University of Alberta

Constraining 3D Petroleum Reservoir Models to Petrophysical Data, Local Temperature Observations, and Gridded Seismic Attributes with the Ensemble Kalman Filter (EnKF)

by

Yevgeniy Zagayevskiy

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science

in

Mining Engineering

Department of Civil and Environmental Engineering

Spring 2012
Edmonton, Alberta

“A man can change his stars”

A Knight’s Tale

To my parents, Vladimir and Nina, my sister, Nina, and my grandparents, Vasiliy and Zoya for their love and support

Abstract

A methodology based on the Ensemble Kalman Filter (EnKF) is proposed for petroleum reservoir characterization, using continuous integration of petrophysical core data, reservoir temperature observations, and gridded time-lapse acoustic impedances. The localization of updating and covariance matrices is performed to assimilate exhaustive seismic data. A shortcut based on propagation of the ensemble mean and co-simulation of the ensemble variations is implemented to reduce computational cost of the forecast step. The integration of additional data from multiple sources and time steps improves the estimates of porosity and permeability. This methodology is applied to a synthetic 2D steam assisted gravity drainage (SAGD) case study to examine the ability of the EnKF to constrain spatial distributions of porosity and permeability. A realistic 3D SAGD case study is used to demonstrate the applicability of this methodology to a real industrial problem. Obtained results show effective application of the EnKF to petroleum reservoir characterization.

Acknowledgments

I would like to express much gratitude to my supervisor, Dr. Clayton V. Deutsch. He was the first person to introduce me to the exciting realm of geostatistics, which has defined my future career path. I want to thank him for his continuous inspiration, motivation, support, and guidance throughout my research, thesis, and above basic schoolwork.

I really appreciate the peaceful learning atmosphere, productive working environment, and financial support offered to me by the Centre of Computational Geostatistics (CCG) and the University of Alberta.

My university years would not have been as effective and full of educational joy without the help of CCG students, graduates, staff, and affiliates. Thanks to my friends Mehran, Hadi, Behrang, and Martha for comments and feedback regarding my research work and general application of geostatistics to petroleum problems. Thanks to John, Brandon, Jeff, and Ryan for sharing their professional knowledge and Canadian experience as senior students. Thanks to Dr. Amir H. Hosseini for giving me an opportunity to apply theoretical knowledge to the real project and for mentoring me throughout it. I am also thankful to other CCG students and my friends for their significant contributions.

I cannot overestimate the value of the friendships I have established with a lot of wonderful people in Canada. Thanks to everyone for your time and regard. Special thanks to Nurlan and Kamshat for making me feeling like being at home.

Last, but not the least, I thank my family. I would not have been able to complete this thesis without their constant encouragement and steadfast support, which helped me through many stressful situations. Although there were thousands of miles between us, thank you for being with me in spirit for these two years!

Table of Contents

List of Tables

List of Figures

List of Abbreviations, Symbols, and Nomenclature

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
4D	four-dimensional
API	American Petroleum Institute (density)
BOPD	barrels of oil per day
CC	correlation coefficient
CMG	Computer Modelling Group Ltd
COP	cumulative oil production
CSOR	cumulative steam oil ratio
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
EnKS	Ensemble Kalman Smoother
ERCB	Energy Resources Conservation Board of Alberta
FORTRAN	software: The IBM Mathematical Formula Translating System
GOR	gas-to-oil ratio
GSLib	software: Geostatistical Software Library
IHS	inclined heterolythic strata
KF	Kalman Filter
KL	Karhunen-Loeve
KLKF	Karhunen-Loeve Kalman Filter
LMC	linear model of coregionalization
MAE	mean absolute error
MATLAB	software: Matrix Laboratory
MCS	Monte Carlo Simulation
MSD	mean standard deviation
NAGD	naphtha assisted gravity drainage
PF	Particle Filter
PUNQ-S3	Production Forecasting with Uncertainty Quantification project
PVT	pressure-volume-temperature

RML	Randomized Maximum Likelihood
RMR	reservoir management report
RMSE	root mean squared error
RSE	root squared error
SAGD	steam assisted gravity drainage
SD	standard deviation
SEEK	Singular Evolutive Extended Kalman
SEIK	Singular Evolutive Interpolated Kalman
SGS	Sequential Gaussian Simulation
SIS	Sequential Indicator Simulation
SPF	Sigma Point Filter
SSC	Sequential Self-Calibration
STARS	software: Steam, Thermal, and Advanced Processes Reservoir Simulator
UKF	Unscented Kalman filter

^{0}C	degrees of Celsius
a	analysis
b	brine water
bbl	barrels
dg	dissolved gas
est	estimated value
f	forecast
g	gas
NS	normal scores
o	oil
orig	original value
pr	pseudoreduced quantity
P-wave	primary wave
RC	reservoir conditions
s	solid rock or phase
SC	surface conditions

sim	simulated value
stand	standardized value
S-wave	secondary wave
target	target value
true	base case or true value
V-shale	volume of shale
w	water

$\vec{v}$	Darcy velocity
$\vec{\nabla}$	divergence operator
$\vec{\Psi}$	energy flux
$\vec{\Phi}$	flux of a fluid phase
∇	gradient operator
$\vec{h}$	lag vector
$\vec{\vec{K}}$	permeability tensor
$\hat{\boldsymbol{C}}$	sample covariance matrix
μ	dynamic viscosity of a fluid phase
μ_K	kinematic viscosity of a fluid phase
a	intercept of linear regression model
a	semivariogram model range
B	formation volume factor
b	slope of linear regression model
B_x	length of model grid or size of model grid in *X* direction
b_x	size of block in X direction
B_y	length of model grid or size of model grid in *Y* direction
b_y	size of block in *Y* direction
B_z	length of model grid or size of model grid in *Z* direction
b_z	size of block in *Z* direction
$\boldsymbol{C}^a$	covariance matrix of analyzed ensemble
$\boldsymbol{C}^f$	covariance matrix of forecasted ensemble
C_P	specific heat capacity at constant pressure

C_V	specific heat capacity at constant volume
$C_X(h)$	covariance function of a variable X
$C_{XY}(h)$	cross-covariance function of variables X and Y
d_t	data vector
$\boldsymbol{D_t}$	discretized data matrix
$\boldsymbol{d_t}$	discretized data vector
$E[X]$	expected value of a variable X
$\boldsymbol{EM}$	model error matrix
$\boldsymbol{em}$	model error vector
F	model operator
f	volume fraction
G	data operator
g	magnitude of vector of gravitational acceleration
G	specific gas gravity or ratio of gas density to air density
H	enthalpy
h	lag distance
$\boldsymbol{H}$	observation matrix
i	index of an order
i_b	order number of a block in model grid
i_{dt}	index of a data in a model grid
i_m	index of a model parameter in a model grid
i_r	index of a realization
i_s	index of a state variable in a model grid
i_x	order number of a model block in X direction
i_y	order number of a model block in Y direction
i_z	order number of a model block in Z direction
k	effective elastic bulk modulus
$\boldsymbol{K}$	Kalman gain matrix
K	magnitude of permeability in any direction
K_r	relative permeability of a fluid phase
K_{xx}	magnitude of permeability in X direction
K_{yy}	magnitude of permeability in Y direction

K_{zz}	magnitude of permeability in Z direction
$\boldsymbol{L}$	localization matrix
$\boldsymbol{M}$	discretized model parameter matrix
$\boldsymbol{m}$	discretized model parameter vector
m	model parameter vector
N	discretized model size
n	number of samples
n_p	number of paired random variables separated by lag distance h
N_b	number of blocks in a model grid
N_{bx}	number of blocks in X direction
N_{by}	number of blocks in Y direction
N_{bz}	number of blocks in Z direction
$N_{d,t}$	number of data sampled at time step t
N_e	number of realizations or ensemble size
N_{fp}	number of follow-up seismic surveys
N_m	number of model parameters
N_s	number of state variables
N_{se}	number of selected realizations for propagation at forecast step
N_{sx}	number of blocks of smoothing window in X direction
N_{sy}	number of blocks of smoothing window in Y direction
N_{sz}	number of blocks of smoothing window in Z direction
N_t	number of time steps
N_v	number of model variables
P	pressure
$p(X)$	probability density function of a random variable X
$p(X\|Y)$	conditional probability density function of a random variable X relative to a variable Y
q	injection or production mass rate of a phase
q_E	energy source or sink
R	mass rate of a fluid phase
r	rate of generation of a phase
R	residual variable

R_G	gas to oil ratio
R_{gas}	gas constant
S	saturation of a fluid
s_0	initial conditions of a state variable
S_b	brine water saturation
$s_{b,t}$	boundary conditions of a state variable
S_g	gas saturation
S_o	oil or bitumen saturation
$\boldsymbol{S_t}$	discretized state variable matrix
$\boldsymbol{s_t}$	discretized state variable vector
s_t	state variable vector
S_w	water saturation
T	temperature
t	time
t'	dimensionless variable in time
$\boldsymbol{U_t}$	discretized dynamic model matrix
$\boldsymbol{u_t}$	discretized dynamic model vector
u_t	dynamic model vector
V	volume
$VAR[X]$	variance of a random variable X
V_p	primary wave acoustic (seismic) velocity
V_s	secondary wave acoustic (seismic) velocity
W	concentration of a fluid phase
w	weight in objective function
X	first of three principal directions
x	value of a random variable X
x	x coordinate
x'	dimensionless variable in space
Y	second of three principal directions
y	y coordinate
y'	dimensionless variable in space
Z	third of three principal directions

z	z coordinate
Z_p	primary wave acoustic (seismic) impedance
Z_s	secondary wave acoustic (seismic) impedance
α	thermal diffusivity
β	Biot coefficient
$\gamma_X(h)$	semivariogram model of a variable X
$\gamma_{XY}(h)$	cross-semivariogram model of variables X and Y
Δ	difference
ε	measurement error
ϑ	heat capacity ratio of gas
κ_T	total thermal conductivity
λ	weight in an interpolation method
λ_R	relative mobility
ν	perturbation
ξ	compressibility factor
Π	external work done by a system
π	facies type or rock type
ρ	density
$\boldsymbol{\Sigma}$	covariance matrix of data
ς	salinity of brine water
σ	standard deviation of a random variable
σ^2	variance of a random variable
τ	effective elastic shear modulus
υ	parameter in SAGD flow
ϕ	porosity
χ	transmissibility of a fluid phase
ψ	mean of a random variable
Ω	internal energy per unit mass of a fluid phase

Chapter 1 – Introduction

Oil, natural gas, and their by-products are an integral part of modern society, which cannot be substituted in full capacity with other energy alternatives or chemical by-products at present. These natural resources are unrecoverable and, therefore, petroleum reserves should be carefully and efficiently developed for our needs and those of future generations. This task is led by geologists, engineers, earth scientists, and other professionals, who form reservoir management teams (Wilson and Burgh, 2008).

Reservoir management designs a dynamic strategy for optimal development of the reservoir that is in compliance with economic, political, social, technical, and environmental requirements (Fahimuddin, 2010). The numerical reservoir model is a key engineering tool in making effective management decisions where the optimization problem is solved. It helps to analyze reservoir behaviour and to predict future performance. High quality plausible petroleum reservoir models help to minimize cost mismanagement and facilitate field development, which ultimately improve oil recovery (Satter and Thakur, 1994).

The accuracy of the quantification of natural phenomena and the prediction of reservoir performance depends on the selected modeling approach, the available data, and their quality. Even though the modeling technique is important, the data, which are used to understand geological architecture and reservoir behaviour, determine the overall information about a field. The data may come from various sources and time steps sampled at different resolutions with changing measurement errors and, thus, they represent different aspects of a reservoir. It is claimed that incorporating additional data into a model should improve its quality (Deutsch, 2002). For this reason a modeling technique that is able to assimilate multiple continuous data with various characteristics is sought.

1.1 Problem Definition

Plausible numerical models of a petroleum reservoir are required for effective oil and gas production. A petroleum reservoir is a dynamic system determined by two main components: static model parameters and dynamic state variables, where the behaviour of the dynamic variables depends on the distribution of static parameters, and initial and boundary conditions (Gu, 2006). Therefore, the proper quantification of the static part of a dynamic system leads to a plausible model of good quality, which can be used to predict reservoir performance. In reservoir characterization, geology, PVT, and relative permeability curves constitute the static part. Time variant states, such as production rates and reservoir pressure, form the dynamic part of the model (Deutsch, 2002). Therefore, if PVT and relative permeability curves are known, only geological properties of a reservoir should be estimated for better prediction of reservoir performance using a conceptual geological interpretation in conjunction with available data.

Petroleum reservoir models are built using various estimation and simulation modeling methods that often involve history matching, where estimates of models are conformed to dynamic data by adjusting their static properties. Among these properties, porosity and permeability distributions are the most vital (Oliver et al, 2001). Usually well production rates and bottomhole pressure are used to constrain petroleum reservoir models. However, frequently sampled reservoir temperature and time-lapse seismic attributes can also be used to constrain the models of thermal reservoirs in northern Alberta operated by the SAGD heavy oil extraction method (Butler, 1991; Oliver and Chen, 2010; Husky Energy, 2010). When new data are acquired, the geological model is rebuilt to history-match all production observations (including old and new measurements) and honour data of static parameters. Thus, modeling a petroleum reservoir is a continuous process. The most common history-matching techniques are based on optimization methods that utilize a gradient approach. Some examples of deterministic modeling methods are Gauss-Newton (Oliver, 1994), Levenberg-Marquardt (Li et al, 2001), and the steepest descent method (Makhlouf et al, 1993). The

problem with these methods is that they cannot quantify uncertainty. On the other hand, stochastic modeling techniques, such as the Sequential Self-Calibration (SSC) (Hernandez, 1997), find the best model estimates and report estimation uncertainty simultaneously. However, petroleum reservoir governing equations have to be known in order to use any of the mentioned history-matching methods. The equations are solved analytically to derive the sensitivity matrix, which is used to adjust model parameters and history-match production data. Thus, the applications of gradient-based modeling techniques are limited by understanding the fluid flow mechanism. Also it is hard to couple these methods with commercial flow simulators, since their internal equations are not available to the public.

Therefore, the inverse modeling technique Ensemble Kalman Filter (EnKF) based on multiple realizations is proposed to continuously and simultaneously constrain petroleum reservoir models to observations of dynamic variables and measurements of static parameters with uncertainty quantification (Evensen, 2009). Spatial distributions of porosity and permeability are the most important for reserve estimation and fluid flow prediction and, thus, estimated first (Deutsch, 2002). The EnKF does not depend on the internal mechanism of a flow simulator and treats it as a black box. Only input and output of the simulator with proper operational parameters are required for history-matching with the EnKF. The proposed modeling method is claimed to be simple to implement, efficient in generating results, and computationally cheap, if modifications are applied.

1.2 Objectives of the Thesis

The principal objective of the thesis is to develop a methodology for continuous constraining spatial distributions of geological components of a petroleum reservoir presented by porosity and permeability to all available hard core data, soft reservoir temperature observations, and time-lapse seismic attributes with the EnKF. The objective does not explicitly impose any requirements to match production data from all time steps, but rather to estimate spatial distributions of the model parameters. The

methodology is developed to improve the ability to forecast the reservoir's performance. The importance of assimilating additional data should be assessed. A comparison of the benefits of integrating different data types for reservoir characterization with the EnKF should be conducted. The proposed modeling technique's characteristics and implementation details should be examined. The conventional EnKF algorithm should be modified to improve estimation accuracy and decrease computational time. Finally, the methodology should be applied to a realistic case study to evaluate its practical significance.

1.3 Proposed Solution

The methodology for continuous data assimilation based on EnKF is built over the conventional data assimilation algorithm, which utilizes the concept of history-matching (Carlson, 2003). The proposed algorithm is demonstrated on thermally operated petroleum reservoirs. The steam assisted gravity drainage (SAGD) method is widely implemented in northern Alberta to extract heavy oil and bitumen reserves. Thus, it is selected as a production mechanism for examined thermal operations (Butler, 1994).

The EnKF characteristics, the significance of additional assimilated data, and the benefits of the contributions of each data type are worked out on the 2D SAGD case study. The examined hard core data are porosity and horizontal permeability sampled from vertical surveillance (observation) wells. The simple probabilistic porosity-log permeability model is devised to relate model parameters to each other. Continuous soft data are presented by reservoir temperature from thermal couples installed along wellbores of the same surveillance wells and time-lapse differences in wave acoustic impedances from 4D seismic surveys, which are accumulated in large amounts during field development. Over time, the CMG's thermal flow simulator, Steam, Thermal, and Advanced Processes Reservoir Simulator (STARS), simulates the change in spatial distributions of the reservoir temperature and other production variables, such as reservoir pressure, water, oil, and gas saturations. It has been shown that in thermally operated oil fields, the temperature's spatial distribution is determined by the

distribution of permeability (Duru and Horne, 2009 and 2010; Li and Zhu, 2010 and 2011). Thus, temperature observations can be used to estimate permeability. A pressure- and temperature-dependent Gassmann's fluid substitution model is proposed for the petroelastic model to generate synthetic time-lapse primary (P-) and secondary (S-) wave seismic velocities and associated acoustic impedances (Gassmann, 1951; Batzle and Wang, 1992; Kumar, 2006). The petroelastic model is a function of the porosity and production variables mentioned above. The theoretical and practical findings support the evident inverse relationship between the porosity and seismic attributes (Johnston et al, 1992; Eastwood et al, 1994; Lumley, 2001; Yuh, 2003; Zhang et al, 2005; Dong et al, 2006; Roste and Husby, 2006; Skjervheim, 2007; Myrseth, 2007; Mavko et al, 2009; Fahimuddin, 2010). Hence, acoustic impedance can be used to estimate porosity. All model parameters and state variables are linked together through the proposed data assimilation algorithm.

The following measures are devised to minimize the computational time associated with the EnKF-based data assimilation algorithm. The localization of updating and covariance matrices is proposed to (1) assimilate exhaustive seismic data, (2) decrease ensemble size, and (3) increase estimation quality. The shortcut based on the ensemble mean and co-simulation of state variables conditional to model parameters is examined to decrease the number of flow simulation runs to one. The proposed methodology is also applied to the realistic 3D thermal SAGD case study adapted from the Tucker thermal project (Husky Energy, 2010).

1.4 Thesis Outline

This thesis consists of seven chapters and one appendix. Chapter Two is devoted to an overview of deterministic numerical modeling and, more specifically, to modeling the petroleum reservoir. The basis of numerical modeling of dynamic reservoir systems is highlighted. The systems' mass, momentum, and energy governing equations are discussed thoroughly. The SAGD mechanism and temperature-dependent petroelastic

model based on Gassmann's theory are shown. The importance and properties of data available at the field are also shown.

Chapter Three describes a probabilistic approach to petroleum reservoir modeling, which differs from the deterministic approach in that it is capable reporting estimation uncertainty. A geostatistical framework that is very common in most mapping techniques of geological properties is presented. The theoretical background of the proposed inverse modeling technique EnKF and its general algorithm are recalled. The chapter includes a literature review on the application of the method in various fields and specifically in petroleum reservoir model characterization. EnKF's most frequently encountered modifications are summarized.

Chapter Four proposes a methodology for integrating hard porosity and permeability data and continuous soft temperature observations and time-lapse differences in acoustic impedances to reservoir models by means of EnKF. Conventional and proposed algorithms are compared. There is a discussion of the advantages and disadvantages of both algorithms. The chapter also includes validation measures used to assess the quality of estimates or accuracy of the estimated model.

Chapter Five comprises implementation details of EnKF carried out on the synthetic 2D SAGD example. A series of sensitivity analysis studies is conducted to examine the characteristics of the inverse modeling technique. These studies focus on which model types are most suitable for the EnKF-based data assimilation method, importance of the initial ensemble, and how it should be set up. The studies also consider the required relationship between the number of realizations in the ensemble and the number of integrated data, as well as solutions for decreasing the computational overburden (localization of updating and covariance matrices at the updating step and application of the shortcut at the flow simulation step). The influence of measurement error and grid effect on estimation results is briefly mentioned. It is found that this methodology with proposed alteration techniques for reducing computational time is effective and leads to better estimates. Incorporating any relevant additional data improves estimation results. Additional soft data may significantly improve understanding of the reservoir

geology. Differences in acoustic impedances bear more information about the reservoir than temperature data, due to larger spatial coverage. As a result, the differences in acoustic impedances produce better estimates. Porosity and permeability estimates derived from hard data are usually more accurate than the estimates from soft data. In some cases soft data produce better results. At any rate, the trade-off between model quality and sampling cost should be taken into account.

Chapter Six shows the successful application of proposed data assimilation methodology to a realistic 3D SAGD case study. The production scheme is adapted from the publicly available reservoir management report (RMR) on the Tucker thermal project prepared by Husky Energy Company for the Energy Resources Conservation Board (ERCB) (Husky Energy, 2010). It is found that assimilating additional temperature data improves estimates even when the shortcut is applied. But porosity and permeability data are redundant for specific observation well configurations, and one of them should be omitted in the estimation.

Finally, Chapter Seven presents the conclusion of the conducted study, discusses possible modifications to the proposed algorithm, and sets out future work. The EnKF may be successfully applied for characterizing categorical variables including facies, whose understanding is vital for predicting fluid flow.

A bibliography is presented at the end of all chapters. The Appendix contains descriptions of FORTRAN programs and associated parameter files written and used for this research.

Chapter 2 – Conventional (Deterministic) Approach to Petroleum Reservoir Modeling

This chapter covers numerical modeling and its place in petroleum reservoir characterization. A deterministic modeling approach is presented. Constituent parameters and states of the petroleum reservoir model are discussed. The relationship between these variables is presented analytically. The chapter also shows the governing equations of conventional fluid flow, an additional equation of energy conservation for thermal oil recovery, and the mechanism of the steam assisted gravity drainage (SAGD) method for heavy oil/bitumen extraction. The chapter includes a description of the framework of a pressure- and temperature-dependent petroelastic model. This model is based on Gassmann's fluid substitution theory. It is employed to relate the seismic response of the earth's interior to geological, petrophysical, thermal, and elastic properties of a rock and its constituent minerals. This chapter ends with a discussion of the characteristics of hard and soft data, data quality determined by scale and resolution, and measurement error, all of which are important in estimation and data assimilation.

2.1 Fundamentals of Numerical Modeling

Reservoir management heavily depends on high quality numerical models of petroleum reservoirs. The models' contribution to reservoir management is valuable and leads to better decision-making and improved oil recovery. Thus, plausible but simple models are sought. Associated modeling uncertainty should be quantified in some fashion as well (Satter and Thakur, 1994).

A petroleum reservoir is a dynamic system that is characterized by static geological, petrophysical, physical, chemical, and geomechanical properties, which are collectively called model parameters. The system is also characterized by dynamic flow and pressure states of the model, called state variables. Static model parameters and dynamic state variables are related to each other through a set of governing equations. It is not

possible for an investigator to know the full extent of true spatial distribution of reservoir geology and mechanism of fluid flow through rock medium (Deutsch, 2002). Analytical and/or numerical models may be constructed in order to mimic reality and characterize the petroleum reservoir. An analytical representation of natural phenomena is usually cultivated in deterministic fashion, which cannot adequately capture uncertainty in variations of reservoir parameters and state variables in time and space. Numerical modeling, on the other hand, discretizes the system into blocks, treating each block separately as a part of whole system that helps to capture geological heterogeneity and predict reservoir performance (Aziz and Settari, 1979). Both analytical and numerical models are constrained to all available information, which is obtained in the form of measurements and observations, conceptual models, previously constructed models, professional experience, etc., in order to be plausible and useful for future applications. The methodology for integrating data into a model is called a modeling method or technique (Carlson, 2003).

Numerical models are used more extensively than analytically derived models due to the associated ease of conducting a simulation study and sensitivity analysis (Carlson, 2003). Traditionally, a deterministic approach to numerical modeling is used to characterize a petroleum reservoir, where a single realization of a petroleum reservoir model is studied and all predictions are made using only this single model (Lake, 1989). In order to assess uncertainty in the estimation and prediction, a probabilistic or stochastic approach to reservoir modeling has been proposed, in which the ensemble of model realizations are derived based on Monte Carlo Simulation (MCS) method (Metropolis and Ulam, 1949; Deutsch, 2002). However, the benefit of additional information derived from multiple realizations and used for risk quantification is sacrificed against the dramatic increase of computational time. Full physics may be replaced with a proxy model to decrease the computational cost. To sum up, stochastic modeling provides greater insight into the petroleum reservoir and, hence, it should provide better support for reservoir management, if time is not an issue.

2.1.1 Analytical Modeling

Analytical models are based on analytical equations and form a basis for numerical modeling of dynamic systems. A dynamic system is a sophisticated object that evolves in time and space. It consists of a set of two types of spatial variables mathematically presented by vectors: model parameters m, which presumably do not vary in time (Equation (1)), and state variables s_t, which do vary in time (Equation (2)) (Aanonsen et al, 2009). There are N_m types of model parameters and N_s types of state variables. These spatial variables are a function of location in the space, which is denoted by x, y, and z coordinates in the Cartesian coordinate system. A petroleum reservoir is a good example of the dynamic system. Facies, porosity, and permeability represent model parameters; and fluid saturation, production rates, and seismic response are state variables. Model operator F relates all state variables s_t to model parameters m and constrains state variables to initial s_0 and boundary conditions $s_{b,t}$ (Equation (2)). Governing equations of petroleum reservoirs represent the model operator. For thermally operated oil reservoirs, which may be additionally characterized by temperature and seismic data, the model operator F is represented by flow equations, mass balance, energy conservation laws, and the wave propagation mechanism (also known as the petroelastic model). Values of model parameters and state variables are constrained to available data d_t sampled at time t, whose vector size $N_{d,t}$ x 1 varies in time (Equation (3)). The data is a measurement of either model parameter or state variable. Therefore, it can be related to model parameters and boundary and initial conditions through data operator G, as shown in the Equation (3). The data are characterized by mean value and measurement error, which can be used for uncertainty quantification. Thus, measurement error is not accounted for in deterministic modeling approaches. When point data are used to constrain the distribution of spatial variables, inverse modeling is considered. Any history-matching modeling techniques should produce such estimates of model parameters and state variables that honour data d_t, $t = t_0, \ldots, t_{Nt}$, from all time steps and should minimize any mismatch between estimates and truth. Model parameters, state variables, and data vectors can be combined together to one vector u_t shown in Equation (4), which represents the entire dynamic model. This model u_t, constrained to the data, aims to assist reservoir management. Size N_v of model

u_t is denoted by the sum of rows in vectors of model parameters and state variables (Equation (5)).

$$m = \left[m(x, y, z) \right]_{N_m \times 1} \tag{1}$$

$$\begin{cases} s_t = \left[s(x, y, z; t) \right]_{N_s \times 1} = \left[F(m, s_0, s_{b,t}; t) \right]_{N_s \times 1} \\ s_0 = \left[s(x, y, z; t_0) \right]_{N_s \times 1} \\ s_{b,t} = \left[s_b(x, y, z; t) \right]_{N_s \times 1} \end{cases} \tag{2}$$

$$d_t = \left[d(x, y, z; t) \right]_{N_{d,t} \times 1} = \left[G(m, s_0, s_{b,t}; t) \right]_{N_{d,t} \times 1} \tag{3}$$

$$u_t = \begin{bmatrix} m \\ s_t \\ d_t \end{bmatrix}_{(N_m + N_s + N_{d,t}) \times 1} = \begin{bmatrix} m \\ F(m, s_0, s_{b,t}; t) \\ G(m, s_0, s_{b,t}; t) \end{bmatrix}_{(N_m + N_s + N_{d,t}) \times 1} \tag{4}$$

$$N_v = N_m + N_s \tag{5}$$

2.1.2 Numerical Modeling

A numerical model of a dynamic system is the discretized representation of its analytical form, where spatial variables are partitioned into a finite number of blocks N_b in three-dimensional space, or into cells in two-dimensional space, or into intervals in one-dimensional space. Blocks form a model grid, which is used to illustrate the spatial distribution of any variable at specific time *t*. Model variables can be continuous, discrete, or categorical (Deutsch, 2002). Discretized values of the variables are assigned to the center of each block. Thus, every block contains a set of model parameters and state variables. Governing equations are applied to every single block to relate discrete values of model parameters and state variables (Aziz and Settari, 1979). Data is integrated into the model based on a chosen modeling technique, and the entire model is updated. Data locations are adjusted to the closest block centers. Support of the data

should be brought to the support of the model, if scale of the data and scale of the model are different.

Blocks are the smallest units of a model grid. Depending on the regularity of the block's shape and size, the grid may be structured or unstructured. A structured model grid with regular blocks and parallel layers is used in this thesis. The schematic is shown in Figure 1 with the Cartesian coordinate system. The coordinate system's axes are aligned with the model's lateral surfaces. All the blocks have the same regular size, which is represented by length b_x, width b_y, and height b_z. The spatial location of a block is characterized through three Cartesian coordinates (x_{ib}, y_{ib}, z_{ib}) of its center (black dot in Figure 1) or through the block order number i_b. A block's order number is counted from the origin of the model grid (green dot) in the following manner: first, from the negative to the positive poles of the *X* axis; second, from the negative to the positive poles of the *Y* axis; and third, from the negative to the positive poles of the *Z* axis of the coordinate system as in GSLib convention (Deutsch and Journel, 1998).

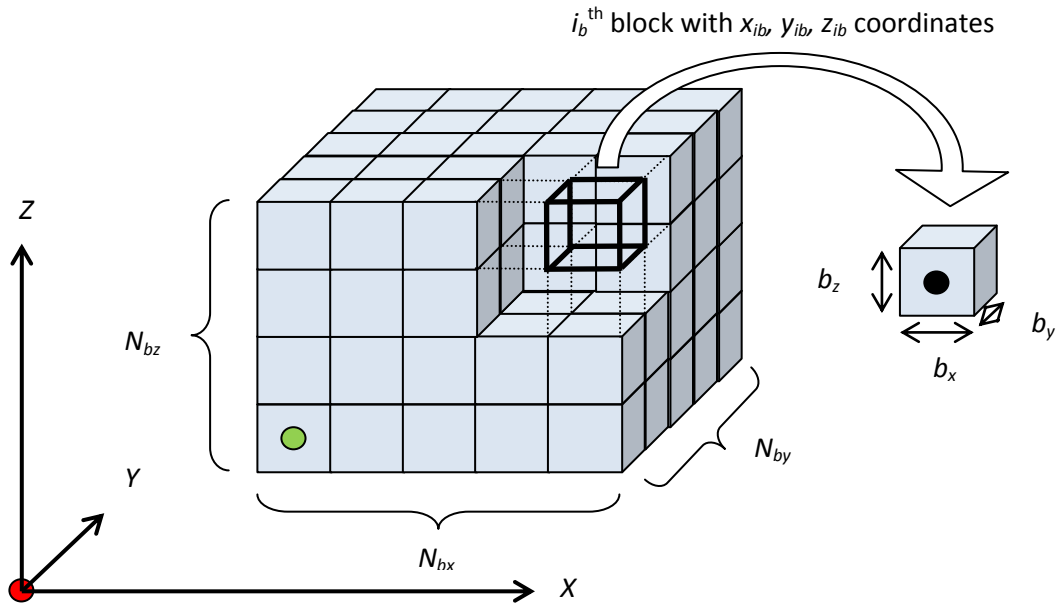


Figure 1: Schematic of the model grid used for numerical modeling throughout this thesis

In order to establish a relationship between the block coordinates and the block order number (Equation (6)), it is necessary to know the Cartesian coordinates (x_0, y_0, z_0) of origin in the coordinate system, which is shown as a red dot, and the numbers of blocks N_{bx}, N_{bx}, and N_{bz} in *X*, *Y* and *Z* directions respectively. The total number of blocks N_b can be computed as shown in Equation (7).

$$i_b = 1 + \left(\frac{x_{i_b} - x_0}{b_x} \right) + N_{by} \cdot \left(\frac{y_{i_b} - y_0}{b_y} \right) + N_{bx} \cdot N_{by} \cdot \left(\frac{z_{i_b} - z_0}{b_z} \right) \quad (6)$$

$$N_b = N_{bx} \cdot N_{by} \cdot N_{bz} \quad (7)$$

In the light of numerical modeling, continuous model parameters $\boldsymbol{m}$ and state variables $\boldsymbol{s}_t$ are discretized according to the established grid. Discretized values of the variables are stored in column vectors in predefined order, where all values of one variable are followed by values of another variable in the same order as the blocks organized in the model grid. The column vectors in this section have larger numbers of rows than the column vectors in the previous section. If the number of types of model parameters used in modeling is N_m and number of types of state variables is N_s, then $\boldsymbol{m}$ and $\boldsymbol{s}_t$ will have vector forms as shown in Equations (8) and (9). The point data of number $N_{d,t}$, sampled at time *t,* are stored in column vector $\boldsymbol{d}_t$ as in Equation (10). They are assigned to centers of relevant blocks. Note that $N_{d,t} \leq N_m \cdot N_b + N_s \cdot N_b$. Although the data vector is not complete, the data values are placed in a vector in the same order as the model parameters and the state variables in accordance with the block order in the model grid.

$$\boldsymbol{m}^T = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,N_b} & m_{2,1} & m_{2,2} & \cdots & m_{2,N_b} & \cdots & m_{N_m,1} & m_{N_m,2} & \cdots & m_{N_m,N_b} \end{bmatrix}_{1 \times N_m \cdot N_b} \quad (8)$$

$$\boldsymbol{s}_t^T = \begin{bmatrix} s_{t,1,1} & s_{t,1,2} & \cdots & s_{t,1,N_b} & s_{t,2,1} & s_{t,2,2} & \cdots & s_{t,2,N_b} & \cdots & s_{t,N_m,1} & s_{t,N_m,2} & \cdots & s_{t,N_s,N_b} \end{bmatrix}_{1 \times N_s \cdot N_b} \quad (9)$$

$$\boldsymbol{d}_t^T = \begin{bmatrix} d_{t,1} & d_{t,2} & \cdots & d_{t,N_{d,t}} \end{bmatrix}_{1 \times N_{d,t}} \quad (10)$$

where ^{T} is the matrix transpose operator; $\boldsymbol{m}_{im,ib}$ is the value of i_m^{th} model parameter at i_b^{th} block (i_m = 1, …, N_m, i_b = 1, …, N_b); $\boldsymbol{s}_{t,is,i}$ is the value of i_s^{th} state variable at i_b^{th} block at time t (i_s = 1, …, N_s, i_b = 1, …, N_b); $\boldsymbol{d}_{t,idt}$ is the i_{dt}^{th} value of measured model parameter or state variable at time t (i_{dt} = 1, …, $N_{d,t}$).

Finally, all the variables and data are combined into one column vector called the state vector $\boldsymbol{u}_t$ (augmented vector or model vector) as shown in Equation (11) (Aanonsen et al, 2009), which is similar to the analytical representation of the model from Equation (4). The difference lies in the increased number of rows. Even though model parameters and state variables are still function of coordinates, their spatial positions are preferably specified by block order number i_b. State vector $\boldsymbol{u}_t$ fully characterizes the discretized numerical model of a dynamic system at certain time t. The set of state vectors representing the model at different time steps with the ability to derive state vector values for the next time step can be called a numerical model of a dynamic system. Assimilated data update the entire model $\boldsymbol{u}_t$ and the relationship between model parameters and state variables is re-established through governing equations. Since the number of data $N_{d,t}$ changes with time, the model size also varies in time accordingly.

$$\boldsymbol{u}_t = \boldsymbol{u}(\boldsymbol{m}, \boldsymbol{s}_0, \boldsymbol{s}_{b,t}; t) = \begin{bmatrix} \boldsymbol{m} \\ \boldsymbol{s}_t \\ \boldsymbol{d}_t \end{bmatrix}_{\left(N_m \cdot N_b + N_s \cdot N_b + N_{d,t}\right) \times 1} = \begin{bmatrix} \boldsymbol{m} \\ F(\boldsymbol{m}, \boldsymbol{s}_0, \boldsymbol{s}_{b,t}; t) \\ G(\boldsymbol{m}, \boldsymbol{s}_0, \boldsymbol{s}_{b,t}; t) \end{bmatrix}_{\left(N_m \cdot N_b + N_s \cdot N_b + N_{d,t}\right) \times 1} \tag{11}$$

2.2 Model Variables: Parameters and States

Most frequently encountered model parameters and state variables of a petroleum reservoir model are summarized in Table 1 by the author. Some may be either static or dynamic depending on the application purpose of these variables. Overall, this classification is subjective, and can be easily violated.

Table 1: List of general model parameters and state variables of petroleum reservoir models

<table>
<tr><th>Model parameters</th><th>State variables</th><th>Model parameters and state variables</th></tr>
<tr><td rowspan="2">Geological properties:
facies
porosity
permeability</td><td rowspan="2">Production parameters:
instantaneous and cumulative production fluid rate
bottomhole pressure
instantaneous and cumulative steam-oil ratio
reservoir temperature
gas-to-oil ratio
water cut</td><td rowspan="4">Seismic response (3D or 4D):
seismic velocity
acoustic impedance
amplitude
seismic profile
amplitude versus offset</td></tr>
<tr><td>Physical properties:
PVT curves
relative permeability curve</td></tr>
<tr><td>Geomechanical or elastic properties:
bulk and shear moduli</td><td>Petrophysical properties:
fluid saturations
water-oil, gas-oil or gas water contacts</td></tr>
</table>

Model parameters may be presented by geological, physical and geomechanical properties of the reservoir. In this thesis, physical and geomechanical properties are assumed to be known, and only geological properties are estimated. Geological properties frame the model that constrains and predicts the values of state variables. Thus, they are deemed the most essential element in petroleum reservoir modeling (Aanonsen et al, 2009). The most frequently encountered representatives of geological properties are facies or rock type (π), porosity (ϕ), and permeability tensor ($\vec{\vec{K}}$), which are related to each other in some fashion. Facies classify rock types with distinct geological properties (Deutsch, 2002). Therefore, the distribution of facies plays a

determinative role in distribution of porosity and permeability. While facies is a categorical spatial variable, porosity and permeability are continuous ones. Porosity is a scalar and determines the ratio of void space in rock to its bulk volume. Its theoretical boundaries are between fractions of 0.0 and 1.0, but practically its range varies between 0.0 and 0.4. Effective porosity, which is a fraction of void space that contributes to fluid flow, is more important in understanding reservoir behaviour than total porosity, which includes some isolated pores that do not contribute to fluid flow. In this thesis, porosity is referred to as an effective porosity. Permeability is an inherent characteristic of rock and describes how easily fluid passes through a rock. Permeability is a tensor (two errors above letter *K*); i.e., its value changes with principal directions (*X*, *Y*, or *Z*). A gradient of pressure in one direction can induce flow in other directions. In this work, the value of permeability is calculated in one (usually *X*) direction. A pre-established relationship based on geological information is used to infer values of permeability in the remaining directions. A lower boundary of permeability values is restricted by 0.0 mD (millidarcy). An upper boundary is less constrained and can be up to several hundred thousand Darcy (Chen, 2007). Porosity is strongly related to the local distribution of grain size, while permeability is a strong function of a local pore size (Lake, 1989). Thus, there is a certain relationship between these two geological properties. To conclude, once geological properties of a model are estimated honouring available data, it can be claimed that the most important static part of the petroleum reservoir model is estimated as well.

Physical properties, such as a phase transition diagram or PVT curves and relative permeability curves for water, oil, and gas, may be treated as model parameters, but omitted in geomodeling and assumed to be known. Geomechanical (elastic and plastic) properties of rock and constituent minerals are also examples of model parameters. They are considered in fracture modeling, hydro fracture procedure for flow stimulation and interpretation and inversion of seismic signal. A grain size analysis and mineralogical study may be conducted as well, in order to better understand the geomechanical properties of rock. Various dimensional seismic surveys such as 1D, 2D, and 3D seismic surveys help to reveal geological structure. Thus, a reservoir's acoustic properties may represent static parameters of a model (Aziz and Settari, 1979; Carlson, 2003).

State variables are usually presented by, but not limited to, production parameters that vary in time, for example production rate, cumulative oil production (COP), cumulative steam oil ratio (CSOR), gas-to-oil ratio, reservoir pressure, in-situ temperature, and petrophysical properties such as fluid saturations or water-oil contact. Repeated time-lapse or 4D seismic surveys are able to track changes in fluid saturation. Hence, time-lapse seismic properties of a reservoir are a source of dynamic information, which can be treated as a state of a reservoir (Lumley, 2001).

2.3 Model Operator

Model operator *F* relates state variables to model operators, and initial and boundary conditions. The flow simulator is an example of model operators used in petroleum reservoir engineering. Governing equations of fluid flow in a reservoir are at the heart of a petroleum reservoir's model operator. Model operator *F* in the proposed EnKF-based data assimilation methodology for characterization of thermally operated petroleum reservoirs, consists of three parts: a geological part that relates permeability to porosity values, a thermal flow simulation part, and a petroelastic model. The geological component is discussed later. This section deals with equations for conventional and thermal oil recoveries. SAGD fundamentals are summarized, since all case studies are built within its framework. SAGD's importance and wide application to heavy oil and bitumen extraction in northern Alberta stipulate its study. The pressure- and temperature-dependent petroelastic model based on Gassmann's theory for generating synthetic seismic attributes is presented here as well.

2.3.1 Conventional Oil Recovery

In order to describe how fluid flows through a petroleum reservoir, it is important to understand the reservoir's structure. A petroleum reservoir consists of an immovable component, which is reservoir rock, and movable fluid components. The fluid components consist of accumulated in-situ oil, gas, and water, and injected alien steam, water, and chemicals including solvents and polymers. A combination of fluids can be

found in three main phases – water or aqueous, oil or oleic, and gaseous. For sake of the simplicity, a "black oil" model is assumed, where fluids flowing through petroleum reservoir pores are either water, oil, or gas. It is further assumed that some fractions of water and oil cannot be displaced and, hence, are referred to as connate water and oil. Thus, the petroleum reservoir model can be thought of as a dynamic multiphase flow system consisting of three fluid phases (aqueous, oleic, and gaseous) and a solid rock phase (Lake, 1989).

Fluid flow governing equations are based on three laws. They are (1) the law of mass conservation; (2) the law of momentum conservation or Darcy`s law; (3) and the law of energy conservation (Carlson, 2003). The first two laws are implemented when modeling the conventional recovery processes, and are described in this section. The last law of energy conservation is discussed in the section devoted to thermal recovery. It is assumed that the phase transition diagram or PVT curves and relative permeability curves are well-established in multiphase flow systems. The contribution of a solid rock phase into mass balance and flow equation is ignored, since the rock phase is assumed to be immovable (Aziz and Settari, 1979; Lake, 1989).

The petroleum reservoir's governing equations are embedded in any flow simulator. CMG's commercial thermal flow simulator, STARS (Steam, Thermal, and Advance Processes Reservoir Simulator), is used in this thesis to predict reservoir performance. The initial and boundary conditions, oil production scheme, geological, petrophysical, physical, chemical, and geomechanical properties of rocks and fluids are specified in STARS to run simulation. The simulation's output is presented in the form of reservoir pressure and temperature, water, oil, and gas saturations. The output is analysed later on in order to characterize the reservoir and better predict its performance. The laws of conservation of mass and momentum are described below (Aziz and Settari, 1979; Lake, 1989).

Mass conservation or mass balance for the i^{th} fluid phase for the isothermal process (at constant temperature) can be simply described as follows: a rate of accumulation of the i^{th} phase in a grid block of volume V has to be equal to the difference between the rate

of the i^{th} phase transported into the block and the rate of the i^{th} phase transported from a block plus the rate of generation of the i^{th} phase in a block. It can be expressed in the following form:

$$\frac{\partial W_i}{\partial t} = -\vec{\nabla}\vec{\Phi}_i + (R_i + q_i), \quad i = w, o, g \tag{12}$$

$$\vec{\nabla}\vec{\Phi}_i = \frac{\partial \Phi_{i,x}}{\partial x} + \frac{\partial \Phi_{i,y}}{\partial y} + \frac{\partial \Phi_{i,z}}{\partial z} \tag{13}$$

where W_i is the overall concentration of the i^{th} fluid phase in units of the mass of the i^{th} phase per unit bulk volume; $\vec{\nabla}$ is the divergence operator; $\vec{\Phi}_i$ is the flux of the i^{th} fluid phase in units of the mass of the i^{th} phase per surface area-time; R_i is the mass rate of the i^{th} fluid phase generation due to chemical and other processes in units of the mass of the i^{th} phase per bulk volume-time ($R_i > 0$ stands for phase generation, and $R_i < 0$ means phase destruction); q_i is the injection (source, $q_i > 0$) or production (sink, $q_i < 0$) mass rate of the i^{th} phase per unit volume per unit time; and t is the time.

The concentration W_i of the i^{th} phase can be described through effective reservoir porosity ϕ, i^{th} phase saturation S_i (which is defined as a fraction of void porous space occupied by i^{th} phase), and i^{th} phase density ρ_i (which depends both on the reservoir pressure and temperature):

$$W_i = \phi \cdot S_i \cdot \rho_i \tag{14}$$

Flux $\vec{\Phi}_i$ of the i^{th} phase can be expressed through density ρ_i and superficial or Darcy velocity $\vec{v}_i$ of this phase:

$$\vec{\Phi}_i = \rho_i \cdot \vec{v}_i \tag{15}$$

And the mass rate of i^{th} phase generation R_i can be expressed through effective reservoir porosity ϕ, i^{th} phase saturation S_i and the rate of generation of the i^{th} phase r_i ($r_i > 0$ – phase generation, $r_i < 0$ – phase destruction):

$$R_i = \phi \cdot S_i \cdot r_i \tag{16}$$

Constraints are imposed on phase saturations and the mass rate of phase generations:

$$\sum_{i=w,o,g} S_i = 1.0 \tag{17}$$

$$\sum_{i=w,o,g} R_i = 0.0 \tag{18}$$

The pressure-dependent formation volume factor B is introduced to define how the volume of phase changes when conditions are changed from reservoir to standard (surface or tank). It is assumed that oil and gas are immiscible and that the gaseous phase can be present only with the oil phase. When that happens, the formation volume factor for every phase has the following form:

$$B_w = \frac{(V_w)_{RC}}{(V_w)_{SC}} = f(P_w) \tag{19}$$

$$B_o = \frac{(V_o + V_{dg})_{RC}}{(V_o)_{SC}} = f(P_o) \tag{20}$$

$$B_g = \frac{(V_g)_{RC}}{(V_g)_{SC}} = f(P_g) \tag{21}$$

where the subscript $_{dg}$ is the dissolved gas; subscript $_{RC}$ is the reservoir conditions; and subscript $_{SC}$ is the standard conditions.

The dissolved gas-to-oil ratio can be defined as follows:

$$R_{dg} = \left(\frac{V_{dg}}{V_o} \right)_{SC} = f\left(P_o\right) \tag{22}$$

If formation volume factors *B* and the gas-to-oil ratio are used to establish the relationship between densities at reservoir and standard conditions, the following expression will be found:

$$\rho_w = \frac{1}{B_w}\left(\rho_{wSC}\right) \tag{23}$$

$$\rho_o = \bar{\rho}_o + \bar{\rho}_{dg} = \frac{1}{B_o}\left(\rho_{oSC}\right) + \frac{R_{dg}}{B_o} \cdot \left(\rho_{gSC}\right) \tag{24}$$

$$\rho_g = \frac{1}{B_g}\left(\rho_{gSC}\right) \tag{25}$$

where ρ_i is the density of the i^{th} phase at reservoir conditions; and ρ_{iSC} is the density of the i^{th} phase at standard or surface conditions.

Thus, mass conservation equations for every phase have the following forms:

$$\frac{\partial}{\partial t}\left[\frac{\phi \cdot S_w}{B_w}\right] = -\vec{\nabla}\left[\frac{\vec{v}_w}{B_w}\right] + q_{V,w} \tag{26}$$

$$\frac{\partial}{\partial t}\left[\frac{\phi \cdot S_o}{B_o}\right] = -\vec{\nabla}\left[\frac{\vec{v}_o}{B_o}\right] + q_{V,o} \tag{27}$$

$$\frac{\partial}{\partial t}\left[\frac{\phi \cdot R_{dg} \cdot S_o}{B_o} + \frac{\varphi \cdot S_g}{B_g}\right] = -\vec{\nabla}\left[\frac{R_{dg} \cdot \vec{v}_o}{B_o} + \frac{\vec{v}_g}{B_g}\right] + q_{V,g} \tag{28}$$

where $q_{V,i}$ is the relative volume flow rate ($q_{V,I} > 0$ for production and $q_{V,I} < 0$ for injection) of the i^{th} phase at standard conditions per unit reservoir volume-time, whose expression for the i^{th} phase is shown below:

$$q_{V,i} = \frac{q_i}{\rho_{iSC}} \tag{29}$$

Finally, if Equations (26) – (28) are combined and summed up over all phases, the resulting equation of total mass conservation or the equation of continuity has the following form:

$$\begin{aligned}\frac{\partial}{\partial t}\left[\frac{\phi \cdot S_w}{B_w} + \frac{\phi \cdot S_o}{B_o} + \left(\frac{\phi \cdot R_{dg} \cdot S_o}{B_o} + \frac{\phi \cdot S_g}{B_g}\right)\right] = \\ -\vec{\nabla}\left[\frac{\vec{v}_w}{B_w} + \frac{\vec{v}_o}{B_o} + \left(\frac{R_{dg} \cdot \vec{v}_o}{B_o} + \frac{\vec{v}_g}{B_g}\right)\right] + q_{V,w} + q_{V,o} + q_{V,g}\end{aligned} \tag{30}$$

The second governing equation is Darcy's law, which defines the relationship between flow rate and pressure gradient for every phase (Aziz and Settari, 1979). The four main mechanisms that initiate fluid flow in porous media are viscous forces justified by pressure, gravity forces initiated by density, dispersion or diffusion caused by concentration gradients, and capillary forces due to curved boundaries between different phases (Lake, 1989). All of them can be approximately described by Darcy's law, which is also known as the flow equation. Its multiphase form is shown below:

$$\vec{v}_i = -\lambda_{Ri} \cdot \vec{\vec{K}} \cdot \left(\nabla P_i - \rho_i \cdot \frac{g}{g_c} \cdot \nabla z\right), \quad i = w, o, g \tag{31}$$

where λ_{Ri} is the relative mobility of the i^{th} phase, whose expression is shown in Equation (32); $\vec{\vec{K}}$ is the tensor representation of permeability in Darcy; K_{Ri} is the relative permeability of the i^{th} phase; P_i is the pressure within the continuous i^{th} phase in Pascal; g is the magnitude of the vector of gravitational acceleration pointing toward the centre of the Earth in m^2/s; g_c is the conversion constant; ∇ is the gradient operator; z is the vertical downward direction in m; K_{Ri} is the relative permeability of the i^{th} phase; and μ_i is the dynamic viscosity of the i^{th} phase in Pascal times second or Poise.

$$\lambda_{Ri} = \frac{K_{Ri}}{\mu_i} \tag{32}$$

The mass conservation Equations (26) – (28) and Darcy's law in Equation (31) are combined together to approximately describe the multiphase flow of fluids in the isothermal reservoir environment as a function of reservoir pressure. Basic multiphase flow equations for every phase can be written as:

$$\frac{\partial}{\partial t}\left[\frac{\phi \cdot S_w}{B_w}\right] = \vec{\nabla}\left[\chi_w \cdot \left(\nabla P_w - \rho_w \cdot \frac{g}{g_c} \cdot \nabla z\right)\right] + q_{V,w} \tag{33}$$

$$\frac{\partial}{\partial t}\left[\frac{\phi \cdot S_o}{B_o}\right] = \vec{\nabla}\left[\chi_o \cdot \left(\nabla P_o - \rho_o \cdot \frac{g}{g_c} \cdot \nabla z\right)\right] + q_{V,o} \tag{34}$$

$$\begin{aligned} &\frac{\partial}{\partial t}\left[\frac{\phi \cdot R_{dg} \cdot S_o}{B_o} + \frac{\phi \cdot S_g}{B_g}\right] = \\ &\vec{\nabla}\left[R_{dg} \cdot \chi_o \cdot \left(\nabla P_o - \rho_o \cdot \frac{g}{g_c} \cdot \nabla z\right) + \chi_g \cdot \left(\nabla P_g - \rho_g \cdot \frac{g}{g_c} \cdot \nabla z\right)\right] + q_{V,g} \end{aligned} \tag{35}$$

where χ_i is the transmissibility of the i^{th} phase defined as below:

$$\chi_i = \frac{K_{Ri}}{\mu_i \cdot B_i} \cdot \vec{\vec{K}} \tag{36}$$

These resulting flow equations with boundary and initial conditions describe the performance of the reservoir in time. The examined volume of interest is discretized for numerical modeling. Flow equations are applied to every block. Partial derivatives are replaced with finite differences and equations are solved numerically and not analytically (Carlson, 2003). Thus, the behaviour of the multiphase fluid flow in the reservoir can be described as only approximately trying to mimic actual reservoir behaviour honouring all available data.

2.3.2 Thermal Oil Recovery

Thermal recovery methods are used to extract viscous oil or bitumen from underground and are widespread in northern Alberta (Husky Energy, 2010). In thermal recovery, the viscosity of oil is reduced by inducing heat into the reservoir (Lake, 1989). For this reason, the fundamentals of thermal recovery are reviewed and related equations are discussed in this section. The most popular heavy oil extraction technique, SAGD, is highlighted in detail later in the chapter. Several case studies are built on it in order to examine the proposed inverse modeling technique EnKF for continuous data integration.

In modeling the thermal recovery processes, the law of energy conservation is taken into account in addition to laws of mass and momentum conversation. Energy balance equations are based on the first law of thermodynamics, in which the rate of energy accumulated in a grid block of volume V equals the difference between the rate of energy transported into a block and the rate of energy taken from a block, plus the rate of energy production in the block (Lake, 1989). Here, the contribution of the solid rock phase cannot be ignored. Mathematically, the energy balance equation for every grid block of a small volume dV can be presented as:

$$\int_V \left\{ \frac{\partial}{\partial t} \left[\left(\phi \cdot \sum_{i=w,o,g} \rho_i \cdot S_i \cdot \Omega_i + (1-\varphi) \cdot \rho_s \cdot \Omega_s \right) + \frac{1}{2} \cdot \sum_{i=w,o,g} \rho_i \cdot |\vec{v}_i|^2 \right] + \vec{\nabla} \cdot \vec{\Psi} \right\} \cdot dV = \Pi + \int_V (q_{E,g} - q_{E,l}) \cdot dV \quad (37)$$

where the first term under the integral represents energy concentration per unit volume; the second term stands for kinetic energy per unit volume; $\vec{\Psi}$ is the energy flux; Ω_i is the internal energy per unit mass of the i^{th} phase, which is shown in Equation (38) (Chen, 2007); Π is the external work done by the system against the reservoir pressure field (work Π_P – Equation (41)) and gravitational forces (work Π_G – Equation (42)) due to the system's expansion; $q_{E,g}$ and $q_{E,l}$ are the energy source (heat gain) and sink (heat loss to overburden or underburden) per unit volume-time; and subscript $_s$ indicates the solid or rock phase. All other energy sources, such as energy coming from an internal reaction or vaporization, are implicitly depicted on the left-hand side of Equation (37).

$$\Omega_i = C_{Vi} \cdot T, \quad i = w,o,g,s \quad (38)$$

$$H_i = C_{Pi} \cdot T, \quad i = w,o,g,s \quad (39)$$

$$H_i = \Omega_i + \frac{P_i}{\rho_i}, \quad i = w,o,g,s \quad (40)$$

where Ω_i and H_i are the internal energy and enthalpy of the i^{th} phase per unit mass, respectively; C_{Vi} and C_{pi} are the specific heat capacity at constant volume and constant pressure of the i^{th} phase, respectively.

$$\Pi_P = -\int_V \sum_{i=w,o,g} \vec{\nabla} \cdot (P_i \cdot \vec{v}_i) \cdot dV \quad (41)$$

$$\Pi_G = \int_V \sum_{i=w,o,g} \rho_i \cdot \vec{v}_i \cdot \frac{g}{g_c} \cdot \nabla z \cdot dV \quad (42)$$

Energy flux $\vec{\Psi}$ represents heat transfer between phases and can be decomposed into flux initiated by convection (fast heat transfer due to flowing phases), conduction (slow heat transfer due to physical contact between relatively stationary phases), and radiation:

$$\vec{\Psi} = \vec{\Psi}_{convection} + \vec{\Psi}_{conduction} + \vec{\Psi}_{radiation} = \sum_{i=w,o,g} \rho_i \cdot \vec{v}_i \cdot \left(\Omega_i + \frac{1}{2} \left| \vec{v}_i \right|^2 \right) - \kappa_T \cdot \nabla T + \vec{\Psi}_{radiation} \quad (43)$$

where κ_T is the total thermal conductivity, which can be expressed through fluid and rock thermal conductivities:

$$\kappa_T = \kappa_f^{\varphi} \cdot \kappa_s^{1-\varphi} \quad (44)$$

In most thermal recovery methods, the contribution of radiation flux $\vec{\Psi}_{radiation}$ is comparatively small and, hence, it may be excluded from the equation.

Equations (37) – (44) are combined to produce a final energy conservation equation applied to every grid block of the reservoir model (Lake, 1989):

$$\frac{\partial}{\partial t} \left[\phi \cdot \sum_{i=w,o,g} \rho_i \cdot S_i \cdot \Omega_i + (1-\phi) \cdot \rho_s \cdot \Omega_s + \frac{1}{2} \sum_{i=w,o,g} \rho_i \cdot \left| \vec{v}_i \right|^2 \right] + \vec{\nabla} \cdot \left[\sum_{i=w,o,g} \rho_i \cdot \vec{v}_i \cdot \left(\Omega_i + \frac{1}{2} \left| \vec{v}_i \right|^2 \right) \right] - \vec{\nabla} \cdot \left[\kappa_T \cdot \nabla T \right] + \sum_{i=w,o,g} \left(\vec{\nabla} \cdot \left(P_i \cdot \vec{v}_i \right) - \rho_i \cdot \vec{v}_i \cdot \frac{g}{g_c} \cdot \nabla z \right) = q_{E,g} - q_{E,l} \quad (45)$$

The notion of enthalpy H_i of the i^{th} phase can be introduced into the Equation (44) by substituting Ω_i with $H_i - P_i/\rho_i$ (Equation (40)).

2.3.3 Steam Assisted Gravity Drainage (SAGD)

The SAGD method is a steam-based enhanced oil-recovery technique developed to extract heavy oil or bitumen (Butler, 1991). It is a special type of steam flooding that is considered a tertiary recovery method (Carlson, 2003; Lake, 1989). Its wide application is found in areas with stacked fluvial channels and point bars located in northern Alberta, Canada. In this case, oil is not pushed by steam, but rather melted by steam and naturally flows to a well. Its main advantage over conventional steam flooding is a decrease in fingering occurrence (Butler, 1991).

Butler (1991) has thoroughly developed and studied the SAGD production mechanism. In this process, steam is generated at the surface facilities (also known as steam generators) at a high quality of close to 100%, and pumped into the reservoir through injection wells. Effectively using the heat stored in the steam is crucial to developing surface facilities and piping systems. The quality of steam decreases slightly when it reaches the bottom of the reservoir. The heat escapes at surface facilities and through vertical parts of the wells to the overburden. The well configuration for the SAGD oil extraction method uses a horizontal well pair, which consists of injection and production wells of 500.0 – 1000.0 m length and with 177.8 mm diameter located close to the reservoir bottom (Husky Energy, 2010). The injection well is placed approximately 5.0 m above the production well. In some cases the injection well may be replaced by a series of vertical wells drilled above a horizontal production well. Figure 2 shows a primitive schematic of a SAGD well pair. Several well pairs are grouped together in pads drilled from the same surface platform in order to cover a larger region of the reservoir, with well spacing varying between 50 and 110 m (Ito et al., 2000). Vertical surveillance wells are drilled to sample reservoir parameters such as porosity and permeability. These wells are also used to observe a change of dynamic variables in time, such as bottomhole pressure and reservoir temperature. Figure 3 shows a schematic plan view of a SAGD pad, horizontal well pairs, and vertical observation wells. Given sufficient reservoir pressure, oil with decreased viscous can be carried to the surface by a natural gas lift. If the reservoir pressure is not high enough, an artificial lift is implemented instead. Conventional sucker-rod pumps and electrical submersible pumps are less

desirable in SAGD, since the configuration of the former is not quite suitable for horizontal wells, and the latter consume a significant amount of electricity.

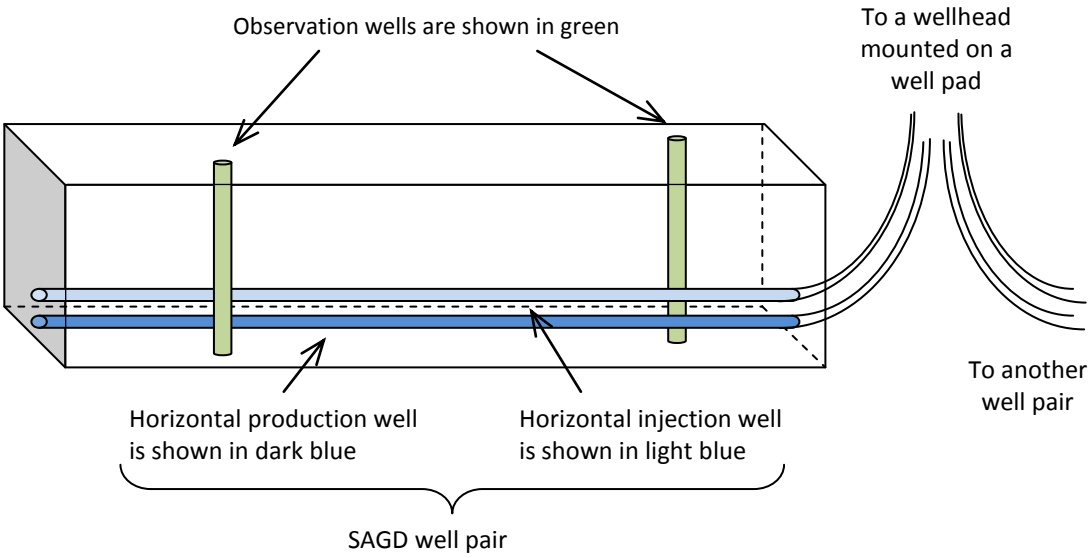


Figure 2: Schematic of a SAGD well pair and observation wells around it

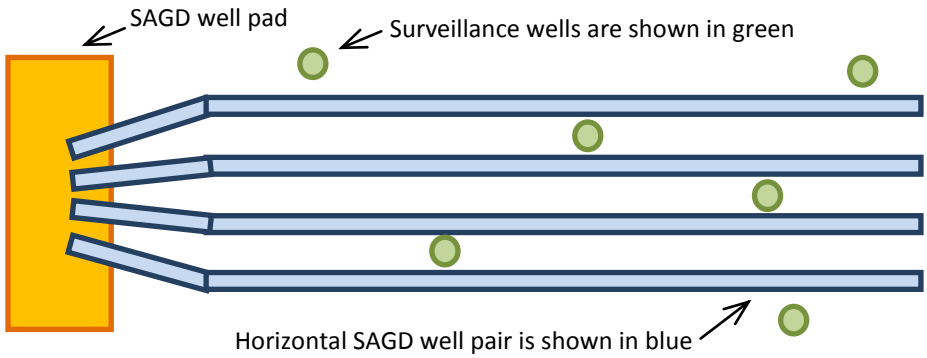


Figure 3: Schematic of plan view of a SAGD well pad, horizontal SAGD well pairs, and vertical observation (surveillance) wells

Figure 4 shows SAGD's flow mechanism in a cross section of a well pair (Butler, 1991). The preheating stage is conducted first. During that time, steam circulation heats the area around the well pair for about 30 – 90 days. The steam may be injected down

through the annulus and come back through tubing or vice versa. Once the bitumen in the region between two wells is warmed up sufficiently, steam circulation in the lower well is stopped. Steam injection into the upper well is continued, unless it is deemed economically infeasible to extract oil any further. Usually at the initial stage of steam injection into the reservoir, high pressure is applied to facilitate fluid flow to the production well. In SAGD, heated oil is replaced with steam due to natural gravitational forces. Because steam has a tendency to rise, it slowly replaces the heated bitumen. A zone of the reservoir occupied by the steam is called the steam chamber. It grows mostly upwards through the permeable rock, such as sand, in the beginning of the process, heating up and replacing more viscous oil until it hits the impermeable top of reservoir, such as shale. Then the chamber starts growing sideways, increasing the zone affected by the steam flooding. Steam condenses on the interface of cold oil and the affected zone. Heavier steam condensate and heated oil flow along the interface to the production well, mainly due to the force of gravity. This oil extraction process is maintained, preferably at constant pressure, until it is deemed no longer feasible. The SAGD mechanism can be simply summarized in four steps (Butler, 1991):

- Steam condenses at the interface of the steam chamber
- Oil and condensate drain to the production well at the bottom along the interface
- Flow is caused by gravity forces
- The chamber grows upwards and sideways at constant pressure

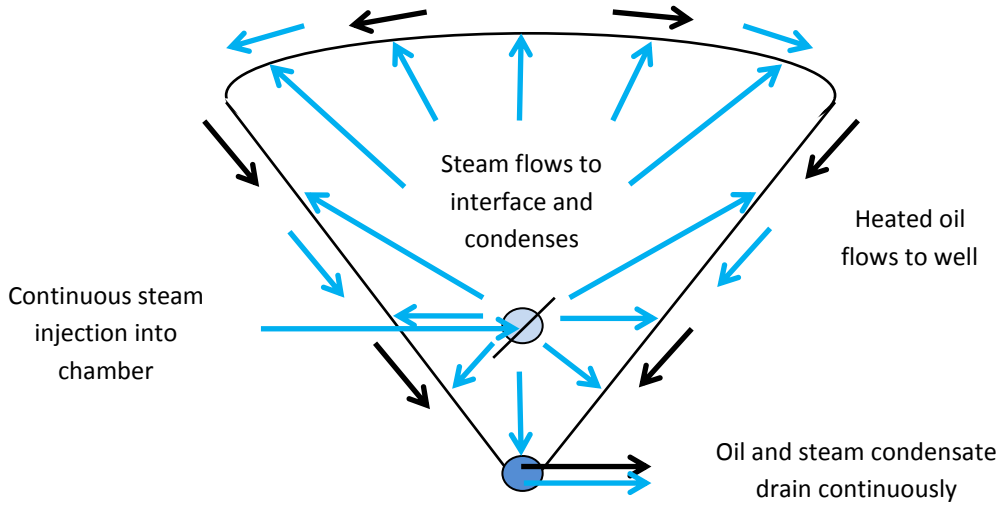


Figure 4: Mechanism of SAGD shown in the cross section of a well pair (Butler, 1991). Blue arrows represent steam or its condensate, black arrows are heated bitumen

The drainage rate from one side of a steam chamber in a homogeneous reservoir can be approximately expressed as shown in Equation (46), if heat is transferred only by means of conduction (Butler, 1991), which is a fair assumption. It is good to know that the bitumen drainage rate does not depend on the shape of the interface and depends only on the elevation *h*.

$$q = \sqrt{\frac{2 \cdot \phi \cdot \Delta S_o \cdot K \cdot g \cdot \alpha \cdot h}{\vartheta \cdot \mu_K(T)}} \tag{46}$$

where q is the drainage rate from one side of the steam chamber, $m^3/m*day$, in order to calculate the total drainage rate, q should be doubled; ϕ is the reservoir porosity; $\Delta S_o = S_{oi} - S_{or}$ is the difference between the initial and residual oil saturation; K is the magnitude of average reservoir permeability, 10^{-12} m^2 = 1 Darcy; g is the standard gravity, 7.32×10^{10} m/day^2; α is the thermal diffusivity, m^2/day; h is the vertical reservoir height, m; υ is the parameter with typical values between 3 and 4; and μ_K is the kinematic viscosity of the oil at the temperature T of the steam, m^2/day.

The position of the interface at a specific time can be roughly calculated as (Butler, 1991):

$$y' = 1 - \frac{1}{2}\left(\frac{t'}{x'}\right)^2 \tag{47}$$

where *y'* and x' are the dimensionless variables describing the interface's position in the vertical and horizontal directions of the cross section y' = *y*/*h*, x' = *x*/*h*; *t'* is the dimensionless time, the expression of which is shown below (Butler, 1991):

$$t' = \frac{t}{h}\sqrt{\frac{K \cdot g \cdot \alpha}{\phi \cdot \Delta S_o \cdot \vartheta \cdot \mu_K \cdot h}} \tag{48}$$

Figure 5 from Butler (1991) shows the possible propagation of interface surfaces in time in the homogeneous reservoir for a producer's varying position.

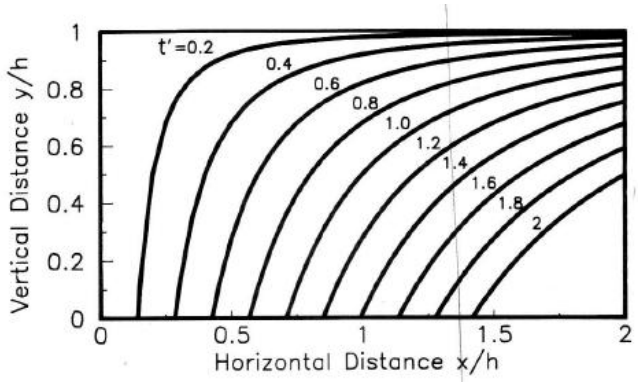


Figure 5: Movement of hot steam-cold bitumen interface curves in time for half of a simple SAGD model (Butler, 1991)

The idealistic growth of steam chambers in a homogeneous reservoir for a set of well pairs grouped in a single pad is shown in Figure 6, wherein each color line represents a succession of the growth in time. At the initial stage of field development, well pairs do not interact with each other and, thus, steam chambers can be studied separately. After

a while, the individual chambers intermingle and form a steam layer at the top of the reservoir close to the cap rock. The resulting effect of individual well pairs should lead to almost complete coverage of the reservoir volume. It should be noted that the shape of steam chambers may significantly differ between theory and practice because of reservoir heterogeneity, such as layering and discontinuities.

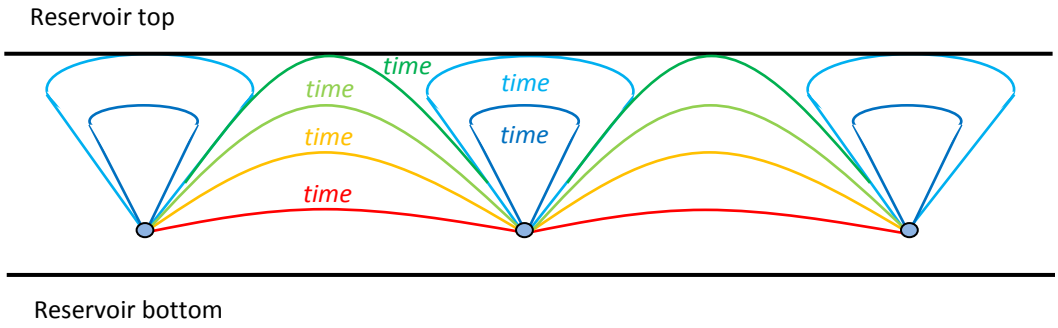


Figure 6: Schematic in cross section of growth of a series of steam chambers above adjacent and parallel horizontal well pairs drilled from a single well pad (Butler, 1991)

In this thesis, the SAGD simulation is carried out on the STARS thermal flow simulator, the governing equations of which have been described in previous sections of this chapter.

The influence of the presence of an aquifer and gas cap on SAGD recovery is studied as well. Water below the reservoir is not a problem for bitumen production, but water above the reservoir affects oil production. Gas above and inside the reservoir also affects bitumen production in some manner. The presence of a gas cap can substantially increase oil recovery. However, the concentration of dissolved gas is naturally very low in bitumen, and in some cases the gaseous phase may be ignored (Butler, 1991; Carlson, 2003).

Some work has been done in optimizing SAGD. One suggestion has been to add noncondensible gas (Butler, 1999; Jiang et al, 2000a, b) or gaseous additives such as methane (Ito et al, 2001) into the steam injection stream in order to increase bitumen

recovery and decrease energy used to generate steam. Palmgren and Edmunds recommend replacing steam with heated naphtha. They call this process naphtha assisted gravity drainage (NAGD), which is used as a diluent for bitumen pumping (Palmgren and Edmunds, 1995).

2.3.4 Pressure and Temperature Dependent Petroelastic Model

The framework of the petroelastic model is described here in detail. Recall that this model is a component of petroleum reservoir model operator *F*. The petroelastic part is based on pressure and temperature Gassmann's theory, which is adequate for SAGD applications.

Time-lapse seismic surveys, also known as 4D surveys, are employed as a reservoir management tool to a great extent (Lumley, 2001). These 4D seismic surveys are a set of repeated 3D or 2D seismic surveys over calendar time. While the first three dimensions are intended to characterize acoustic properties of saturated rock in space, the fourth dimension captures time. Properly acquired and interpreted time-lapse seismic data provide extensive spatial knowledge about the reservoir. The subsurface's geological characteristics, such as formation tops, faults, folds, flow barriers, and baffles, can be inferred. With the help of geological advances that improve the reservoir management process, it is possible to detect a reservoir's dynamic features, which include fluid flow, flood fronts, and thief zones. Time-lapse seismic attributes may capture the change in fluid saturations, reservoir pressure and temperature (Nur and Simmons, 1967; Wang and Nur, 1988). The main advantage of 4D seismic surveys over 3D is that the combined acoustic properties of rock and fluids that are difficult to differentiate can be separated in time-lapse seismic data. As a result, fluid flow can be predicted and managed. Interpreted seismic attributes are used to estimate the static part of the reservoir, such as porosity. The dynamic characteristics, such as the variation of saturations, are inferred from the difference in seismic attributes from the baseline survey and subsequent ones. It is important for the time increment in 4D seismic acquisition to be long enough to track changes of fluid saturations. Therefore, on average, the 4D seismic survey is conducted every one or two years (Husky Energy, 2010).

Time-lapse seismic data can be used for the qualitative and quantitative description of a reservoir. It is helpful to refer to one of the earliest implementations of 4D seismic data for qualitative reservoir characterization (Wayland and Lee, 1986; Cooper and Thorogood, 1999; Lumley, 1999) where bypassed oil has been tracked. More recent work shows that time-lapse seismic data has been used to identify fluid flow and communications through faults (Roste et al, 2006). Also it has been shown that trace of injection of a displacement agent in thermal oil recovery processes, such as the interface between virgin reservoir zones and the steam front, can be easily monitored in unconsolidated high-porosity sands by repeated seismic prospecting (Zhang et al, 2005). Thus, it is important to understand the nature of elastic properties of saturated rock and seismic wave propagation in order to properly extract information about reservoir geology and fluid flow. The pressure and temperature dependent petroelastic model based on Gassmann's fluid substitution model is proposed for the generation of synthetic seismic attributes. The model is used to predict flow baffles and barriers and to characterize the SAGD oil extraction method, a vital part of which is forecasting steam chamber growth and the temperature front's propagation.

In the academic community, a few papers exist that describe the generation of synthetic seismic attributes. Most are based on Gassmann's fluid substitution model (Gassmann, 1951; Hong et al, 2006; Mavko et al, 2009). Fahimuddin has recently implemented the low frequency Gassmann theory to obtain seismic data in order to characterize the reservoir using the modeling technique Ensemble Kalman Filter (Fahimuddin, 2010). Equations for a relationship between acoustic properties of saturated rock and reservoir pressure and temperature can be found in the paper by Batzle and Wang (1992). The Kumar's paper (2006) documents the MATLAB implementation framework of Gassmann's fluid substitution model. The proposed petroelastic model for generating temperature and pressure dependent seismic attributes is based on the mentioned papers. The FORTRAN program fluidsub_tp.exe is written for a generation of synthetic seismic attributes, the description of which can be found in the Appendix. It is slightly similar to the program fluidsub.exe by Hong and Deutsch (2008), but differs from its predecessor in taking into account the non-isothermal environment for acoustic wave

propagation. In other words, in the new program, the elastic and physical properties of fluids and saturated rock are expressed as a function of the reservoir's thermodynamic state. There is an assumption that constituent minerals of a rock are insensitive to pressure and temperature variations.

Synthetic seismic attributes should be a function of the reservoir pressure, temperature, and elastic, physical properties of fluids and rock in order to define the reservoir architecture and monitor the flow of the injected heating agent in thermally operated oil extraction processes. Any change in fluid composition and temperature should be reflected in repeated synthetic seismic attributes. Also, noise should be incorporated into synthetic attributes to resemble the complexity associated with the acquisition of real seismic data. In the proposed petroelastic model, there is an assumption that artificial noise follows the normal distribution with zero mean and specified standard deviation in a percentage of true seismic value.

Among the wide diversity of seismic attributes, P- and S-wave velocities and corresponding acoustic impedances (V_P, V_S, Z_P, and Z_S) are chosen and derived from Gassmann's fluid substitution model because they can detect change in fluid saturations and reservoir temperature (Zagayevskiy and Deutsch, 2011a). Figure 7 shows a schematic workflow of the derivation of pressure and temperature dependent seismic velocities and acoustic impedances. The following text explains the mathematical elaboration.

The relationship between seismic velocity and acoustic impedance can be expressed as follows (Mavko et al, 2009):

$$Z_P = \rho_{sat} \cdot V_P \tag{49}$$

$$Z_S = \rho_{sat} \cdot V_S \tag{50}$$

where V_P and V_S are the P- and S-wave seismic velocities, m/s; Z_P and Z_S are the P- and S-wave characteristic acoustic impedances in Pa*s/m; and ρ_{sat} is the density of saturated rock (rock with pore fluid, which can be brine water, oil, gas, or their mixture) in kg/m^3.

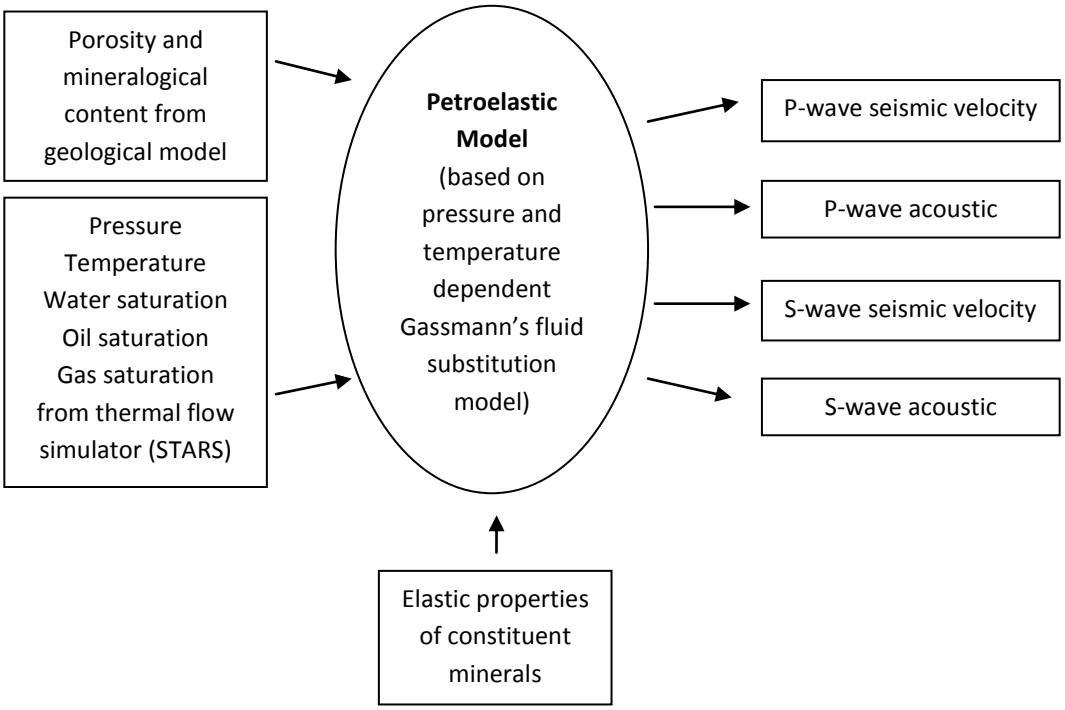


Figure 7: Diagram of petroelastic Gassmann's fluid substitution model used to generate pressure and temperature dependent seismic attributes

Seismic velocities in an isotropic, homogenous, elastic medium can be computed as shown in Equations (51) and (52) (Mavko et al, 2009).

$$V_P = \sqrt{\frac{k_{sat} + \frac{4}{3}\tau_{sat}}{\rho_{sat}}} \tag{51}$$

$$V_S = \sqrt{\frac{\tau_{sat}}{\rho_{sat}}} \tag{52}$$

where, k_{sat} is the effective elastic bulk modulus of saturated rock, Pa; τ_{sat} is the effective elastic shear modulus of saturated rock, Pa.

The bulk and shear moduli of saturated rock can be computed from the low-frequency Gassmann theory (Gassmann, 1951; Mavko et al, 2009):

$$\frac{k_{sat}}{k_{matrix}-k_{sat}}=\frac{k_{dry}}{k_{matrix}-k_{dry}}+\frac{k_{fluid}}{\phi\cdot(k_{matrix}-k_{fluid})} \tag{53}$$

$$k_{sat}=k_{dry}+\frac{\left(1-k_{dry}/k_{matrix}\right)^2}{\phi/k_{fluid}+\left(1-\phi\right)/k_{matrix}-k_{dryt}/k_{matrix}^2} \tag{54}$$

$$\tau_{sat}=\tau_{dry} \tag{55}$$

where k_{dry} and τ_{dry} are the effective elastic bulk and shear moduli of dry rock (no presence of pore fluid), Pa; k_{matrix} is the bulk modulus of the mineral material of the rock, Pa; and ϕ is the porosity of the rock medium.

The bulk modulus of dry rock is derived from laboratory measurements, empirical relationships, or wireline log data, and can be approximately described as follows (Krief et al, 1990):

$$k_{dry}=k_{matrix}(1-\beta) \tag{56}$$

$$\tau_{dry}=\tau_{matrix}(1-\beta) \tag{57}$$

where β is the Biot coefficient defined as the ratio of pore-volume change ΔV_p to bulk-volume change ΔV at constant pore pressure.

The empirical equation of the Biot coefficient can be presented as follows (Krief et al, 1990):

$$(1-\beta) = (1-\phi)^{\frac{3}{1-\phi}} \tag{58}$$

Since it is intended to apply synthetic seismic attributes to characterize northern Alberta oil fields, the lithology of a reservoir can be approximated to local formations. Thus, several classes can represent lithofacies of the McMurray formation. Frequently encountered facies are sand, shale, sandy and muddy inclined heterolythic strata (IHS), breccia, mudstone, and carbonate. Most of them comprise two distinct lithological sequences: permeable sand and impermeable shale. Their presence in facies is expressed through a volume fraction coefficient called *V*-shale, which is the ratio of rock volume occupied by shale to the total volume of rock specimen (*V*-shale = V_{shale}/V_{total}). Hence, the lithology of any facies can be simply described by shale content, where the *V*-shale value determines the volume of shale and *1.0* – *V*-shale determines volume of sand in a facies. The mineralogy of these two sequences mainly consists of two minerals: quartz and clay. While quartz is more dominant in sand, shale is mostly made of clay particles. On average it can be assumed that the volume fraction of quartz f_{quartz} is 0.3 and f_{clay} is 0.7 ($f_{clay} + f_{quartz} = 1.0$). The described mineralogical model provides grounds for calculating elastic moduli and the density of the rock matrix.

The bulk and shear moduli of the rock matrix can be approximated by the average of Hashin-Shtrikman bounds for the isotropic linear elastic composite consisting of only two constituent minerals (Equations (59) and (60)) (Mavko et al, 2009).

$$k_{matrix} = \frac{k_{matrix}^{upperHS} + k_{matrix}^{lowerHS}}{2} \tag{59}$$

$$\tau_{matrix} = \frac{\tau_{matrix}^{upperHS} + \tau_{matrix}^{lowerHS}}{2} \tag{60}$$

$$k_{matrix}^{upperHS/lowerHS} = k_{1/2} + \frac{f_{2/1}}{\left(k_{2/1} - k_{1/2}\right)^{-1} + f_{1/2} \cdot \left(k_{1/2} + \frac{4}{3}\tau_{1/2}\right)^{-1}} \quad (61)$$

$$\tau_{matrix}^{upperHS/lowerHS} = \tau_{1/2} + \frac{f_{2/1}}{\left(\tau_{2/1} - \tau_{1/2}\right)^{-1} + 2 \cdot f_{1/2} \cdot \frac{\left(k_{1/2} + 2 \cdot \tau_{1/2}\right)}{5 \cdot \tau_{1/2} \cdot \left(k_{1/2} + \frac{4}{3}\tau_{1/2}\right)}} \quad (62)$$

where $k^{upperHS}_{matrix}$ and $k^{lowerHS}_{matrix}$ are the upper and lower Hashin-Shtrikman bounds for the bulk modulus of two constituent isotropic elastic media, Pa; $\tau^{upperHS}_{matrix}$ and $\tau^{lowerHS}_{matrix}$ are the upper and lower Hashin-Shtrikman bounds for shear modulus, Pa; k_1 and k_2 are the bulk moduli of individual constituents (quartz and clay), Pa; τ_1 and τ_2 are the shear moduli of individual constituents, Pa; f_1 and f_2 are the volume fractions of individual minerals; and a slash separates the order of constituents in the calculation of the upper and lower bounds.

It has been mentioned that the fluid component of a reservoir may be represented as single brine water, oil, and gas, or a mixture of any of the three. Thus, the fluid bulk modulus k_{fluid} can be derived using a Reuss (isostress) average of the fluid mixture, taking into account the bulk modulus of every fluid type (Mavko et al, 2009):

$$\frac{1}{k_{fluid}} = \frac{S_{brine}}{k_{brine}} + \frac{S_{oil}}{k_{oil}} + \frac{S_{gas}}{k_{gas}} \quad (63)$$

where k_{brine}, k_{oil} and k_{gas} are the the bulk modulus of brine water, oil, and gas respectively, Pa; S_{brine}, S_{oil} and S_{gas} are the brine water, oil, and gas saturations respectively, they have to add up to 1.0 ($S_{brine} + S_{oil} + S_{gas} = 1.0$).

The density of saturated rock can be expressed as follows (Mavko et al, 2009):

$$\rho_{sat} = (1-\phi)\cdot\rho_{matrix} + \phi\cdot\rho_{fluid} \quad (64)$$

where ρ_{matrix} and ρ_{fluid} are the densities of a rock matrix and fluid mixture of brine water, oil, and gas respectively, kg/m^3.

The density of the fluid mixture is (Mavko et al, 2009):

$$\rho_{fluid} = S_{brine}\cdot\rho_{brine} + S_{oil}\cdot\rho_{oil} + S_{gas}\cdot\rho_{gas} \quad (65)$$

where ρ_{brine}, ρ_{oil} and ρ_{gas} are the densities of brine water, oil, and gas at reservoir pressure and temperature respectively, kg/m^3.

The matrix density can be also expressed through the density and volume fraction of quartz and clay minerals (Mavko et al, 2009):

$$\rho_{matrix} = f_{clay}\cdot\rho_{clay} + f_{quartz}\cdot\rho_{quartz} \quad (66)$$

The relationship between elastic properties and densities of fluids, temperature, and pressure are summarized below for every fluid component. It is assumed that properties of rock are independent of temperature or pressure and, therefore, are constant. Table 2 summarizes the elastic and seismic properties of quartz and clay. The same properties of fluid components are presented in the Table 3. Note that transverse S-waves cannot be generated in liquid or gas phases.

Table 2: Elastic and seismic properties of clay and quartz minerals (Mavko et al, 2009)

Mineral	Density (g/cm^3)	Bulk modulus (GPa)	Shear modulus (GPa)	V_p (km/s)	V_s (km/s)
Clay (kaolinite)	1.58	15.0	6.0	1.44	0.93
Quartz	2.65	37.0	44.0	6.05	4.09

Table 3: Elastic and seismic properties of pure water, bitumen and natural gas at standard conditions (15.6 ^{0}C, 0.1 MPa) (Mavko et al, 2009)

Fluid	Density (g/cm^3)	Bulk modulus (GPa)	V_p (km/s)
Pure water	0.9991	2.20	1.497
Bitumen	1.0100	1.66	1.500
Natural gas	0.0008	0.01	0.450

Properties of brine water are discussed below. The elastic bulk modulus *k* of liquid or gas can be found from the equation of velocity of sound in a homogeneous substance as shown below (Kinsler et. al., 2000):

$$V_p = \sqrt{\frac{k}{\rho}} \tag{67}$$

where V_p is the P-wave sound velocity in a homogeneous liquid or gaseous medium, m/s; k is the bulk modulus of the medium, Pa; and ρ is the density of the medium, kg/m^3.

The bulk modulus can be easily determined from Equation (67), in which a homogeneous liquid or gas substance's density and acoustic velocity are known at a specific pressure and temperature. Thus, Equation (66) is used to compute the bulk moduli of brine water, oil, and gas.

$$k = \rho \cdot V_p^2 \tag{68}$$

It has been derived empirically that the acoustic velocity of brine water free of gas can be approximately expressed as (Batzle and Wang, 1992):

$$
\begin{aligned}
V_{p,brine} &= V_{p,water} + \varsigma \cdot (1170 - 9.6 \cdot T + 0.055 \cdot T^2 \\
&- 8.5 \times 10^{-5} T^3 + 2.6 \cdot P - 0.0029 \cdot T \cdot P - 0.0476 \cdot P^2) \\
&+ \varsigma^{1.5} \cdot \left(780 - 10 \cdot P + 0.16 \cdot P^2\right) - 820 \cdot \varsigma^2
\end{aligned} \tag{69}
$$

where V_{brine} is the acoustic velocity in brine water, m/s; V_{water} is the acoustic velocity in fresh water, m/s; P and T are the pressure and temperature of brine water, MPa and ^{0}C; and ς is the salinity of brine water or the weight fraction of dissolved salt in water, ppm/1000000.

Water is divided into four main salinity groups based on the amount of dissolved salts in it (Table 4) (Lewis, 1980).

Table 4: Salinity grade ς of water based on amount of dissolved salts

Water type	Salinity of water (ppm)
Fresh water	< 500
Brackish water	500 – 30000
Saline water	30000 – 50000
Brine	> 50000

The acoustic velocity of pure water can also be presented through in-situ pressure and temperature (Batzle and Wang, 1992):

$$
\begin{aligned}
V_{p,water} &= \sum_{i=0}^{4} \sum_{j=0}^{3} w_{ij} T^i P^j \\
&= 1402.85 + 1.524 \cdot P + 3.437 \times 10^{-3} \cdot P^2 - 1.197 \times 10^{-5} \cdot P^3 + \\
&+ 4.871 \cdot T - 0.0111 \cdot T \cdot P + 1.739 \times 10^{-4} \cdot T \cdot P^2 - 1.628 \times 10^{-6} \cdot T \cdot P^3 - \\
&- 0.04783 \cdot T^2 + 2.747 \times 10^{-4} \cdot T^2 \cdot P - 2.135 \times 10^{-6} \cdot T^2 \cdot P^2 + 1.237 \times 10^{-8} \cdot T^2 \cdot P^3 + \\
&+ 1.487 \times 10^{-4} \cdot T^3 - 6.503 \times 10^{-7} \cdot T^3 \cdot P - 1.455 \times 10^{-8} \cdot T^3 \cdot P^2 + 1.327 \times 10^{-10} \cdot T^3 \cdot P^3 - \\
&- 2.197 \times 10^{-7} \cdot T^4 + 7.987 \times 10^{-10} \cdot T^4 \cdot P + 5.230 \times 10^{-11} \cdot T^4 \cdot P^2 - 4.614 \times 10^{-13} \cdot T^4 \cdot P^3
\end{aligned} \tag{70}
$$

Densities of brine and pure water can be approximately computed as follows (Batzle and Wang, 1992):

$$\rho_{brine} = \rho_{water} + \varsigma \cdot (0.668 + 0.44 \cdot \varsigma + 1\times10^{-6}\left[300 \cdot P - 2400 \cdot P \cdot \varsigma + T \cdot \left(80 + 3 \cdot T - 3300 \cdot \varsigma - 13 \cdot P + 47 \cdot P \cdot \varsigma\right)\right]) \quad (71)$$

$$\rho_{water} = 1 + 1\times10^{-6}(-80 \cdot T - 3.3 \cdot T^2 + 0.00175 \cdot T^3 + 489 \cdot P - 2 \cdot T \cdot P + 0.016 \cdot T^2 \cdot P - 1.3\times10^{-5} \cdot T^3 \cdot P - 0.333 \cdot P^2 - 0.002 \cdot T \cdot P^2) \quad (72)$$

where ρ_{brine} and ρ_{water} are the densities of brine and pure water respectively, g/cm^3.

The pressure- and temperature-dependent bulk modulus of brine water is computed using Equation (68).

Oil properties are shown below from Batzle and Wang (1992). Elastic properties of bitumen are assumed to be close to properties of heavy oil. The bulk modulus of oil saturated with gas is computed as shown in Equation (68). The P-wave seismic velocity of pseudo live oil at reservoir conditions can be computed as shown below:

$$V_{p,oil} = 2096\sqrt{\frac{\rho_{oil,sat,pseudo\_rc}}{2.6 - \rho_{oil,sat,pseudo\_rc}}} - 3.7 \cdot T + 4.64 \cdot P + 0.0115 \cdot \left[4.12\sqrt{\frac{1.08}{\rho_{oil,sat,pseudo\_rc}} - 1} - 1\right] \cdot T \cdot P \quad (73)$$

where $\rho_{oil,sat,pseudo\_rc}$ is the density of synthetic live oil saturated with a gas component at simulated reservoir conditions, g/cm^3.

The density of pseudo-saturated oil at reservoir conditions can be expressed through the density of oil measured at surface conditions (101325 Pa ≈ 0.1 MPa and 15.6 ^{0}C):

$$\rho_{oil,sat,pseudo\_rc} = \frac{\rho_{oil,sc}}{B_0 \cdot (1 + 0.001 \cdot R_G)} \tag{74}$$

where $\rho_{oil,sc}$ is the density of oil at surface conditions, g/cm^3; B_0 is the oil formation volume factor, which represents the ratio of the volume of oil and gas at reservoir conditions to the volume of oil at surface or standard conditions; R_G is the gas-to-oil ratio (GOR) or the ratio of gas volume volatilized from saturated oil to the volume of the oil at surface conditions, Liters/Liter.

The oil formation volume factor B_0 and GOR R_G can be expressed as shown below:

$$B_0 = 0.972 + 0.00038 \cdot \left[2.4 R_G \cdot \left(\frac{G}{\rho_{oil,sc}} \right)^{1/2} + T + 17.8 \right]^{1.175} \tag{75}$$

$$R_G = 0.02123 \cdot G \cdot \left[P \cdot \exp\left(\frac{4.072}{\rho_{oil,sc}} - 0.00377 \cdot T \right) \right]^{1.205} \tag{76}$$

where G is the specific gas gravity or the ratio of the gas density to air density at surface conditions, its value usually varies between 0.56 and 1.80.

The combined effect of pressure, temperature, and gas content on oil density can be expressed as follows:

$$\rho_{oil} = \frac{\rho_{oil,sat,rc} + \left(0.00277 \cdot P - 1.71 \times 10^{-7} P^3\right)\left(\rho_{oil,sat,rc} - 1.15\right)^2 + 3.49 \times 10^{-4} P}{0.972 + 3.81 \times 10^{-4} \left(T + 17.78\right)^{1.175}} \tag{77}$$

where $\rho_{oil,sat,rc}$ is the density of live oil saturated with a gas component at reservoir conditions, g/cm^3.

$$\rho_{oil,sat,rc} = \frac{\rho_{oil,sc} + 0.012 \cdot G \cdot R_G}{B_0} \tag{78}$$

Finally, the bulk modulus for live oil at reservoir conditions is computed using Equation (68).

Gas properties are presented below from Batzle and Wang (1992). The bulk modulus of hydrocarbon gas is derived in a different manner than the bulk moduli of brine water and saturated oil are derived. The bulk modulus of gas under adiabatic conditions can be expressed as follows:

$$k_{gas} \cong \frac{P}{\left(1 - \frac{P_{pr}}{\xi}\frac{\partial \xi}{\partial P_{pr}}\right)_T}\theta \tag{79}$$

where P_{pr} is the dimensionless pseudoreduced pressure; ξ is the compressibility factor (Equation (80)); θ is the heat capacity ratio (ratio of gas heat capacity at constant pressure to gas heat capacity at constant volume), whose expression is shown in Equation (80).

$$\begin{aligned}\xi = &\left[0.03 + 0.00527 \cdot (3.5 - T_{pr})^3\right] \cdot P_{pr} + \left(0.642 \cdot T_{pr} - 0.007 \cdot T_{pr}^4 - 0.52\right) + \\ &+ 0.109 \cdot \left(3.85 - T_{pr}\right)^2 \exp\left\{-\left[0.45 + 8 \cdot \left(0.56 - 1/T_{pr}\right)^2\right] \cdot \frac{P_{pr}^{1.2}}{T_{pr}}\right\}\end{aligned} \tag{80}$$

Thus,

$$\begin{aligned}\left(\frac{\partial \xi}{\partial P_{pr}}\right)_{T=T_{pr}} &= 0.03 + 0.00527 \cdot \left(3.5 - T_{pr}\right)^3 - \\ &- 0.1308 \cdot \left(3.85 - T_{pr}\right)^2 \frac{P_{pr}^{0.2}}{T_{pr}}\left[0.45 + 8 \cdot \left(0.56 - \frac{1}{T_{pr}}\right)^2\right] \cdot \\ &\exp\left\{-\frac{P_{pr}^{1.2}}{T_{pr}}\left[0.45 + 8 \cdot \left(0.56 - \frac{1}{T_{pr}}\right)^2\right]\right\}\end{aligned} \tag{81}$$

$$\theta = 0.85 + \frac{5.6}{P_{pr} + 2} + \frac{27.1}{\left(P_{pr} + 3.5\right)^2} - 8.7 \cdot \exp\left[-0.65\left(P_{pr} + 1\right)\right] \quad (82)$$

where T_{pr} is the pseudoreduced temperature, ^{0}C.

Pseudoreduced pressure and temperature can be expressed through absolute pressure *P* and temperature *T* of the gas and its specific gravity *G*:

$$P_{pr} = \frac{P}{P_{pc}} = \frac{P}{4.892 - 0.4048 \cdot G} \quad (83)$$

$$T_{pr} = \frac{T + 273.15}{T_{pc}} = \frac{T + 273.15}{94.72 + 170.75 \cdot G} \quad (84)$$

The density of in-situ gas is calculated as shown below:

$$\rho_{gas} \cong \frac{28.8 \cdot G \cdot P}{\xi \cdot R_{gas} \cdot (T + 273.15)} \quad (85)$$

where *R* is the gas constant, 8.314472 J/mol*K.

Once all the equations are combined, synthetic seismic attributes based on Gassmann's theory may be generated. It has been found that examined, error-free methodology leads to seismic attributes that are inversely related to porosity values and, thus, can be used to estimate porosity distribution (Zagayevskiy and Deutsch, 2011a). Permeability is derived from the porosity-permeability relationship. Gas saturation has a negative correlation coefficient with all seismic attributes and their change in time except for a positive correlation with difference in S-wave velocity. Steam chamber growth may be predicted poorly by using P-wave velocity, P-wave acoustic impedance, and the difference in P-wave velocity. However, the difference in S-wave velocity and difference

in both P- and S-wave acoustic impedances can be used to accurately predict steam chamber growth. The difference in P-wave acoustic impedances can be used to estimate porosity and predict steam chamber growth. The difference in P-wave acoustic impedances in EnKF-based data assimilation algorithms should be used to estimate porosity and permeability. Gas saturation and temperature values have same influence on seismic attributes. Hence, they may not be differentiated from each other using seismic data (Zagayevskiy and Deutsch, 2011a). However, reservoir temperature and gas saturation are closely related to each other and have similar spatial distributions in SAGD. Therefore, temperature data can be used to predict steam chamber growth. It should be kept in mind that the amount of information inherent in seismic attributes depends on associated measurement, processing, interpretation errors, and resolution (fuzziness) of seismic tools, which can be introduced into the examined petroelastic model by white noise (Zagayevskiy and Deutsch, 2011a). White noise is a random variable with zero mean and user-defined standard deviation that follows the Gaussian distribution.

2.4 Data

The importance and characteristics of the data are highlighted. The data are vital in modeling. They should be treated with caution, because they constrain distributions of model parameters and states of the petroleum reservoir model and bring model estimates closer to reality. Observed data usually come from various sources sampled at different scales and resolutions. Locations in space at which data are measured are called the sampled locations. The model regions at which the data are not obtained, but estimates are desirable, are called the unsampled locations (Deutsch, 2002). Even though measurements of variables are available at sampled locations, true values of variables are not known precisely due to disturbances in readings and limitations of measurement gauges. Fatigue, lack of experience, poor skills, and reduced attentiveness lead to additional bias in the data. Thus, the quality of the data is subject to presence and value of error. If possible, the errors should be corrected before modeling in order to extract valuable information. If that is not possible, the error in the modeling

technique should be accounted for (Deutsch et al, 2010). Redundant data, outliers, and extreme values should be examined further for possible removal from the data sets (Devore and Peck, 2005).

The author refers to data describing the spatial distribution of model parameters of reservoir models as measurements or static data. The author refers to data describing the behaviour of state variables as observations or dynamic data. Core data, well logs, 1D, 2D, or 3D seismic data are examples of static data. Usually core data are obtained from vertical surveillance wells. Well logs are data derived by geophysical means. They characterize the area around the well bore only at short distances. Seismic data cover larger regions of the reservoir and usually represent an extensive source of information about underground structure (Deutsch, 2002). Production data, well test readings, continuous temperature observations and 4D seismic data fall into the class of dynamic data. The most common observations that characterize production data are oil, gas, and water production rates, water cut, cumulative oil production, steam-to-oil ratio, gas-to-oil ratio, and bottomhole pressure. Production data are observed not only in surveillance wells, but also along production and injection wells. Piezometers, extensometers, thermocouples, and inclinometers are used to assess the reservoir's performance. First type of gauges is used to measure reservoir pressure. The extensometers monitor the spatial change in reservoir's geomechanical properties such as rock stresses, shears, strains, and volumetric expansions. They are adjusted for temperature variations. The inclinometers are installed to measure the dipping of rock beddings. Thermocouples and fibre optics are essential gauges to successfully implement the SAGD oil recovery method. These gauges measure the PVT condition of a reservoir. They are used to predict steam chamber growth, indicate barriers and baffles, and assess overall reservoir heterogeneity (Butler, 1991; Dake, 1998; Deutsch, 2002; Husky Energy, 2010).

Model parameters data can be divided further into hard data and soft data depending on whether they are direct or indirect measurements of examined variables (Deutsch, 2002). Hard data correspond to direct core samples, from which facies, and porosity and permeability measurements are derived in a laboratory. Well log data, production data,

PVT well tests, seismic attributes, and other data types that contain information about model parameters, but are not direct measurements of them, are called soft data. It is deemed that soft data support hard data and improve model estimates. Outcrops, conceptual models, geological interpretations, and data from petroleum reservoirs with similar geology are another example of indirect measurements of model parameters that provide extra information for reservoir modeling (Deutsch, 2002).

The data's scale and resolution are two similar characteristics of its spatial sampling. Scale characterizes volume, from which data have been sampled, and may be either point or block scale relative to the model scale. Resolution represents the quality of the details inherent in the gridded data. The model scale depends on the purpose for which the model is constructed. In micro-modeling, where local spatial distribution of laminated sequences is modeled, the model scale is small (Hosseini et al, 2008). In cases in which oil reserves are to be estimated, a coarser scale is selected to decrease computational time. An intermediate scale is used for flow simulation, which captures flow processes (Deutsch, 2002). Finer scale models lead to more detailed estimates, but at the same time computational expenses increase dramatically. So the trade-off between model quality and computational cost should be found. The block size of the model grid is directly related to the model scale. The block size used in thermal simulations should be less than that used in conventional simulations. Time steps should be also smaller in thermal simulations (Carlson, 2003).

Usually the scale varies for different data sets. So upscaling (averaging finer scale data) or downscaling (splitting coarser scale data) are sometimes required to bring all the data to a unique target model scale in a pre-determined fashion (Isaaks and Srivastava, 1989). Usually hard data are sampled at a finer scale in a vertical direction, which is different from a reservoir model scale or a scale used in a flow simulation. It is good practice to upscale hard data to a coarser target scale of the model that preserves all of the important information. Production data are characterized by a lower scale. The well log scale falls between the core and production scales. Seismic attributes describe the model at larger scales with varying resolution. On average, seismic data have from low

to medium resolution in a vertical direction and high resolution in a horizontal direction (Fahimuddin, 2010).

Later chapters will focus on a methodology proposed for continuous data integration based on EnKF. This methodology simultaneously assimilates measurements of model parameters and observations of model states and accounts for inherent measurement error. Before data assimilation, the data should be brought to the scale of the model grid represented by block size.

Chapter 3 – Probabilistic Approach to Petroleum Reservoir Modeling

The previous chapter reviewed the fundamentals of analytical and numerical approaches to deterministic petroleum reservoir modeling and the properties of data appropriate for assimilation. This chapter is devoted to the probabilistic or stochastic approach to reservoir modeling by geostatistical means. The main advantage of this approach is that it assesses the uncertainty associated with estimation results. The inverse modeling technique Ensemble Kalman Filter (EnKF) is proposed to continuously assimilate data into a reservoir model and to estimate model parameters and predict state variables. This chapter presents the method's theoretical background and extensions. Chapter 5 includes discussions about its characteristics and implementation details.

3.1 Geostatistical Framework of Data Integration

Geostatistics is a branch of applied statistics, the main goal of which is to estimate the distributions of spatial variables (Isaaks and Srivastava, 1989). It comprises mapping estimation methods, simulation techniques, and data analysis tools. While geostatistical estimation methods derive the best estimates and estimation variance maps, geostatistical simulation methods use the Monte Carlo Simulation (MCS), in which a set of equally probable estimates, also known as realizations, is generated in order to investigate different probable scenarios of the variable's distribution (Metropolis and Ulam, 1949). The mean and variance of the simulated realizations may be used to assess the best estimate and to quantify uncertainty. Estimation methods lead to unique and smooth estimates compared to simulation methods, which produce realizations that might be quite different from each other and not so smooth. Thus, simulation is more appropriate for modeling flow systems (Deutsch, 2002).

Most of the geostatistical techniques assume that a random function is stationary. The random function $\boldsymbol{X}$ is a generalized form of the spatial random variable $\boldsymbol{X}$, which

discretizes the variable into finite pieces over the modeling region, treating them as dependent random variables *X*s related to each other through covariance function. Thus, the covariance function stores patterns of spatial distribution of a discretized variable. It is used to derive weights in mapping techniques and simulation methods to compute estimates at unsampled locations. If two or more variables are used together in an estimation, a linear model of co-regionalization (LMC) should be constructed, in which cross-covariance functions between variables are required for modeling. The covariance function $C_X(\vec{h})$ of random function ***X*** can be calculated as shown in Equation (86). The cross-covariance function $C_{XY}(\vec{h})$ of spatial random variables ***X*** and ***Y*** is shown in Equation (87). The lag vector $\vec{h}$ represents the separation distance vector in space between any two pieces X_i and X_j of a discretized spatial random variable ***X*** or pieces *X* and *Y* of discretized spatial random variables ***X*** and ***Y***. Thus, the assumption of stationarity implies two statements. First, the expected value of any random function is the constant throughout the modeling region and represents the global mean. Second, covariance functions (and cross-covariance functions) of any random functions are true throughout the modeling region (Deutsch and Journel, 1998). These statements refer to the assumption of first and second order stationarity.

$$C_X(\vec{h}) = E\left[\left(X_i - E\left[X_i\right]\right)\cdot\left(X_j - E\left[X_j\right]\right)\right] \approx \frac{\sum_{k=1}^{n_p}\left(x_{i,k} - \bar{x}_i\right)\cdot\left(x_{j,k} - \bar{x}_j\right)}{n_p - 1} \qquad (86)$$

$$C_{XY}(\vec{h}) = E\left[\left(X_i - E\left[X_i\right]\right)\cdot\left(Y_j - E\left[Y_j\right]\right)\right] \approx \frac{\sum_{k=1}^{n_p}\left(x_{i,k} - \bar{x}_i\right)\cdot\left(y_{j,k} - \bar{y}_j\right)}{n_p - 1} \qquad (87)$$

where $E[X]$ is the expected value of a random variable *X* or the mean of a random variable *X*; $x_{i,k}$ is the data value of any random variable X_i, which makes the k^{th} pair with any random variable X_j separated by a lag $\vec{h}$; n_p is the number of pairs of any random variables X_i and X_j that are separated by a lag $\vec{h}$.

Instead of computing the covariance function $C_X(\vec{h})$ and cross-covariance function $C_{XY}(\vec{h})$ from the data, it is more practical to compute the experimental semivariogram $\gamma_X(\vec{h})$ for each spatial random variable ***X***, and the experimental cross-semivariogram $\gamma_{XY}(\vec{h})$ for each pair of spatial random variables ***X*** and ***Y*** for several lag distances. Experimental semivariogram and cross-semivariogram are shown in Equations (88) and (89) respectively. The continuous variogram models are fit to these experimental variograms. The covariance functions are computed from the fit models (Equations (90) and (91)). Note that continuous semivariogram models have to be positive-definite. The most common semivariogram models are nugget, spherical, exponential, Gaussian, hole effect, and dampen hole effect. Several semivariogram models may be combined to form a nested structure, if stipulated by the data (Deutsch and Journel, 1998).

$$2 \cdot \gamma_X(\vec{h}) = E\left[\left(X_i - X_j\right)^2\right] \approx \frac{\sum_{k=1}^{n_p}\left(x_{i,k} - x_{j,k}\right)^2}{n_p} \tag{88}$$

$$2 \cdot \gamma_{XY}(\vec{h}) = E\left[\left(X_i - X_j\right) \cdot \left(Y_i - Y_j\right)\right] \approx \frac{\sum_{k=1}^{n_p}\left(x_{i,k} - x_{j,k}\right) \cdot \left(y_{i,k} - y_{j,k}\right)}{n_p} \tag{89}$$

$$\gamma_X(\vec{h}) = C_X(0) - C_X(\vec{h}) \tag{90}$$

$$\gamma_{XY}(\vec{h}) = C_{XY}(0) - C_{XY}(\vec{h}) \tag{91}$$

where $C_X(0)$ and $C_{XY}(0)$ are the covariance and cross-covariance functions at the zero lag distance $\vec{h}$ = 0 or global variance and covariance respectively.

The most popular conventional geostatistical estimation technique is kriging, and the most popular simulation technique is Sequential Gaussian Simulation (SGS) for continuous variables or Sequential Indicator Simulation (SIS) for categorical variables

(Isaaks and Srivastava, 1989). The equation for the kriging estimate is shown in Equation (92), when data of one variable type is used. Kriging weights are derived from Equation (93) through minimizing error variance. The error is treated as a random variable, which represents the mismatch between truth and the estimated value of the random variable at an unsampled location. Kriging produces very smooth estimation maps.

$$x_{est,0} - E[X_0] = \sum_{i=1}^{n} \lambda_i \cdot \left(x_i - E[X_i]\right) \tag{92}$$

where X_0 is the random variable at the unsampled location *0*, whose true value x_0 is unknown and to be estimated; $x_{est,0}$ is the estimate of the random variable X_0; X_i is the random variable at the sampled location *i*, whose value x_i is known (i = 1, …, n); n is the number of variables or data used in the estimation; and λ_i is the i^{th} kriging weight, which corresponds to the random variable X_i.

$$\sum_{i=1}^{n} \lambda_i \cdot C(\vec{h}_{ij}) = C(\vec{h}_{0j}),\ j = 1,\ldots,n \tag{93}$$

where $\vec{h}_{ij}$ is the lag (separation) vector between random variables X_i and X_j located at sampled locations *i* and *j* (i = 1, …, n, j = 1, …, n); $\vec{h}_{0j}$ is the lag (separation) distance vector between random variables X_0 and X_j located at unsampled location *0* and sampled location *j*.

The variance of a kriged variable X_0 at estimation location is lower than the variance at data locations by the value of kriging variance (Equations (94) and (95)), which varies from location to location (Deutsch, 2002). This is due to smoothing effect of the spatial regression. Simulation techniques are devised in a way that diminishes presence of the smoothing effect and makes global variance of a random function constant throughout the modeling region (second order stationarity). Also, simulation results provide some additional insight into reality through a set of possible, but different, realizations (Deutsch, 2002).

$$VAR[X_0] = C(0) - \sigma_{k,0}^2 \tag{94}$$

$$\sigma_{k,0}^2 = C(0) - \sum_{j=1}^{n} \lambda_j \cdot C(\vec{h}_{0j}) \tag{95}$$

where $\sigma_{k,0}^2$ is the kriging variance at unsampled location *0*.

In Sequential Gaussian Simulation, data are transformed to normal scores first. The normal scores are values that follow normal distribution with zero mean and unit variance. Then, kriging is performed in the Gaussian space using a semivariogram model that honours spatial distribution of the normal scored data. In the next step, the residual value of random variable *R*, which follows normal distribution with zero mean and kriging variance (Equation (95)), is added in normal scores to every single kriging estimate at the unsampled location (Equation (96)). This additive value represents a single realization of the simulated random variable. Every simulated value at the unsampled location is added continuously to the data set in order to honour the covariance function in a simulation of the random variable's value at another unsampled location. The path of the simulation locations is chosen randomly. The procedure is performed for the entire modeling region and repeated for every realization. The values of random variables X_0 and *R* are simulated independently from each other. In this set up of SGS, the variance of simulated random variable $X_{sim,0}$ is equal to the global variance, and the assumption about the second order stationarity is not violated (Equation (97)). Once all of the realizations are computed, they can be back-transformed to their original units (Deutsch, 2002).

$$x_{sim,0}^{NS(i_r)} = x_{est,0}^{NS} + r^{NS(i_r)} \tag{96}$$

where $x_{sim,0}^{NS(i_r)}$ is the i_r^{th} simulated value of random variable X_0 at the unsampled location *0* in normal scores; $x_{est,0}^{NS}$ is the kriging estimate of random variable X_0 in normal scores;

$r^{NS(i_r)}$ is the i_r^{th} value of the residual random variable *R*, which is distributed normally with zero mean and kriging variance $\sigma_{k,0}^2$.

$$VAR[X_{sim,0}] = C^{NS}(0) \tag{97}$$

In the context of a stochastic estimation and simulation vectors of model parameters $\boldsymbol{m}$, state variables $\boldsymbol{s}_t$, and data $\boldsymbol{d}_t$ of the reservoir model are extended to the matrix forms $\boldsymbol{M}$, $\boldsymbol{S}_t$ and $\boldsymbol{D}_t$, respectively (Equations (98) – (102)), preserving the overall structure of the model $\boldsymbol{U}_t$ of a dynamic system (Equation (98)). Here, every column corresponds to a single realization of model variables. Values stored in columns represent gridded model parameters, state variables, and scattered data. A user specifies the total number of realizations N_e. The ratio of the number of data to the number of realizations should be around 1:10 to have a sufficient number of degrees of freedom (Fahimuddin, 2010). The data is represented in a stochastic fashion. To achieve this, perturbations $v_{t,i}^{(i_r)}$ are added to the true value of data $d_{t,i}$ to artificially account for the error in addition to the unknown measurement error $\varepsilon_{t,i}$. A user specifies the distribution of perturbations, which is usually normal with zero mean and a small standard deviation (Evensen, 2009).

$$\boldsymbol{U}_t = \begin{bmatrix} \boldsymbol{M} \\ \boldsymbol{S}_t \\ \boldsymbol{D}_t \end{bmatrix}_{(N_m \cdot N_b + N_s \cdot N_b + N_{d,t}) \times N_e} = \begin{bmatrix} \boldsymbol{M} \\ F(\boldsymbol{M}, \boldsymbol{S}_0, \boldsymbol{S}_{b,t}; t) \\ G(\boldsymbol{M}, \boldsymbol{S}_0, \boldsymbol{S}_{b,t}; t) \end{bmatrix}_{(N_m \cdot N_b + N_s \cdot N_b + N_{d,t}) \times N_e} \tag{98}$$

$$\boldsymbol{M} = \begin{bmatrix} m_{1,1}^{(1)} & m_{1,1}^{(2)} & \dots & m_{1,1}^{(N_e)} \\ m_{1,2}^{(1)} & m_{1,2}^{(2)} & \dots & m_{1,2}^{(N_e)} \\ \dots & \dots & \dots & \dots \\ m_{1,N_b}^{(1)} & m_{1,N_b}^{(2)} & \dots & m_{1,N_b}^{(N_e)} \\ m_{2,1}^{(1)} & m_{2,1}^{(2)} & \dots & m_{2,1}^{(N_e)} \\ m_{2,2}^{(1)} & m_{2,2}^{(2)} & \dots & m_{1,2}^{(N_e)} \\ \dots & \dots & \dots & \dots \\ m_{2,N_b}^{(1)} & m_{2,N_b}^{(2)} & \dots & m_{2,N_b}^{(N_e)} \\ \dots & \dots & \dots & \dots \\ m_{N_m,1}^{(1)} & m_{N_m,1}^{(2)} & \dots & m_{N_m,1}^{(N_e)} \\ m_{N_m,2}^{(1)} & m_{N_m,2}^{(2)} & \dots & m_{N_m,2}^{(N_e)} \\ \dots & \dots & \dots & \dots \\ m_{N_m,N_b}^{(1)} & m_{N_m,N_b}^{(2)} & \dots & m_{N_m,N_b}^{(N_e)} \end{bmatrix}_{N_m \cdot N_b \times N_e} \tag{99}$$

$$\boldsymbol{S}_t = \begin{bmatrix} s_{t,1,1}^{(1)} & s_{t,1,1}^{(2)} & \dots & s_{t,1,1}^{(N_e)} \\ s_{t,1,2}^{(1)} & s_{t,1,2}^{(2)} & \dots & s_{t,1,2}^{(N_e)} \\ \dots & \dots & \dots & \dots \\ s_{t,1,N_b}^{(1)} & s_{t,1,N_b}^{(2)} & \dots & s_{t,1,N_b}^{(N_e)} \\ s_{t,2,1}^{(1)} & s_{t,2,1}^{(2)} & \dots & s_{t,2,1}^{(N_e)} \\ s_{t,2,2}^{(1)} & s_{t,2,2}^{(2)} & \dots & s_{t,1,2}^{(N_e)} \\ \dots & \dots & \dots & \dots \\ s_{t,2,N_b}^{(1)} & s_{t,2,N_b}^{(2)} & \dots & s_{t,2,N_b}^{(N_e)} \\ \dots & \dots & \dots & \dots \\ s_{t,N_m,1}^{(1)} & s_{t,N_m,1}^{(2)} & \dots & s_{t,N_m,1}^{(N_e)} \\ s_{t,N_m,2}^{(1)} & s_{t,N_m,2}^{(2)} & \dots & s_{t,N_m,2}^{(N_e)} \\ \dots & \dots & \dots & \dots \\ s_{t,N_m,N_b}^{(1)} & s_{t,N_m,N_b}^{(2)} & \dots & s_{t,N_m,N_b}^{(N_e)} \end{bmatrix}_{N_s \cdot N_b \times N_e} \tag{100}$$

$$\boldsymbol{D}_t = \begin{bmatrix} d_{t,1}^{(1)} & d_{t,1}^{(2)} & \dots & d_{t,1}^{(N_e)} \\ d_{t,2}^{(1)} & d_{t,2}^{(2)} & \dots & d_{t,2}^{(N_e)} \\ \dots & \dots & \dots & \dots \\ d_{t,N_d}^{(1)} & d_{t,N_d}^{(2)} & \dots & d_{t,N_d}^{(N_e)} \end{bmatrix}_{\cdot N_{d,t} \times N_e} \tag{101}$$

$$d_{t,i}^{(i_r)} = d_{t,i} + \nu_{t,i}^{(i_r)} = d_{true,t,i} + \varepsilon_{t,i} + \nu_{t,i}^{(i_r)} \tag{102}$$

where $m_{i,i_b}^{(i_r)}$ is the value of the i_r^{th} realization of the i^{th} model parameter at the i_b^{th} block (i = 1, …, N_m, i_b = 1, …, N_b, i_r = 1, …, N_e); $m_{t,i,i_b}^{(i_r)}$ is the value of the i_r^{th} realization of the i^{th} state variable at the i_b^{th} block at the time t (i = 1, …, N_s, i_b = 1, …, N_b, i_r = 1, …, N_e); $d_{t,i}^{(i_r)}$ is the value of the i_r^{th} realization of the i^{th} data at the time t (i = 1, …, $N_{d,t}$, i_r = 1, …, N_e); $\varepsilon_{t,i}$ is the unknown measurement error of the i^{th} datum value sampled at the time step t; and $v_{t,i}^{(i_r)}$ is the i_r^{th} perturbation value added to the i^{th} datum at time t, which artificially forms the measurement error matrix $\boldsymbol{\Sigma}_t$.

3.2 Ensemble Kalman Filter (EnKF)

Co-kriging based methods use measurements of primary and secondary variables to get estimates and are good for static systems, but their usage is limited for dynamic systems. Instead, data assimilation techniques can be used for dynamic systems, which account for continuous data integration into a model (Huang et al, 2009). However, most dynamic data integration methods are not easy to implement (Oliver et al, 2001). Some of them require knowledge about sensitivity or the Hessian matrix, such as the master point method Sequential Self-Calibration (SSC) (Gomez-Hernandez et al., 1997). The analytical equations governing the flow process should be understood and solved analytically for the entire reservoir. Gradient-based methods such as conjugate gradient or steepest descent require a large number of iterations, which is computationally expensive. Methods based on eigenvectors are computationally expensive as well. The Gauss-Newton method is fast and good for large systems, but is limited to a small number of data, which can be assimilated (Oliver et al, 2001). For this reason, the Ensemble Kalman Filter (EnKF) is proposed for continuous data assimilation, which is easy to implement and computationally fast (Evensen, 2009). There is no need to derive a sensitivity matrix. The EnKF treats the flow simulator as a black box (Aanonsen et al, 2009).

The Ensemble Kalman Filter is a modification of the Kalman Filter (Kalman, 1960). The Kalman Filter (KF) is one of the inverse methods of sequential data assimilation, which recursively provides dynamic calibration of variables' distributions as new data are becoming available in time (Huang et al, 2009). In the inverse theory, an inverse problem means the reconstruction of the distribution of a spatial variable given some data sampled at limited number of point locations (Bennett, 1992). It is deemed that there is always relationship between the variable and its data, which is unknown precisely and should be found. The inverse problem is usually ill posed, i.e. there is always measurement error in data that makes derivation of true distribution of the spatial variable more challenging. Also, the problem implies non-uniqueness of a solution, which means that several solutions may equally honor the data and none of the solutions are better than any other. The inverse method seeks possible plausible solution to the inverse problem in order to infer true distribution of the spatial variable that honors available observations. In the light of the KF and the EnKF, these inverse methods integrate both direct (hard) and indirect (soft) measurements of model parameters to estimate model parameters and honor observations of model states sampled at multiple time steps. Thus, estimates of model parameters that honor both static and dynamic data are not unique, and this non-uniqueness is represented by ensemble realizations.

The KF is based on the Bayesian update and tunes values of model parameters and state variables in order to improve a model's quality. Initially, the KF was devised to eliminate the noise and other inaccuracy in the measurements in signal processing, but later its applications extended to other fields. In this thesis, filter means estimation and prediction algorithm and is weakly related to the error filter in signal processing. KF is named after its developer, Rudolf E. Kalman, a Hungarian-American professor, engineer, and mathematician. The Kalman Filter was applied to several petroleum problems, especially in a geophysical and drilling context, but because of its limitations handling large-scale problems it was not widely used. Eisenman et al. (1994) and Corser et al. (2000) present some details in their papers. It is worth mentioning that the KF is more suitable for cases with a smaller number of variables, where a linear relationship between variables exists. However, such cases are very rare for reservoir models.

The KF is based on the Bayesian theory. The Bayesian update problem's objective is to obtain the model's corrected or updated (posterior) distribution $p(\boldsymbol{u}|\boldsymbol{d})$, combining the model's preliminary (prior) probability distribution $p(\boldsymbol{u})$ and its data likelihood $p(\boldsymbol{d}|\boldsymbol{u})$, represented by available measurements of its parameters and observations of state variables (Mandel, 2006). The equation related to the calculation of prior distribution is named the forecast step. The equation devoted to calculating the posterior distribution is called the analysis step, at which data are assimilated.

Let $\boldsymbol{u}$ be a vector consisting of first-guess (initial) N_v model variables that represent the system. The model variables consist of model parameters and state variables. Let $\boldsymbol{C}^f$ be an N_v x N_v prior covariance matrix and $\boldsymbol{\Sigma}$ be an N_d x N_d covariance matrix of measurements of model parameters and observations of state variables. The data are stored in the vector $\boldsymbol{d}$. Note that the observed values are not true values of state variables because of natural noise, but they are the closest ones before the filter is applied. Thus, the solution to the Bayesian update problem is expressed mathematically as the Bayesian theorem (Press, 2003):

$$p(\boldsymbol{u}|\boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{u}) \cdot p(\boldsymbol{u}) \tag{103}$$

where $\propto$ denotes proportion; $p(\boldsymbol{u})$ is the prior probability density function of model $\boldsymbol{u}$; $\boldsymbol{u}$ is the model column vector of size N_V; $\boldsymbol{d}$ is the data or measurement vector of length N_d sampled at one time step, and it is not necessary that $N_d = N_v$; $p(\boldsymbol{d}|\boldsymbol{u})$ is the likelihood function of model $\boldsymbol{u}$; $p(\boldsymbol{u}|\boldsymbol{d})$ is the posterior probability distribution of model $\boldsymbol{u}$ and is the current best estimate of the model until new observations are obtained.

The prior Gaussian distribution $p(\boldsymbol{u})$ is defined in the following form for the Gaussian distribution (Evensen, 2009), where the initial forecasted covariance and mean value matrices are generated preserving all prior information on model $\boldsymbol{u}$:

$$p(\boldsymbol{u}) = p(\boldsymbol{u}^f) \propto \exp\left(-\frac{1}{2}\left(\boldsymbol{u}^f - E[\boldsymbol{u}^f]\right)^T (\boldsymbol{C}^f)^{-1}\left(\boldsymbol{u}^f - E[\boldsymbol{u}^f]\right)\right) \quad (104)$$

where $E[\boldsymbol{u}^f]$ is the vector of the prior mean values of N_v model variables $\boldsymbol{m}$ and $\boldsymbol{s}$ at the forecast step, jointly denoted by $\boldsymbol{u}$; $\boldsymbol{C}^f$ is the corresponding N_v x N_v covariance matrix at the forecast step, which is obtained from prior information on the model distribution.

The following form defines the likelihood function $p(\boldsymbol{d}|\boldsymbol{u})$ or distribution of data $\boldsymbol{d}$ conditioned to the model vector $\boldsymbol{u}$ that follows normal distribution (Evensen 2009):

$$p(\boldsymbol{d} \mid \boldsymbol{u}) \propto \exp\left(-\frac{1}{2}\left(\boldsymbol{H}\cdot\boldsymbol{u} - E[\boldsymbol{d}]\right)^T (\boldsymbol{\Sigma})^{-1}\left(\boldsymbol{H}\cdot\boldsymbol{u} - E[\boldsymbol{d}]\right)\right) \quad (105)$$

where $\boldsymbol{\Sigma}$ is the N_d x N_d diagonal measurement error covariance matrix of data $\boldsymbol{d}$, measurement errors ($\varepsilon_i \sim N(0,\Sigma_{ii})$) are assumed to be independent from each other; $\boldsymbol{H}$ is the N_d x N_v observation matrix-operator consisting of 0s and 1s, and it relates model variables to corresponding data values. If data exists for a certain state variable or model parameter, then the corresponding element of $\boldsymbol{H}$ is 1; otherwise it is 0. The 1 can be present only once in each row; other elements in the row are zeros. Below is an example for ***H:***

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}_{N_d \times N_v} \quad (106)$$

The posterior distribution $p(\boldsymbol{u}|\boldsymbol{d})$ for the Gaussian model $\boldsymbol{u}^a$ can be expressed as follows (Evensen, 2009):

$$p(\boldsymbol{u} \mid \boldsymbol{d}) \approx p(\boldsymbol{u}^a) \propto \exp\left(-\frac{1}{2}\left(\boldsymbol{u}^a - E[\boldsymbol{u}^a]\right)^T (\boldsymbol{C}^a)^{-1}\left(\boldsymbol{u}^a - E[\boldsymbol{u}^a]\right)\right) \quad (107)$$

where superscript ^{a} denotes updated values.

If prior, likelihood, and posterior distributions are combined together through the Bayesian theorem (Equation (103)), and the probability of the posterior distribution function is maximized, the updated mean value vector $E[\boldsymbol{u}^{a}]$ and updated covariance matrix $\boldsymbol{C}^{a}$ will have the following forms:

$$E[\boldsymbol{u}^{a}] = E[\boldsymbol{u}^{f}] + \boldsymbol{K} \cdot (E[\boldsymbol{d}] - \boldsymbol{H} \cdot E[\boldsymbol{u}^{f}]) \tag{108}$$

$$\boldsymbol{C}^{a} = \left(\boldsymbol{I} - \boldsymbol{K} \cdot \boldsymbol{H}\right) \cdot \boldsymbol{C}^{f} \tag{109}$$

where $\boldsymbol{K}$ stands for the Kalman gain matrix:

$$\boldsymbol{K} = \boldsymbol{C}^{f} \cdot \boldsymbol{H}^{T} \cdot (\boldsymbol{H} \cdot \boldsymbol{C}^{f} \cdot \boldsymbol{H}^{T} + \boldsymbol{\Sigma})^{-1} \tag{110}$$

Equations (108) and (109) are called KF equations, and Equation (110) is called the Kalman gain. Also, Equation (108) is referred to as the analysis step of the Kalman Filter.

Thus, the EnKF solution of Bayesian problem changes the form of prior distribution of the spatial variable, expressed through cumulative distribution function (CDF) or histogram, and its covariance structure $\boldsymbol{C}^{f}$ according to spatial distribution of integrated data to obtain posterior (analyzed) distribution of the variable. The rate of form and covariance structure change depends on prior (initial) distribution of the variable and the number of integrated data. Updated spatial distribution of the variable inherits both prior information and spatial distribution of integrated data as shown in Equation (109). The semivariogram γ of the variable changes similarly to covariance structure change. Data integration may be stopped, when there is no significant change in CDF or γ brought by integration of additional data: $||CDF^{f} - CDF^{a}||$ < some tolerance.

The relationship between the updated mean vector and the forecasted mean vector at the future time step can be expressed through time series Equation (111). This relationship is called the forecast step.

$$E[\boldsymbol{u}_t^f] = F \cdot E[\boldsymbol{u}_{t-1}^a] + \boldsymbol{em} \tag{111}$$

where *F* is the model operator, which linearly relates state variables to model parameters; ***em*** is the model error, which is always close to zero or equals zero.

The KF relies on several assumptions. All state variables and model parameters in vector ***u*** are distributed normally (according to Gaussian distribution). The relationship between the updated model $\boldsymbol{u}^a$ and its forecasted $\boldsymbol{u}^f$ and measured ***d*** values is linear (see Equation (108)). Also, the advancement of the system's covariance matrix in time is linear as shown in Equation (109). Thus, KF is dedicated to dealing with linear dynamical systems (Equation (111)), which follow the Gaussian distribution. To resolve the limitation of the KF application to nonlinear systems, the Extended Kalman Filter (EKF) was devised (Jazwinski, 1970). It is claimed that EKF is computationally costly and cannot characterize highly nonlinear models very well (Leng and Yeh, 2003). The Ensemble Kalman Filter (EnKF) was developed to deal with large-scale nonlinear systems that follow distributions close to Gaussian (Evensen, 1994; Evensen, 2003). The advantage of EnKF over EKF can be found in the paper by Reichle et al. (2002), who compared both methods when estimating a soil moisture profile. The EnKF can be effectively applied to non-Gaussian systems, if modifications are applied (Wen and Chen, 2006; Gu, 2006). EnKF has a relatively easy implementation algorithm that does not depend on any dynamical model mechanism. EnKF treats the model operator (e.g., flow simulator) as a black box, only input and output of which with operational features should be known (Burgers et al., 1998).

The EnKF is an inverse modeling technique that integrates data into the model as soon as the data become available (Evensen, 2009). The technique is based on an ensemble of realizations, and thus, the best estimate and estimation variance are reported. EnKF is

a numerical approach to the KF, where the covariance matrix is replaced with a sample covariance matrix obtained from realizations. The EnKF estimates static variables (model parameters) and predicts dynamic variables (state variables) in compliance with available data. The relationships between model parameters and state variables should exist. An estimation based on an EnKF approach possesses identical properties of kriging (Zagayevskiy et al, 2010a). EnKF's computational algorithm is simple and relatively cheap at the analysis step, but might be computationally expensive at the forecast step. Chapter 5 includes implementation details about EnKF.

The EnKF technique relies on the initial ensemble (first-guess values) and is good for large-scale Gaussian systems. However, its application for non-Gaussian and highly nonlinear models is restricted. The recursive integration of the data may be applied to resolve this issue (Wen and Chen, 2006; Gu, 2006). The method consists of a two-step procedure: a nonlinear prediction and linear update based on the covariance structure between all examined variables. These steps are called forecast (forward) and analysis (update) steps, the expressions of which are shown below. Together, the two steps form one step called the assimilation step. First-guess (initial) values of the model parameters are generated first, using all available prior information. Then the EnKF forecast step is performed (Equation (112)), in which the model operator is usually represented by a flow simulator. Since explicit knowledge about the flow simulators' mechanisms is not required, EnKF can be easily adapted for any simulator. In comprarison, it is hard to couple a flow simulator with other modeling techniques such as SSC. The analysis step is performed after the forecast step, at which time data are integrated into a model (Equation (113)).

$$\boldsymbol{U}_t^f = F(\boldsymbol{U}_{t-1}^a) + \boldsymbol{EM} \text{ - forecast step} \tag{112}$$

$$\boldsymbol{U}_t^a = \boldsymbol{U}_t^f + \boldsymbol{K}_t \cdot (\boldsymbol{D}_t - \boldsymbol{H}_t \cdot \boldsymbol{U}_t^f) \text{ - analysis step} \tag{113}$$

$$\boldsymbol{K}_t = \hat{\boldsymbol{C}}^f \cdot \boldsymbol{H}_t^T \cdot (\boldsymbol{H}_t \cdot \hat{\boldsymbol{C}}^f \cdot \boldsymbol{H}_t^T + \boldsymbol{\Sigma}_t)^{-1} \text{ - Kalman gain} \tag{114}$$

$$\hat{\boldsymbol{C}}_t^f = \frac{\left(\boldsymbol{U}_t^f - \overline{\boldsymbol{U}}_t^f\right)\cdot\left(\boldsymbol{U}_t^f - \overline{\boldsymbol{U}}_t^f\right)^T}{N_e - 1} \text{ - sample covariance} \quad (115)$$

$$\overline{\boldsymbol{U}}_t^f = \frac{\left(\boldsymbol{U}_t^f \cdot e\right)\cdot e^T}{N_e} \text{ - ensemble mean} \quad (116)$$

where $\boldsymbol{U}$ is the $(N_m \cdot N_b + N_s \cdot N_b)$ x N_e ensemble matrix or model matrix, whose rows consist of N_m model parameters (static characteristic of a model) and N_v state variables (dynamic characteristic of a model) and columns comprise N_e realizations (note that data vector is excluded from the model for calculation purposes); subscript $_t$ is the time or assimilation step; superscript ^{f} is the forecast step; superscript ^{a} is the analysis step; F is the model operator that relates the analyzed matrix $\boldsymbol{U}^a_{t-1}$ from previous time to the forecasted matrix $\boldsymbol{U}^f_t$ at the current time t; $\boldsymbol{EM}$ is the model error matrix usually set to 0; $\boldsymbol{K}_t$ is the Kalman gain matrix whose expression is shown in Equation (114); $\boldsymbol{D}_t$ is the $N_{d,t}$ x N_e data matrix consisting of available measurements of model parameters and observations of state variables at time t; $\boldsymbol{H}_t$ is the $N_{d,t}$ x $(N_m \cdot N_b + N_s \cdot N_b)$ observation matrix at time t that relates the ensemble matrix to the data matrix; $\hat{\boldsymbol{C}}^f$ is the $(N_m \cdot N_b + N_s \cdot N_b)$ x $(N_m \cdot N_b + N_s \cdot N_b)$ sample covariance matrix whose values are calculated as shown in Equation (115); $\boldsymbol{\Sigma}_t$ is the $N_{d,t}$ x $N_{d,t}$ diagonal measurement error matrix consisting of perturbation variances; and e is the N_e x 1 vector consisting of 1s.

The first application of data assimilation with the EnKF is found in meteorology and in oceanography. Later its use is spread to hydrology and geophysics. The main goal of data assimilation in these areas is to increase the predictability of weather forecasts (Daley, 1991), ocean dynamics (Bennett, 1992), soil moisture (Houser et al, 1998), soil temperature profiles (Huang et al, 2008), contaminant propagation in the groundwater (Huang et al, 2009), and characterization of a medium heterogeneity (Christakos, 2005). EnKF has recently gained popularity in characterizing petroleum reservoirs as well (Nævdal et al, 2002a, 2002b; Gu et al, 2006; Aanonsen et al, 2009; Oliver and Chen, 2010).

The first attempt to apply EnKF for data integration into a petroleum reservoir was made by Geir Nævdal et al. (2002a and 2002b). In their first paper, they used EnKF to update a near-well reservoir model by using well measurements, which comprise both the model parameters and state variables. In that paper, only a simple synthetic example is worked out effectively. In the next paper, Nævdal et al. (2002b) applied EnKF to a more complicated semi-synthetic 2D field model from the North Sea, where a larger number of state variables is continuously updated. The measurements consist of dynamic data, such as bottom-hole pressures, water cuts, and gas-oil ratios; and static data, including permeability. The filter's performance is examined thoroughly for reservoir engineering applications. The results show significant consistency between measured and predicted values.

Later Yaqing Gu and Dean Oliver compared the EnKF estimation and prediction method with other traditional data integration methods in a case study of the PUNQ-S3 reservoir model (Production forecasting with Uncertainty Quantification) (Gu and Oliver, 2005). The PUNQ-S3 is a small-scale synthetic reservoir engineering model build based on the Elf Exploration Production S.A.S. Company's data. Gu proves the superiority of EnKF over other methods in predicting oil production; however, permeability and porosity maps are not matched well. In consequent papers, Gu tries to resolve this issue by modifying initial variables to make them follow a normal distribution (Gu and Oliver, 2006). Also Gu finds out that although the updated model is different from the model's earlier ensembles, the forecasts are very consistent with real data and that the constructed model honours all measured data. However, the EnKF has one small but unavoidable drawback: while the number of data to be assimilated increases, the size of the ensemble augments as well. This leads to higher computational costs.

Yan and Zhang examined EnKF's robustness in updating a hydraulic conductivity field in 2D and 3D models using synthetic examples (Yan and Zhang, 2005). Also, the authors conducted a sensitivity analysis of the methodology, in which they varied the ensemble size, type, and sampling timing of data and prior statistics. They found that prior knowledge of the data distribution plays an important role in forecasting state variables. If the prior statistics (ensemble mean, variance, etc.) are chosen far apart from actual

one, the estimated values might never converge to measurements and observations. Yan and Zhang argue that dynamic observations are more essential for predictions than static ones.

In their paper, Xian-Huan Wen et al. added a confirming option, which leads to better forecast precision and a diminishing uncertainty range in estimating non-Gaussian systems (Wen and Chen, 2006). The confirmation option is intended to assure that updated static parameters and dynamic states are consistent with the flow equations at the current assimilation step. Huseby et al. discuss incorporating natural and artificial traces in reservoir models with the EnKF (Huseby et al, 2009). They have shown that tracking and integrating information about tracers make it possible to obtain more accurate estimates. Tracers are presented by geochemical isotropic concentration variations in injected and formation waters and by concentration observations from inter-well tracer tests. A general result, which is found in many other publications, states that additionally assimilated data improves the reservoir model's predictability and robustness (Zagayevskiy and Deutsch, 2011b).

Dongxiao Zhang et al. went further and suggested using the efficient dimension-reduced KF, which is based on the Karhunen-Loeve (KL) orthogonal decomposition, to integrate reservoir data (Zhang et al, 2007). Here, the covariance of the reservoir properties is approximated by a small set of eigen-values and eigen-functions using KL decomposition to reduce a system's dimensionality. The covariance matrix can be reconstructed whenever needed. Comparing the EnKF and Karhunen-Loeve-based Kalman filter (KLKF) carried out in a synthetic 2D example shows that KLKF achieves better forecasts, but is not good for very large-scale systems.

There are two up-to-date review articles about using the EnKF application in reservoir engineering. The articles cover the most recent findings on the EnKF application, with detailed explanations of its theoretical background (Aanonsen et al, 2009; Oliver and Yan, 2010).

In highlighted research works and papers, little attention is paid to integrating reservoir temperature observations and time-lapse seismic attributes into SAGD reservoir models. A study about incorporating these data should not be omitted, since the data have great potential for characterizing thermally operated petroleum reservoirs. It is believed that incorporating additional soft temperature and seismic observations significantly improves knowledge about reservoir architecture (Zagayevskiy and Deutsch, 2011b). Three-dimensional seismic surveys provide extensive spatial knowledge about reservoir geology. Four-dimensional seismic data, which are repeated measurements of the 3D seismic response in time over a fixed domain, may even be used to track fluid flow in a reservoir (Lumley, 2001), to predict front displacement of an injected agent (Eastwood et al, 1994; Landrø, 2001; Yuh, 2003), to forecast steam chamber growth in SAGD applications (Zhang et al, 2005), and to define spatial distribution of flow baffles and barriers (Zagayevskiy and Deutsch, 2011a). Steam is used as a displacement agent in these oil recovery processes and, hence, is closely tied to spatial distribution of temperature profiles (Butler, 1991; Zagayevskiy and Deutsch, 2011a). Temperature data might be used to understand reservoir geology, predict steam chamber growth, and manage depletion processes. Temperature is continuously measured by downhole permanent gauges installed in vertical surveillance wells and along horizontal well pairs in SAGD fields. As a result, the observed temperature data reflect dynamic properties of the reservoir (Husky Energy, 2010).

Seismic data is a valuable source of geological information. However, integrating seismic attributes with EnKF is challenging, because a highly extensive dataset may lead to the collapse of an ensemble of realizations that represent a model. Some alterations are introduced to EnKF in order to assimilate large datasets. The decomposition of the covariance matrix (Skjervheim et al, 2005), the hierarchical model (Myrseth, 2007), the localization of the updating matrix (Dong et al, 2006), the localization of the covariance matrix (Fahimuddin, 2010) are implemented to continuously assimilate vast seismic data.

Even though the literature may contain information about using temperature data to estimate various geological measures and flow rates, this research topic is relatively new

in petroleum reservoir geomodeling. Duru and Horne apply deconvolution with the Bayesian inversion concept to predict the flow rate in a conventional oil reservoir using temperature data from permanent downhole gauges (Duru and Horne, 2009). They show that temperature is a function of the well product flow rate and thermal properties of fluid and rock surrounding a well and, therefore, it is an additional constraint for reservoir geology. In Duru and Horne's subsequent paper, temperature is utilized to estimate porosity and log-permeability distributions through a quasilinear Bayesian inversion and EnKF, in which flow and thermal models for a conventional oil reservoir are coupled together (Duru and Horne, 2010). The authors argue that including temperature data improves the estimation of geological properties in comparison to assimilating only production data (bottom-hole pressure, and oil and water production rates). The Li and Zhu paper touches on a topic of predicting the water profile in a horizontal well using bottom-hole pressure and temperature data measured by downhole gauges and fibre optic sensors (Li and Zhu, 2010). The Li et al. paper shows that incorporating production data and bottom-hole pressure with temperature data into the geological model helps in predicting permeability distribution and leads to optimized control of the horizontal well performance (Li et al., 2010). To be more precise, non-isothermal reservoir model and temperature data are used to derive a coarse-scale permeability field, which is downscaled by block kriging and calibrated to permeability distribution derived from production data.

There are also integrated modeling approaches in which, in addition to temperature and production data, other soft data are integrated in a reservoir model. For instance, temperature, hydrologic, pneumatic, and geochemical data are used to characterize flow and transport mechanisms that occur in fractured reservoirs (Wu et al, 2007).

Attempts have been made to use seismic and temperature data simultaneously to constrain the reservoir model. Temperature dependent seismic velocities from 4D seismic surveys have been employed to predict steam flood distribution in a 5-spot pattern at the South Casper Creek oil field in Wyoming (Johnston et al, 1992). To sum up, this thesis is aimed at applying EnKF to characterize petroleum reservoir models through integrating continuous temperature measurements and time-lapse seismic

attributes at SAGD operated oil fields, where spatial distributions of porosity and permeability are estimated. This topic is not studied well in the literature, but it represents a challenging and vital research direction.

A study comparing temperature data integration with EnKF and other modeling techniques, including the SSC, found that EnKF's objective function minimization rate was faster (Zagayevskiy and Deutsch, 2010).

3.3 Extensions of the EnKF

As was mentioned earlier, EnKF has evolved from KF as a numerical extension. However, other KF modifications are present and summarized in Table 5. Figure 8 shows the chart with the most essential KF modifications. Table 6 contains a tabulation of recent field case studies using EnKF. Corresponding papers are shown in the last column of the table. All authors applied EnKF to large-scale models using different data sources.

Table 5: Summary of the evolution of the Kalman Filter (Aanonsen et al, 2009)

Method	Description
Original method: Kalman Filter	
It is an efficient filter that estimates the state of a linear dynamical system from a series of noisy measurements. It consists of two main equations: forecast and analysis (Kalman, 1960; Maybeck, 1979; Stengel, 1994; Cohn, 1997).	
For nonlinear models: Extended Kalman Filter (EKF) and Sigma Point Filter	
Extended Kalman Filter (EKF)	In the forecast equation of the KF, linear model operator *F*, which describes the dynamics of a system, is replaced with a suitable nonlinear function *F* in EKF. The estimates of EKF for *highly nonlinear models* and for *large-scale systems* are not accurate (Eppstein and Dougherry, 1996; Hantush and Marin; 1997; Eigbe et al., 1998; Leng and Yeh, 2003).
Sigma Point Filter (SPF) or Unscented Kalman Filter (UKF)	It is devised for non-linear models. The difference with EKF is in the way uncertainty information propagates. Here the covariance matrices are updated on the basis of the sample points. Thus, SPF is more robust in

	comparison to EKF. But it is also not applicable to *large-scale systems* (Julier et al., 2000; Wan and Nelson, 2001; Lefebvre et al., 2004).
For large-scale systems: Ensemble Kalman Filter (EnKF) and Reduced Order Filters	
Ensemble Kalman Filter (EnKF)	The covariance matrix is propagated only partially. The models are updated based on information from first- and second-order moments (mean and covariance matrix) and their estimates are not fully correct, if the systems do not follow the Gaussian distribution (Evensen, 1994; Burgers et al, 1998; Houtekamer and Mitchell, 1998; Houtekamer and Mitchell, 2001; Whitaker and Hamill, 2002; Anderson, 2001; Evensen, 2003; Ehrendorfer, 2007; Evensen, 2009).
	EnKF applications within the petroleum industry (Lorentzen et al, 2001a and 2001b; Lorentzen et al, 2003; Skjervheim et al., 2007; Nævdal et al., 2001a, 2001b, 2005; Gu and Oliver, 2005).
	Application of EnKF in closed-loop reservoir management. Here EnKF is suitable for monitoring and finding optimal waterflooding strategy using adjoint-based optimization (Brower, 2004; Overbeek et al., 2004; Lorentzen et al, 2006; Nævdal et al., 2006; Chen et al., 2008; Wang et al., 2009).
Ensemble Kalman Smoother (EnKS)	Similar technique to EnKF, but it uses all accumulated observations from the past, so it does not have a continuous assimilation feature (Evensen, 2009).
Reduced order filters: SEEK, SEIK, KLKF	
Singular Evolutive Extended Kalman (SEEK)	Here the covariance matrix is approximated by the deterministic selection of eigen-values and eigen-vectors (Pham, 2001; Rozier et al, 2007).
Singular Evolutive Interpolated Kalman (SEIK)	Here the covariance matrix is approximated by combining a deterministic selection with some Monte Carlo drawings (Liang et al., 2007).
Filter based on Karhunen-Loeve decomposition and polynomial expansions (KLKF)	This method is not applicable when a large number of eigenvectors is considered (Zhang et al., 2007).

To avoid high non-Gaussian behaviour in EnKF: Particle Filter (PF) and Randomized Maximum Likelihood (RML)	
Particle Filter (PF)	The empirical distribution from samples and additional information from observations are combined to obtain posterior distribution. It is not applicable for large-scale systems (Gordon et al., 1993; Doucet et al., 2001; Kivman, 2003).
Randomized Maximum Likelihood (RML)	It gives a correct sample for a linear problem and an approximate solution for nonlinear problems. It requires the computation of gradients and matching of all data at once (Kitandis, 1995; Oliver et al., 1996; Gao et al., 2006; Zafari and Reynolds, 2007).

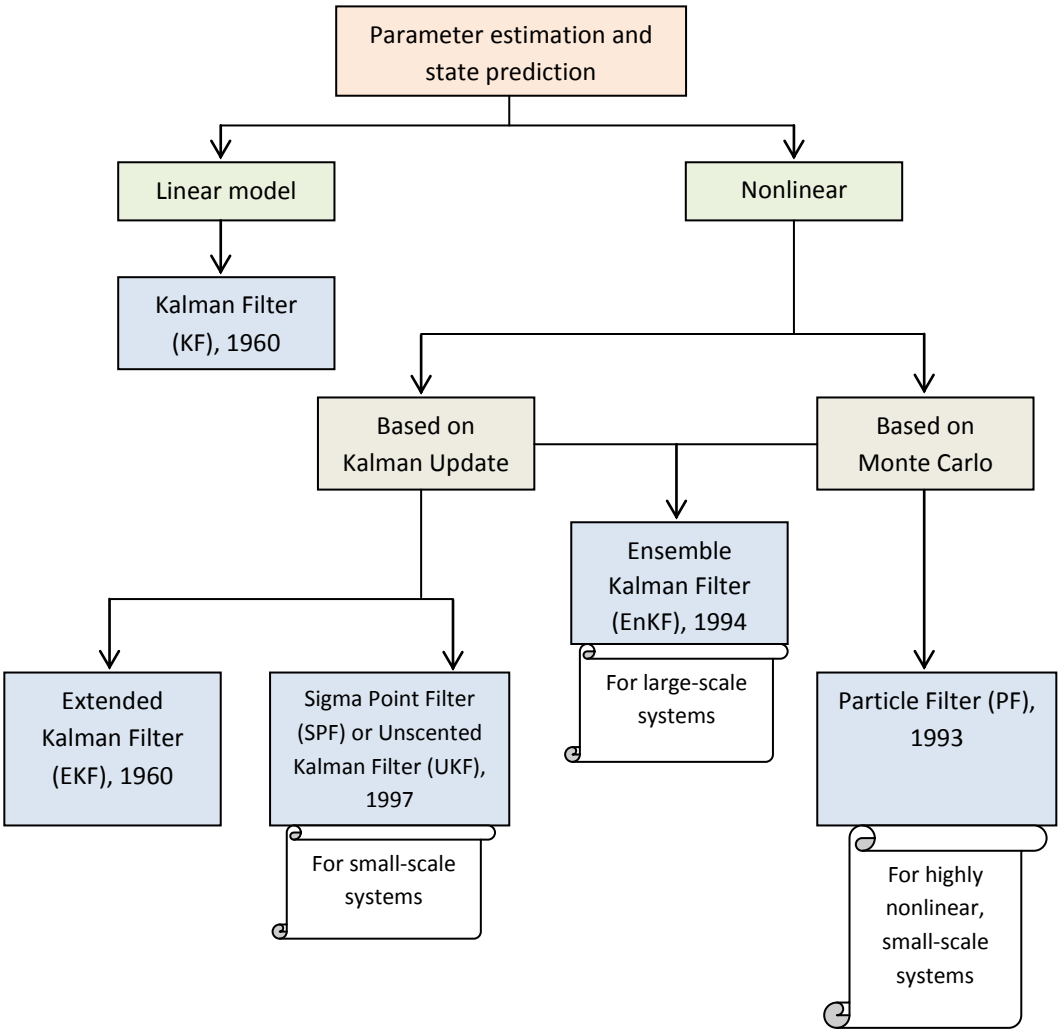


Figure 8: Kalman Filter evolution (Chitralekha, 2010)

Table 6: Summary of EnKF applications for characterizing real petroleum reservoirs (Oliver and Yan, 2010)

Model Size (# blocks)	Wells	Data	Model Parameters	Source
29580	5 producers 2 water injectors 1 gas injector	12 years: Water cut Gas-oil ratio (4D seismic)	Permeability Porosity	Skjervheim et al., 2007
45000	4 producers 2 gas injector	5 years: Bottomhole pressure Oil production Water cut Gas-oil ratio	Permeability Porosity	Haugen et al., 2008
25669	2 producers	3 years: Bottomhole pressure Oil production Water cut Gas-oil ratio	Permeability Porosity	Bianco et al., 2007
30740	9 producers	11 years: Water cut	Permeability	Arroyo-Negrete et al., 2008
95379	6 producers 2 water injectors	6 years: Water cut Gas-oil ratio Fluid production Bottomhole pressure	Permeability Porosity Trend coefficient	Zhang and Oliver, 2010
60000	4 producers 1 water injector	6 years: Oil production Water cut Gas-oil ratio	Permeability Porosity Water-oil contact Gas-oil contact Fault trans Vertical trans Relative permeability	Evensen et al., 2007; Seiler et al., 2009
416240	3 producers 1 water injector	3 years: Oil production Water production (Static pressure)	Permeability Porosity	Cominelli et al., 2009
44550	54 producers 30 water injectors	20 years: Oil production Water production, Bottomhole pressure (4D seismic)	Horizontal permeability Vertical permeability Porosity Net-to-gross ratio Relative permeability Water-oil contact	Chen and Oliver, 2010

Chapter 4 – Methodology for Data Assimilation with the EnKF

The beginning of this chapter highlights the objective of the research work. Then, the data integration algorithm based on EnKF is compared to the conventional data assimilation algorithm in order to find similar and distinct features of every approach. Methodology for integrating continuous temperature data and time-lapse seismic attributes is presented next. The end of the chapter includes validation measures for assessing estimation results.

4.1 Problem Formulation

This research work is devoted to characterizing geological properties of SAGD operated petroleum reservoirs. Specifically, the work focuses on estimating porosity and permeability distributions by means of EnKF, conditional to all available hard data (porosity and permeability core samples) and continuous soft observations (temperature data from vertical surveillance wells and exhaustive seismic data from time-lapse surveys). Understanding the model parameters of the petroleum reservoir model should lead to better predictions of reservoir performance. It is claimed that integrating additional data leads to better estimates of porosity and permeability fields. Two case studies are worked out in later chapters to investigate EnKF properties, show its implementation details, present benefits of the proposed data assimilation methodology, and modify the existing EnKF algorithm to make it more efficient and less computationally expensive.

4.2 Conventional and EnKF-based Data Integration Algorithms

Before proceeding to the proposed data integration algorithm based on EnKF for SAGD operated oil fields, the conventional history-matching algorithm is reviewed and compared to the general EnKF-based data assimilation algorithm. Figure 9 shows a diagram of the conventional data assimilation algorithm for reservoir characterization (Carlson, 2003). Figure 10 shows a diagram of the proposed EnKF-based data

assimilation algorithm. Solid black arrows indicate workflow, while dot arrows represent information flow between any two stages. In conventional modeling, history-matching of dynamic data is used as an objective function to optimize the model estimate. A reservoir model is constrained to all available dynamic data from all time steps. Once new data are obtained, history-matching and the data assimilation algorithm are repeated from the beginning to honour old and new data, which is computationally expensive. In the proposed data assimilation algorithm based on EnKF, the objective function takes into account both dynamic and static data. Also, when new data become available, there is no need to integrate all previously assimilated data. Instead, only new data should be integrated into the previously updated model. However, for highly nonlinear modeling systems, a recursive algorithm similar to conventional history-matching may be required.

The following description of the conventional data assimilation algorithm comes from Carlson (2003). The first stage in a conventional simulation algorithm is a geological review. Basics of reservoir geology should be understood and taken into account in future modeling. Faults, folds, fractures, and petrophysical and lithological properties of the reservoir should be defined. The next stage is a reservoir performance review. Here, production process characteristics (bottomhole pressure, water cut, gas-to-oil ratio, etc.), PVT tests, well logs, and drive mechanisms should be determined. The third stage is data gathering and screening. It is one of the most important steps in numerical modeling. Reliable data should be selected with minimized measurement error, and outliers and extreme values should be removed. It is desirable to bring all data to a single scale of the model. Thus, upscaling or downscaling may be required. These three stages are referred to as preliminary steps of data assimilation. The model construction and initialization stage is performed next. Model parameters are generated based on all prior knowledge. They are usually represented by the model's geological properties. Then the flow mechanism is understood properly in order to define the state variables' initial and boundary conditions. State variables are defined from model parameters and boundary and initial conditions by means of a model operator, which is usually presented through a flow simulation mechanism. The flow simulation for thermal recovery utilizes fluid transport, mass balance, and energy conservation equations. The

entire model is initialized and defined at this point. At the history-matching stage, the selected modeling technique integrates the data into the model. Different history-matching techniques are available in the literature (Oliver et al, 2001). Model parameters are tuned in order to match dynamic data and state variables. The objective function value determines how good the match is. Generally, the sum of the squared difference between predicted values from the reservoir model and dynamic data for all time steps constitutes an objective function. Thus, if the value of the objective function is less than a predefined small termination value, the model is deemed adequate and the history match is stopped. Predicted values of state variables at future time steps are made. In other cases, when the computed objective function value is larger than the predefined criterion, the model is iteratively tuned to honour all available data and to minimize the objective function. If new data are obtained, the process should be repeated from the model construction and initialization stage to honour all new and old data, which may be very computationally expensive. Finally, when all data are integrated, the reservoir management team reports, documents, and later uses the optimized estimates of model parameters and predictions of the state variables.

In contrast to the conventional data assimilation algorithm presented above, the proposed algorithm utilizes feature of continuous data integration for linear models (Figure 10). Also, model parameters and state variables are tuned simultaneously to honour both static and dynamic data. Here, all stages are similar to the previous algorithm for highly nonlinear systems. To honour old and new data when newly acquired data are integrated into a reservoir, EnKF must be recursively rerun from the beginning of flow simulation to the last updating step, using the current model as the initial one. But if linear or close-to-linear systems are modeled, the algorithm is slightly modified. Once new data become available, they can be easily integrated into the current updated model. There is no need to rerun the simulations from the beginning in order to re-establish relationships between model parameters and state variables, since these relationships are honoured automatically for linear systems. Predictions are made using a model operator and can be validated later on, at the prediction time step, as data become available. It should be recalled that EnKF is ensemble-based estimation and prediction technique. Therefore, multiple models are constructed instead of a single

one, and they are used for uncertainty assessment. Hence, compared to the conventional technique, the proposed data assimilation technique has one major drawback: its computational cost is usually higher and depends on the number of realizations in the ensemble matrix. If data from all time steps are assimilated at once, the Ensemble Kalman Smoother (EnKS) is considered (Evensen and van Leeuwen, 2000). The EnKS is exactly the same as the conventional data assimilation algorithm (Figure 9). It is good and effective for small models, but more computationally expensive than the EnKF. The EnKS honours static and dynamic data from all time steps and preserves relationships between model parameters and state variables.

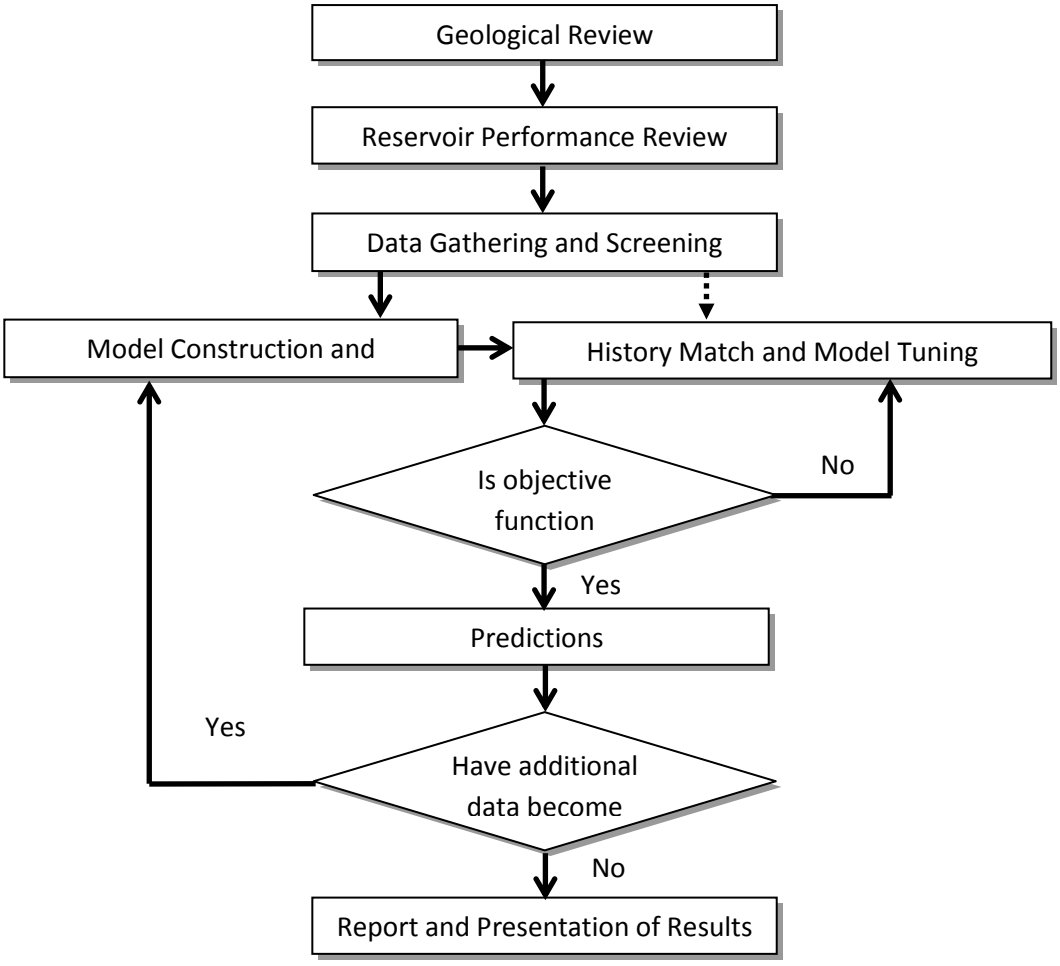


Figure 9: Conventional data assimilation algorithm for petroleum reservoir characterization (Carlson, 2003)

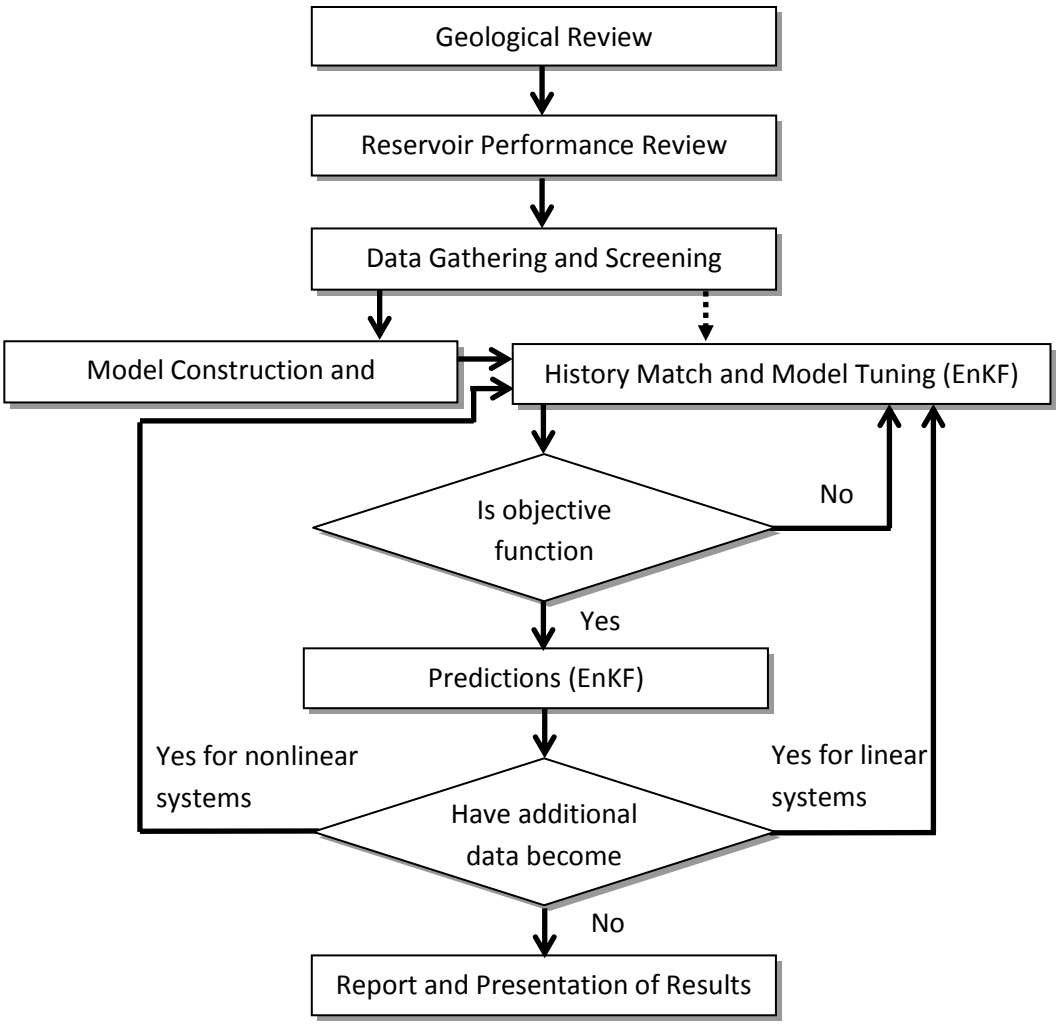


Figure 10: Proposed EnKF-based data assimilation algorithm for petroleum reservoir characterization

4.3 Proposed Methodology for Data Integration with the EnKF

Figure 11 summarizes the proposed methodology for integrating hard data, such as porosity and permeability, and continuous soft data, such as temperature from thermocouples and seismic attributes from geophysical surveys through EnKF into the petroleum reservoir model to estimate distribution of porosity and permeability.

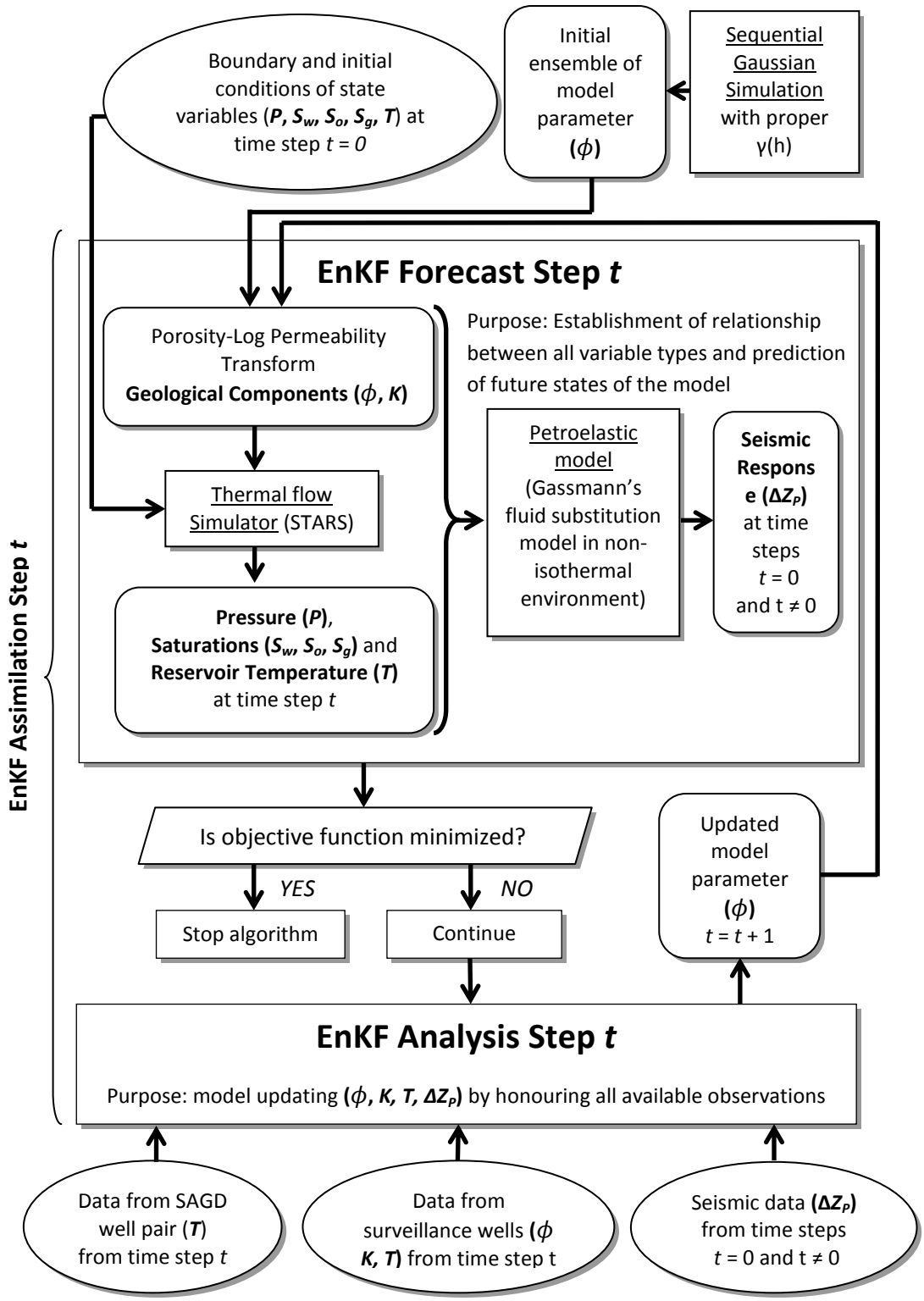


Figure 11: A diagram of the proposed methodology for continuous integration of core porosity and permeability data, temperature observations, and time-lapse seismic attributes with EnKF

Model parameters are presented by porosity and permeability variables, state variables are reservoir temperature and difference between baseline acoustic impedance and any subsequently observed acoustic impedances. The initialization stage is performed after the geology, reservoir performance, and data are reviewed. To initiate the EnKF algorithm, it is necessary to have the initial ensemble of the porosity model parameter ϕ along with production parameters and initial and boundary conditions of state variables such as fluid saturations, operating pressure, temperature of injected steam, PVT, and relative permeability curves. This stage is referred to as time step t = 0. A Sequential Gaussian Simulation (SGS) (Deutsch, 2002) is used with a proper semivariogram model derived from the porosity data and an understanding of reservoir heterogeneity to generate the initial porosity ensemble ϕ, consisting of a required number of realizations. There are two ways to generate the initial ensemble: conditional and unconditional to the data. The global mean and variance of the initial unconditional porosity ensemble should match the mean and variance of the data.

The model operator consists of three sub-models: the probabilistic porosity-log permeability transform, the thermal flow simulator, and the petroelastic model. Once the initial porosity field is generated, it can be incorporated into a reservoir model through the EnKF forecast step. The porosity-log permeability transform relates porosity values to horizontal permeability K_{xx} as shown in Equation (117). Other principal components of permeability are defined through Equation (118). For simplicity's sake, it is assumed that permeability is a diagonal tensor.

$$\log(K_{xx}) = a + b \cdot \phi + \varepsilon \tag{117}$$

$$\begin{aligned} K_{xx} &= K_{xx} \\ K_{yy} &= \alpha(K_{xx}) \\ K_{zz} &= \beta(K_{xx}) \end{aligned} \tag{118}$$

where K_{xx}, K_{yy}, and K_{zz} are the diagonal components of the permeability tensor $\vec{\vec{K}}$, all other non-diagonal components are assumed to be zero; a and b are the regression

coefficients of the linear model between porosity and logarithmic permeability derived from the data; and ε is the error in the regression model, which introduces the probabilistic feature to this linear regression model.

CMG`s thermal simulator STARS is chosen as a thermal flow simulator. Parameters that describe the production scheme and initial porosity and permeability ensembles come directly to the thermal flow simulator as input. Mineralogical content, elastic, and physical properties of saturated rock should be specified in the petroelastic model as well. Reservoir pressure and temperature, water, oil, and gas saturations are reported as an output of the simulator. These output values and porosity ensemble are employed in the pressure- and temperature-dependent petroelastic model based on Gassmann`s fluid substitution model to generate synthetic seismic attributes, such as P- and S-wave acoustic velocities (V_p and V_s) and corresponding impedances (Z_p and Z_s). The difference in P-wave acoustic impedance has been chosen as a seismic attribute. Chapter 2 contains a description of this petroelastic model's framework. The seismic attributes contain information about coupled geology and reservoir behaviour, such as steam chamber growth (Zagayevskiy and Deutsch, 2011). In order to monitor the change in fluid saturations and temperature distribution, it is necessary to calculate the difference in seismic attributes. The first geophysical survey, which corresponds to the time step when production has not yet begun, is called the baseline survey. Seismic attributes from this survey are subtracted from seismic attributes from any subsequent geophysical survey to find the difference in seismic attributes. At this stage, the entire model is initialized for time step t = 1 and the relationship between all variables is established. The analysis step can be performed to assimilate all available data at time t, if an objective function is not satisfied. The general form of the objective function O is shown in Equation (119) (Gomez-Hernandez et al, 1997). If the objective function is satisfied, all forecasted values of pressure, temperature, and other state variables represent predicted reservoir behaviour at a future time step, for which data are not available. The estimated spatial distribution of porosity and permeability represent the geological model's best up-to-date estimate.

$$O = \sum_{i=1}^{N_{d,t}} \sum_{i_r=1}^{N_e} \sum_{t=1}^{N_t} w_{i,i_r,t} \cdot \left(u_{i,i_r,t} - d_{i,t}\right)^2 \quad (119)$$

where O is the objective function, which has to be minimized; $u_{i,ir,,t}$ and $d_{i,t}$ are the i_r^{th} realization of model estimate and associated datum values at i^{th} location at time t; $w_{i,ir,t}$ is the weight of difference; N_t is the number of time steps from which data are coming; and $N_{d,t}$ is the number of data assimilated at time t.

Data from different sources are incorporated into the ensemble. The entire model is updated based on the covariance structure between all variables. Data may come from vertical surveillance wells, SAGD well pairs, and seismic surveys at thermally operated SAGD oil fields. Once the model is updated, the forecast step is performed to re-establish a relationship between analyzed model parameters and state variables and to predict future model states. If the model is nonlinear, data should be recursively integrated several times into the currently updated reservoir model from the beginning of flow simulation until the objective function is minimized.

The author claims that the described approach leads to a high quality plausible estimation of porosity and permeability fields. Note that SGS, the thermal flow simulator, and the petroelastic model are used as a black box in this data assimilation approach based on EnKF. An analytical solution to optimizing the problem is not required. The only requirements to running EnKF for estimating model parameters and predicting state values are input values of model parameters and a specified production mechanism.

4.4 Validation of Estimation Results

Validation of estimation or prediction results can be carried out in two ways (Gu, 2006). If a synthetic case study is examined, in which there is a known base case or reference case representing a variable's true distribution, results can be validated using the root mean squared error (*RMSE*) concept and linear correlation coefficient (*CC*). *RMSE* is the

average root squared difference between the model realizations and the base case. The root squared error (RSE_{ir}) between a single i_r^{th} realization and a base case can be expressed as shown in Equation (120). The *RMSE* can be found by averaging *RSE*s of all realizations as in Equation (121). Sometimes in petroleum applications, there is more interest in the overall estimation quality and not in reproducing the data locally. For this reason, the local estimate $u^{est}_{i_x,i_y,i_z,i_r}$ is replaced with a smoothed value $\bar{u}^{est}_{i_x,i_y,i_z,i_r}$, which is the average of all estimated values of the i_r^{th} realization falling into a moving smoothing window of size N_{sx} x N_{sy} x N_{sz} with a center at the (i_x, i_y, i_z) block.

$$RSE_{i_r} = \sqrt{\frac{\sum_{i_x=1}^{N_{bx}}\sum_{i_y=1}^{N_{by}}\sum_{i_z=1}^{N_{bz}}\left(u^{true}_{i_x,i_y,i_z} - \bar{u}^{est}_{i_x,i_y,i_z,i_r}\right)^2}{N_{bx}\cdot N_{by}\cdot N_{bz}}} \tag{120}$$

$$RMSE = \sqrt{\frac{\sum_{i_r=1}^{N_e}\left(RSE_{i_r}\right)^2}{N_e}} \tag{121}$$

where RSE_{ir} is the root squared error value computed for the i_r^{th} realization; *RMSE* is the root mean squared error of all *RSE*s over realizations; N_{bx}, N_{by}, and N_{bz} are the number of blocks of a model in *X*, *Y*, and *Z* directions; N_e is the number of realizations or the ensemble size; $u^{true}_{i_x,i_y,i_z}$ is the true or base case value of a variable in the (i_x, i_y, i_z) block of a model; and $\bar{u}^{est}_{i_x,i_y,i_z,i_r}$ is the average of the i_r^{th} estimated realization values over blocks falling into a smoothing window of size N_{sx} x N_{sy} x N_{sz} with a center at the (i_x, i_y, i_z) block.

The correlation coefficient (*CC*) between the base case and the mean of the EnKF estimates may be used as well to validate the model. Equation (122) shows the mathematical expression of *CC.*

$$CC = \frac{\sum_{i_x=1}^{N_{bx}} \sum_{i_y=1}^{N_{by}} \sum_{i_z=1}^{N_{bx}} \left(\overline{u}_{i_x,i_y,i_z}^{est} - \overline{u}^{est} \right) \cdot \left(u_{i_x,i_y,i_z}^{true} - \overline{u}^{true} \right)}{\sqrt{\sum_{i_x=1}^{N_{bx}} \sum_{i_y=1}^{N_{by}} \sum_{i_z=1}^{N_{bx}} \left(\overline{u}_{i_x,i_y,i_z}^{est} - \overline{u}^{est} \right)^2 \cdot \sum_{i_x=1}^{N_{bx}} \sum_{i_y=1}^{N_{by}} \sum_{i_z=1}^{N_{bx}} \left(u_{i_x,i_y,i_z}^{true} - \overline{u}^{true} \right)^2}} \tag{122}$$

where $\overline{u}^{est}$ is the average of model estimates over all realizations and grid blocks; and $\overline{u}^{true}$ is the average of true model values over grid blocks.

The second way to validate estimation and prediction results is to use a mean standard deviation (*MSD*), when the base case is not known. For an individual block, the equation for standard deviation $SD_{ix,iy,iz}$ is:

$$SD_{i_x,i_y,i_z} = \sqrt{\frac{\sum_{i_r=1}^{N_e} \left(u_{i_x,i_y,i_z,i_r}^{est} - \overline{u}_{i_x,i_y,i_z}^{est} \right)^2}{N_e}} \tag{123}$$

The mean value of individual standard deviations over the model can be computed as:

$$MSD = \sqrt{\frac{\sum_{i_x=1}^{N_{bx}} \sum_{i_y=1}^{N_{by}} \sum_{i_z=1}^{N_{bz}} SD_{i_x,i_y,i_z}^2}{N_{bx} \cdot N_{by} \cdot N_{bz}}} \tag{124}$$

Smaller values of *RMSE* show a better match between the base case and the estimate. Absolute values of *CC* closer to 1.0 represent a better match as well. Smaller values of *MSD* show a better convergence of estimated realizations to true distribution.

Chapter 5 – Implementation Details

Chapter 5 shows that it is important in characterization of geological properties of the reservoir to assimilate additional static and dynamic data into petroleum reservoir model by means of EnKF. The data integration is demonstrated with synthetic 2D SAGD case study. The benefit of each data type is distinguished. The chapter also presents characteristics of the proposed data assimilation method and discusses implementation details associated with conducting EnKF. It also proposes and examines EnKF modifications to decrease the computational overburden.

5.1 Main Characteristics of the EnKF

A synthetic 2D SAGD case study demonstrates the value of additionally assimilated data and the benefit of each data type for reservoir characterization using an EnKF-based data integration algorithm. A two-dimensional model is selected for two reasons: first, it is easier to explore the problem on a smaller model; and, second, in SAGD most of the fluid flow occurs in cross sections perpendicular to horizontal well pairs (Butler, 1991). The examined model comprises only one SAGD well pair and can be easily extended to three-dimensional large scale models that cover the area around a single well pad or an entire oil field. The model grid is 50 x 1 x 30 blocks of 1.0 m x 100.0 m x 1.0 m. Figure 12 shows a schematic of the reservoir model grid. The injecting well is placed in block (25, 1, 7) above the producing well, which is in block (25, 1, 3), with a separation distance of 4.0 m. Two vertical observation wells are placed symmetrically relative to the SAGD well pair.

Figure 13 shows base cases of porosity, horizontal permeability, temperature, and the difference in P-wave acoustic impedances for time steps 1 and 2 (90 and 180 days after the beginning of the bitumen production respectively) with observation locations. Data are sampled from gridded base cases and, thus, represent the same scale. It is clear that porosity and permeability fields consist of two zones of low and high values respectively.

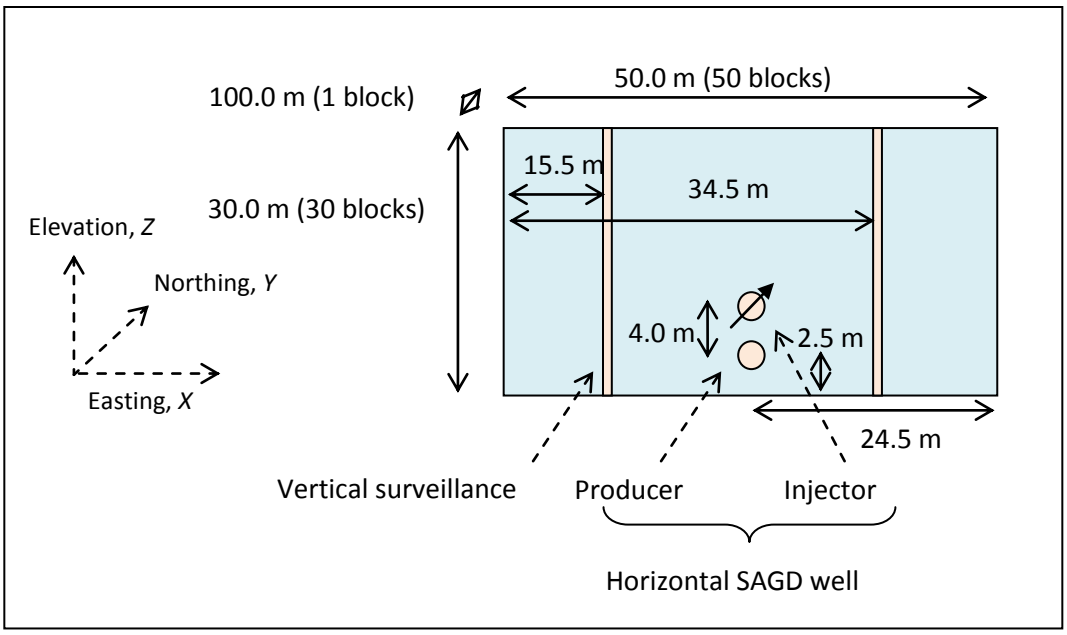


Figure 12: Schematic and configurations of 2D SAGD petroleum reservoir model

The reservoir is preheated for 90 days in a region around SAGD horizontal wells before steam injection and production begin. It is assumed that the SAGD production scheme and regime are known to a modeler. The porosity and permeability data are sampled in a co-locative fashion from two vertical surveillance wells before any warm-up procedure. Hence, 14 data of porosity and 14 data of permeability from each vertical well are available. Twenty-eight permanent temperature gauges are located along vertical surveillance wells and in the SAGD well pair. The temperature is measured three times: just before production begins (time step 0 – 90.0 days), and after 90 and then 180 days of production (time steps 1 and 2 – 180.0 and 270.0 days respectively). A geophysical baseline survey is conducted at time step 0 and two more surveys are conducted at time steps 1 and 2. Reservoir coverage of the seismic is even and extensive with 250 observation locations. Figure 14 shows histograms of base cases. Figure 15 shows histograms of data. The histograms of hard porosity and horizontal permeability data sampled from a single eastern well do not depict true distributions of porosity and permeability. But the same distributions are depicted by porosity and permeability data coming from two observation wells. Temperature data barely follow base case

distributions for all time steps. Since seismic data are an exhaustive dataset, their distributions are close to base case for both time steps.

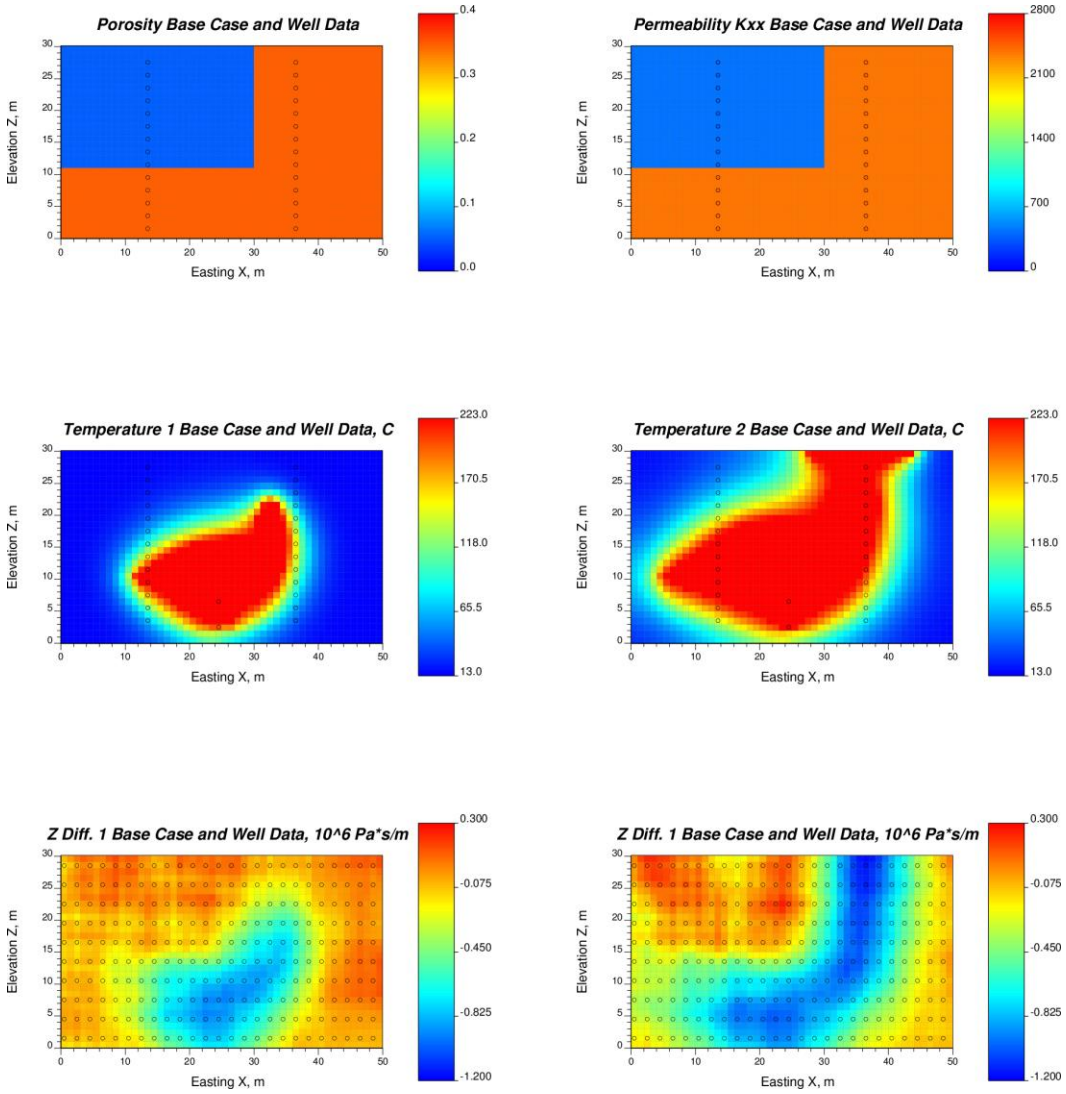


Figure 13: Base case maps of porosity, horizontal permeability, temperature, and difference in P-wave acoustic impedances from two time steps. Corresponding observation locations of every data type are shown as circles

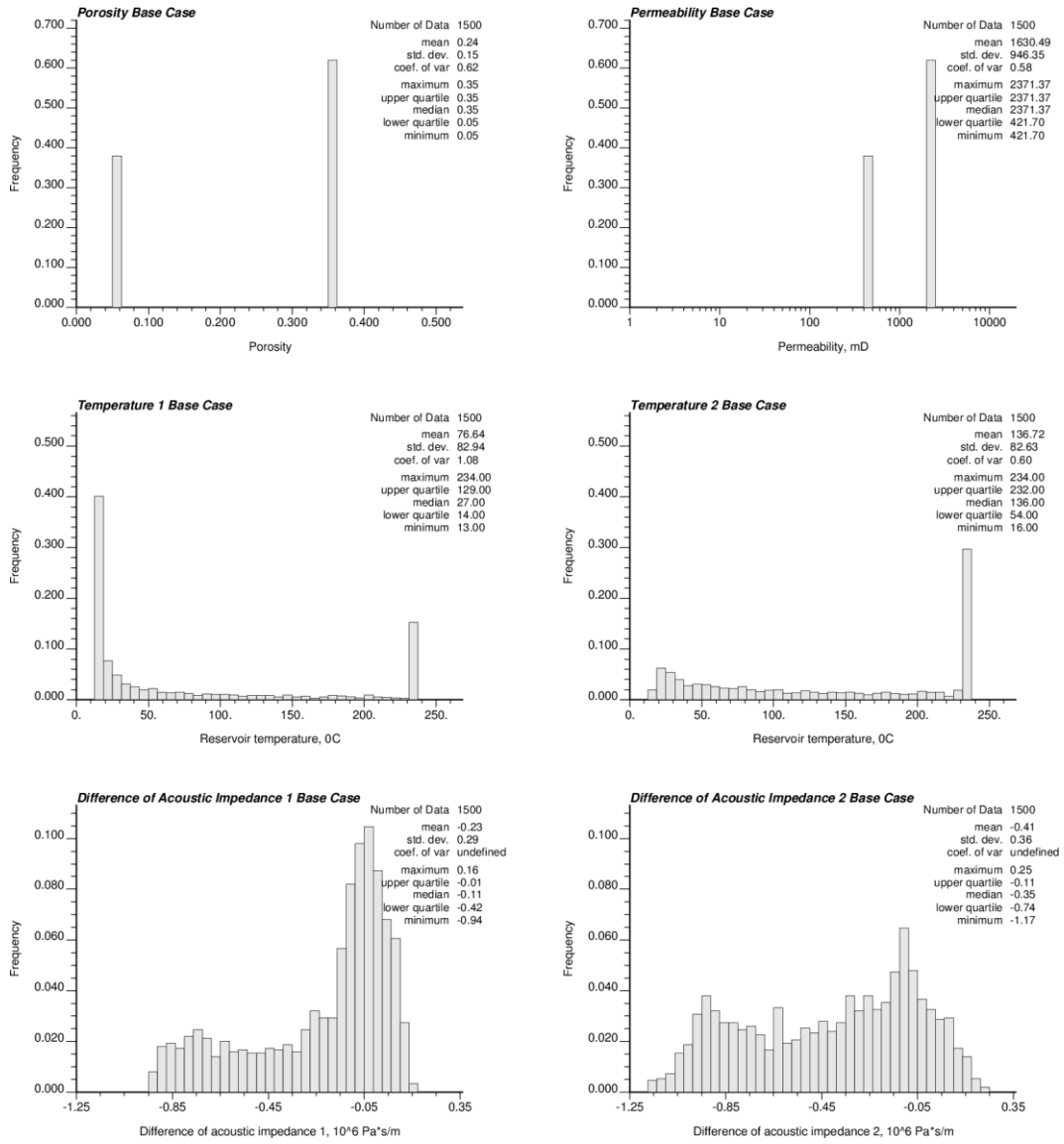


Figure 14: Histograms of base case porosity, horizontal permeability, temperature, and difference in P-wave acoustic impedances from two time steps

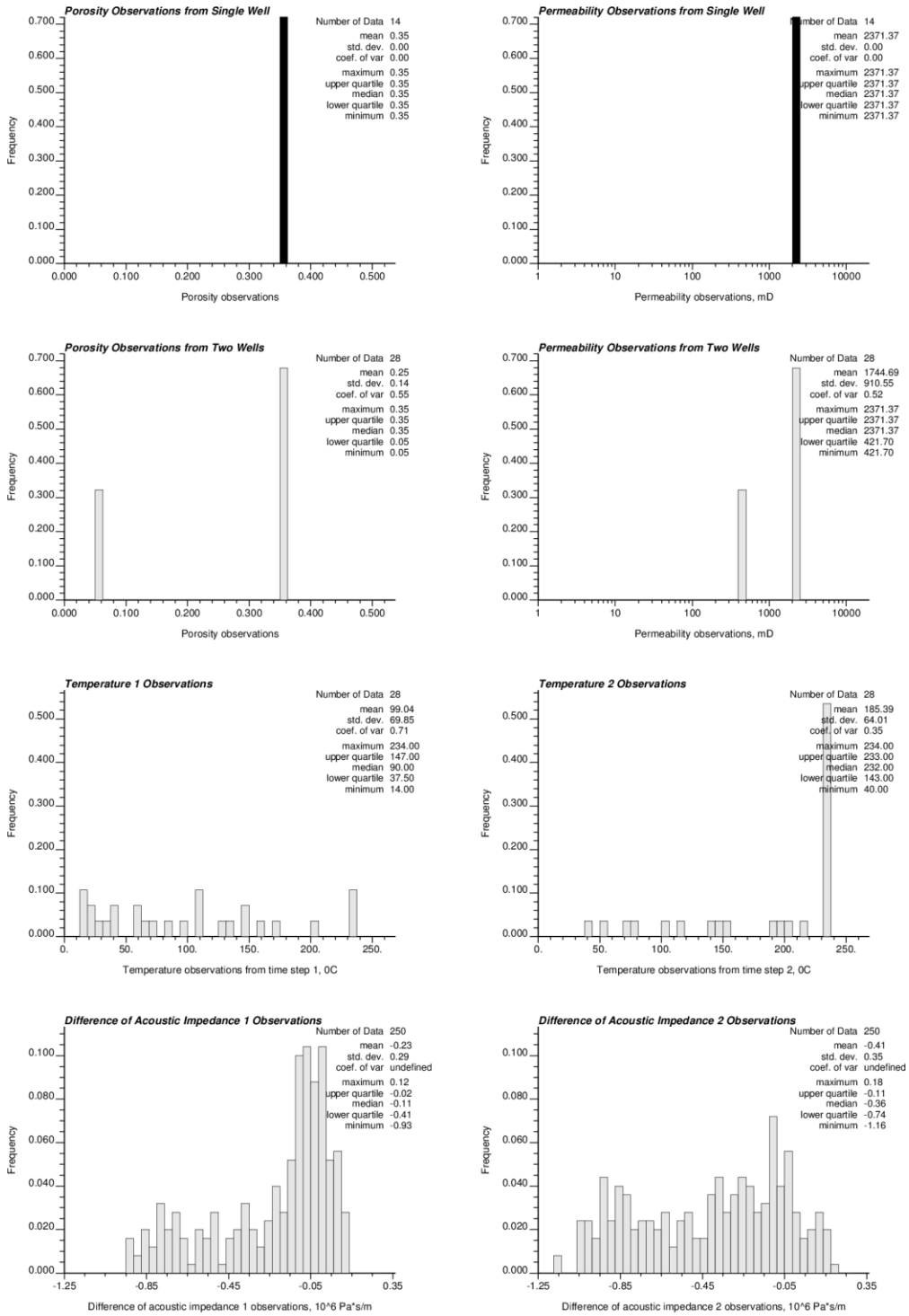


Figure 15: Histograms of observations of porosity and horizontal permeability data sampled from one and two vertical surveillance wells, temperature data from two time steps sampled from vertical wells and SAGD well pair, and difference in seismic attributes from two time steps sampled extensively over the field

Ten different cases of estimating spatial distributions of porosity and permeability are examined. The proposed EnKF-based continuous data assimilation technique with covariance localization is employed to integrate data and estimate model parameters. Localization techniques are discussed later in this chapter. It is worthwhile to mention now that covariance localization leads to better, smoother, and more stable estimation results with a smaller number of realizations (Fahimuddin, 2010). The value of additional data and the benefit of each data type integrated into the model are shown. The case studies are briefly described below:

1. No data are used in estimation; isotropic initial ensembles of porosity and permeability are the best estimates in this case.
2. 14 porosity data from a single eastern (on right) surveillance well are used, which do not capture the low value zone of the base case.
3. 28 porosity data from both surveillance wells are used.
4. 28 porosity and 28 permeability data from both surveillance wells are used. A comparison of cases 3 and 4 reveals redundancy of porosity and permeability data.
5. 14 porosity data from a single eastern surveillance well and 28 temperature measurements from all wells at time step 1 (180.0 days) are used.
6. 28 temperature observations from time step 2 (270.0 days) are integrated in an ensemble from the previous case.
7. 14 porosity data from a single eastern surveillance well and 250 differences in acoustic impedances from time step 1 (180.0 days) are used.
8. 250 differences in acoustic impedances from time step 2 (270.0 days) are integrated in an ensemble from the previous case.
9. 14 porosity data from a single eastern surveillance well, 28 temperature measurements from time step 1 (180.0 days), and 250 differences in acoustic impedances from time step 1 (180.0 days) are used.
10. 28 temperature measurements from time step 2 (270.0 days) and 250 differences in acoustic impedances from time step 2 (270.0 days) are integrated in an ensemble from the previous case.

The initial ensemble of porosity consists of 1000 realizations to avoid ensemble collapse. A 1:10 ratio of the number of data to the number of realizations is satisfied approximately (Fahimuddin, 2010). The initial porosity ensemble is generated as follows: first, an ensemble of normal scores is generated using the nested isotropic semivariogram model shown in Equations (125) and (126). It is believed that this semivariogram model is an adequate representation of the true spatial distribution of porosity in normal scores. Then, the generated normal scores are modified to obtain the initial porosity ensemble in original units of porosity. Normal scores are back-standardized to normal distribution with a mean of 0.25 and variance of 0.14, which roughly correspond to the mean and variance of porosity data from both surveillance wells. Note that negative values of porosity are present in the initial ensemble and are not removed to preserve the covariance function.

$$\gamma(h) = 0.01 + 0.99 \cdot Exp_{a=60.0m}(h) \tag{125}$$

$$Exp_a(h) = 1.0 - \exp\left(-\frac{3h}{a}\right) \tag{126}$$

where $Exp(h)$ is the positive-definite exponential semivariogram model (Deutsch and Journel, 1998); a is the semivariogram range; and h is the lag distance.

The relationship between porosity and permeability is established using the porosity-log permeability model discussed in Chapter 4 (Equation (117)). It is assumed that these two variables are perfectly correlated, and that one parameter can be certainly estimated, when the value of the second parameter is known at the same location. Other components of the permeability tensor can be found by knowing relationship with horizontal permeability (Equation (118)). In this case study, horizontal permeability is assumed to be the isotropic and its value is double of vertical permeability (Equation (127)). Because of the perfect correlation between porosity and permeability and the well-understood relationship between them, porosity and permeability data are deemed to be redundant, which is shown later.

$$K_{xx} = K_{xx}$$
$$K_{yy} = K_{xx} \quad (127)$$
$$K_{zz} = 0.5 \cdot K_{xx}$$

CMG's thermal flow simulator STARS is used to simulate reservoir temperature, pressure, and fluid saturations, which are used later on in the petroelastic model based on Gassmann's fluid substitution model to generate synthetic seismic attributes (Zagayevskiy and Deutsch, 2011a). Physical and elastic properties of constituent minerals and fluids are assumed to be known.

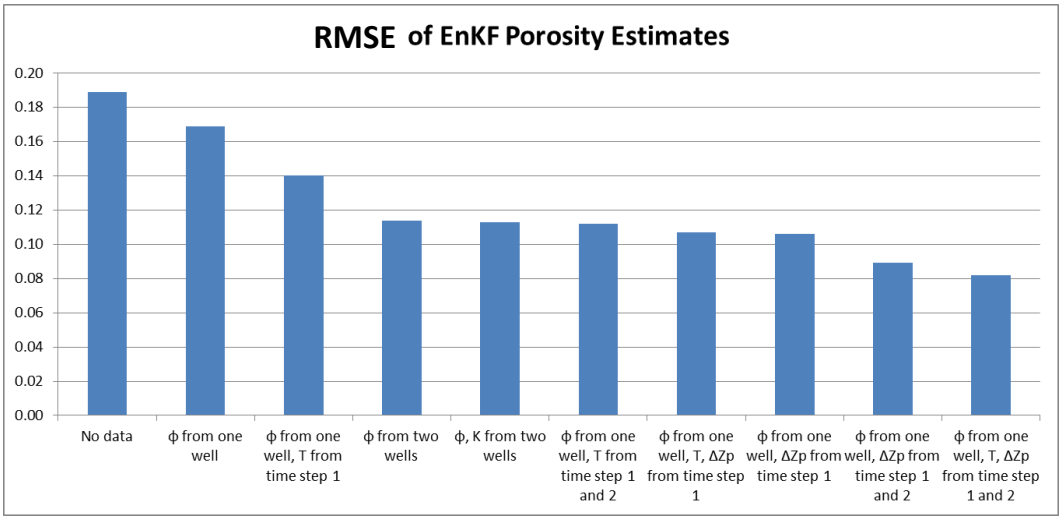
EnKF is implemented with the FORTRAN program enkf.exe. Its parameter file, with a description, is in the Appendix. Table 7 contains a summary and comparison of estimation results from ten cases through *RMSE*, *CC*, and *MSD* validation measures. The bar chart (Figure 16) shows values of *RMSE* validation measures for every case. Figure 17 - Figure 23 show the means of EnKF-estimated realizations of porosity and horizontal permeability with associated estimation variance maps. Vertical permeability can be computed from horizontal permeability by dividing its values by two. The results show that incorporation of additional data from different sources and time steps improves the estimation quality of both porosity and permeability fields. The permeability measurements are not included in most of the studied cases, because they are redundant with the porosity data (compare estimation results from the case, in which porosity and permeability are used, with the case, in which only porosity data are assimilated). Additional temperature and seismic data improve porosity and permeability estimates. Incorporating dynamic data from both time steps improves estimates even further. It might be argued that exhaustive seismic data brings more information to the model than temperature measurements, because of a denser sampling pattern. However, the seismic data's informational value should be assessed against the sampling cost, and additional hard data may be preferred instead. Also, uncertainty in estimation drops as more data are integrated into the model. Note that additional constrains are applied to porosity and permeability values, which state that porosity estimates can only fall between zero and one and that permeability estimates

cannot be negative. These constraints decrease estimation uncertainty a little bit more, especially for the left upper corner of permeability estimates. To sum up, flow zones and barriers are predicted fairly well in this synthetic 2D example by integrating hard and soft data to the petroleum model. The information value of dynamic soft data is undoubtedly significant. However, note that if the estimated spatial distribution of porosity and permeability are used to generate values of dynamic variables at sampled locations, it is not necessary to have a perfect match between the dynamic variables' produced values and the observed data. This is because of the modeled thermal petroleum reservoir system's nonlinear nature. Recursive re-assimilation of the data would be required for a very large mismatch.

Table 7: Comparison of the EnKF porosity and permeability estimates derived using different data sets – 2D SAGD case study

Case	Data	Number of data	Porosity			Permeability		
			RMSE	***CC***	***MSD***	***RMSE***	***CC***	***MSD***
1	No data	0	0.189	0.254	0.142	1618	0.272	1789
2	ϕ data from one observation well	14	0.169	0.506	0.115	1410	0.485	1485
3	ϕ data from two obs. wells	28	0.114	0.835	0.089	884	0.861	982
4	ϕ and *K* data from two obs. wells	56	0.113	0.835	0.089	881	0.861	978
5	ϕ and *T* from one time step	42	0.140	0.800	0.096	1182	0.707	1324
6	ϕ and *T* from two time steps	70	0.112	0.864	0.085	865	0.808	1010
7	ϕ and ΔZ_p from one time step	264	0.106	0.855	0.090	857	0.838	969
8	ϕ and ΔZ_p from two time steps	514	0.089	0.899	0.076	691	0.887	772
9	ϕ, *T* and ΔZ_p from one time step	292	0.107	0.856	0.086	840	0.855	989
10	ϕ, *T* and ΔZ_p from two time steps	570	0.082	0.914	0.070	629	0.898	711

ϕ – porosity; *K* – permeability; *T* – temperature; ΔZ_p – difference in P-wave acoustic impedances



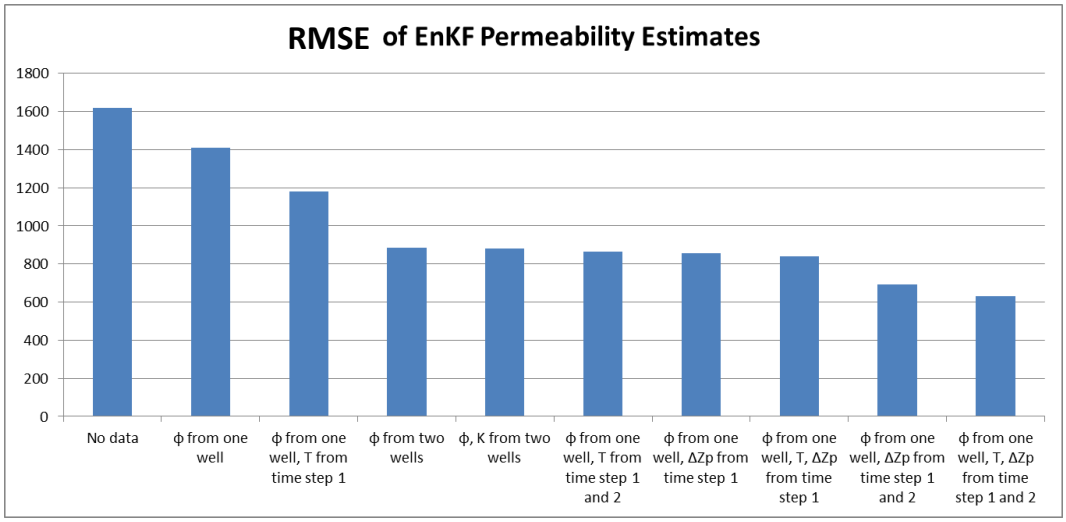


Figure 16: Comparison of *RMSE* values of EnKF porosity and permeability estimates (top and bottom charts respectively)

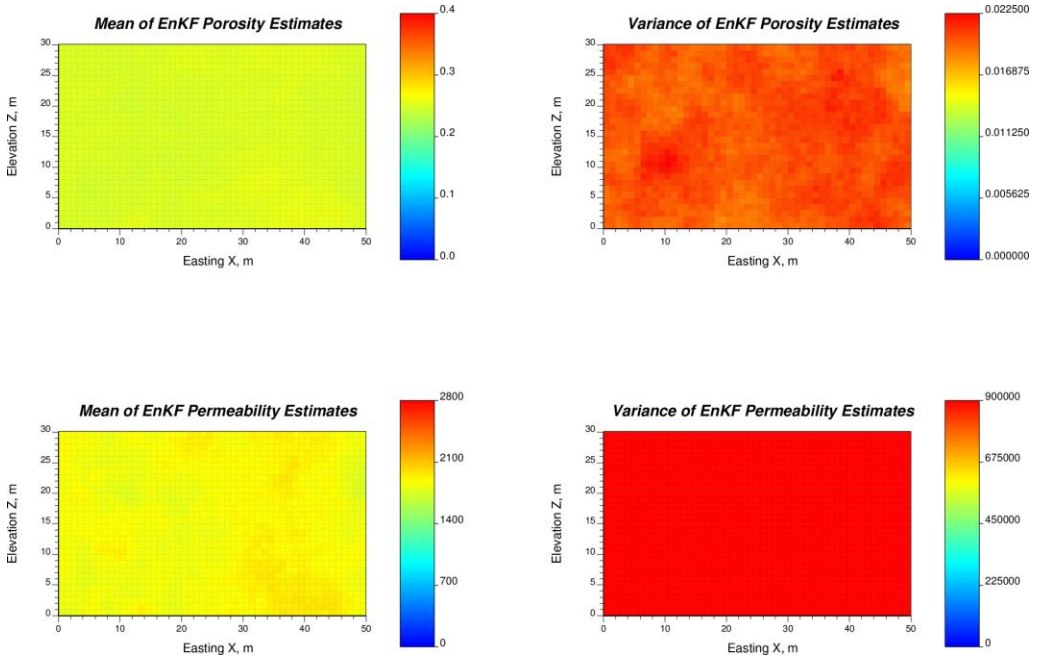


Figure 17: Maps of mean and variance of realizations of EnKF porosity and permeability estimates derived using no data (initial ensembles)

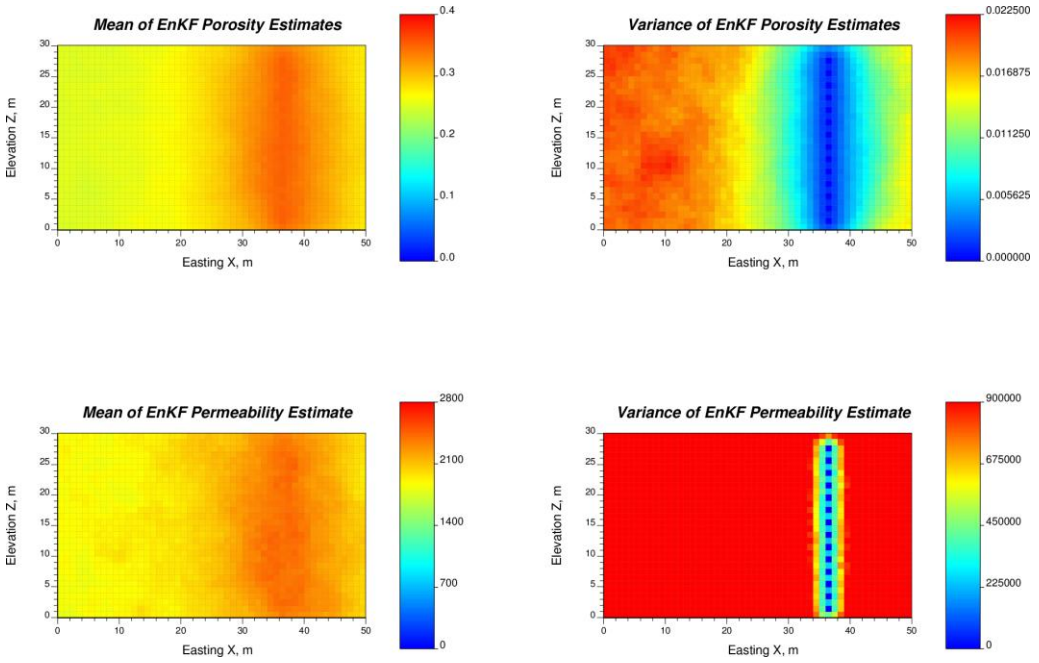


Figure 18: Maps of mean and variance of realizations of EnKF porosity and permeability estimates derived using porosity data from single surveillance well

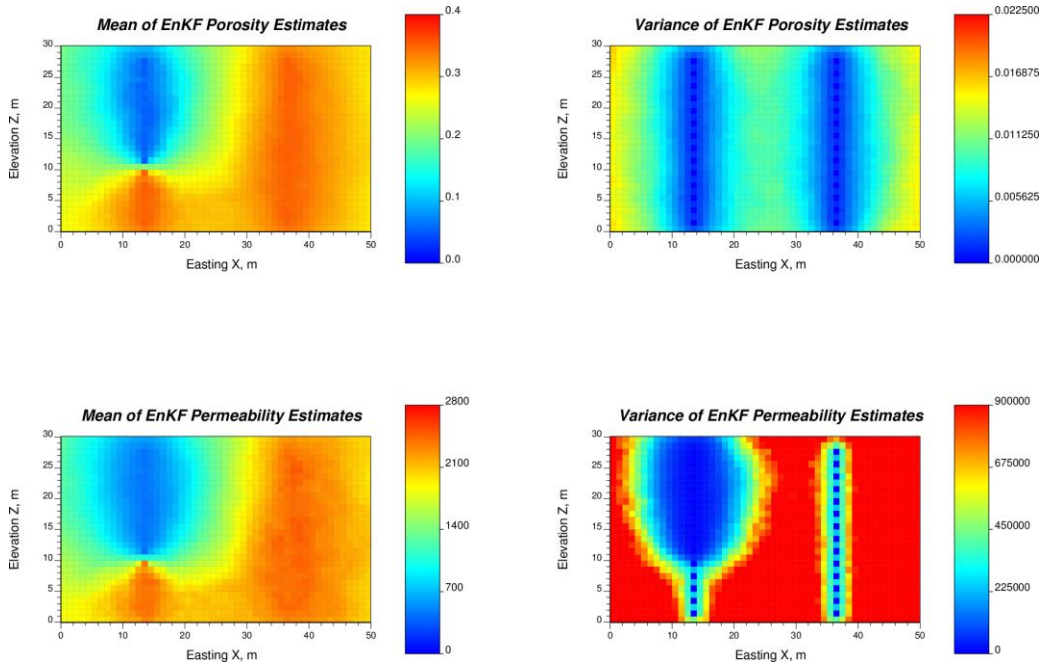


Figure 19: Maps of mean and variance of realizations of EnKF porosity and permeability estimates derived using porosity data from both surveillance wells

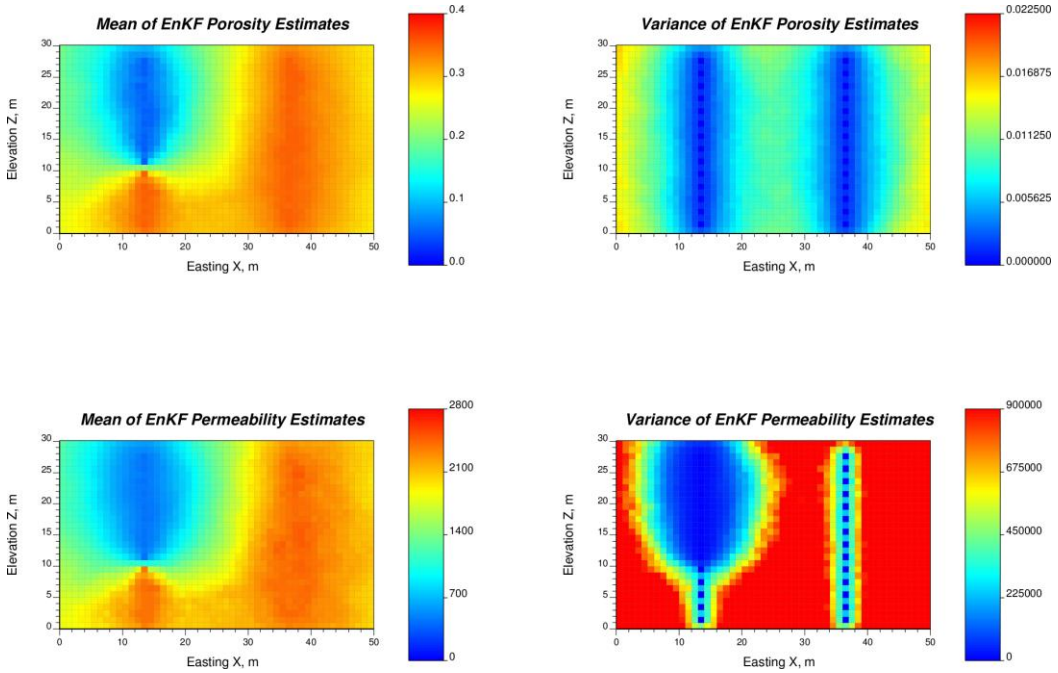


Figure 20: Maps of mean and variance of realizations of EnKF porosity and permeability estimates derived using porosity and permeability data from both surveillance wells

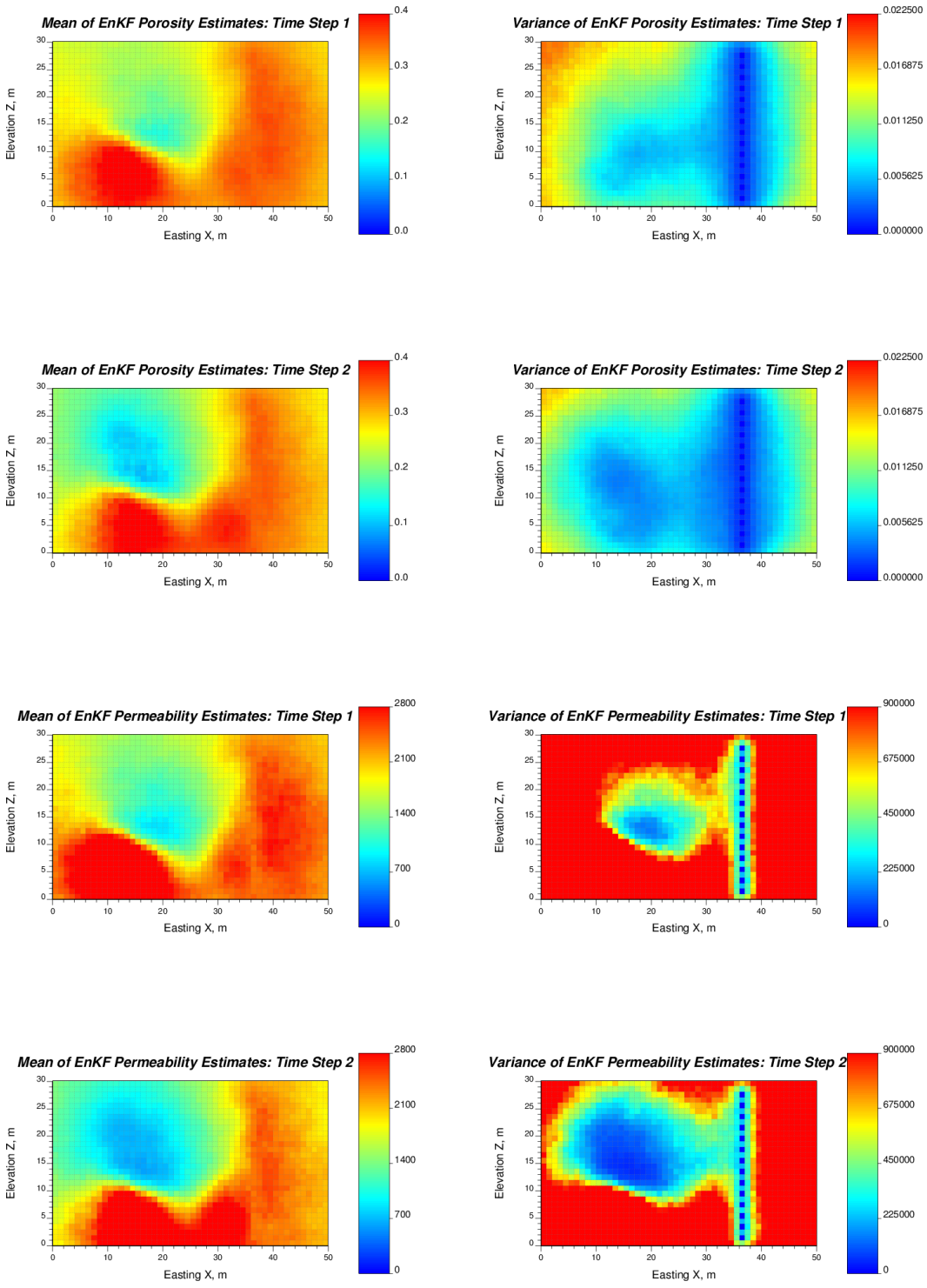


Figure 21: Maps of mean and variance of realizations of EnKF porosity and permeability estimates derived using porosity data from a single surveillance well and temperature observations from two time steps

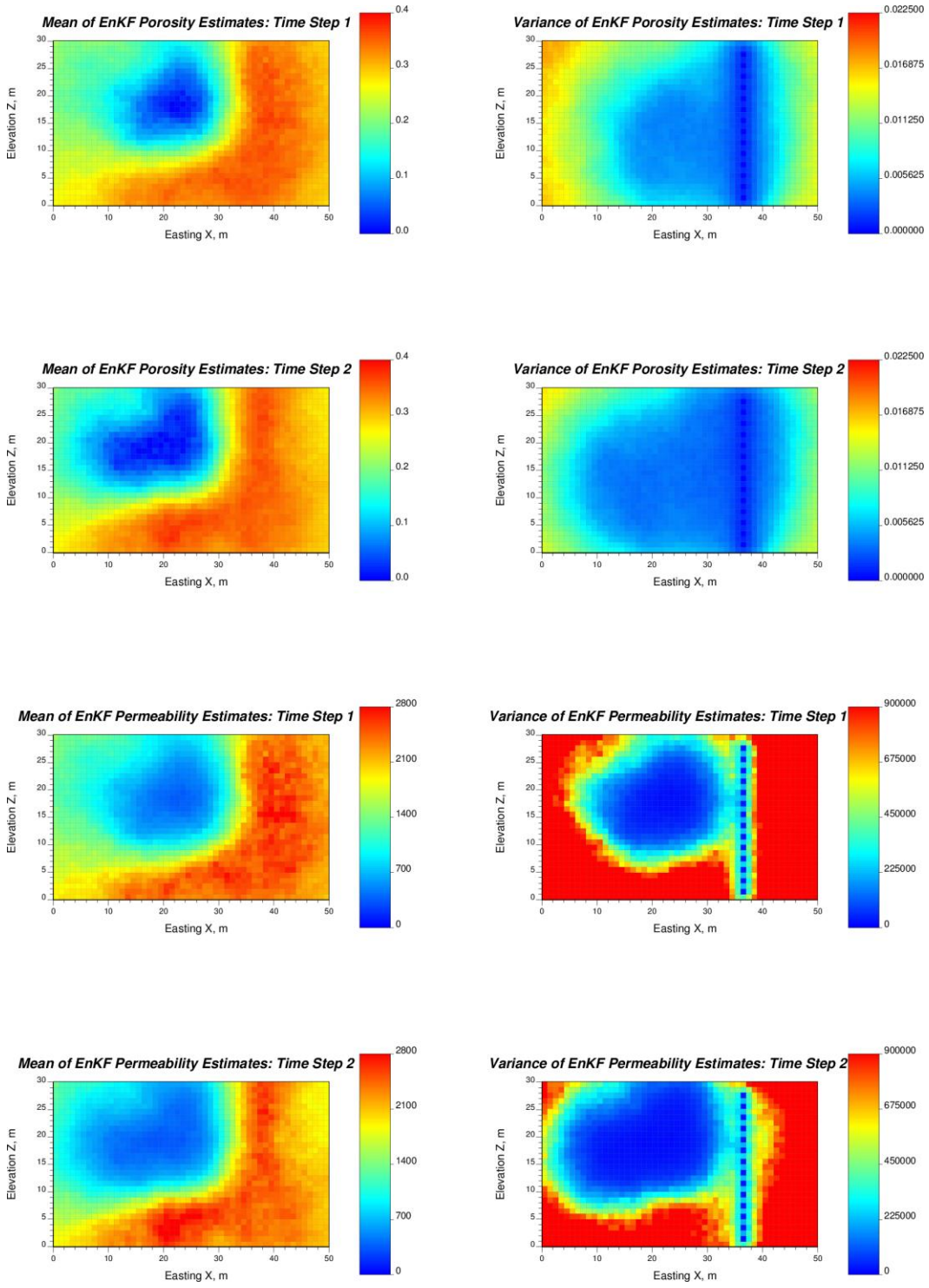


Figure 22: Maps of mean and variance of realizations of EnKF porosity and permeability estimates derived using porosity data from a single surveillance well and differences in acoustic impedances from two time steps

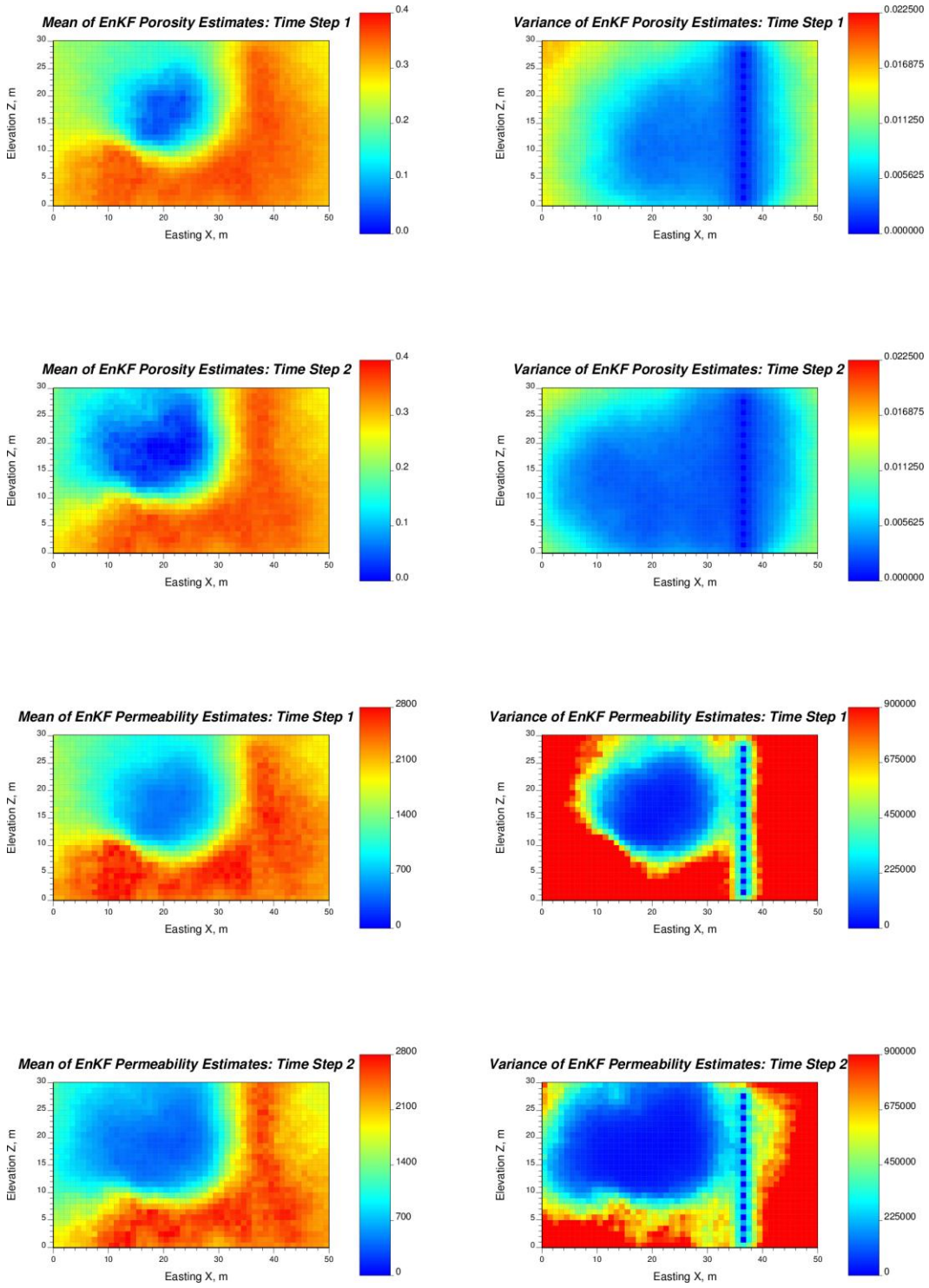


Figure 23: Maps of mean and variance of realizations of EnKF porosity and permeability estimates derived using porosity data from a single surveillance well, temperature observations and differences in acoustic impedances from two time steps

5.2 Additional Considerations

Implementation details and characteristics of EnKF are summarized here. Also modifications to EnKF are proposed and shown that lead to better estimates with less computational cost.

5.2.1 Model Properties

Even though EnKF finds its best application in linear and Gaussian systems, it can also be effectively used for nonlinear and non-Gaussian systems, if modifications are applied. For nonlinear systems, an iterative and recursive procedure for integrating data through EnKF can be implemented. This is called the recursive confirmation step (Wen and Chen, 2006). Another option is to replace EnKF with EnKS, in which all data from different time steps are assimilated at once (Evensen and van Leeuwen, 2000). A normal score transformation can be used for systems that do not follow the Gaussian distribution (Deutsch, 2002). There is no need to modify or standardize in any way units of model variables in EnKF, if they follow distributions close to normal. The forecast step requires input values in original units and generates output variables in original units as well. The Kalman gain in the analysis step is the only component that might be sensitive to a choice of units. However, because of its nature (see Equation (114)), the Kalman gain is a dimensionless matrix and, thus, it does not depend on units.

5.2.2 Initial Ensemble

Generating the initial ensemble is a vital procedure, since the stored covariance structure inside it influences the estimation of model parameters and the prediction of state variables at all subsequent time steps. It is recommended to generate an initial ensemble of model parameters that follows normal distribution. The initial ensemble of state variables should be derived from the initial ensemble of model parameters and the known relationship between them, which is expressed in the form of a model operator. Values from different realizations have to be spatially correlated to each other. An SGS with a proper semivariogram model $\gamma(\boldsymbol{h})$ derived from the data is one of the best ways

to generate the initial ensemble of model parameters (Deutsch and Journel, 1998). If the generated initial or any other ensembles have a distribution that is too far from the Gaussian, a normal score transformation may be applied to a variable of interest. Also, patterns of spatial continuity of variables stored in the initial ensemble should depict their realistic configuration. For instance, spatial distribution of temperature should be tied with well configurations, geology of the reservoir, and spatial distribution of gas saturation in SAGD. Otherwise, very weak estimates are obtained (Zagayevskiy and Deutsch, 2011b).

5.2.3 Ensemble Size and Number of Data

It is claimed that the ensemble size should be at least ten times greater than the number of assimilated data (Fahimuddin, 2010); otherwise, spurious long-range covariances occur that lead the ensemble to collapse. The reason lies in an insufficient number of degrees of freedom and loss of the matrix rank. Therefore, if an exhaustive dataset, such as 4D seismic attributes, is integrated into numerical models, a very large number of realizations is required to get a reliable estimate, which causes computational time to increase. Applying localization of the updating matrix or covariance matrix may help to solve this issue. Localization leads to a local update of the ensemble matrix with possible artifacts depending on the updating window size or selected smoothing function and, thus, it is an approximation of a global update.

Localization of the updating matrix implies that only ensemble members lying in a region around the assimilated data are updated at the analysis step. The rest of the ensemble is kept unchanged. The updating region is specified by a modeler as an updating window and is intended to eliminate the long-range spurious correlation between model realizations. Mathematically local updates based on specified window size can be expressed as shown in Equation (128), where only a Kalman gain is modified. One-datum-at-time assimilation is implemented in the program enkf.exe, using a rectangular updating window, the size of which is specified by a user.

$$\boldsymbol{K}_t^{l,u} = \hat{\boldsymbol{C}}_t^f \cdot \boldsymbol{H}_t^T \cdot \left[\boldsymbol{L}_t^u\right] \cdot \left(\boldsymbol{H}_t \cdot \hat{\boldsymbol{C}}_t^f \cdot \boldsymbol{H}_t^T + \boldsymbol{\Sigma}_t\right)^{-1} \quad (128)$$

where $\boldsymbol{L}^u{}_t$ is the N_{dt} x N_{dt} updating localization matrix consisting of zeros and ones that determine part of the ensemble to be updated at time step t; N_{dt} is the number of integrated data.

Localization of the covariance matrix is slightly different from localization of the updating matrix. In examples shown below, it is argued that the covariance matrix localization produces results with higher accuracy. Localization of the covariance can be expressed as shown in Equation (129), where only Kalman gain is modified, keeping other EnKF equations unchanged. Symbol ◦ stands for an element-wise product of two matrices of same size (Equation (130)). Localization of the covariance matrix is implemented through a smoothing positive definite matrix $\boldsymbol{L}^c{}_t$ eliminating the influence of a long-range relationship at the analysis step and, thus, preserving the matrix's rank. A fifth-order correlation function of Gaspari and Cohen is widely used as a covariance localization matrix $\boldsymbol{L}^c{}_t$ (Fahimuddin, 2010).

$$\boldsymbol{K}_t^{l,c} = \left[\boldsymbol{L}_t^c \circ \hat{\boldsymbol{C}}_t^f\right] \cdot \boldsymbol{H}_t^T \cdot \left(\boldsymbol{H}_t^T \cdot \left[\boldsymbol{L}_t^c \circ \hat{\boldsymbol{C}}_t^f\right] \cdot \boldsymbol{H}_t^T + \boldsymbol{\Sigma}_t\right)^{-1} \quad (129)$$

$$\left[\boldsymbol{L}_t^c \circ \hat{\boldsymbol{C}}_t^f\right]_{i,j} = \left[\boldsymbol{L}_t^c\right]_{i,j} \cdot \left[\hat{\boldsymbol{C}}_t^f\right]_{i,j} \quad (130)$$

where $\boldsymbol{L}^c{}_t$ is the covariance localization matrix of size N_v x N_v represented by a continuous function; ◦ stands for the element-wise matrix product or Hadamard product.

Figure 24 shows an example of localization of the updating matrix for an exhaustive data set such as seismic data. The figure also shows a comparison of global update and update with localized covariance matrix. A base case of a 1500 grid block 2D model (50 blocks in X direction and 30 in Z direction) represents a difference in acoustic impedances from baseline and subsequent surveys. Two-hundred and fifty data are

integrated in the petroelastic model to restore the variable's full distribution. A total of 10 realizations is used as an ensemble size. The updating window size is 3 x 3 blocks. Exponential weighting (Equation (131)) is chosen to localize the covariance matrix. The same localization function was used in a 2D SAGD case study. Figure 24 shows that the mean of EnKF estimates of the difference in acoustic impedances obtained by global integration of data does not match the base case well, because there are not enough degrees of freedom. Most of the data is not reproduced. However, when localization is applied, seismic estimates reproduce the base case fairly well. In this example, localizing the covariance matrix leads to better results. Localizing the updating matrix produces some artifacts.

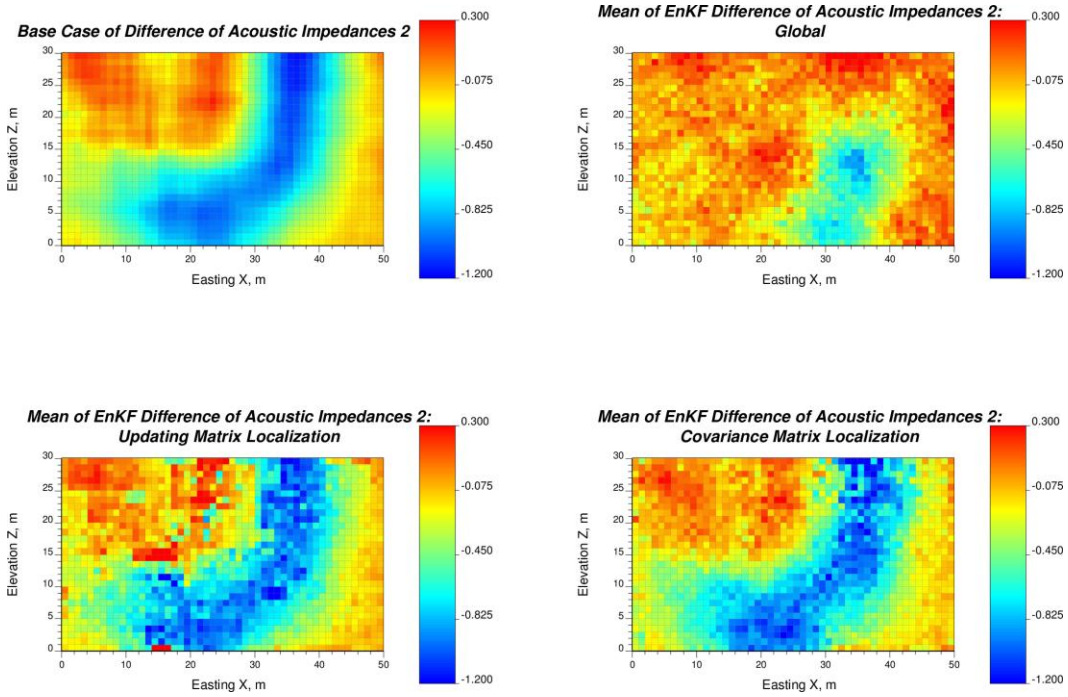


Figure 24: EnKF global and local estimates of the difference in acoustic impedances ΔZ_p using an exhaustive data set of 250 data and a small ensemble of 10 realizations: ΔZ_p base case, mean of realizations of ΔZ_p global estimates, mean of realizations of ΔZ_p estimates with localized updating matrix, and mean of realizations of ΔZ_p estimates with localized covariance matrix

$$L_t^c(h) = \exp\left(-3 \cdot \frac{h}{\sqrt{B_x^2 + B_y^2 + B_z^2}} \right) \tag{131}$$

where h is the distance between assimilated datum and the estimated location; B_x, B_y and B_z are the length, width, and height of the model grid in *X*, *Y,* and *Z* directions.

Figure 25 provides another example of exponential weighting (Equation (131)). What is provided here is a comparison of permeability field estimates derived using conventional EnKF, a localized updating matrix and localized covariance matrix. One hundred realizations are conditioned to 24 permeability data from both vertical surveillance wells described in a synthetic 2D SAGD case study. The *RMSE* values of permeability estimates obtained through global data integration, local data integration with localized updating matrix, and local data integration with localized covariance matrix are 1170.0, 987.5, and 885.9 respectively. The *RMSE* of initial permeability ensemble is 1680.0. Based on *RMSE* and graphical representation of permeability estimates (Figure 25), it is argued by the author that even though all estimates improved in comparison to initial ensemble, localizing the covariance matrix again leads to better estimation results. It just unintentionally happens that randomly selected size of updating window in localization of updating matrix leads to fair good reproduction of horizontal boundary between low and high value zones. Note that such boundary is actually an artifact of the localization technique caused by updating window size and particular spatial distribution of the base case.

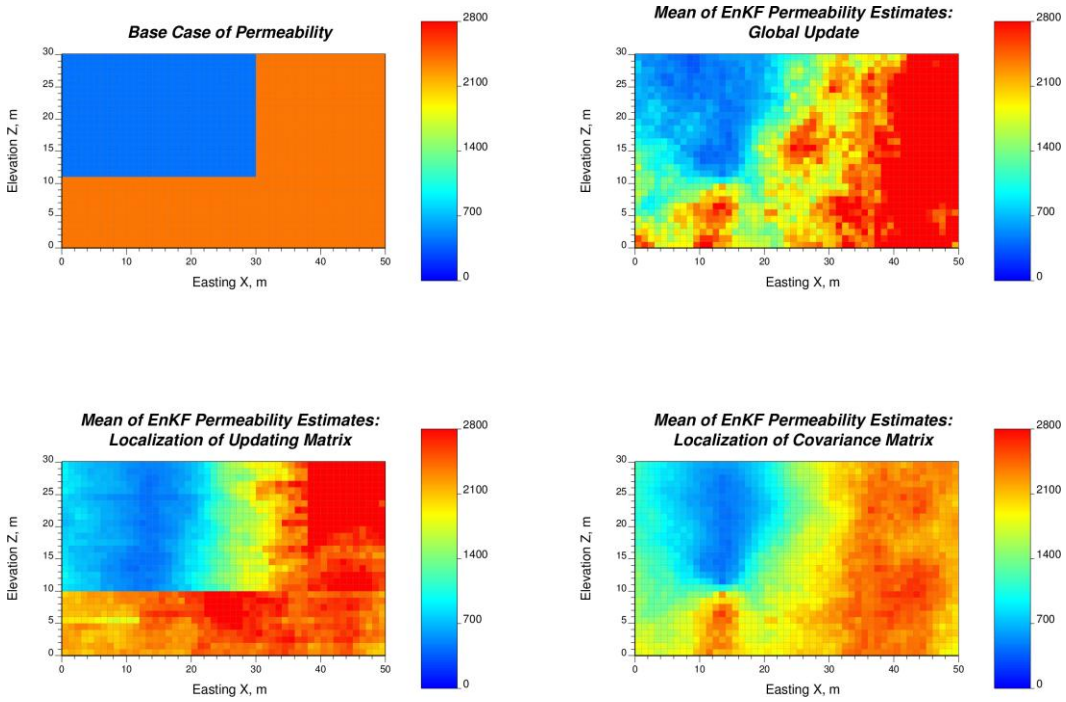


Figure 25: EnKF global and local estimates of horizontal permeability K_{xx} using 100 realizations: K_{xx} base case, mean of realizations of K_{xx} global estimates, mean of realizations of K_{xx} estimates with localization of an updating matrix, and mean of realizations of K_{xx} estimates with localization of a covariance matrix

5.2.4 Computational Time

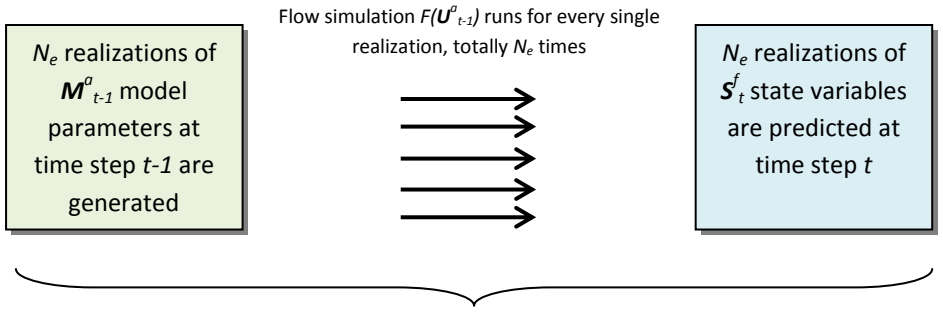
In addition to the model size, two main factors determine the computational cost of the EnKF. They are the ensemble size (number of realizations) at the analysis step, and the number of the flow simulator runs at the forecast step. The number of the flow simulator runs comprises the larger portion of computational overburden. Attempts are made to decrease computational time and increase the model's estimation quality. As was mentioned, one possible and effective solution to decrease the number of realizations is to localize the updating or covariance matrices. But there is no single answer by how much exactly the localization techniques reduce the computational time. The CPU time is tightly related to number of realizations. The required number of realizations is selected based on desired quality of estimates and number of data to be

integrated. Larger ensemble size leads to better estimates. Also integration of larger datasets requires more realizations in order to avoid ensemble collapse. The localization techniques can reduce required number of realizations for integration of large datasets with decreased estimation quality. This decrease in N_e depends on parameters of the localization techniques, such as window size in localization of updating matrix or type of smoothing function in localization of covariance matrix. Therefore, ability of a localization technique to estimate distribution of a variable should be examined on a case-by-case basis in order to understand by how much ensemble size can be reduced to properly assimilate available data. From the author's experience, mentioned localization techniques helped to reduce required number of realizations from thousand to thirty-twenty realizations preserving relatively the same estimation quality of model parameters.

Computational cost of the forecast step is determined by product of flow simulation run time of single realization and total number of realizations. This computational cost can be minimized by reducing number of flow simulation runs, i.e. by decreasing number of realizations that are used in flow simulation. One of the possibilities is to rank the realizations, for instance, according to distribution's quantiles, and use only N_{se} selected realizations (quantile values) in flow simulation. Rest of the realizations should be modified at the forecast step in accordance to change of chosen realizations, which are used in flow simulation. In this case, reduction of computational cost is approximately N_{se}/N_e, where N_{se} is the number of selected realizations from ranking, and N_e is the ensemble size or original number of realizations. Additional little computational time is required to adjust rest of the realizations. Another option is to propagate ensemble mean by the flow simulator and add variations to the forecasted mean. Corresponding reduction in computational cost is the highest and equals to $1/N_e$. The drawback of such approach is that mean of realizations is not necessarily depicts realistic spatial distribution of the variable of interest in comparison to realizations themselves, and spatial distributions of predicted state variables might be unrealistic. The alternative approach to this mean-based approach is to use realization that corresponds to distribution's median (P50) instead of ensemble mean, but reconstruction of other realizations may be problematic with forecasted median. However, despite mentioned

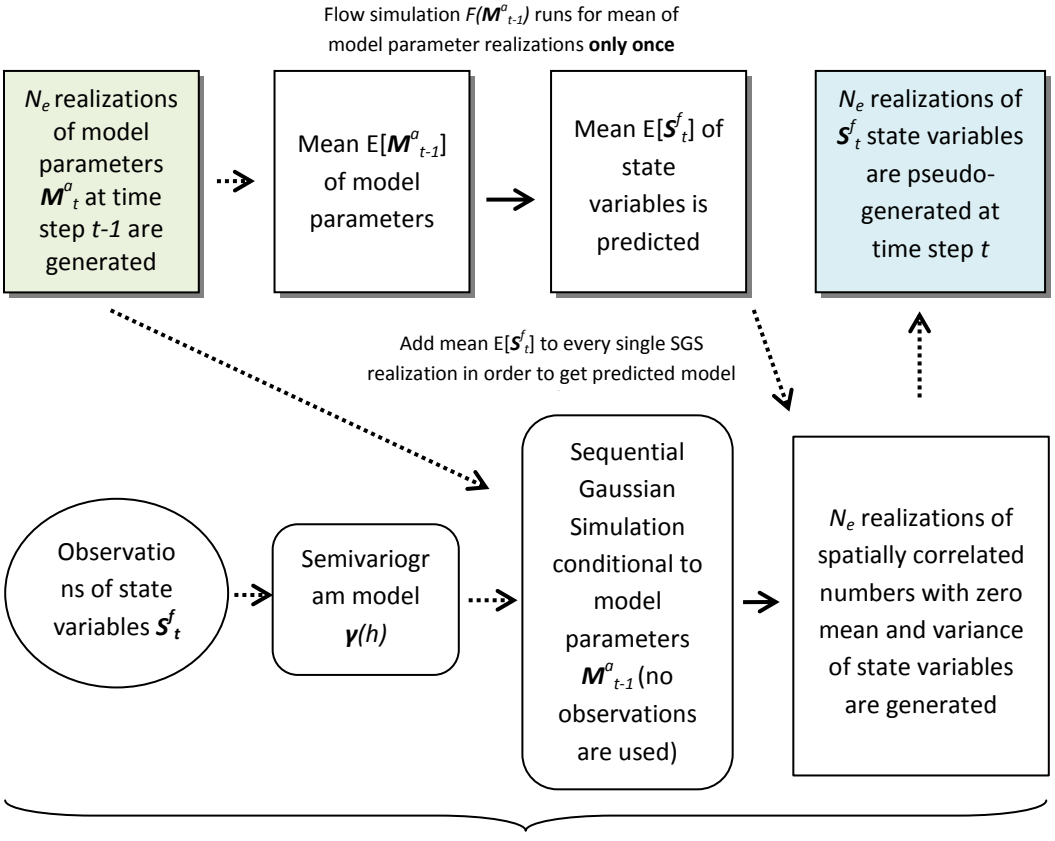
advantages of using ranked realizations, the shortcut based on the concept of the mean of realizations and co-simulation of state variables conditional to model parameters is proposed in this thesis in order to minimize the number of flow simulation runs. This approach is easy in implementation, and generated pseudo-realizations at forecast step average out to plausible forecasted mean. This shortcut is described in detail below. While the conventional forecast step is depicted schematically in Figure 26, the proposed shortcut is shown in Figure 27. In a conventional procedure, the flow simulator uses every single realization of model parameters to obtain a corresponding realization of state variables. When the ensemble size is large and a reservoir model is not simple, applying EnKF to data integration becomes impractical due to the very high computational cost. If the proposed shortcut is applied, flow simulation runs only once, which can save considerable time. The shortcut's proposed algorithm is as follows: the mean of model parameters at time step *t-1* is used to get the mean of state variables at time step *t* by running flow simulation once. At the same time N_e realizations of spatially correlated random numbers are co-simulated with realizations of model parameters by means of Sequential Gaussian Simulation unconditional to data. A proper semivariogram model obtained from observations of state variables is used in co-simulation. Thus, generated SGS realizations should have zero mean and variance equaled to the variance of the state variables' observations. The reason to constrain realization to model parameters lies in linear nature of EnKF updating. If model parameters and state variables were uncorrelated, soft secondary data that represent state variables would not contribute to estimates of model parameters. The next step is to add the mean of predicted state variables to every SGS realization to get corresponding pseudo-realizations of the state variable. Finally N_e pseudo-realizations of state variables are obtained and the entire reservoir model is defined at time *t*. Time is sacrificed against estimation quality, which unfortunately drops in comparison to results from the conventional EnKF procedure. In the proposed shortcut, realizations of state variables are generated only by means of geostatistical tools in order to supply a sufficient number of degrees of freedom and, thus, should not be examined individually for estimating spatial distributions of state variables. Only the mean of predicted state variables has a meaningful interpretation. This shortcut is an adaptation of geostatistical co-simulation for the EnKF (Deutsch, 2002). Note that even though the number of flow

simulation runs is reduced to one, geostatistical simulation should be performed instead to create an ensemble of realizations. Therefore, this concept should be applied only when the computational cost of SGS is lower than the cost of running flow simulations.



Entire model $\boldsymbol{U}^f_t$ is defined at time step *t*. Note that $\boldsymbol{M}^f_t = \boldsymbol{M}^a_{t-1}$.

Figure 26: Conventional forecast step of EnKF



Entire model $\boldsymbol{S}^f_t$ is defined at time step *t*. Note that $\boldsymbol{M}^f_t = \boldsymbol{M}^a_{t-1}$.

Figure 27: Proposed forecast step of the EnKF that is based on the ensemble mean and co-simulation of model states conditional to model parameters

A comparison of the EnKF's conventional implementation with a proposed algorithmic shortcut is demonstrated with a 2D SAGD example discussed earlier. In this case study, the porosity field is estimated using porosity data from a single well, and temperature observations from two surveillance wells and one SAGD pair coming from one time step. The ensemble size consists of one thousand realizations. The base case of porosity is shown in Figure 28, which comprises two distinct low and high value zones.

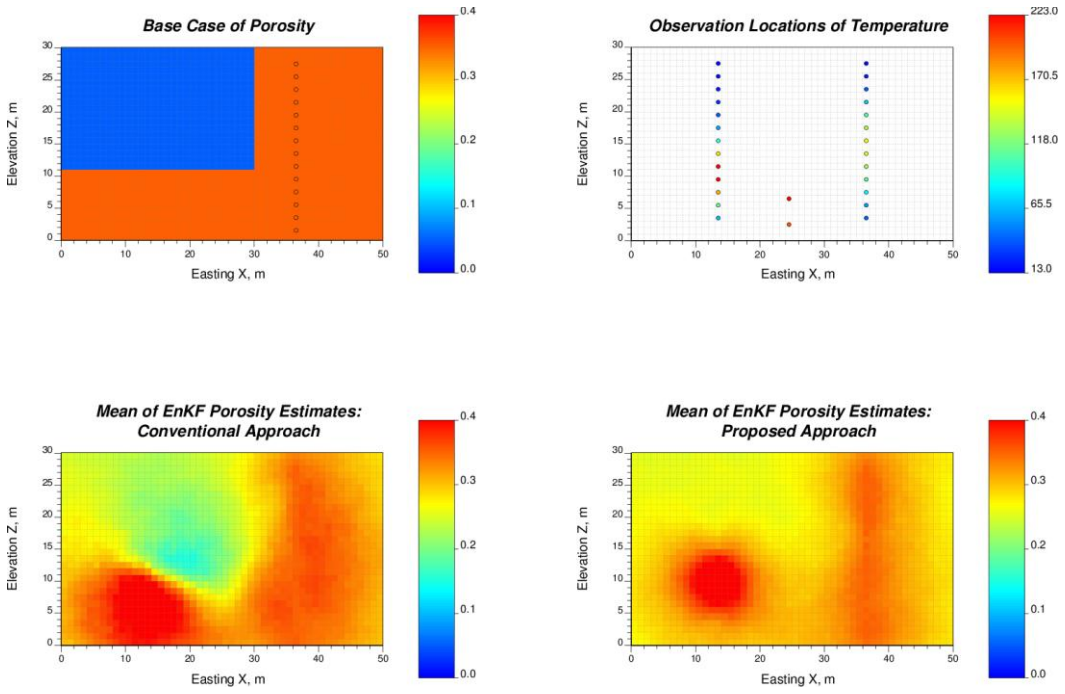


Figure 28: Porosity base case with observation locations, temperature observation locations and mean of realizations of EnKF porosity estimates derived from the conventional approach and the proposed shortcut-based approach respectively

Figure 28 contains locations of porosity and temperature observations. Porosity data are sampled from a single vertical surveillance well, while temperature gauges are allocated along two vertical wells and in the SAGD well pair. The figure also presents the EnKF porosity best estimates obtained from conventional and shortcut-based approaches. Corresponding *RMSE* values of initial ensemble, conventionally updated ensemble with the EnKF, and ensemble updated with the proposed shortcut are 0.189, 0.140, and 0.161. Also, visual estimation results and *RMSE* values indicate that conventional EnKF

produces better porosity estimates than the proposed shortcut (Figure 28). The low value zone is slightly highlighted in the shortcut-based estimate. However, porosity estimate, which is derived using the shortcut, is not as good as porosity estimate, which is derived using the conventional approach. In all cases, the data are matched fairly well. Even though the shortcut does not have the same estimation quality as the conventional approach, it can be still applied to estimate the distribution of static variables when computational overburden is an issue.

5.2.5 Measurement Error

It is important to properly account for measurement error, along with artificially introduced perturbation noise, which is necessary for avoiding ensemble collapse and artificial construction of the measurement error's covariance matrix $\boldsymbol{\Sigma}_t$. Incorporating measurement error has a more significant effect on estimation variance, rather than on the best estimate of realizations (Deutsch et al, 2010). In a paper, Zagayevskiy et al. (2010b) examined measurement error properties.

5.2.6 Grid and Support Effects

This thesis does not examine the grid effect, but the grid effect definitely affects estimation results (Nguyen et al, 2001). All data should be brought to the same support level in order to avoid bias in estimates. A smaller block size should be selected for thermal flow simulation in comparison with conventional flow simulation to obtain results with high accuracy (Lake, 1989).

5.3 Discussion

This chapter presents a 2D synthetic SAGD case study that successfully assimilates hard data (porosity and permeability) and continuous soft data (temperature and difference in acoustic impedances) into petroleum reservoir models through the EnKF in order to constraint spatial distributions of porosity and permeability fields. Any additional data somehow related to model parameters improves estimation of static porosity and

permeability fields. Exhaustively sampled seismic data bring more information to the model than temperature data from surveillance wells and the SAGD well pair. Both temperature and seismic soft data enhance the understanding of porosity and permeability, but their sampling cost should be taken into account. Additional hard data may be used instead, which are direct measures of estimated variables, and thus contain vital information about reservoir geology. To avoid ensemble collapse when exhaustive seismic data were assimilated, and to improve estimation accuracy, a large number of realizations was used and localization of updating and covariance matrices was carried out. The chapter also discussed other implementation details and shortcuts to save computational time, which is based on the combined effect of propagating the mean of state variables at the forecast step, and introducing variations through the Sequential Gaussian Simulation conditional to model parameters. It has been shown that with modifications, the EnKF has significant potential for estimating model parameters of large nonlinear non-Gaussian systems.

Chapter 6 – Realistic 3D SAGD Case Study

Chapter 6 presents an example of EnKF being applied to a realistic case study adapted from the Tucker thermal project operated by Husky Energy (Husky Energy, 2010). Porosity and permeability static variables were constrained to available core data and temperature observations to estimate reservoir heterogeneity and better predict steam chamber growth with an uncertainty assessment.

6.1 Overview of Oil Field

The Tucker reservoir is located in northern Alberta, Canada. The closest settlement is Lloydminster, 233 km east of Edmonton on the Saskatchewan border (Figure 29). The same figure shows the lease area, which is not developed entirely. The red lines form a border around the northwestern section that indicates developing area. To date, horizontal well pairs have been drilled from three pads (A, B, and C) with a varying number of wells: 8, 12, and 20 wells respectively (Figure 30). With daily production of 30,000 BOPD, 350 million bbl. of oil may be recovered from the field. The estimated ultimate recovery factor is about 55%. Oil is represented by bitumen with high viscosity and a density of 9-10^0 API. For this reason, 95% quality steam at 280 ^{0}C is injected into the reservoir in SAGD fashion. Production began in 2006 after a three-month preheating procedure.

The reservoir is deposited in the Clearwater formation of the Manville Group, approximately 500.0 m below the surface. Figure 31 shows the reservoir's stratigraphy. It is argued that the reservoir's geology is formed by three stacked, incised valleys created in a deltaic environment and partially in a marine environment. The average thickness of net pay is 45.0 m. The average porosity, horizontal permeability, and oil saturations are 0.31, 3,000 mD, and 0.56 respectively. Figure 32 shows the bivariate relationship between porosity and horizontal permeability as a scatter plot. The cutoff value for permeability has been implemented at 10,000.0 mD. The relationship between

these two variables is slightly different for different depositional environments, but the correlation is generally good.

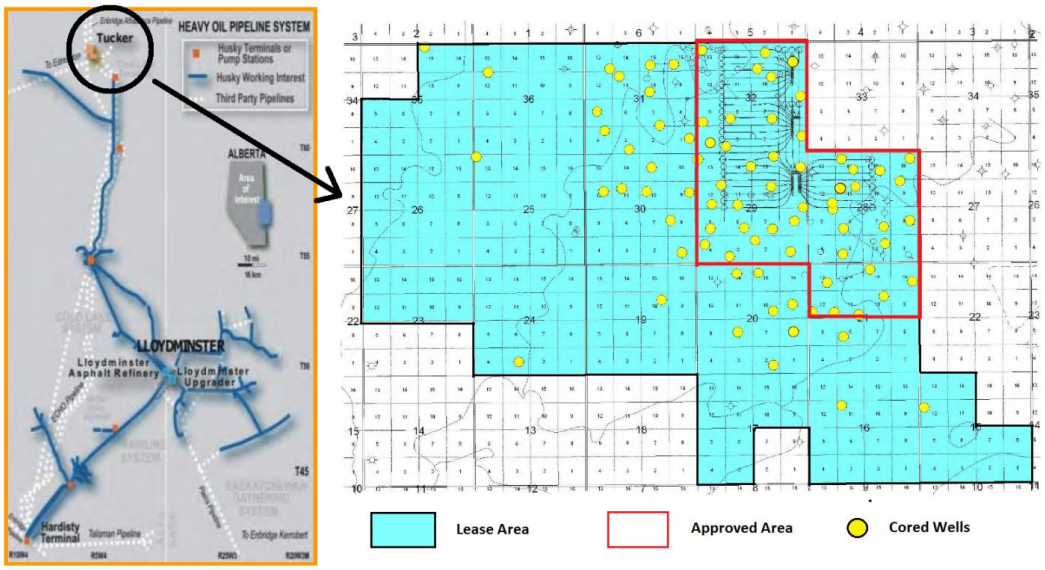


Figure 29: Location map and lease area of the Tucker Thermal Project (Husky Energy, 2010)

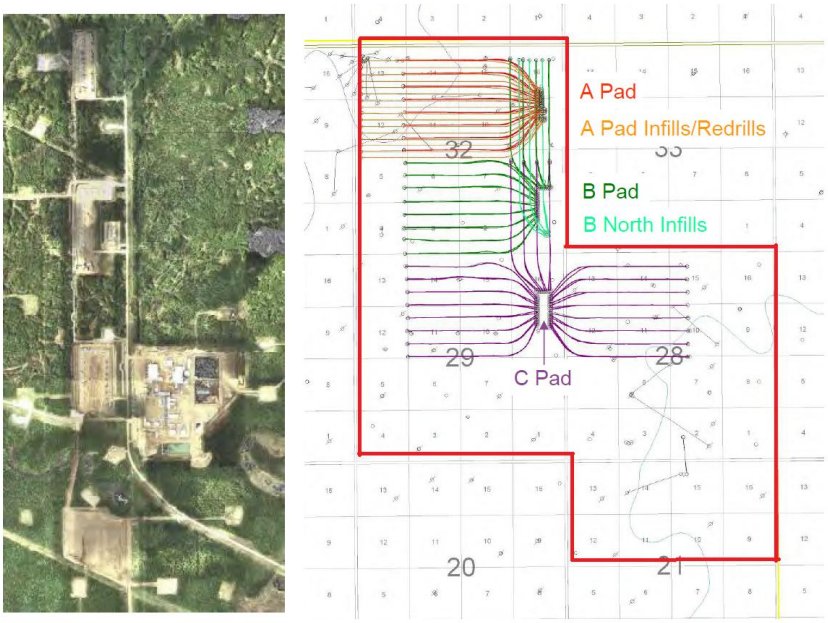


Figure 30: Aerial photo and approved production area of the Tucker Thermal Project with pads (Husky Energy, 2010)

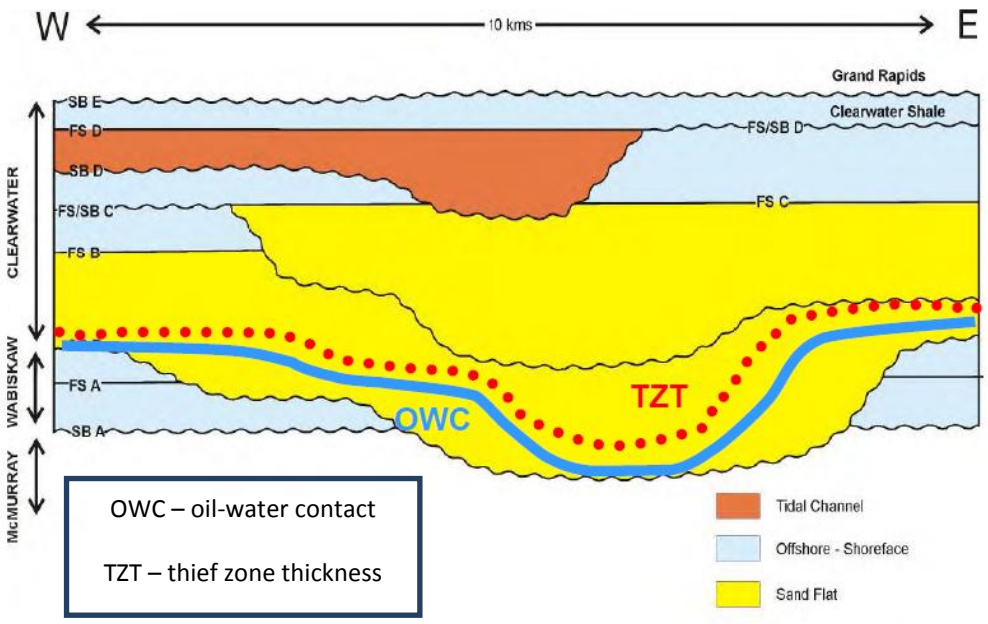


Figure 31: Stratigraphy of the Tucker reservoir (Husky Energy, 2010)

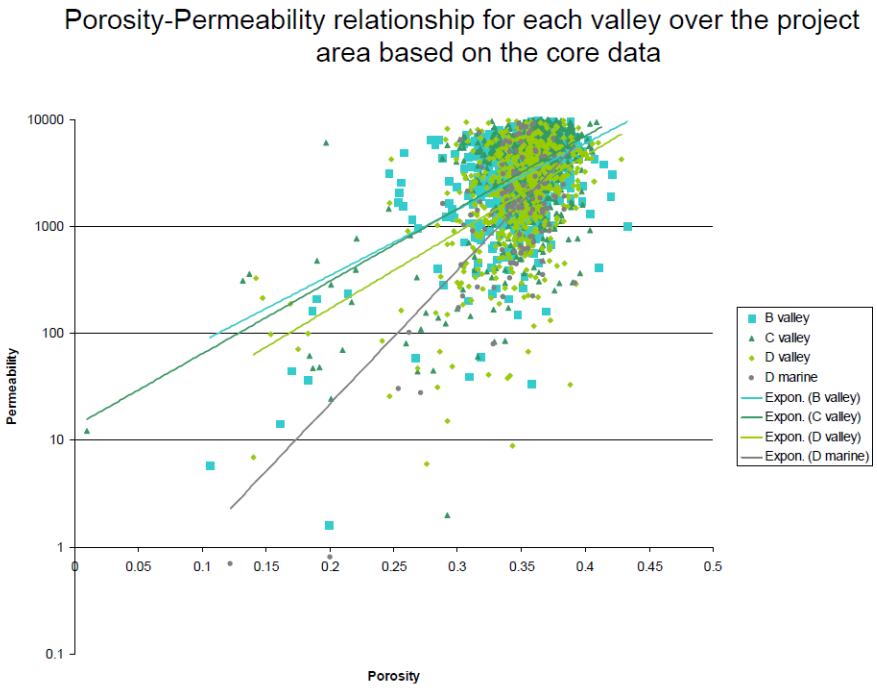


Figure 32: Porosity – horizontal permeability relationship from core analysis at the Tucker reservoir (Husky Energy, 2010)

6.2 3D SAGD Example

This realistic 3D SAGD case study estimates spatial distributions of porosity and permeability. The thesis's modeling region was selected to enclose SAGD well pairs. The surveillance wells around the well pairs indicate significant variations of temperature with time. An analysis of temperature data from the observation wells shows a temperature change over time at only several wells, denoted by a check mark (Figure 33). Because the temperature change is greatest at the two well pairs at the C pad with three observation wells around them, these wells have been chosen for the thermally operated reservoir's modeling region. Figure 33 shows a schematic of the selected region. Figure 34 shows variations of observed real temperature in time for selected surveillance locations. The temperature changes only in surveillance wells #2 and #3, not in #1. Based on this information, the 3D petroleum reservoir model's base case is built to mimic the spatial distribution of temperature observations in time. Figure 35 shows the model grid. The grid configurations are adapted from the RMR on the Tucker thermal project (Husky Energy, 2010). The facies model was built first and consists of permeable sand and impermeable shale sequences, which are typical for northern Alberta reservoir formations. Porosity and permeability models are built for the sand and shale sequence separately. Later they are merged together according to a previously constructed facies model (Deutsch, 2002). The final base cases of porosity and permeability are in compliance with the real data. Note that the relationships between components of the permeability tensor are the same as in the 2D SAGD case study from Chapter 5 (Equation (127)). Porosity and permeability measurements and temperature observations after 540.0 days (18.0 months) of bitumen extraction are sampled at surveillance well locations to estimate porosity and permeability distributions. Figure 36 and Figure 37 show base case and data histograms, respectively. Note that the base case porosity and permeability scatter plot should reproduce the observed true relationship between the same variables (compare Figure 32 and Figure 36). Figure 38 shows temperature observations along well bores of surveillance wells for a single time step (540.0 days). Note that even though synthetic observations reproduce a real spatial temperature distribution, they are smaller than actual observed temperature observations (compare Figure 34 and Figure 38).

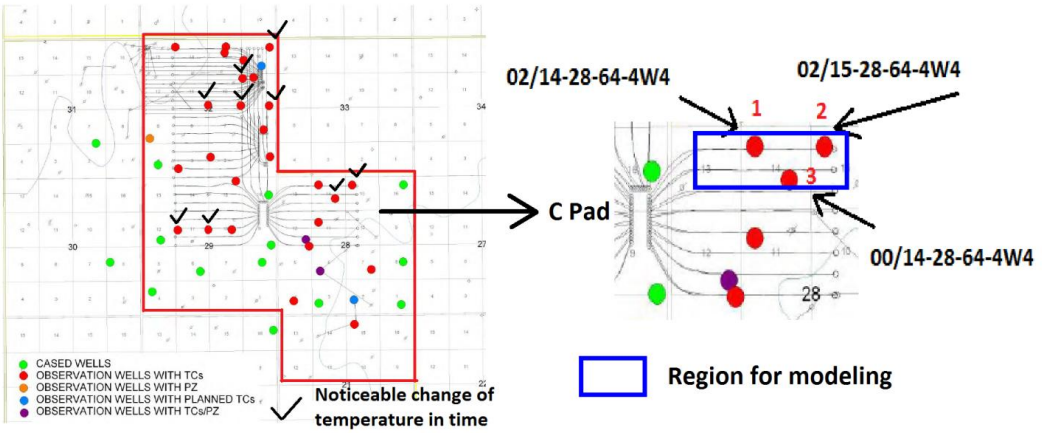


Figure 33: Locations of observation wells and modeling region (Husky Energy, 2010)

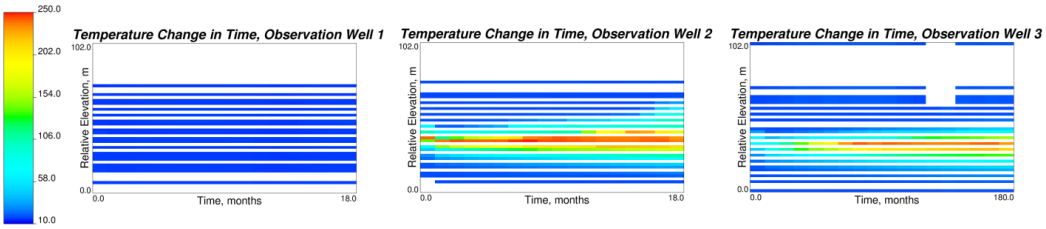


Figure 34: Observed change of temperature along observation wells 1 (02/14-28-64-4W4), 2 (02/15-28-64-4W4), and 3 (00/14-28-64-4W4) respectively (Husky Energy, 2010)

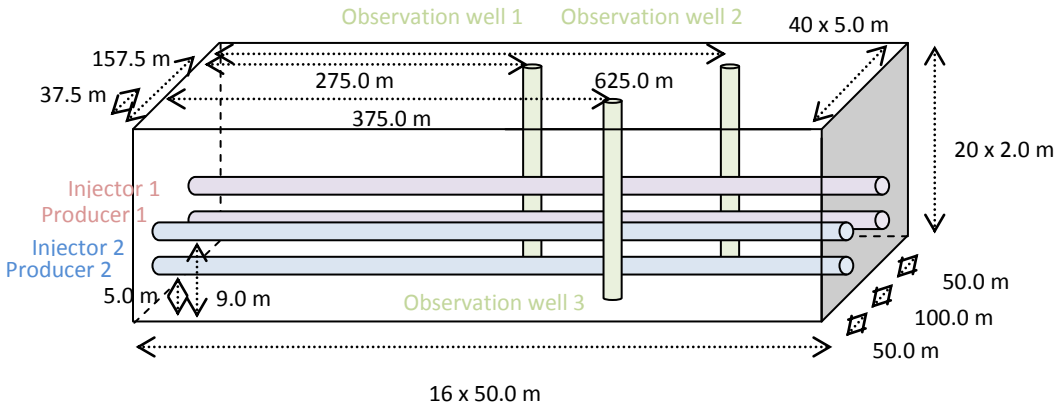


Figure 35: Schematic model grid with model configurations of the SAGD petroleum reservoir model and location of SAGD well pairs 1, 2 and observation wells 1, 2, and 3

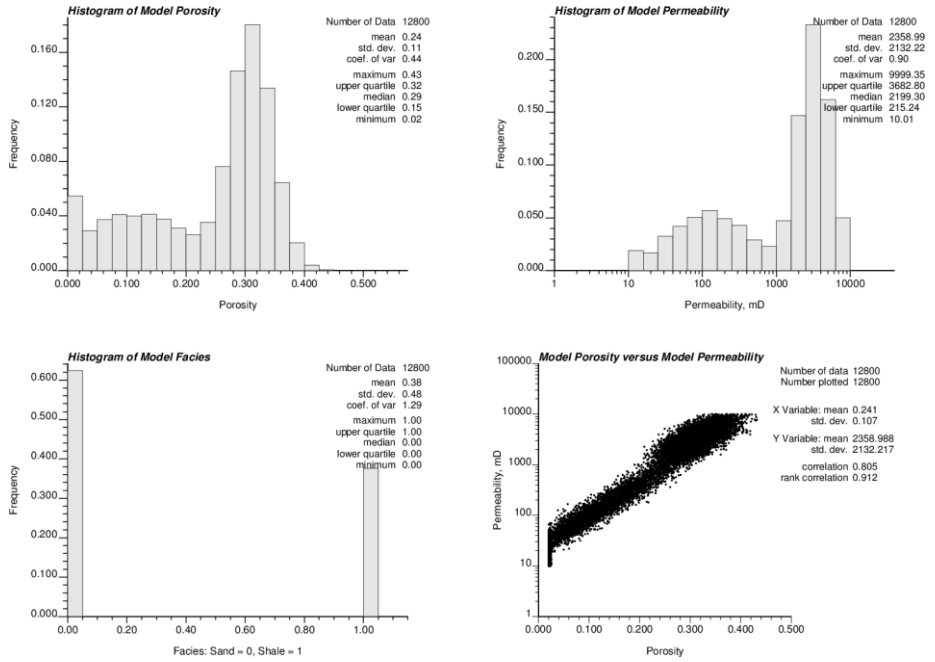


Figure 36: Histograms of realistically simulated base case porosity, horizontal permeability, and facies and scatter plot between porosity and horizontal permeability

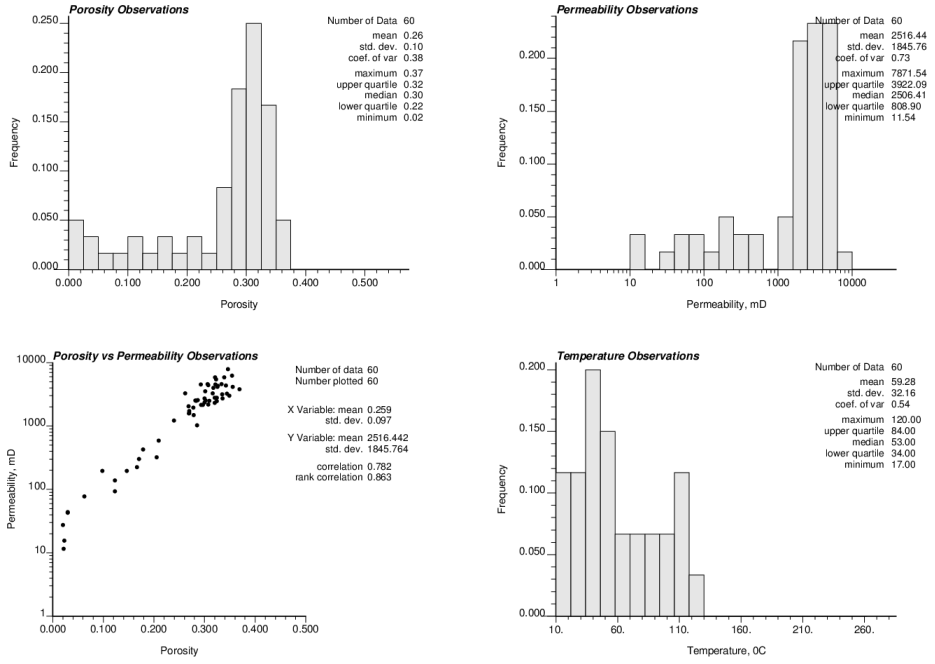


Figure 37: Histograms of porosity and horizontal permeability data, reservoir temperature observations and scatter plot between porosity and horizontal permeability data

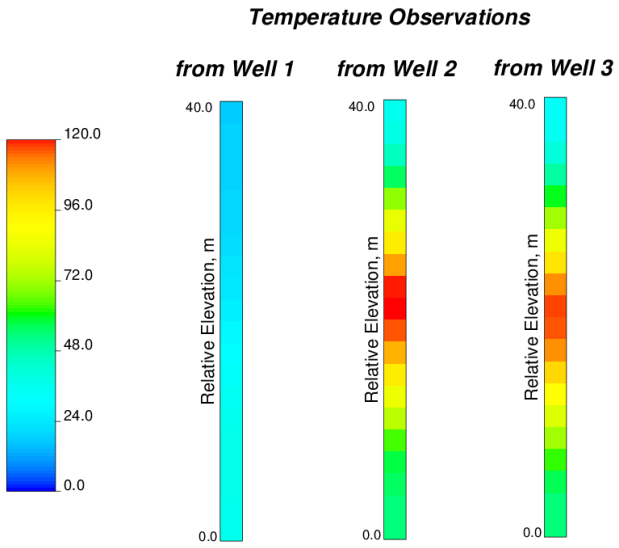


Figure 38: Temperature observations from surveillance wells used in modeling porosity and permeability

Eight different case studies are worked out in order to investigate the influence of data incorporation of different types on estimation quality of porosity and permeability fields. Examined cases are:

1. No data are used in the estimation. The initial ensemble, which is generated based on the best knowledge of the porosity and permeability variables' spatial continuity, represents the best estimate.
2. Porosity and permeability data from surveillance wells are integrated into the model simultaneously.
3. Permeability data from surveillance wells are integrated into the model.
4. Porosity data from the same surveillance wells are assimilated into the model.
5. Temperature data measured after 540.0 days of production are assimilated.
6. Permeability and temperature data measured after 540.0 days of production are used to estimate porosity and permeability distributions.
7. Porosity and temperature data measured after 540.0 days of production are used to estimate porosity and permeability distributions.
8. Porosity, permeability, and temperature data measured after 540.0 days of production are used to estimate porosity and permeability distributions.

Note that temperature data are integrated into the reservoir using the proposed shortcut. One thousand realizations are used for the ensemble size. No localization of updating or covariance matrices is applied. The initial porosity ensemble is generated using a standardized semivariogram model shown in Equation (132), which describes the data's approximate spatial distribution. The standardized initial porosity distribution's variance is adjusted to the data's variance.

$$\gamma(h) = 0.05 + 0.95 \cdot Sph_{\substack{a_1=800.0m \\ a_2=800.0m \\ a_3=10.0m}}(h) \tag{132}$$

$$Sph_a(h) = \begin{cases} 1.5 \cdot \frac{h}{a} - 0.5 \cdot \left(\frac{h}{a}\right)^3, & if \quad h \le a \\ 1.0, & if \quad h \le a \end{cases} \tag{133}$$

Figure 39 shows the EnKF horizontal permeability estimates' mean for all eight cases for slice eight of the model. Figure 40 presents corresponding histograms. Other permeability components can be found from Equation (127). The estimated porosity distribution closely follows the estimated horizontal permeability distribution (see high correlation coefficient between porosity and permeability base cases in Figure 36) and, thus, the porosity estimates are not shown. The maps of EnKF mean estimates in Figure 39 show that the low and high value zones of horizontal permeability are captured fairly well for slice eight. Incorporating any available data improves the estimate's distribution for the high value zone, but not for the low value zone. This is because there are more high value samples, which can be seen on histograms and the scatter plot of porosity and permeability core data (Figure 37). Integrating soft temperature observations alone improves permeability estimates, but not as well as assimilation of core permeability data. Permeability data with temperature observations together produce the best estimate among all cases. Here, porosity data are redundant, but when they are integrated into the model alone, they produce an estimate, which is better than the estimation derived from temperature observations. Note that all assimilated data are history-matched with corresponding estimates.

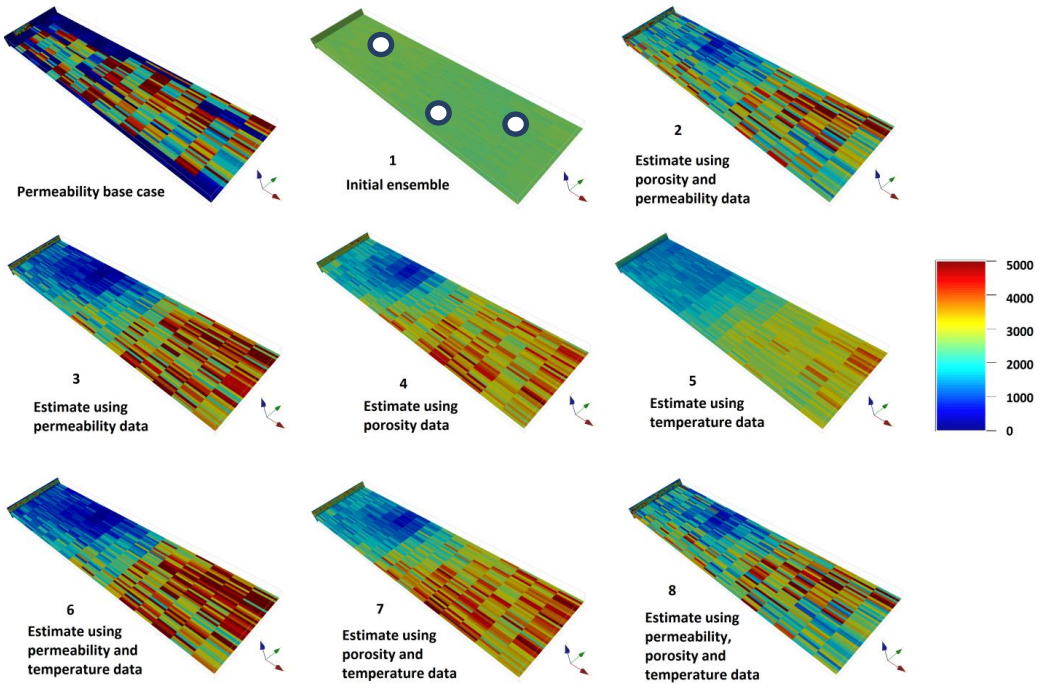


Figure 39: Horizontal permeability base case and mean of realizations of the EnKF horizontal permeability estimates derived using no data (initial ensemble), permeability and porosity data, permeability data, porosity data, temperature data, permeability and temperature data, porosity and temperature data, and all data assimilated at once. Slice 8 of the K_{xx} permeability model is shown. White dots are observation wells.

The estimation results look slightly different if histograms of the means of realizations (best estimates) are compared (see Figure 40). The homogenous initial ensemble is slightly improved by temperature observations. Incorporating porosity data captures the base case horizontal permeability distribution's high value tail. Permeability data lead to the best estimate and may slightly predict the low value tail. Incorporating additional temperature observations to permeability or porosity data does not significantly change the horizontal permeability's estimated distribution. Using all three data types simultaneously does not produce the best estimate, because of data redundancy. Recall that the shortcut for temperature data assimilation was implemented. A conventional EnKF approach might have produced better results, but computational time would increase dramatically.

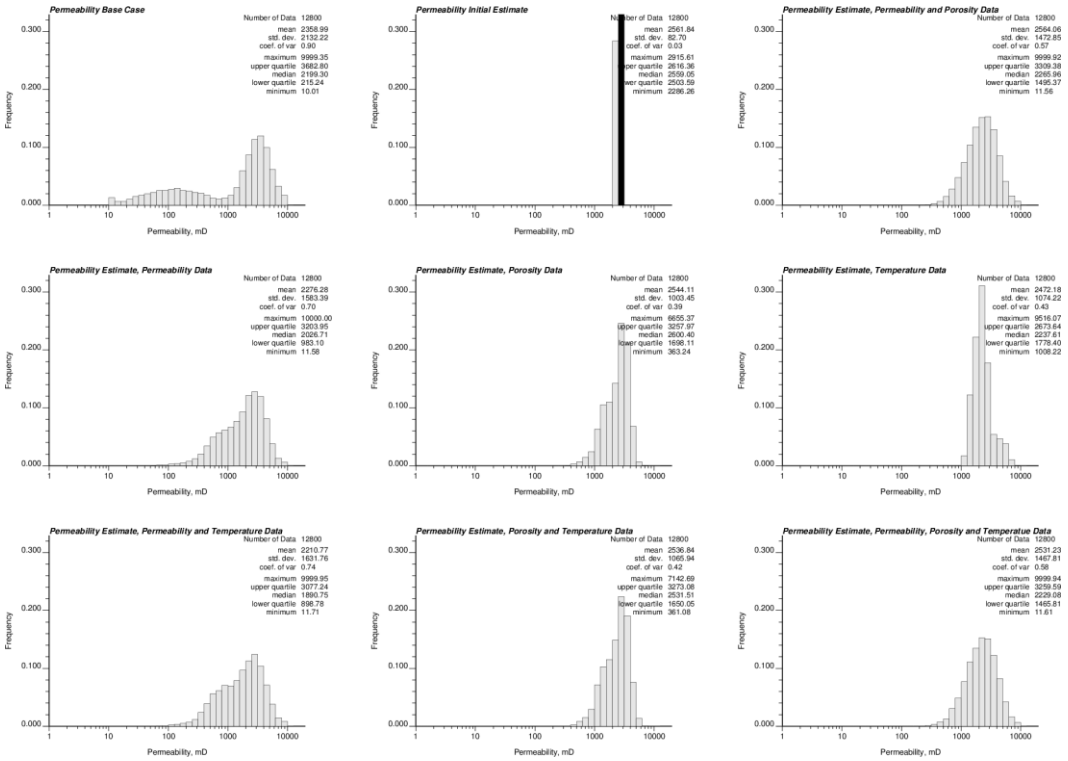


Figure 40: Histograms of horizontal permeability base case and mean of realizations of the EnKF horizontal permeability estimates derived using no data (initial ensemble), permeability and porosity data, permeability data, porosity data, temperature data, permeability and temperature data, porosity and temperature data and all data assimilated at once.

The estimation results from the cases are also compared through *RMSE*, *CC,* and *MSD* validation measures (Equations (120) – (122)). These measures show slightly different findings, since all slices and realizations are studied in comparison to a single slice eight and mean of the EnKF horizontal permeability estimates discussed in the previous paragraph. Comparison results are tabulated in Table 8, which shows that on average, more assimilated data lead to better estimates and a faster convergence of realizations. However, this is not always the case due to the initial ensemble's subjectivity, the representative level of available data, and the locations of observation wells. Redundant data such as porosity and permeability should not be integrated into the model together, since redundancy leads to worse estimates. Here, integrating permeability data along with porosity data worsens estimates of horizontal permeability and porosity. Incorporating temperature data into the reservoir model along with porosity or

permeability data sampled from the same wells improves these model parameters' distribution. Using all data simultaneously produces the best estimates over examined cases, if *RMSE* or *MSD* are compared for porosity estimates. Porosity and temperature data assimilated together lead to the best porosity and permeability estimates, if only the correlation coefficient *CC* is taken into account for porosity estimates and all validation measures are considered for permeability estimates. The discrepancy is caused by redundant porosity and permeability data, and approximations in the porosity-permeability transform (Equation (117)). Also, it is shown in a practical manner that kriging and the EnKF are mathematically identical (see correlation for kriging and the EnKF when only porosity data are assimilated) (Zagayevskiy et al., 2010a). The overall application of the EnKF is found to be effective for estimating porosity and permeability spatial distributions.

Table 8: Comparison of the EnKF porosity and permeability estimates derived using different data sets – 3D SAGD case study

Data Availability	Quality of Porosity Estimates			Quality of Permeability Estimates		
	RMSE	*CC*	*MSD*	*RMSE*	*CC*	*MSD*
No data	0.1275	-0.083	0.0904	2447.8	-0.189	2136
Porosity data	0.1111	0.318	0.0674	2164.1	0.247	1743.5
Kriging with porosity data	-	0.318	-	-	-	-
Permeability data	0.1351	0.31	0.0693	2263.4	0.231	1860.4
Temperature data	0.1224	0.196	0.0796	2382.8	0.173	1992.2
Porosity and permeability data	0.1115	0.277	0.0615	2194.7	0.174	1817.6
Porosity and temperature data	0.1104	0.325	0.0616	2146.2	0.254	1686.3
Permeability and temperature data	0.1346	0.321	0.0666	2245.8	0.243	1771.9
Porosity, permeability and temperature data	0.1096	0.286	0.0596	2155.7	0.185	1723.1

6.3 Discussion

This chapter has presented the influence of the assimilated data's type and number on the quality of porosity and permeability estimates for the realistic case study based on the Tucker thermal project data. In general, extra assimilated hard or soft data have improved estimation results. But surprisingly, the opposite of this conclusion has been also found, when an increase in the number of assimilated data has worsened the estimation results. This occurs because of the subjectivity of the initial ensemble and validation measures (*RMSE*, *MSD*, and *CC*), the representation of available data, sampled locations, and data redundancy. To sum up, the following conclusions are made regarding applying the EnKF to the characterization of the realistic thermally operated SAGD petroleum reservoir:

- The initial ensemble, choice of data type, and observation locations are crucial for estimating model parameters and predicting state variables.
- Assimilating any relevant data improves the estimation accuracy of porosity and permeability distribution in comparison to the initial ensemble, and captures low and high value zones.
- Assimilating hard porosity and permeability data improves the estimation of model parameters better than assimilating soft temperature observations.
- Additional assimilation of soft temperature observations improves the estimation of model parameters.
- Redundant data should be removed. In this case porosity and permeability are redundant. This is one of the reasons porosity and temperature data produce the best estimates of permeability and constrain the permeability distribution shape fairly well in comparison to permeability estimates derived from assimilating all data at once. However, note that incorporating permeability data brings proper extreme values to the distribution as shown in the histograms (Figure 40).
- The relationship between all variable types, such as parameters of porosity-horizontal permeability transform, should be properly preserved to obtain better estimates with less recursive steps.

Chapter 7 – Conclusion and Future Work

This is last chapter. Conclusions are made regarding integrating core porosity and permeability data, and continuous temperature observations and seismic attributes into SAGD-operated petroleum reservoir models in order to constrain porosity and permeability spatial distributions using the EnKF algorithm. The chapter also discusses possible directions for future work.

This thesis develops a methodology based on the inverse modeling technique EnKF for continuous integration of core porosity and permeability measurements, soft temperature observations, and differences in acoustic impedances into thermally operated petroleum reservoir models for reservoir characterization. A proposed algorithm is presented and thoroughly discussed. The EnKF characteristics are examined, and implementation details are summarized through a series of sensitivity studies. It has been shown on synthetic 2D and realistic 3D SAGD case studies that EnKF is good for estimating model parameters (static variables such as porosity and permeability) and predicting a dynamic model's states (dynamic variables such as temperature and acoustic impedance). Uncertainty associated with estimates can be easily reported because of the EnKF's ensemble nature. The proposed approach for continuous data assimilation works quite well for models that follow the Gaussian distribution. If a non-Gaussian highly nonlinear system is encountered, a confirmation step can be added (Wen and Chen, 2006) or EnKF can be replaced with the EnKS (Evensen and van Leeuwen, 2000). Localizing updating and covariance matrices is another option to extend EnKF's applicability for characterizing non-Gaussian and nonlinear dynamic systems. Since an estimate's quality depends on the initial ensemble, ensemble size, number of data integrated, and available data quality, setting up the EnKF is tedious. The EnKF estimates are very sensitive to the initial (first-guess) ensemble, which should follow a normal distribution and have a mean close to the data average. A variable's spatial distribution should be depicted in the initial ensemble as well. The ensemble size should be at least ten times larger than the number of integrated data (Fahimuddin, 2010). Redundant data should be removed and proper observation locations should be selected. In other words, the data should bear spatial information about a reservoir, and

continuous data should show dynamics in subsequent measurements. In general, incorporating additional hard and soft data into reservoir improves model quality. The method might be computationally expensive, if a conventional algorithm is applied, but at the same time significant estimation accuracy is achieved. In cases in which localization of updating or covariance matrices is applied at the analysis step, the number of integrated data can be fairly large like exhaustive seismic attributes. If the shortcut based on the ensemble mean and co-simulation is implemented, it is possible to omit the computational overburden at the EnKF forecast step. In the conventional EnKF setting, the flow simulation should be performed as many times as the number of realizations. The shortcut requires only one run of the flow simulator, which saves computational time. But estimation accuracy drops unavoidably as well. A modeler should be able to find a balance between quality and cost. Parallel programming is another option to decrease computational time. Overall, the EnKF has proven itself to be a relatively simple but effective data assimilation technique to integrate different data types including core measurements and continuous temperature and exhaustive seismic observations from thermally operated SAGD fields. Other data types such as bottomhole pressure, cumulative oil rate, and log data can be also integrated through the EnKF into a reservoir model. The literature includes many discussions of this type of integration (Gao et al., 2006; Evensen et al., 2007; Aanonsen et, 2009; Huseby et al., 2009; Seiler et al., 2009; Fahimuddin, 2010, etc). The EnKF is deemed to be a reliable technique for uncertainty quantification.

As for future work, the EnKF can be applied to characterize categorical variables, a good example of which is facies modeling. The localization of the covariance matrix may be studied further in order to find an optimal number of realizations and keep estimation accuracy unchanged or even improve it. EnKF modifications, such as PF and RML stochastic estimation methods, can be examined in more detail as well. The EnKF should be applied to more complex case studies in order to assess its commercial value in numerical modeling of real petroleum reservoirs.

Bibliography

Aanonsen, S.I., Naevdal, G., Oliver, D.S., Reynolds, A.C., and Valles, B. (2009). The Ensemble Kalman Filter in Reservoir Engineering - a Review. *SPE Journal*, 393 – 412

Anderson, J.I. (2001). An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, 129(12), 2884 – 2903

Arroyo-Negrete, E., Devegowda, D., Datta-Gupta, A. and Choe, J. (2008). Streamline-assisted Ensemble Kalman Filter for Rapid and Continuous Reservoir Model Updating. *SPE Reservoir Evaluation & Engineering*, 11(6), 1046 – 1060

Aziz, K. and Settari, A. (1979). *Petroleum Reservoir Simulation.* Applied Science Publishers Ltd, London, 476

Batzle, M. and Wang, Z. (1992). Seismic properties of pore fluids, *Geophysics*, 64, 1396 – 1408

Bennett, A.F. (1992). *Inverse methods in physical oceanography.* New York: Cambridge University Press, 346

Bianco, A., Cominelli, A. , Dovera, L., Nævdal, G. and Valles, B. (2007). History Matching and Production Forecast Uncertainty by Means of the Ensemble Kalman Filter: A Real Field Application. *SPE Europec/EAGE Annual Conference and Exhibition,* London, UK

Brower, D.R., Nævdal, G., Jansen, J.-D., Vefring, E.H., and van Kruijsdijk, C.P. (2004). Improved Reservoir Management Through Optimal Control and Continuous Model Updating. *SPE Annual Technical Conference and Exhibition,* Houston

Burgers, G., van Leeuwen, P. J. and Evensen, G. (1998). Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review,* 126(6), 1719 – 1724

Butler, R. (1991). *Thermal Recovery of Oil and Bitumen*. Prentice Hall, Englewood Cliffs, New Jersey, 524

Butler, R. (1999). The Steam and Gas Push (SAGP). *Journal of Canadian Petroleum Technology*, 38(3), 54 – 61

Butler, R.M., Jiang, Q., and Yee, C.T. (2000). The Steam and Gas Push (SAGP) – 3: Recent Theoretical Developments Laboratory Results. *Journal of Canadian Petroleum Technology*, 39(8), 51 – 60

Carlson, M.R. (2003). *Practical reservoir simulation: using, assessing, and developing results*. PennWell Corporation, Tulsa, Oklahoma, 575

Chen, Y. and Oliver, D.S. (2010). Ensemble-based Closed-loop Optimization Applied to Brugge Field. *SPE Reservoir Evaluation & Engineering*, 13(1), 56 – 71

Chen, Y. and Zhang, D. (2005). Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Advances in Water Resources,* 29, 1107 – 1122

Chen, Y. Oliver, D.S. and Zhang, D. (2008). Efficient Ensemble-Based Closed-Loop Production Optimization. *SPE/DOE Symposium on Improved Oil Recovery,* Tulsa

Chen, Z. (2007). *Reservoir Simulation: Mathematical Techniques in Oil Recovery*. SIAM, Philadelphia, 248

Chitralekha, S. (2010). *Seminars on State and Parameter Estimation*, the University of Alberta, Edmonton

Christakos, G. (2005). Methodological developments in geophysical assimilation modeling. *Reviews of Geophysics,* 43

Cohn, S.E. (1997). An Introduction to Estimation Theory. *Journal of the Meteorological Society of Japan* 75(1B), 257 – 288

Cominelli, A., Dovera, L., Vimercati, S. and Naevdal, G. (2009). Benchmark Study of Ensemble Kalman Filter Methodology: History Matching and Uncertainty Quantification for a Deep-water Oil Reservoir. *International Petroleum Technology Conference 2009,* Doha, Qatar

Cooper, M.E. and Thorogood, A. (1999). Foinhaven active reservoir management: The time lapse signal. *SEG Extended Abstracts*

Corser, G.P., Harmse, J.E., Corser, B.A., Weiss, M.W. and Whitflow, G.L. (2000). Field test results for a real-time intelligent drilling monitor. *IADC/SPE Drilling Conference,* New Orleans, Louisiana

Dake, L.P. (1998). *Fundamentals of Reservoir Engineering*, 17th Edition, Elsevier, The Netherlands, 498

Daley, R. (1991). *Atmospheric data analysis.* New York: Cambridge University Press, 472

Deutsch, C.V. (2002). *Geostatistical Reservoir Modeling*. Oxford University Press, New York, 376

Deutsch, C.V. and Journel, A.G. (1998). *GSLIB: Geostatistical Software Library and User's Guide*, 2nd Edition. Oxford University Press, New York, 369

Deutsch, J.L., Boisvert, J.B. and Deutsch, C.V. (2010). A New Dimension to Account for Data Error and Volume Support, *Centre for Computational Geostatistics*, 12, 308-1 – 308-16

Devore, J. and Peck, R. (2005). *Statistics. The Exploration and Analysis of Data*, 5th Edition, Thomson, Brooks/Cole, Canada, 763

Dong, Y., Gu, Y. and Oliver, D.S. (2006). Sequential Assimilation of 4D Seismic Data for Reservoir Description Using the Ensemble Kalman Filter. *Journal of Petroleum Science and Engineering,* 53, 83 – 99

Doucet, A., de Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice.* New York: Statistics for Engineering and Infromation Science, Springer, 553

Duru, O.O. and Horne, R.N. (2009). Simultaneous Interpretation of Pressure, Temperature and Flowrate Data for Improved Model Identification and Reservoir Parameter Estimation. *2009 SPE Annual Technical Conference and Exhibition*, New Orleans, 1 – 19

Duru, O.O. and Horne, R.N. (2010). Joint Inversion of Temperature and Pressure Measurements for Estimation of Permeability and Porosity Fields. *SPE Annual Technical Conference and Exhibition*, Italy, 1 – 11

Eastwood, J., Lebel, P., Dilay, A. and Blakeslee, S. (1994). Seismic Monitoring of Steam-based Recovery of Bitumen. *The Leading Edge*, 242 – 251

Ehrendorfer, M. (2007). A Review of Issues in Ensemble-Based Kalman Filtering. *Meterologische Zeitschrift* 16(6), 795-818

Eigbe, U., Beck, M.B., Wheater, H. S. and Hiran, F. (1998). Kalman Filtering in Groundwater Glow Modelling: Problems and Prospects. *Stochastic Hydrology and Hydraulics,* 12(1), 15 – 32

Eisenmann, P. and Juchereau, B. (1994). Improved Rxo measurements through semi-active focusing. *SPE Annual Technical Conference and Exhibition,* New Orleans, Louisiana

Eppstein, M.J., and Dougherry, D.E. (1996). Simultaneous Estimation of Transmissivity Values and Zonation. *Water Resources Research,* 32(11), 3321 – 3336

Evensen, G. (1994). Sequential Data Assimilation with a Nonlinear Quasi-geostrophic Model Using Monte Carlo Methods for Forecast Error Statistics. *Journal of Geophysical Research*, 99(C5), 143 – 162

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. *Ocean Dynamics,* 53, 343 – 367

Evensen, G. (2007). *Data Assimilation: The Ensemble Kalman Filter.* Springer Verlag, Berlin, 305

Evensen, G. (2009). *Data assimilation: The Ensemble Kalman Filter,* 2nd Edition. Bergen, Springer, 307

Evensen, G. and van Leeuwen, P.J. (2000). An Ensemble Kalman Smoother for Nonlinear Dynamics. *Monthly Weather Review,* 128(6), 1852 – 1867

Evensen, G., Hove, J., Meisingset, H.C., Reiso, E., Seim, K.S. and Espelid, Ø. (2007). Using the EnKF for Assisted History Matching of a North Sea Reservoir Model. *Proceedings of the 2007 SPE Reservoir Simulation Symposium*

Fahimuddin, A. (2010). 4D Seismic History Matching Using the Ensemble Kalman Filter (EnKF): Possibilities and Challenges. *PhD Thesis*, University of Bergen, Norway, 116

Fanchi, J. (2006). *Principles of Applied Reservoir Simulation*, 3rd Edition. Elsevier, Burlington, 528

Gao, G., Zafari, M. and Reynolds, A.C. (2006). Quantifying Uncertainties for the PUNQ-S3 Problem in Bayesian Setting with RML and EnKF. *SPE Journal*, 11(4), 506 – 515

Gassmann, F. (1951). Über die elastizität poröser medien: Vierteljahrsschrift der Naturforschenden Gesellschaft in Zurich, 96, 1-23. The English version of this paper can be accessed at http://sepwww.stanford.edu/sep/berryman/PS/gassmann.pdf

Gomez-Hernandez, J.J., Sahuquillo, A. and Capilla, J.E. (1997). Stochastic Simulation of Transmissivity Fields Conditional to Both Transmissivity and Piezometric Data, 1. The theory. *Journal of Hydrology*, 203, 162 – 174

Gordon, N.J., Salmond, D.J. and Smith, A.F. (1993). Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation. *IEE Proceedings,* 140(2), 107 – 113

Gu, Y. (2006). History Matching Production Data using the Ensemble Kalman Filter. *PhD Thesis*, University of Oklahoma, Norman, Oklahoma, 183

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 – 224

Gu, Y. and Oliver, D.S. (2006). The ensemble Kalman filter for continuous updating of reservoir simulation models. *Journal of Energy Resources Technology*, 128, 79 – 87

Hantush, M.M., and Marin, M.A. (1997). Estimation of Spatial Variable Aquifer Hydraulic Properties Using Kalman Filtering. *Journal of Hydraulic Engineering*, 123(11), 1027 – 1035

Haugen, V., Nævdal, G., Natvik, L.-J., Evensen, G., Berg, A.M. and Flornes, K.M. (2008). History Matching Using the Ensemble Kalman Filter on a North Sea Field Case. *SPE Journal*, 13(4), 382 – 391

Hong, S. and Deutsch, C.V. (2008). Fluid Substitution Model to Generate Synthetic Seismic Attributes: FluidSub.exe. *Centre for Computational Geostatistics*, 10, 204-1 – 204-8

Hong, S., Leuangthong, O. and Deutsch, C.V. (2006). A Geostatistical Approach to Stochastic Seismic Inversion. *Centre for Computational Geostatistics*, 8, 201-1 – 201-22

Hosseini, A.H., Leuangthong, O. and Deutsch, C.V. (2008). An Integrated Approach to Permeability Modeling Using Micro-Models, *International Thermal Operations and Heavy Oil Symposium*, Calgary

Houser, P.R, Shuttleworth, W.J., Famiglietti, J.S., Gupta, H.V., Syed, K.H. and Goodrich, D.C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, 34(12), 3405 – 3420

Houtekamer, P.L. and Mitchell, H.L. (1998). Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*, 126(3), 796 – 811

Houtekamer, P.L. and Mitchell, H.L. (2001). A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, 129(1), 123 – 137

Huang, C., Hu, B.X., Li, X., and Ye, M. (2009). Using Data Assimilation Method to Calibrate a Heterogeneous Conductivity Field and Improve Solute Transport Prediction with an Unknown Contamination Source. *Stochachstic Environmental Resources Risk Assessment*, 23, 1155 – 1167

Huang, C., Li, X. and Lu, L. (2008). Retrieving soil temperature profile by assimilating MODIS LST products with Ensemble Kalman Filter. *Remote Sensing of Environment*, 112(4), 1320 – 1336

Huseby, O., Valestrand, R., Nævdal, G. and Sagen, J. (2009). Natural and conventional tracers for improving reservoir models using the EnKF approach. *EUROPEC/EAGE Conference and Exhibition.* Amsterdam, The Netherlands

Husky Oil Operations Limited. (2010). *Annual Performance Presentation for Tucker Thermal Project (Commercial Scheme #9835).* Energy Resources Conservation Board

(ERCB), Calgary, 153, retrieved from http://www.ercb.ca/portal/server.pt/gateway/PTARGS_0_0_321_0_0_43/http%3B/ercbContent/publishedcontent/publish/ercb_home/industry_zone/industry_activity_and_data/in_situ_progress_reports/2010/

Isaaks, E.H. and Srivastava, R.M. (1989). *Applied Geostatistics.* Oxford University Press, New York, 561

Ito, Y., Ichikawa, M. and Hirata, T. (2000). The Growth of the Steam Chamber during the Early Period of the UTF Phase B and Hangingstone Phase 1 Projects. *Canadian International Petroleum Conference*

Ito, Y., Ichikawa, M., and Hirata, T. (2001). The Effect of Gas Injection on Oil Recovery during SAGD Projects. *Journal of Canadian Petroleum Technology*, 40(1), 20 – 23

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory.* Academic Press, New York, 376

Jiang, Q., Butler, R.M., and Yee, C.T. (2000). The Steam and Gas Push (SAGP) – 2: Mechanism Analysis and Physical Model Testing. *Journal of Canadian Petroleum Technology*, 39(4), 52 – 61

Johnston, P.F., Andersen, G.R., Wachi, N., Lee, D.S., Martens, F.G. and Han, D.H. (1992). Integration of Seismic Monitoring and Reservoir Simulation Results for a Steamflood at South Casper Creek Oil Field, Wyoming. *67th Annual Technical Conference and Exhibition of the SPE*, Washington, DC, 513 – 519

Julier, S., Uhlmann, J. and Durrant-Whyte, H.F. (2000). A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators. *IEEE Transactiions on Automatic Control*, 45(3), 477 – 482

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic* Engineering (Citeseer), 82(1), 35 – 45

Kinsler, L.E., Frey, A.R., Coppens, A.B. and Sanders, J.V. (2000). *Fundamentals of Acoustic*, 4th Edition. John Wiley and Sons, New York, 552

Kitandis, P.K. (1995). Quasi-Linear Geostatistical Theory for Inversing. *Water Resources Research*, 31(10), 2411 – 2419

Kivman, G.A. (2003). Sequential Parameter Estimation for Stochastic Systems. *Nonlinear Processes in Geophysics*, 10, 253 – 259

Krief, M., Garat, J., Stellingwerff, J. and Ventre, J. (1990). A Petrophysical Interpretation Using the Velocities of P and S Waves (Full-Waveform Sonic). *The Log Analyst*, 355 – 369

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

Lake, L.W. (1989). *Enhanced Oil Recovery*. Prentice Hall, Englewood Cliffs, New Jersey, 550

Landrø, M. (2001). Discrimination between Pressure and Fluid Saturation Changes from Time-lapse Seismic Data. *Geophysics,* 66(3), 836 – 844

Lefebvre, T., Bruyninckx, H. and De Schutter, J. (2004). Kalman Filter for Non-Linear Systems: a Comparison of Performance. *International Journal of Control*, 77(7), 639 – 653

Leng, C.H., and Yeh, H.D. (2003). Aquifer Parameter Identification Using the Extended Kalman Filter. *Water Resources Research*, 39(3), 1062 – 1075

Lewis, E.L. (1980). The Practical Salinity Scale 1978 and Its Antecedents. *IEEE Journal of Oceanic Engineering*, 5(1), 3 – 8

Li, R., Reynolds, A.C. and Oliver, D.S. (2001). History Matching of Three-Phase Flow Production Data. *SPE Reservoir Simulation Symposium*, Texas

Li, Z. and Zhu, D. (2010). Predicting Flow Profile of Horizontal Well by Downhole Pressure and Distributed Temperature Data for Waterdrive Reservoir. *SPE Annual Technical Conference and Exhibition*, New Orleans, 296 – 304

Li, Z., Yin, J., Zhu, D. and Datta-Gupta, A. (2010). Using Downhole Temperature Measurements to Assist Reservoir Characterization and Optimization, *SPE International Oil and Gas Conference and Exhibition,* China

Liang, B., Alpak, F.O., Sepehrnoori, K. and Delshad, M. (2007). A Singular Evolutive Interpolated Kalman Filterfor Rapid Uncertainty Quantification. *SPE Reservoir Simulation Symposium,* Houston

Lorentzen, R. J., Berg, A.M., Nævdal, G. and Vefring, E.H. (2006). A New Approach for Dynamic Optimization of Waterflooding Problems. *Intelligent Energy Conference and Exhibition,* Amsterdam

Lorentzen, R. J., Fjelde, K.K., Frøyen, J., Lage, V.M., Nævdal, G. and Vefring, E.H., (2001a). Underbalanced and Low-Head Drilling Operations: Real Time Interpretation of Measured Data and Operational Support. *SPE Annual Technical Conference and Exhibition,* New Orleans

Lorentzen, R. J., Fjelde, K.K., Frøyen, J., Lage, V.M., Nævdal, G. and Vefring, E.H., (2001b). Underbalanced Drilling: Real Time Data Interpretation and Decision Support. *SPE/IADC Drilling Conference. Amsterdam*

Lorentzen, R. J., Nævdal, G. and Lage, V.M. (2003). Tuning of Parameters in a Two-Phase Flow Model Using an Ensemble Kalman Filter. *International Journal of Multiphase Flow*, 29(8), 1283 – 1309

Lumley, D. (1999). Meren field Nigeria: A 4D case study. *SEG Extended Abstracts,* 1628

Lumley, D. (2001), Time-Lapse Seismic Reservoir Monitoring. *Geophysics*, 66(1), 50 – 53

Makhlouf, E.M., Chen, W.H., Wasserman, M.L. and Seinfeld, J.H. (1993). A General History Matching Algorithm for Three-Phase, Three-Dimensional Petroleum Reservoirs. *SPE Advanced Technology Series*, 1(2), 83 – 92

Mandel, J. (2006). Efficient implementation of the ensemble Kalman filter. *University of Colorado at Denver and Health Sciences Center*, Denver, Colorado, 9

Mavko, G., Mukerji, T. and Dvorkin, J. (2009). *The Rock Physics Handbook: Tools for Seismic Analysis in Porous Media*, 2nd Edition. Cambridge University Press, New York, 525

Maybeck, P.S. (1979). *Stochastic Models, Estimation, and Control,* Volume 1. New York, Mathematics in Science and Engineering, Academic Press, 423

Metropolis, N. and Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247), 335 – 341

Myrseth, I. (2007). Ensemble Kalman Filter Adjusted to Time-Lapse Seismic Data. *Petroleum Geostatistics Conference, Portugal,* European Association of Geoscientists and Engineers

Nævdal, G., Brouwer, D.R. and Jansen, J.-D. (2006). Waterflooding Using Closed-Loop Control. *Computational Geosciences*, 10(1), 37 – 60

Nævdal, G., Johnsen, L.M., Aanonsen, S.I. and Vefring, E.H. (2005). Reservoir monitoring and continuous model updating using ensemble Kalman filter. *SPE Journal*, 10(1), 66 – 74

Nævdal, G., Mannseth, T. and Vefring, E.H. (2002a). Instrumented Wells and Near-Well Reservoir Monitoring Through Ensemble Kalman Filter. *8th European Conference on the Mathematics of Oil Recovery. Freiberg*, Germany

Nævdal, G., Mannseth, T. and Vefring, E.H. (2002b). Near-Well Reservoir Monitoring Through Ensemble Kalman Filter. *SPE/DOE Improved Oil Recovery Symposium*, Tulsa, Oklahoma

Nguyen, H, Deutsch, C.V. and Bentsen, R. (2001). Quantification of Grid Discretization Effects in Presence of Realistic Geological Heterogeneities, *Centre for Computational Geostatistics*, 3, 1 – 17

Nur, A., and Simmons, G. (1967). The effect of saturation on velocity in low porosity rocks. *Earth and Planetary Sciences Letters*, 7(2), 183 – 193

Oliver, D.S. (1994). Incorporation of Transient Pressure Data into Reservoir Characterization. *In Situ*, 18(3), 243 – 275

Oliver, D.S., and Chen, Y. (2010). Recent Progress on Reservoir History Matching: a Review. *Computational Geosciences*, 15(1), 185 – 221

Oliver, D.S., He, N. and Reynolds, A.C. (1996). Conditioning Permeability Fields to Pressure Data. *5th European Conference on the Mathematics of Oil Recovery,* Leoben, Austria

Oliver, D.S., Reynolds, A.C., Bi, Z. and Abacioglu, Y. (2001). Integration of Production Data into Reservoir Models. *Petroleum Geoscience*, 7(S), 65 – 73

Overbeek, K.M., Brouwer, D.R., Nævdal, G., Jansen, J.-D. and van Kruijsdijk, C.P. (2004). Use of a Virtual Asset to Demonstrate Closed-Loop Reservoir Management. *9th European Conference on the Mathematics of Oil Recovery,* Cannes, France

Palmgren, C., and Edmunds, N.R. (1995). High Temperature Naphtha (SiC) to Replace Steam in the SAGD Process. *International Heavy Oil Symposium,* Proceedings SPE 30294, Calgary, Alberta

Pham, D.T. (2001). Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems. *Monthly Weather Review*, 129(5), 1194 – 1207

Press, S.J. (2003). *Subjective and Objective Bayesian Statistics: Princeiples, Models, and Applications*, 2nd Edition. Wiley Series in Probabilistic and Stataistics, New Jersey, 591

Reichle, R.H., Walker, J.P., and Koster, R.D. (2002). Extended versus ensemble Kalman filtering for land data assimilation. *Journal of Hydrometeorology*, 3(6), 728 – 740

Remy, N., Boucher, A. and Wu, J. (2009). *Applied Geostatistics with SGeMS: A User's Guide*. Cambridge University Press, New York, 264

Roste, T. and Husby, O. (2006). Using 4D seismic to monitor fluid flow in a heterogeneous and compartmentalized reservoir – cases from the Heidrun field. *EAGE Annual Conference & Exhibition*, Vienna, Austria

Rozier, D., Birol, F., Cosme, E., Brasseur, P., Brankart, J.M. and Verron, J. (2007). A Reduced-Order Kalman Filter for Data Assimilation in Physical Oceanography. *SIAM Review*, 49(3), 449 – 465

Satter, A. and Thakur, G.C. (1994). *Integrated Petroleum Reservoir Managemnet: A Team Approach*. PennWell Books, Oklahoma, 335

Seiler, A., Evensen, G., Skjervheim, J.A., Hove, J. and Vabo, J.G. (2009). Advanced Reservoir Management Workflow using an EnKF based Assisted History Matching Method. *SPE Reservoir Simulation Symposium,* The Woodlands, Texas, USA

Skjervheim, J.-A., Evensen, G., Aanonsen, S.I. and Johansen, T.A. (2007). Incorporating 4D Seismic Data in Reservoir Simulation Model Using Ensemble Kalman Filter. *SPE Journal*, 12(3), 282 – 292

Stengel, R.F. (1994). *Optimal Control and Estimation.* Mineola, New York, Dover Publications, 639

Wan, E.A., and Nelson, A.T. (2001). *The Unscented Kalman Filter: Chapter 7*. Wiley Online Library, retrieved from http://onlinelibrary.wiley.com/doi/10.1002/0471221546.ch7/summary

Wang, C., Li, G. and Reynolds, A.C. (2009). Production Optimization in Closed-Loop Reservoir Management. *SPE Journal*, 14(3), 506 – 523

Wang, Z., and Nur, A. (1988). Effect of temperature on wave velocities in sands and sandstones with heavy hydrocarbons. *SPE Reservoir Engineering Journal*, 3(1), 158 – 164

Wayland, J. and Lee, D. (1986). *Seismic mapping of EOR processes.* The Leading Edge, 5(12), 36 – 40

Wen, X.-H. and Chen, W.H. (2006). Real-time reservoir model updating using ensemble Kalman filter with confirming option. *SPE Journal*, 11(4), 431 – 442

Whitaker, J.S., and Hamill, T.M. (2002). Ensemble Data Assimilation Without Perturbed Observations. *Monthly Weather Review*, 130(7), 1913 – 1924

Wilson, J.R. and Burgh G. (2008). *Energizing Our Future – Rational Choices for the 21st Century*. A John Wiley & Sons, New Jersey, 419

Wu, Y-S., Lu, G., Zhang, K. and Bodvarsson, G.S. (2007). An Integrated Modeling Approach for Characterizing Multiphase Flow, Chemical Transport, and Heat Transfer in Fractured Reservoirs. *SPE Europec/EAGE Annual Conference and Exhibition*, London, 1 – 10

Yuh, S.H. (2003). Time-Lapse Seismic Monitoring of Subsurface Fluid Flow. *PhD Dissertation*, Texas A&M University, 126

Zafari, M. and Reynolds, A.C. (2007). Assessing the Uncertainty in Reservoir Desription and Performance Predictions with the Ensemble Kalman Filter. *SPE Journal*, 12(3), 382 – 391

Zagayevskiy, Y.V. and Deutsch, C.V. (2010). Constraining Geostatistical Realizations to Temperature Data with an EnKF. *Centre for Computational Geostatistics*, 12, 205-1 – 205-12

Zagayevskiy, Y.V. and Deutsch, C.V. (2011a). Temperature and Pressure Dependent Gassmann's Fluid Substitution Model for Generation of Synthetic Seismic Attributes. *Centre for Computational Geostatistics*, 13, 210-1 – 210-19

Zagayevskiy, Y.V. and Deutsch, C.V. (2011b). Temperature and Seismic Data Integration to a Petroleum Reservoir Model by Means of Ensemble Kalman Filter (EnKF). *Centre for Computational Geostatistics*, *13*, 209-1 – 209-24

Zagayevskiy, Y.V. and Deutsch, C.V. (2011c). Application of Ensemble Kalman Filter (EnKF) to Tucker Oil Sands Project. *Centre for Computational Geostatistics*, *13*, 211-1 – 211-8

Zagayevskiy, Y.V., Hosseini, A.H. and Deutsch, C.V. (2010a). *A Guidebook on Ensemble Kalman Filter for Geostatistical Applications.* Centre for Computational Geostatistics, Edmonton, 82

Zagayevskiy, Y.V., Hosseini, A.H. and Deutsch, C.V. (2010b). Characteristics of the EnKF for Geostatistical Problems. *Centre for Computational Geostatistics*, 12, 125-1 – 125-20

Zhang, D., Lu, Z. and Chen, Y. (2007). Dynamic reservoir data assimilation with an efficient, dimension-reduced Kalman filter. *SPE Journal*, 12(1), 108 – 117

Zhang, W., Youn, S. and Doan, Q. (2005). Understanding Reservoir Architectures and Steam Chamber Growth at Christina Lake, Alberta, by Using 4D Seismic and Croswell Seismic Imaging. *SPE/PS-CIM/CHOA International Thermal Operations and Heavy Oil Symposium*

Zhang, Y., and Oliver, D.S. (2010). History Matching Using a Multiscale Stochastic Model with the Ensemble Kalman Filetr: A Field Case Study. *SPE Journal*, 16(2), 307 – 317

Appendix

The Appendix includes a description of FORTRAN programs used throughout this thesis, and corresponding parameter files.

A.1 Program enkf.exe

The program enkf.exe was devised to implement a proposed methodology based on the EnKF to integrate core porosity and permeability data and dynamic temperature and difference in acoustic impedances observations into SAGD petroleum reservoir models.
In order to run the program enkf.exe, its parameter file should be defined first and several supplementary executable files have to be stored in the folder, enkfsupexe, which must be present in the program directory. The supplementary programs are poroperm.exe, starstogslib.exe, fluidsub_tp.exe, seisdif.exe, which are discussed later, and sgsim.exe from GSLib book by Deutsch and Journel (1998). The first supplementary program derives log permeability values from porosity. The second program converts STARS thermal simulator`s output file to GSLib format. The third and fourth programs are devoted to a generation of synthetic seismic attributes, which are a function of reservoir pressure and temperature. The last executable implements the Sequential Gaussian Simulation. The difference in P-wave acoustic impedances ΔZ_p is chosen as a seismic attribute due to its ability to capture geological features and propagate steam chamber growth (Zagayevskiy and Deutsch, 2011a). A description of the parameter file of program enkf.exe is given below. Figure 41 shows a sample parameter file.

```
                  Parameters for ENKF
                  ********************

            START OF PARAMETERS:
1-          1                           -EnKF: implementation option (1-conventional, 2-shortcut)
2-          100                         - number of realizations
3-          50   0.5   1.0              - nx, xmin, xsize
4-          10   50.0 100.0             - ny, ymin, ysize
5-          30   0.5   1.0              - nz, zmin, zsize
6-          69069                       - random number generation seed
7-          poro_initial.out            - input file with gridded initial porosity ensemble
```

8-	1	- column for attribute
9-	enkf.out	- output file name for EnKF estimates
10-	0.1	-Observations: standard dev. of measurement error in %
11-	obs_poro.out	-input file name with porosity data
12-	1 1	- avail (1-yes,0-no),integr (1-gl,2-loc up,3-loc cov)
13-	1 2 3 4	- column for X, Y, Z coordinates and attribute
14-	3 3 3	- updating window nxup, nyup, nzup for local integration
15-	obs_perm.out	-input file name with permeability data
16-	1 1	- avail (1-yes,0-no),integr (1-gl,2-loc up,3-loc cov)
17-	1 2 3 4	- column for X, Y, Z coordinates and attribute
18-	3 3 3	- updating window nxup, nyup, nzup for local integration
19-	obs_temp.out	-input file name with temperature data
20-	1 1 270.0	- avail (1-yes,0-no),integr (1-gl,2-loc up,3-loc cov), *t*
21-	1 2 3 4	- column for X, Y, Z coordinates and attribute
22-	5 3 3	- updating window nxup, nyup, nzup for local integration
23-	obs_difz.out	-input file name with impedance difference data
24-	1 3 90.0 270.0	- avail (1-yes,0-no),integr (1-gl,2-loc up,3-loc cov), *t*
25-	1 2 3 4	- column for X, Y, Z coordinates and attribute
26-	3 3 3	- updating window nxup, nyup, nzup for local integration
27-	1.0	-Porosity-perm. model: correlation between Phi and log(K)
28-	2.5	- intercept of regression line, log(mD)
29-	2.5	- slope of regression line
30-	1.58 2.65	-Petroelastic model: densities of mineral 1 & 2 in g/cm3
31-	0.7	- volume fraction of mineral 1
32-	15.0 37.0	- bulk moduli of mineral 1 and 2 in GPa
33-	6.0 44.0	- shear moduli of mineral 1 and 2 in GPa
34-	50000.0	- salinity of brine water in ppm
35-	0.8	- specific gas gravity (ratio of gas to air densities)
36-	0.9	- density of oil at standard (surface) condition
37-	25 1 3	-SAGD: block numbers xloc, yloc, zloc of heel of producer
38-	25 1 3	- block numbers xloc, yloc, zloc of toe of producer
39-	25 1 7	- block numbers xloc, yloc, zloc of heel of injector
40-	25 1 7	- block numbers xloc, yloc, zloc of toe of injector

Figure 41: Parameter file of the program enkf.exe for implementation of the EnKF-based data assimilation algorithm

The parameter file is divided into five distinct sections: EnKF and model specifications, observations, porosity-permeability model, petroelastic model, and SAGD production scheme. The first section is devoted to the EnKF settings and the specification of the examined model grid. The EnKF implementation option is defined on first line. It is either

a conventional approach or a shortcut for reducing the number of flow simulator runs at the forecast step as was described in the thesis. The shortcut is used only to update ensembles of temperature and difference in acoustic impedances. Their initial values are generated unconditional to data, but correlated to initial ensemble of updated porosity values. The next line is reserved for the number of realizations or ensemble size. Lines 3 to 5 determine the model grid. The number of blocks, coordinates of model origin, and block sizes in three principal *X*, *Y*, *Z* directions are entered here (Deutsch and Journel, 1998). A random number generation seed is defined on line 6, which is used for multiple reasons, including to generate permeability values based on porosity or to assign white noise perturbations to observed data at EnKF analysis step. The input file name with the initial porosity ensemble is specified on the next line 7. The SGS is recommended to generate the initial ensemble with the proper semivariogram model. The column for porosity attributes is defined on the next line. The root name for output files is entered on line 9. The program generates several output files. The file names for EnKF estimates of porosity, permeability, temperature, and the difference in acoustic impedances use the prefixes poro_, perm_, temp_t1_, and difz_t0_t1_ respectively, where *t0* stands for time step when the baseline survey is conducted and *t1* represents the time step when temperature and repeated acoustic impedance are measured. Similar output files for the mean and variance of realizations of every variable are generated with the intermediate prefix _mv_.

The second section is devoted to measurements and observations of static and dynamic variables respectively. The standard deviation of data measurement error (perturbation) in the percentage of the true data value is defined on line 10. The next sixteen lines are used to define names of input files with data and type of assimilation procedure. The assimilation procedure is global, local with the localized updating matrix or local with the localized covariance matrix. The input file with porosity measurements is defined on line 11. Flags for porosity data availability and the assimilation procedure are entered on the next line. Line 13 lists the specification of columns with coordinates of porosity data and the attribute. If the local assimilation method is chosen, a half size of the updating window for the porosity data is entered on line 14 in units of blocks. The next four lines are exactly the same as the previous four, but are only used to specify permeability

data. The definition of temperature and the difference in acoustic impedances data is slightly different than for the porosity and permeability data. The observation time step should be indicated for temperature and acoustic impedance and furthermore, the time step of the baseline survey should be specified for acoustic impedance.

The third section of the parameter file is about the porosity-logarithmic horizontal permeability relationship. Forecasted values of the permeability variable are generated based on the linear regression model (Equation (117)) and porosity values. Note that α and β, which are used to define permeability components in directions different from X in the Equation (118), are constant 1.0 and 0.5, respectively. The correlation coefficient between porosity and log permeability, intercept, and slope of regression line are defined on lines 27 – 29.

The fourth section of the parameter file specifies the petroelastic model parameters based on the pressure- and temperature-dependent Gassmann's fluid substitution model. For more details on Gassmann's theory, see Chapter 2 of the thesis or Zagayevskiy and Deutsch (2011a). Densities in the g/cm^3 of a rock's constituent minerals are indicated on line 30. Recall that the assumption implies that a rock consists of two mineral types. Line 31 shows the volume fraction of a mineral. The volume fraction of another mineral is computed automatically as: volume fraction of second mineral = 1.0 – volume fraction of first mineral. Elastic bulk and shear moduli of both minerals in GPa are defined on the next two lines. Lines 34, 35, and 36 are reserved for brine water salinity in ppm, specific gas gravity (gas to oil densities ratio, which is usually between 0.56 and 1.80), and density of oil (or bitumen) at standard conditions (15.6 ^{0}C, 0.1 MPa).

The last section is used to specify the SAGD production scheme and, more specifically, the single well pair location in space. Block numbers of the producer's and injector's heels and toes in three principal X, Y, Z directions are defined on lines 37 – 40.

A.2 Program poroperm.exe

Figure 42 shows a sample parameter file of program poroperm.exe that generates permeability values from porosity based on a probabilistic regression model (Equation (117)). The input file name with porosity values is specified on line 1. The next line is for the porosity column attribute. The correlation coefficient between porosity and logarithmic horizontal permeability is defined on line 3. The intercept in log_{10}(mD) and the slope of the linear regression line are entered on lines 4 and 5 respectively. The random number generation seed is required to draw values of white noise, which is specified on line 6. The last line 7 is devoted to the name of the output file, which contains permeability, logarithmic permeability, and porosity values in three columns.

Parameters for POROPERM
****

	START OF PARAMETERS:	
1-	poro.dat	-input file with porosity data
2-	1	- column for attribute
3-	0.8	-correlation coefficient between poro and log perm
4-	1.5	-intercept of regression line, log(mD)
5-	2	-slope of regression line
6-	69069	-random number generation seed
7-	poroperm.out	-output file for porosity and permeability

Figure 42: Parameter file of the program poroperm.exe for generating permeability values from porosity

A.3 Program fluidsub_tp.exe

Below is a description of a parameter file of program fluidsub_tp.exe for generating pressure- and temperature-dependent synthetic seismic attributes based on Gassmann`s fluid substitution model. See Figure 43 for a sample parameter file of this program.

```
Parameters for FLUIDSUB_TP
*************************

START OF PARAMETERS:
1-    poro.dat                  -input file with porosity
2-    1                         -  column for attribute
3-    pres.dat                  -input file with pressure in kPa
4-    1                         -  column for attribute
5-    temp.dat                  -input file with temperature in oC
6-    1                         -  column for attribute
7-    wsat.dat                  -input file with brine water saturation
8-    1                         -  column for attribute
9-    osat.dat                  -input file with oil saturation
10-   1                         -  column for attribute
11-   gsat.dat                  -input file with gas saturation
12-   1                         -  column for attribute
13-   1                         -  number of realizations
14-   100   100   100           -  nx, ny, nz
15-   1.580      2.650          -densities of mineral 1 and 2 in g/cm3
16-   0.7                       -volume fraction of mineral 1
17-   15.0       37.0           -bulk moduli of mineral 1 and 2 in GPa
18-   6.0        44.0           -shear moduli of mineral 1 and 2 in GPa
19-   50000.0                   -salinity of brine water in ppm
20-   0.80                      -specific gas gravity G (ratio of gas to air densities)
21-   0.900                     -density of oil at standard (surface) conditions
22-   seismic.out               -output file for seismic attributes: Vp, Vs, Zp, and Zs
23-   1                         -  standard deviation of measurement error in seismic, %
24-   69069                     -  random number generation seed for measurement error
25-   5   5   5                 -  size of smoothing window: nxs, nys, nzs
```

Figure 43: Parameter file of the program fluidsub_tp.exe for generating synthetic seismic attributes using petroelastic Gassmann's fluid substitution model

To run the program, it is necessary to know a reservoir's geological and dynamic production properties; mineralogical content; and the constituent minerals' elastic and physical properties, grid size, and some other miscellaneous parameters. Lines 1 to 12 define the input files with porosity, reservoir pressure, temperature, brine water, oil and gas saturation values, and columns in the files for these attributes. The number of realizations is entered on next line. The number of blocks in a model for all three dimensions is specified on line 14. The densities of the constituent two minerals (for

instance, clay and quartz) are defined on the next line with the volume fraction of first mineral on subsequent line. The elastic bulk and shear moduli of the minerals are specified on lines 17 and 18 respectively. Line 19 is reserved for water salinity. Line 20 defines specific gas gravity (ratio of gas to air densities, which is usually between 0.56 and 1.80). The next line is for the density of oil at standard or surface conditions (101325 Pa and 16.6 ^{0}C). Line 22 specifies the output file for four seismic attributes. Generated seismic attributes are P-wave and S-wave seismic velocities, and P-wave and S-wave acoustic impedances. The option for adding a measurement, resolution or inversion error to seismic attributes is embedded to the program. The error is added in the form of white noise with zero mean and specified standard deviation in a percentage of the seismic attributes' true value. Also, to mimic error, the smoothing window concept can be applied to smooth out the seismic attributes' value in adjacent blocks. Both error generation approaches can be combined and applied at the same time.

A.4 Program starstogslib.exe

This program transforms the output of CMG's thermal flow simulator STARS to GSLib format (Deutsch and Journel, 1998). Figure 44 contains a description of the parameter file.

```
                Parameters for STARStoGSLib
                ***************************

          START OF PARAMETERS:
1-        stars.out                    -input file with STARS simulation results
2-        100   100   100              -  nx, ny, nz
3-        3                            -how many times should data be extracted?
4-        120.0  240.0  360.0          -  time for extraction, days
5-        stars_gslib.out              -output file for simulation results
```

Figure 44: Parameter file of the program starstogslib.exe for converting output of thermal flow simulator STARS to GSLib format

STARS's output is specified on the first line. Then, the number of blocks in the model grid for three dimensions is defined. The model may have two or three dimensions. The

number of time steps, for which data should be transformed, is entered on line 3. The next line is used to determine the time steps in days for data extraction. The last 5th line defines the output file, which contains reservoir pressure, temperature, and water, oil, and gas saturations. Error messages may show up if something goes wrong in the calculation procedure, if the STARS' output file format is set up improperly, or if the input file is missing.

A.5 Program seisdif.exe

This program is designed to compute the difference between seismic attributes from a baseline seismic survey and any other subsequent surveys. A file name with attributes from the baseline survey and the column for required attribute are specified on lines 1 and 2 respectively. The number of follow-up surveys N_{fp} is entered on line 3. The next 2*N_{fp} lines are dedicated to file names with follow-up seismic attributes and columns for relevant data. The last line defines the output file name, which contains the difference in seismic attributes (seismic from i^{th} follow-up seismic survey minus seismic from baseline survey).

Parameters for SEISDIF
****

START OF PARAMETERS:

1-	base_survey.out	-input file with seismic attributes from base survey
2-	1	- column for attribute
3-	2	-number of follow-up surveys
4-	1_survey.out	-input file with seismic attributes from 1st survey
5-	1	- column for attribute
6-	2_survey.out	-input file with seismic attributes from 2nd survey
7-	1	- column for attribute
8-	seisdif.out	-output file for difference between seismic attributes

Figure 45: Parameter file of the program seisdif.exe for computing the difference in seismic

A.6 Program extrobs.exe

Figure 46 shows a sample parameter file of the program extrobs.exe. This program is used to extract scattered data from the gridded base case for further use in data assimilation based on the EnKF. The input file name with gridded data is specified on the first line. The column for the attribute is entered on the next line. The model grid specification is defined on lines 3 – 5. The file with desired data locations' *X*, *Y*, and *Z* coordinates and their column numbers are specified on lines 6 and 8. Line 7 shows the number of first *n* observations that must be extracted. If all coordinates are intended to be used to extract the data, a -1 flag should be entered on this line. The output file contains data, their coordinates, and the standard deviations of measurement error in percentage of true values (line 10). The output file name is specified on line 9.

```
                Parameters for ExtrObs
                **********************

        START OF PARAMETERS:
1-      input.dat                   -input file with data
2-      1                           -   column for attribute
3-      10  5.0  10.0               -   nx, xmn, xsiz
4-      10  5.0  10.0               -   ny, ymn, ysiz
5-      10  5.0  10.0               -   nz, zmn, zsiz
6-      obs_loc.dat                 -input file with observation locations
7-      10                          -   number of observation locations, -1 for all
8-      1   2   3                   -   columns for X, Y and Z coordinates
9-      obs.out                     -output file for observations
10-     1                           -   std. err. of observations in % of true value
```

Figure 46: Parameter file of the program extrobs.exe for extracting scattered values at specific locations from a gridded data file

A.7 Program valid.exe

The program valid.exe is used to compute the root mean square error *RMSE* (Equation (121)), mean absolute error *MAE,* linear correlation coefficient CC (Equation (122)), and mean standard deviation *MSD* (Equation (124)) for several realizations of 3D models in

order to compare estimated realizations to the base case or to see how different the realizations are from each other. Figure 47 shows the program's parameter file.

```
                Parameters for VALID
                ********************

          START OF PARAMETERS:
1-        simulation.dat              -input file with simulation data
2-        1                           -  column for attribute
3-        100                         -  number of realizations
4-        100   100   100             -  model grid: nx, ny, nz
5-        base_case.dat               -input file with reference/base case
6-        1                           -  available: 1-yes, 0-no
7-        1                           -  column for attribute
8-        2   2   2                   -half size of moving average window in X, Y, Z directions
9-        valid.out                   -output file for RMSE, MAE, CC, and MSD
```

Figure 47: Parameter file of the program valid.exe for calculating root mean square error (*RMSE*), mean absolute error (*MAE*), linear correlation coefficient (*CC*), and mean standard deviation (*MSD*)

Absolute errors AE_{ir} of every realization and their average *MAE* are computed in a manner similar to RSE_{ir} and *RMSE* as shown in Equations (134) and (135). The smaller value of *MAE* indicates a better match between the base case of the model and its estimate. The name of the input file with simulated data is defined on line 1 with a specified column attribute on the next line. The number of realizations and the model grid size are entered on lines 3 and 4. The name of a file with a gridded base case or reference case values should be indicated on line 5 with a column attribute on line 7. The availability of the base case file is specified on line 6. If the base case file is not present, only *MSD* is reported. Line 8 is reserved for the half size of the smoothing window for averaging estimated values falling into the moving smoothing window. The parameter file's last line specifies the root name of output files for *RMSE*, *MAE*, and *CC* values with prefix rmse_mae_cc_, and for *MSD* value with prefix msd_.

$$AE_{i_r} = \frac{\sum_{i_x=1}^{N_{bx}} \sum_{i_y=1}^{N_{by}} \sum_{i_z=1}^{N_{bz}} \left| u_{i_x,i_y,i_z}^{true} - \bar{u}_{i_x,i_y,i_z,i_r}^{est} \right|}{N_{bx} \cdot N_{by} \cdot N_{bz}} \tag{134}$$

$$MAE = \frac{\sum_{i_r=1}^{N_e} AE_{i_r}}{N_e} \tag{135}$$

A.8 Program standvar.exe

The purpose of the program standvar.exe is to either standardize or back-standardize values of a random variable *U* according to Equations (136) and (137). If a variable *U* is standardized, its mean and standard deviation are set to zero and one respectively. If a variable *U* is back-standardized, its zero mean and unit standard deviation are changed to target mean and standard deviation. Figure 48 shows the program's sample parameter file. The name of input file with data, which have to be altered, is specified on the first line. The column attribute for the data is defined on the next line. Line 3 shows the option flag for standardization or back-standardization. If the back-standardization procedure is selected, the target mean and standard deviation of the final distribution have to be specified on lines 5 and 6. Otherwise they should be skipped. Finally, the output file's name is entered on the parameter file's last line.

$$u^{stand} = \frac{u^{orig} - \psi_u^{orig}}{\sigma_u^{orig}} \tag{136}$$

$$u^{orig} = u^{stand} \cdot \sigma_u^{target} + \psi_u^{target} \tag{137}$$

where u^{stand} is the standardized value of variable *U*; u^{orig} is the original value of variable *U*; ψ_u^{orig} and σ_u^{orig} are the average and standard deviation of the original values of variable *U* respectively; ψ_u^{target} and σ_u^{target} are the target mean and standard deviation for variable *U* respectively.

```
Parameters for STANDVAR
**********************

1-      START OF PARAMETERS:
2-      input.dat                   -input file with data
3-      1                           -   column for attribute
4-      1                           -standardize (1) or back-standardize (2) values?
5-      1000                        -if (2): mean value
6-      100                         -if (2): standard deviation
7-      output.out                  -output file for old and new data
```

Figure 48: Parameter file of the program standvar.exe for standardization or back-standardization of distribution of a random variable