates for Probability Moments. *Mathematics*, 72(10): 3812-14.

Rosi, B., Sykora, J., Pajonk, O., Kucerova, A. and Mattheis, H.G. 2014. Comparison of Numerical Approaches to Bayesian Updating. Institute of Scientific Computing, Braunschweig, Germany.

Rougier, J. 2009. Formal Bayes Methods for Model Calibration with Uncertainty. In Beven, K. and Hall, J. Applied Uncertainty Analysis for Flood Risk Management. Imperial College Press, London, UK.

Sarkarfarshi, M. and Gracie, R. 2013. A Bayesian Approach to Mitigate Parameter Uncertainty in Carbon Dioxide Sequestration Models. *Proceedings of 47th US Rock Mechanics/Geomechanics Symposium*. San Francisco, CA.

Sarkarfarshi, M. and Gracie, R. 2015. Unscented Importance Sampling for Parameter Calibration of Carbon Sequestration Systems. *Stochastic Environmental Research and Risk Assessment*, 29(3): 975-93.

Sarkarfarshi, M., Malekzadeh, F.A., Gracie, R. and Dusseault, M.B. 2014. Parametric Sensitivity Analysis for CO_2 Geosequestration. *International Journal of Greenhouse Gas Control*, 23: 61-71.

Schlumberger, 2013. ECLIPSE Technical Description Version 2013.2.

Seto, C.J. and Mcrae, G.J. 2011. Reducing Risk in Basin Scale CO_2 Sequestration, a Framework for Integrated Monitoring Design. *Enviromental Science and Technology*, 45(3): 845-59.

Seto, C.J. and McRae, G.J. 2011. Reducing Risk in Basin Scale Sequestration, A Bayesian Model Selection Framework for Improving Detection. *Energy Procedia*, 4: 4199-206.

Shu, Q., Kemblowski, M.W. and McKee, M. 2005. An Application of Ensemble Kalman Filter in Integral-Balance Subsurface Modeling. *Stochastic Environmental Research and Risk Assessment*, 19(5): 361-74.

Siegenthaler, U., Stocker, T.F., Monnin, E., Luthi, D., Schwaner, J., Stauffer, B., Raynaud, D., Barnola, J.M., Fischer, H., Masson-Delmotte, V. and Jouzel, J. 2005. Stable Carbon Cycle-Climate Relationship During the Late Pleistocene. *Science*, 310(5752): 1313-17.

Sifuentes, W., Blunt, M.J. and Giddins, M.A. 2009. Modelling CO_2 Storage in Aquifers: Assessing the Key Contributors to Uncertainty. *Proceedings of 2009 SPE Offshore Europe Oil & Gas Conference*. Aberdeen, Scotland.

Smith, M. 1993. Neural Networks for Statistical Modeling. Van Nostrand Reinhold, New York, NY.

Smith, P.J., Shafi, M. and Gao, H. 1997. Quick Simulation: A Review of Importance Sampling Techniques in Communications Systems. *IEEE Journal on Selected Areas in Communications*, 15(4): 597-613.

Sobol, I.M. 2001. Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates. *Mathematics and Computers in Simulation*, 55(1-3): 271-80.

Soreide, I. and Whitson, C.H. 1992. Peng Robinson Predictions for Hydrocarbons, CO_2, N_2 and H_2S with Pure Water and NaCl Brine. *Fluid Phase Equilibria*, 77: 217-40.

Sorooshian, S. and Dracup, J.A. 1980. Stochastic Parameter Estimation Procedures for Hydrologic Rainfall-Runoff Models: Correlated and Heteroscedastic Error Cases. *Water Resour Research*, 16(2): 430-42.

SPE International, 2013. Permeability determination. Available at: http://petrowiki.org/Permeability_determination#Determining_permeability [Accessed 2013].

Stedinger, J.R., Vogel, R.M., Lee, S.U. and Batchelder, R. 2008. Appraisal of the Generalized Likelihood Uncertainty Estimation (GLUE) Method. *Water Reseources Research*, 44(12).

Stein, M.L. 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer , New York, NY.

Surdam, R.C., Jiao, Z. and Bentley, R. 2007. The Rock Springs Uplift: An Outstanding Geological CO_2 Sequestration Site in Southwest Wyoming. Wyoming State Geological Survey, Laramie, WY.

Tamaki, H., Kita, H. and Kobayashi, S. 1996. Multi-Objective Optimization by Genetic Algorithms: A Review. *Proceedings of IEEE International Conference on Evolutionary Computation*. Nagoya, Japan.

Tavakoli, R., Yoon, H., Delshad, M., ElSheikh, A.H., Wheeler, M.F. and Arnold, B.W. 2013. Comparison of Ensemble Filtering Algorithms and Null-Space Monte Carlo for Parameter Estimation and Uncertainty Quantification Using CO_2 Sequestration Data. *Water Resources Research*, 49(12): 8108-27.

The Royal Society, 2014. Climate Change Evidence and Causes. The Royal Society and The US National Academy.

Tsai, C.W. and Franceschini, S. 2005. Evaluation of Probabilistic Point Estimate Methods in Uncertainty Analysis for Environmental Engineering Applications. *Journal of Environmental Engineering*, 131(3): 387-95.

Van Der Merwe, R., Doucet, A., Freitas, N.D. and Wan, E. 2000. The Unscented Particle Filter. Cambridge University, Cambrdige, UK.

Van Leeuwen, P.J. and Evensen, G. 1996. Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation. *Monthly Weather Review*, 124(12): 2898-913.

Vrugt, J.A., ter Braak, C.J.F., Gupta, H.V. and Robinson, B.A. 2009. Equifinality of Formal (DREAM) and Informal (GLUE) Bayesian Approaches in Hydrologic Modeling? *Stochastic Environmental Research and Risk Assessment*, 23(7): 1011-26.

Walton, F.B., Tait, J.C., LeNeveu, D. and Sheppard, M.I. 2004. Geological Storage of CO_2: A Statistical Approach to Assessing Performence and Risk. *Proceedings of 7th International Conference on Greenhouse Gas Control Technologies*. Vancouver, BC.

Wang, G.G. 2003. Adaptive Response Surface Method Using Inherited Latin Hypercube Design Points. *Journal of Mechanical Design*, 125(2): 210-20.

Wang, G.G., Dong, Z. and Aitchison, P. 2002. Adaptive Response Surface Method - a Global Optimization Scheme for Computation-Intensive Design Problems. *Engineering Optimization*, 33(6): 707-33.

Wan, E. and Van Der Merwe, R. 2001. The Unscented Kalman Filter. In Kalman Filtering and Neural Networks. John Wiley & Sons, New York, NY.

Wen, X.H. and Chen, W.H. 2005. Real-Time Reservoir Model Updating Using Ensemble Kalman Filter. *Proceedings of SPE Reservoir Simulation Symposium*. The Woodlands, TX.

Wilson, M. and Monea, M. 2004. IEA GHG Weyburn CO_2 Monitoring and Storage Project Summary Report 2000-2004. *Proceedings of 7th International Conference on Greenhouse Gas Control Technologies*. Vancouver, BC.

Wu, J. and Zeng, X. 2013. Review of the Uncertainty Analysis of Groundwater Numerical Simulation. *Chinese Science Bulletin*, 58(25): 3044-52.

Zhao, H., Liao, X., Chen, Y. and Zhao, X. 2010. Sensitivity Analysis of CO_2 Sequestration in Saline Aquifers. *Petroleum Science*, 7(3): 372-78.

Zhdanov, S.M., Endo, M., Black, N., Spangler, L., Fairweather, S., Hibbs, A., Eiskamp, G.A. and Will, R. 2013. Electromagnetic Monitoring of CO_2 Sequestration in Deep Reservoirs. *First Break*, 31: 71-78.

Appendix A. Escape Mechanisms for UIS

As mentioned in Section 4.5.3, when one of the samples carries all of the weight, i.e., when one W_i is equal to one and all other W_i are equal to zero, N_{eff} becomes equal to one and the situation is known as *degeneracy* or *collapse* of the algorithm (Arulampalam *et al.*, 2002). Assume UIS collapses or is about to collapse in cycle n and sample $\widehat{\boldsymbol{m}}_c$ gains almost all the weight in that cycle so that $W_c \cong 1$. In this case, the posterior distribution becomes equal or very close to a Dirac delta function, i.e., $\pi^n(\boldsymbol{m}) \cong \delta(\boldsymbol{m} - \widehat{\boldsymbol{m}}_c)$. Consequently, the prior and posterior distributions of all cycle beyond n become equal or very close to the same Dirac delta function.

The adaptive nature of UIS, use of the defensive mixture distribution and re-sampling at every cycle from a continuous proposal distribution, make the degeneracy of the UIS unlikely. However, in rare cases that N_{eff} gets very close to one, escape mechanisms are required to avoid the collapse of the samples. According to our tests, this might happen if the posterior distribution is concentrated in a very small area, or when the number of samples is extremely small. Two escape mechanisms are developed and introduced here to avoid the collapse of the UIS in these cases.

The first mechanisms is adaptively adjusting the defensive mixture ratio with N_{eff}. As discussed in Section 4.6.3, a larger defensive mixture ratio diminishes the efficiency of UIS because a large number of samples are drawn from non-important areas of the parameter space. However, when the algorithm is about to collapse, drawing more samples from the rest of the parameter space can be beneficial since the collapse might happen near a suboptimal area and a larger defensive mixture ratio helps searching the parameter space for better samples. Accordingly, we monitor N_{eff} in all cycles. As shown in Figure A-1, If N_{eff} falls below a certain threshold denoted by N_{th}, η will be increased linearly with N_{eff} up to a certain limit, so that the maximum η is reached when $N_{eff} = 1$. N_{th} and the limits of η are tuning parameters and this escape mechanism.

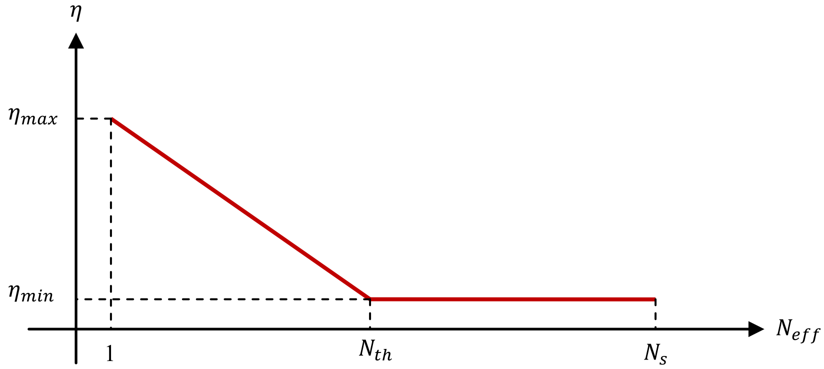


Figure A-1: The adaptive adjustment of the defensive mixture ratio (η) with the effective number of samples (N_{eff}) in order to avoid sample collapse in UIS.

In our tests, this mechanism was effective in the first case, i.e., when posterior distribution is concentrated in a very small area. The recommended values for the tuning parameters based on our tests are $N_{th} = 0.1 \times N_s$, $\eta_{min} = 0.1$ and η_{max}=0.5.

The second escape mechanism is preventing the variance of each parameter in the UKF stage prior distribution, and the proposal distribution for each parameter, to become very close to zero. To do so, we monitor the covariance matrix of the prior distribution of the UKF stage in every cycle and prevent the diagonal members of the covariance matrix to fall below a certain threshold. This threshold is the tuning parameter of this escape mechanism. In our tests, this mechanism was effective in the second case, i.e., when the number of samples is extremely small. The recommended value for the threshold of the diagonal covariance matrix members varies based on the problem. In our tests, a small fraction of the same element of the covariance matrix of the initial prior distribution, e.g., 10^{-4}, responded well.

Appendix B. Polynomial Regression

In this appendix we describe the process of polynomial regression by least square method for the purpose of approximating a complex function with a simple polynomial. Assume $g(\boldsymbol{m})$ is a complex mathematical model with N_m parameters, $\boldsymbol{m} = [m_1, m_2, \ldots, m_{N_m}]$. Assume N_{sp} data points (model runs) of the form $\{[\widehat{\boldsymbol{m}}_i, g(\widehat{\boldsymbol{m}}_i)]: i = 1 \ldots N_s\}$ are available. First, we derive the formulation for a first order polynomial regression and then we extend it to higher order polynomials.

Assume a meta-model, $\hat{g}(\boldsymbol{m})$, is a linear function of m_1 to m_{N_m} and is defined by

$$\hat{g}(\boldsymbol{m}) = \beta_0 + \beta_1 m_1 + \beta_2 m_2 + \cdots + \beta_{N_m} m_{N_m} = \sum_{i=0}^{N_m} \beta_i m_i \tag{B.1}$$

where $m_0 = 1$. The difference between the full model $g(\boldsymbol{m})$ and the meta-model $\hat{g}(\boldsymbol{m})$ at any arbitrary data point $\widehat{\boldsymbol{m}}_j$ can be described with an error, $\hat{\varepsilon}_{mod,j}$,

$$g(\widehat{\boldsymbol{m}}_j) = \sum_{i=0}^{N_m} \beta_i \, \widehat{m}_{i,j} + \hat{\varepsilon}_{mod,j} \qquad j = 1 \ to \ N_{sp} \tag{B.2}$$

where $\widehat{m}_{i,j}$ is the i^{th} parameter of $\widehat{\boldsymbol{m}}_j$. Sum of Squared Errors (SSE) for the N_{sp} data points is obtained by

$$SSE = \sum_{j=1}^{N_{sp}} \left(g(\widehat{\boldsymbol{m}}_j) - \sum_{i=0}^{N_m} \beta_i \widehat{m}_{i,j} \right)^2 \tag{B.3}$$

To minimize the SSE, we obtain partial derivatives of the SSE with respect to β_0 to β_{N_m} and set them equal to zero,

$$\frac{\partial SSE}{\partial \beta_i} = 2 \sum_{j=1}^{N_{sp}} \left(\widehat{m}_{i,j} \left(\sum_{i=0}^{N_m} \beta_i \widehat{m}_{i,j} \right) - \widehat{m}_{i,j} g(\widehat{\boldsymbol{m}}_j) \right) = 0 \qquad i = 0 \ldots N_m \tag{B.4}$$

or in the matrix form,

$$\left(\boldsymbol{A}_m{}^T\boldsymbol{A}_m\right)\boldsymbol{B}_\beta = \boldsymbol{A}_m{}^T\boldsymbol{G}_m \tag{B.5}$$

where,

$$\boldsymbol{A}_m = \begin{bmatrix} 1 & \hat{m}_{1,1} & \hat{m}_{2,1} & \cdots & \hat{m}_{N_m,1} \\ 1 & \hat{m}_{1,2} & \hat{m}_{2,2} & \cdots & \hat{m}_{N_m,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \hat{m}_{1,N_{sp}} & \hat{m}_{2,N_{sp}} & \cdots & \hat{m}_{N_m,N_{sp}} \end{bmatrix} \tag{B.6}$$

and

$$\boldsymbol{B}_\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{N_m} \end{bmatrix} \tag{B.7}$$

and

$$\boldsymbol{G}_m = \begin{bmatrix} g(\hat{\boldsymbol{m}}_1) \\ g(\hat{\boldsymbol{m}}_2) \\ \vdots \\ g\left(\hat{\boldsymbol{m}}_{N_{sp}}\right) \end{bmatrix} \tag{B.8}$$

The solution to the system of equations above is obtained by

$$\boldsymbol{B}_\beta = \left(\boldsymbol{A}_m{}^T\boldsymbol{A}_m\right)^{-1}\boldsymbol{A}_m{}^T\boldsymbol{G}_m \tag{B.9}$$

Now, the output of the full model $g(\boldsymbol{m})$ for an arbitrary parameter set $\boldsymbol{m}$ can be approximated by $\hat{g}(\boldsymbol{m}) = \boldsymbol{B}_\beta{}^T\boldsymbol{m}^T$.

In cases that $g(\boldsymbol{m})$ is highly nonlinear, the first order linear regression could be a poor approximation of $g(\boldsymbol{m})$. Thus, a higher order polynomial can be used instead of the first order polynomial. Here, we introduce the formulation for a second order polynomial regression which is the most commonly used from of meta-model used in RSM (Jin *et al.*, 2003). The formulation for higher order polynomials can be derived in the same fashion.

First, we add the quadratic terms to (B.1),

$$\hat{g}(\boldsymbol{m}) = \beta_0 + \beta_1 m_1 + \cdots + \beta_{N_m} m_{N_m} + \beta_{N_m+1} m_1^2 + \cdots + \beta_{2N_m} m_{N_m}^2 + \beta_{2N_m+1} m_1 m_2 + \cdots + \beta_{2N_m+\binom{N_m}{2}} m_{N_m-1} m_{N_m} \tag{B.10}$$

where, $\binom{N_m}{2} = \frac{N_m(N_m-1)}{2}$. We can re-write (B.10) can as

$$\hat{g}(\boldsymbol{m}) = \sum_{k=0}^{2N_m+\binom{N_m}{2}} \beta_k \psi_k \tag{B.11}$$

where ψ_k is one of the parameters term (m_i, m_i^2 or $m_i m_{j\neq i}$). It can be seen that (B.11) is still a linear function of the coefficients β_k. Thus, the regression formulation for the second order polynomial can be derived similar to the first order polynomial. Linear regression formulation for higher order polynomials can be derived in the same fashion.