Data-Driven Proxy Models for Assisted History Matching of SAGD Reservoirs

by

Tarang Jain

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science
in
Petroleum Engineering

Department of Civil and Environmental Engineering
University of Alberta

Abstract

In reservoir simulation studies, history matching is extensively used for uncertainty reduction and reservoir management. History matching using Ensemble Kalman Filter (EnKF) is a promising approach due to its non-iterative nature and ability to assimilate a large number of model parameters. However, in processing a large set of realizations, this method suffers from high computational time and cost associated with the use of commercial reservoir simulators. Therefore, there is a scope for some improvement in this approach especially in the case of complex thermal recovery process such as steam assisted gravity drainage (SAGD). In this work, the computational cost is reduced significantly by developing proxy models that can substitute the need of reservoir simulator during the assisted history matching process. Different proxy models such as Polynomial Chaos Expansion (PCE) and Artificial Neural Networks (ANN) are tested to represent the outputs of the conventional reservoir simulator. Permeability realizations of the SAGD reservoir are first parameterized using Karhunen- Loeve (KL) series expansion and represented in the form of uncorrelated random variables. The developed proxy models utilize random variables obtained from KL expansion as input parameters. Proxy models are further integrated with EnKF algorithm as a substitute for reservoir simulator. Computational requirement of the proxy model during the development as well as deployment as compared to commercial reservoir simulator is emphasized in this study. The proposed approach is validated using a field-scale SAGD case study of northern Alberta. The observed daily oil rate, cumulative oil production, and cumulative steam

to oil ratio are history matched using the proposed method. Results show that as compared to conventional EnKF, the integration of data-driven proxy models can perform assisted history matching in quick, low-cost manner while maintaining the accuracy of results. This work has a potential to cut down the monetary and time constraints during the assisted history matching process.

Preface

Chapter 3 of this thesis is partially based on a manuscript submitted for publication as T. Jain, R.G. Patel, and J. Trivedi, "Application of Polynomial Chaos Theory as an Accurate and Computationally Efficient Proxy Model for Heterogeneous SAGD Reservoirs," *Journal of Petroleum Science & Engineering*, submitted, March 2016". I was responsible for programming code, carrying out simulations, validation against the simulator, manuscript composition and explanation of results. R.G. Patel was responsible for data collection. Dr. Trivedi was the supervisory author and contributed to manuscript edits.

Chapter 4 of this thesis is partially based on a manuscript submitted for publication as T. Jain, R.G. Patel, and J. Trivedi, "Application of ANN based Proxy Models for Efficient and Fast track Assisted History Matching of SAGD Reservoirs," *in proceedings of World Heavy Oil Congress*, September 2016". I was responsible for programming code, carrying out simulations, validation against the simulator, manuscript composition and explanation of results. R.G. Patel was responsible for data collection and formulation of the initial model. Dr. Trivedi was the supervisory author and contributed to manuscript edits.

Chapter 5 of this thesis is partially based on a manuscript submitted for publication as R.G. Patel, T. Jain and J. Trivedi, "Polynomial-Chaos-Expansion Based Assisted History Matching Workflow for Computationally Efficient Reservoir Characterization: A SAGD Field Case Study," *in SPE Journal of Reservoir Engineering and Evaluation*, December 2016". R.G. Patel was responsible for programming code, data collection, carrying out simulations, manuscript composition and explanation of results. I was responsible for generating of proxy models. Dr. Trivedi was the supervisory author and contributed to manuscript edits.

Acknowledgement

I am sincerely grateful and would like to offer my earnest gratitude to my supervisor, Dr. Japan J Trivedi for his persistent support, guidance and encouragement for my research throughout my Masters'. I pay my deepest appreciation to him and I am deeply indebted for all his efforts and help.

I would also like to thank Rajan G Patel for his valuable guidance in the use of Petrel, CMG and MATLAB software. His valuable suggestions have always been very helpful during the course of my project.

I would like to express my love and gratitude to my family back home, my parents, my brother Aakash and my sister-in-law Rashi for their constant support & endless love, throughout my studies. I would like to thank all my friends in Edmonton Yogesh, Ankit, Nidhi, Hemant, Sahil, Rohtaz, Prashant, Amit, Nikhil for providing me a very friendly atmosphere and sharing their time and joy with me.

I gratefully acknowledge the financial support from MITACS and Total E&P Canada Ltd for this research through grants and scholarship. Software support by Schlumberger and Computer Modelling Group is also greatly appreciated.

Table of Contents

List of Tables

List of Figures

Abbreviations

AHM	Assisted History Matching
ANN	Artificial Neural Network
BFGS	Broyden Fletcher Goldfarb Shanno
CMG	Computer Modeling Group
DCT	Discrete Cosine Transform
EnKF	Ensemble Kalman Filter
GA	Genetic Algorithm
KL	Karhunen Loeve
PCE	Polynomial Chaos Expansion
PCM	Probabilistic Colocation Method
RBF	Radial Basis Function
RMSE	Root Mean Square Error
RSM	Response Surface Models
SAGD	Steam Assisted Gravity Drainage
SOR	Steam Oil Ratio
SGS	Sequential Gaussian Simulation
TPWL	Trajectory Piecewise Linearization

Chapter 1:

General Introduction

1.1 Reservoir Modeling and Simulation

Reservoir modeling and simulation is an art to predict the flow of fluids through subsurface geological formations. It plays a very crucial role in decision-making process and field development strategies. It takes into account various factors significantly contributing to the economic risk and reservoir performance. These factors include a variation of fluid properties, properties and types of rocks, porosity and permeability characteristics, and substantial physical processes taking place within the reservoir. Traditional reservoir simulation techniques consist of analogical, experimental and mathematical approaches (Ertekin et al. 2001).

Analogical methods predict the performance of reservoir based on the information of already developed reservoirs with similar properties. These approaches have vital importance at the early stage and when minimal or no data is available for the target reservoir. However, such methods failed to provide any information if development strategies or production scenarios differ from the scenarios known for the sample reservoir. Experimental methods deal with direct measurements of flow properties. Reservoir rock samples are taken from the field in the form of cores and studied under laboratory environment. Even though the experimental

analysis reflects the reservoir performance but the difficulty lies in scaling up the lab sample features to the actual reservoir scale.

Conventional mathematical methods used to predict the reservoir performance is decline curve analysis and material balance (Dake, 2007). In decline curve analysis, a curve is fitted graphically to the observed production data and the reservoir performance is predicted. Production rate versus the cumulative production is plotted on semi-logarithmic plot and based on a mathematical model (exponential, hyperbolic or harmonic) a curve is fitted to the plot. As no physical interpretation of subsurface flow is involved, and also pressure data is not used in the analysis, decline curve approach predicted the reservoir performance only under certain circumstances and failed when development strategy is changed. Material balance is another approach which is based on the assumption that the reservoir volume is constant and expansion of other components and aquifer influx causes the depletion of the volumetric component. This approach also assumes that the reservoir is homogeneous and reacts spontaneously and equally throughout its volume which is one of its main drawbacks. It is also not possible to determine new development strategies such as the location of the new well using simplified material balance analysis.

Currently, complex mathematical simulation models are probably the most widely used techniques in reservoir simulation. It encompasses discretization of the reservoir into grid blocks, each with a different set of rock and fluid properties. Material balance analysis is conducted on each grid block coupled with initial and boundary conditions of the reservoir, and mathematical equations are solved to

estimate the pressure and flow rates of reservoir fluids at each grid block. Due to a large number of grid blocks and complexity of the system, it is tough to solve the model equations analytically. Hence numerical methods are used for solving mathematical models.

1.2 Reservoir Simulation Process

There is a standard workflow for reservoir simulation process. Below are the major components of a typical reservoir simulation study (Carlson, 2003). The process has also been summarized in Figure 1.

Data Gathering: In this step, all the data is gathered and selected for the input in the simulation model. Collective data regarding relative permeability, capillary pressure, and PVT data is screened. Correlations are applied on various lab data, and geological maps of porosity and net pay are developed and digitized into a grid format.

Initialization: This step is mainly known as data checking step. An initial run is completed, and preliminary calculations are conducted. Calculation of parameters like grid block saturation, original oil in place is the main objective of this step. Results are cross-checked against other available data and results.

History Matching: This step includes the tuning of input data based on the field production data. It involves a series of trial and error runs to obtain the correct values and therefore is the most time-consuming step. The simulation model is run through time with field production data, and the idea is to match the production behavior of the simulator with the actual behavior that happened

within the reservoir. Several methods and approaches are used in petroleum engineering community for efficient and fast history matching.

Predictions: Once the model is history matched, the tuned input parameters are stored to continue the simulation at a later time. Various predictions are made for the reservoir performance using different injector patterns, well locations, operating constraints.

Documentation: In the final stage, all the input and output data is thoroughly studied, and results are documented and presented for field development strategies and decision-making processes.

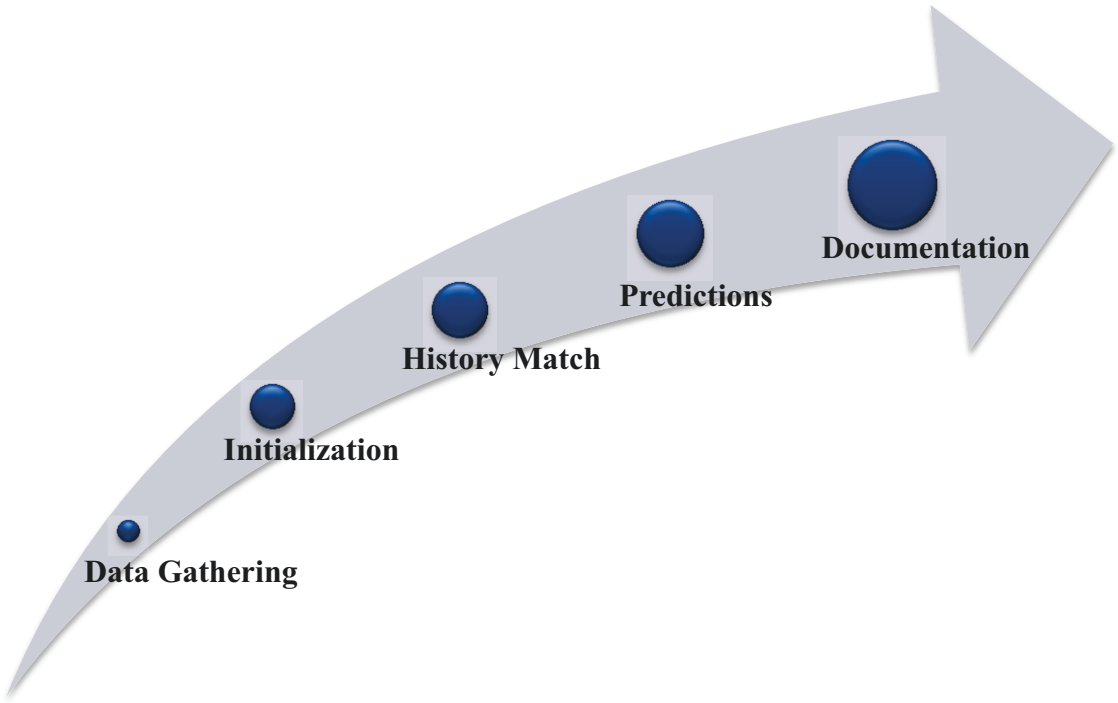


Figure 1: Reservoir Simulation Process

1.3 Uncertainty in Reservoir Properties

An important step of reservoir modeling is to address the uncertainty in geological properties of reservoir rocks. Geological properties such as porosity, permeability and saturations are obtained using different kinds of well logs measured during drilling and exploration phase. Due to the presence of noise and measurement error in well logs, core samples, and seismic data, uncertainty becomes an intrinsic characteristic of any geological model. In almost any case, available data is limited and geological formations are way too complicated to be pronounced on a model level conceptually. Even if all the information is known, computational power would be a constraint for solving large-scale problems. Researchers are developing different approaches to get a reasonable solution for these issues. Models are divided into two categories, deterministic models, and stochastic models. Deterministic models assume that all the parameters and conditions are known hence claim a single best prediction as an actual result. On the other hand, stochastic models use multiple input realizations and provide a range of possible solutions and hence help in quantifying the uncertainty for complex dynamic systems. Uncertainty in the reservoir properties, scale, and complexity of the reservoir are some of the main difficulties during a modeling and simulation study. History matching is an approach for handling the lack of available data and uncertainty in geological information during a simulation study.

1.4 Introduction to History Matching

History matching is a process of adjusting model input parameters to obtain a model output similar to historical/dynamic data. Dynamic data normally includes production data measured at the surface as oil, water, and steam production rate, cumulative oil/water production. This is a non-unique problem since there is more than a solution which can match the available data. Therefore, a unique history matched reservoir model is not sufficient to predict reservoir performance, and a calculation of uncertainty in reservoir model variables becomes the important part of history matching problem.

History matching techniques have been practiced widely in petroleum engineering applications for the last several decades. Traditionally, reservoir models were history matched manually using a good engineering knowledge and judgment. Manual History matching is a trial-and-error approach and is known to be time-consuming and ineffective. It is obviously not the practical approach. Therefore, numerous statistical and mathematical methods have been studied as automatic or assisted history matching algorithms in reducing the time required during the process while maintaining the accuracy of results. Assisted history matching (AHM) is software enabled approach which can calibrate a reservoir model using known data. A detailed discussion on various history matching techniques is presented in the further chapter.

1.5 Problem Definition

Data assimilation and history matching of the large-scale reservoir is a challenging task due to high computational cost and time requirement associated with the use of commercial reservoir simulators. For a conventional reservoir of average size, hundreds or thousands of grid blocks can be employed each one with different properties. During assisted history matching process various realizations of these models are evaluated using commercial reservoir simulators repeatedly at different time steps, resulting in high computational cost and time. Although the growth in computational capabilities in recent years has supported the assisted history matching process, for large scale models and SAGD reservoir models, in particular, the computational cost and time is still a major limiting factor. Various approaches are documented in the literature to reduce the time and number of simulations in the history matching process. One of the approaches to reduce the computational cost and time is by reducing the number of input realizations by application of different screening/ranking methods (Patel et al. 2015). Another approach which is of interest is the use of proxy or surrogate models instead of commercial reservoir simulators. Proxy models are referred as mathematically derived models that imitate the output of a simulation model for selected input parameters. Proxy models are widely applied in different numerical modeling tasks such as sensitivity analysis, probabilistic forecasting, reservoir management and process optimization. In cases where proxy models can adequately represent relevant output parameters, they can be used as an adequate substitution for full reservoir simulations. An integrated workflow is therefore

required which can conglomerate the use of proxy models with data assimilation algorithm for fast and low-cost history matching of large-scale reservoirs while maintaining the accuracy of results.

1.6 Research Objectives

The primary goal of this thesis is to develop an integrated framework using data-driven proxy models with Ensemble Kalman Filter (EnKF) data assimilation approach for efficient and fast-track history matching of large scale complex reservoirs. In this work, data-driven proxy models, which can substitute the need of using reservoir simulator during the assisted history matching process, are developed. Spatial distribution of permeability field for various synthetic and real geological reservoirs is parameterized using Karhunen- Loeve (KL) series expansion and represented in the form of uncorrelated random variables. Proxy models are developed using the random variables obtained from KL expansion as input parameters and used to predict production parameters as outputs. Established models are further integrated into EnKF framework to predict production parameters in forecast step while updating the random variables instead of permeability of each grid block in analysis step of EnKF. Computational requirement of the proxy models during the development as well as deployment as compared to commercial reservoir simulator is emphasized in this study. Primary objectives of this thesis are summarized as below:

- To parameterize the random permeability fields of geological reservoirs in the form of uncorrelated variables using KL series expansion.
- To construct different data driven proxy models based on polynomial chaos expansion and artificial neural networks capable of imitating the outputs of reservoir simulator.
- To compare the efficiency and accuracy of developed proxy models with commercial simulators for different reservoir models.
- To develop a framework combining data-driven proxy models with EnKF approach for efficient and fast-track history matching of geological reservoirs.
- To history match a real 3D SAGD reservoir using data-driven proxy model based EnKF algorithm.

1.7 Thesis Outline

The thesis is organized as follows.

Chapter 1 provides the general overview of reservoir modeling and simulation process. It also explains the objectives and scope of research work.

Chapter 2 provides a literature review on history matching algorithms and various proxy modeling methods and parameterization techniques.

Chapter 3 discusses the KL parameterization method used as a dimensional reduction tool for geological reservoirs in this work. Also, different proxy models are presented in this section. Detailed methodology, principle, and application for a real SAGD field are provided.

Chapter 4 provides an integrated framework for the use of ANN based proxy models with Ensemble Kalman filter for fast and efficient history matching of a real 3D SAGD reservoir. Results are compared with the conventional approach of EnKF.

Chapter 5 provides an integrated framework for the use of PCE based proxy models with Ensemble Kalman filter for a real 3D SAGD reservoir.

Chapter 6 discusses the summary and conclusions of this study along with some insights into future work for application of proxy models in reservoir modeling and simulation.

Chapter 2:

Literature Review

2.1 History Matching

Reservoir Simulation is a vital exercise for decision making and field development planning. An important step of reservoir modeling is to address the uncertainty in geological properties of reservoir rocks. Geological properties such as porosity, permeability and saturations are obtained using different kinds of well logs measured during drilling and exploration phase. Due to the noisy and sparse nature of well logs, core samples, and seismic data, uncertainty becomes an intrinsic characteristic of any geological model. History matching is used to solve this problem and to estimate the spatially varying reservoir properties. History matching is a process of adjusting model parameters to obtain a model output similar to historical/dynamic data. Dynamic data normally includes production data measured at the surface as oil, water, and steam production rate, cumulative oil/water production.

History matching techniques have been practiced widely in petroleum engineering applications for the last several decades. It provides a non-unique solution where different variable sets may result in an equally good match with the observed production history of the reservoir (Oliver & Chen, 2011). Traditionally, reservoir models were history matched manually using a good engineering knowledge and

judgment. Manual History matching is a trial-and-error approach and is known to be time-consuming and ineffective (Romeu, 2010). It may also result in loss of geological sanity (Oliver & Chen, 2011). Agarwal et al. (2003) applied manual history matching for a complex fracture chalk reservoir in the Norwegian North Sea. They stated that it took almost one year of rigorous work to history match the production data manually. It is obviously not the practical approach. Therefore, numerous statistical and mathematical methods have been studied as automatic or assisted history matching algorithms in reducing the time required during the process. Advancement in computer technology has also generated interest in automated history matching techniques.

In assisted history matching, simulated data is compared to historical data by minimizing a misfit function. Various algorithms have been developed to perform history matching in an efficient and reliable way. They can broadly be divided into local and global algorithms. Oliver & Chen (2011) reviewed different approaches like gradient methods, neighborhood algorithm, adjoint methods, etc. used in assisted history matching. Gradient-based algorithms and adjoint methods fall under the category of local algorithms and have the advantage to converge faster than global approaches. However, in the case of multidimensional and non-linear problems, these methods tend to get stuck in local minima. They provide only one solution corresponding to a minimum which depends on the initial set of reservoir parameters. Adjoint methods have to be hard coded in the simulators which reduce their adaptability to different simulators (Cancelliere et al. 2011). On the other hand, global algorithms such as genetic algorithms and evolutionary

strategies provide multiple solutions in a single run and can spurt the issue of local minima efficiently. Evolutionary algorithms are frequently used approach because of their adaptability to different simulators. However, slow rate of convergence decreases their efficiency when dealing with a large number of parameters (Cancelliere et al., 2011). These methods require a large number of simulation runs to evaluate the misfit function which results in high computational cost and time. A brief introduction and limitations of various history matching algorithms have been presented in this chapter.

2.1.1 Gradient-based approach

In the Gradient-based approach for history matching, the unknown model parameters are calculated by minimizing an expression called the objective function. The objective function gradients on model parameters are estimated, and direction of optimization search is then determined. The objective function is typically based on the squared difference of the simulated data as compared to the observed data.

$$O(z) = \|f(z,t) - f_{obs}(z,t)\|^2 \quad \text{..... 2.1}$$

Where $f(z,t)$ is the output of simulated model over model parameter z and $f_{obs}(z,t)$ is the observed data. The task of history matching is to find z which minimize the objective function represented by Eqn.2.1. The loop followed by the process includes running the flow simulator for entire history matching period, evaluating the cost function, updating the static parameters and going back to the first step (Liang, 2007). There are several algorithms available in literature the

which follows the basis of the gradient-based method such as Gauss-Newton, Levenberg-Marquardt, Conjugate gradient, Quasi-Newton, Limited Memory Broyden Fletcher Goldfarb Shanno (BFGS). These methods fall into the category of Gradient-Based Methods. In view of finding the minimum of the objective function O, its gradient concerning z should be zero. Using the Taylor Series expansion, the following expression is derived (Nejadi, 2014):

$$Z_{k+1} = Z_k - H^{-1}(Z_k).\nabla O(Z_k) \quad \text{2.2}$$

Where, k is the iteration step, and $H(Z_k)$ is the Newton Hessian Matrix, represented by $\nabla\left(\left(\nabla O(Z_k)\right)^T\right)$. Prime difficulty with the gradient-based method for history matching is the evaluation of the Hessian matrix.

Another limitation of gradient based approach is that it does not consider the reference statistics of the model parameters (Nejadi, 2014). Although the initial guess is made from reference distribution and statistics but got wrecked during the successive updates of the model parameter. Also, in the case of multidimensional and non-linear problems this approach tends to get stuck in local minima (Liang 2007). It provides only one solution corresponding to a minimum which depends on the initial set of reservoir parameters and may be far away from the global optimum.

2.1.2 Adjoint Method

Adjoint method is the most efficient method for computing the gradient of a function and is practical for history matching problems with a large number of model parameters. As mentioned in the previous subsection that the main drawback with gradient based techniques is to evaluate gradients and Hessian matrix. In the adjoint method, the sensitivity of each production data concerning model parameters is generated in terms of derivatives and adjoint equations (Liang, 2007). The system of adjoint equations is analogous but distinct from the finite difference equations in the reservoir simulator. While the finite difference equations may be nonlinear, the adjoint state variables are the solutions of a linear system of equations. Also, reservoir simulation runs forward in time while the adjoint variables are propagated backward in time that includes information from the simulation results.

Oliver et al., (2008) have presented the application of the adjoint method for the history matching of a one-dimensional problem. The adjoint method was applied to a water-oil two-phase problem by Zhan et al. (1999) and to three-dimensional three-phase problems by Makhlouf et al. (1993). The computational time for adjoint equations is comparatively less than the corresponding time required for reservoir simulator. But for a large number of observed data, this method becomes computationally expensive. In other words, this method is impractical for large-scale multiphase flow due to high computational time and cost associated with the process. Even if due to advancement in computing technologies, the adjoint method becomes feasible for large scale problems, it is still limited to the

implantation into the source code of the reservoir simulator. Adjoint methods have to be hard coded in the simulators which reduce their adaptability as source code for different simulators are very complicated and are not easily reachable (Cancelliere et al. 2011).

2.1.3 Genetic algorithms

Genetic Algorithm (GA) is a randomized search technique, which is based on an analogy to natural selection according to the evolutionary theory of Darwin and the ”survival of the fittest” principle. This method was developed by John Holland at the University of Michigan (Holland, 1975). It starts with an initial sample space of possible solutions, from which specific solutions are selected according to a stochastic process for the best combination of parameters to improve the match and discard the bad candidates. The process is repeated till a convergence is achieved, or specified number of evaluations are performed. Genetic algorithm approach consists of three main steps, designing of sample space which contains the uncertain possible solution in terms of variables or realizations, defining and selecting proper structure to create solutions, and thirdly the operators to generate new solutions. This approach is easy to implement but computationally expensive (Kaleta, 2011). It has been used multiple times in petroleum engineering applications. Genetic algorithms have also been embedded in various commercial software (Begum, 2009). BP’s “Top-Down Reservoir Modelling” approach proposed by (Williams et al., 2004) used the genetic algorithm as a global optimizer in combination with the reservoir simulator to

accomplish history matching tasks. Unfortunately, due to the computation cost arising from the slow convergence, a genetic algorithm is still very limited in real problems (Liang, 2007).

2.1.4 Simulated Annealing

Simulated annealing is a probabilistic procedure for global optimization problems, which locate the global optimum of a given function by implementing certain cooling schedules in a large search space. This approach is analogous to annealing in metallurgy, a technique which involves heating and controlled cooling of the material to alter the size of its crystals and reduce their defects. The heating increases the energy of the atoms within the material. The slow cooling causes the solid material to form a new homogeneous crystalline structure with minimal energy. Metropolis et al., (1953) numerically simulated the molecular behavior when the energy level of the system is altered. Kirkpatrick et al., (1983) further implemented the idea to optimization problems. The idea is to use a control parameter, like temperature, to search for feasible solutions and find the global minimum of the objective function.

There have been numerous applications of simulated annealing in the petroleum industry. The major trouble in the application of this approach is that there is no obvious analogy for the temperature on a free parameter in the combinatorial problem (Begum, 2009). Also, implementation of this algorithm is too expensive for reservoir characterization. Petroleum reservoir models consist of a vast number of grid blocks, optimization of such large model parameter space requires many iterations and significant computational effort (Gomez et al., 2001).

2.1.5 Neighbourhood algorithm

The Neighbourhood Algorithm is a direct search algorithm in the same class of inversion techniques such as Simulated Annealing and Genetic Algorithms. These methods are similar in that they use randomized decisions in exploring the parameter space and require a particular forward model primarily. The algorithm generates history matched models by first randomly generating an initial set of models and then identifying the models with smallest data mismatch. Finally, new models are created by a random walk in the Voronoi cell of each selected model.

One of the advantages of the Neighborhood Algorithm is the simplicity of its two-parameter tuning scheme in contrast to the more complicated adjustment mechanisms of other methods, such as the cooling schedule required by Simulated Annealing (Sambridge, 1999). Christie et al., (2002) applied this approach to history matching and uncertainty quantification. Rather than using a single, lowest misfit model to make inferences about the system, this method retains multiple models that have small data mismatch. The idea is that even poor-fitting models contain information about the system. The application of this approach is although limited to a small number of parameters (Sambridge, 1999).

2.1.6 Ensemble Kalman Filter

In recent years, the research community has shown great interest for Ensemble Kalman filter (EnKF) as a data assimilation technique because of its simplicity and ability to adapt large-scale non-linear systems. As data becomes available, this method sequentially and continuously updates the reservoir model states (saturation, pressure, etc.) and parameters (permeability, porosity, etc.). This approach can be easily combined with any reservoir simulator. Evensen (1994) initially introduced this method for oceanic models. Lorentzen et al. (2001) used this approach for updating both dynamic variables and model parameters for a two-phase well flow model used in underbalanced drilling. Naevdal et al.(2003) showed the application of this method in history matching with encouraging results. It has also been applied to several synthetic cases for parameter estimation and production optimization (Brouwer et al. 2004, Wang et al. 2007). Haugen et al. (2006) presented a study for a North Sea field using EnKF for assimilating production data. Bianco et al. (2007) applied this method to a saturated oil reservoir and examined the influence of ensemble size on history matching results. Chitralekha et al. (2010) used production data in EnKF to characterize, and history match a 3D synthetic steam assisted gravity drainage (SAGD) reservoir. EnKF method for history matching mainly comprises of two steps, the forecast step, and the analysis step. In forecast step, all the realizations in the ensemble are forwarded using numerical simulation from current time step (t_k) to next time step (t_{k+1}). Then uncertainty in production forecast is assessed, and if uncertainty is high, then analysis step is performed in which unknown model

parameters are updated by assimilating production data available from the field at current time step. Again, production forecast of the ensemble with updated model parameters is obtained for the next time step, and uncertainty is measured. This process goes on till the uncertainty is reduced up to a level where realizations can be used further to develop field related strategies.

Differences between traditional history matching and EnKF

Traditional history matching follows the iterative approach to update the static parameters like porosity and permeability (Liang, 2007). Following are some of the characteristics of traditional history matching algorithms:

- Repeated flow simulations with all data when new data are available,
- Not fully automated,
- Not suitable for real-time reservoir model updating,
- Hard for uncertainty assessment.

On the other hand EnKF approach updates the reservoir model sequentially for both static parameters like porosity and permeability as well as dynamic parameters like pressure and saturations. Following are the characteristics of EnKF approach:

- Production data assimilated sequentially in time,
- Fully automated,
- Suitable for updating nonlinear reservoir simulation model on a large scale,
- The uncertainty of prediction is straightforward from the ensemble members.

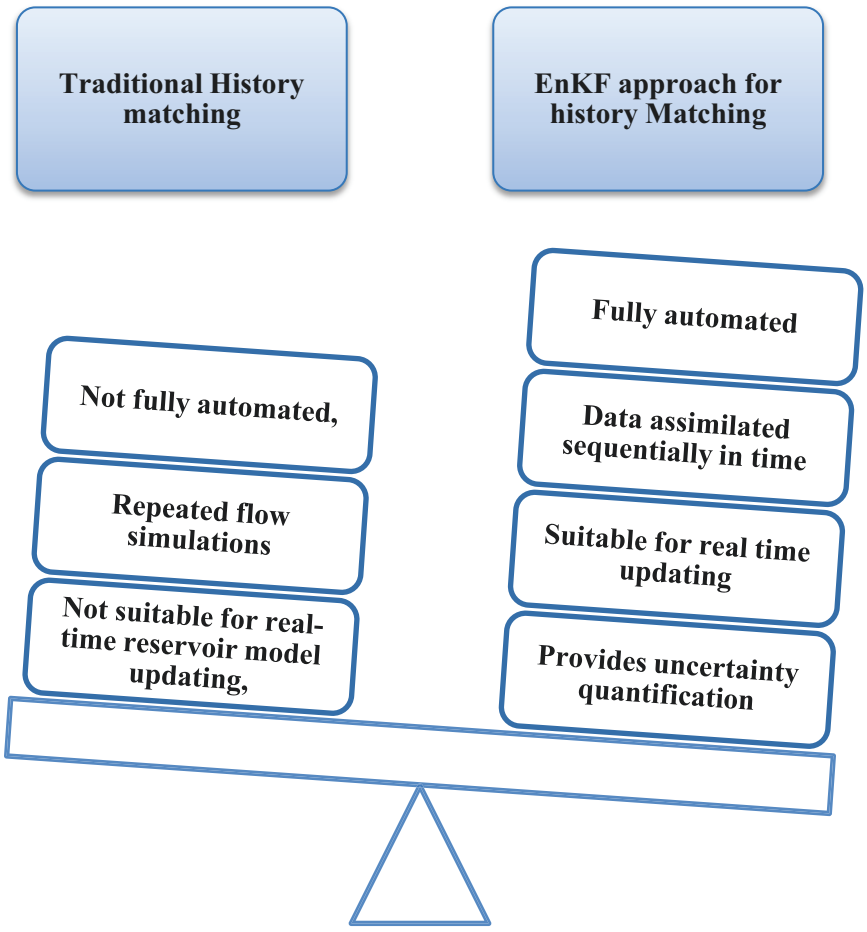


Figure 2: Traditional History Matching vs EnKF approach

Some of the important things to be considered while using EnKF approach for history matching are the determination of Ensemble size and maintaining the ensemble variability.

Ensemble Size

EnKF method relies greatly on the size and the features of the initial ensemble (Mitchell & Houtekamer, 2000). Ensemble should be selected such that it statistically represents the parameter space and characterize the uncertainties of the model. Small ensemble size does not represent model uncertainties' correctly, and the problem is under sampled (Nejadi, 2014). Many realizations in the ensemble lead to high computational cost and time during the assisted history matching process. The problem becomes more complicated with large scale thermal recovery process such as SAGD for which simulation of each ensemble member would take several hours. The ensembles of around 100 realizations are considered to be large enough for practical applications (Mitchell & Houtekamer, 2000).

Ensemble collapse (loss of ensemble variance)

EnKF depends on statistical measures of the ensemble during the analysis (update) step. Small ensemble size and sampling errors poorly estimate the cross-covariance matrix (Nejadi, 2014). Poor cross covariance approximation results in large Kalman gain values and unphysical updates resulting loss in ensemble variance after few assimilation steps. The updated realizations collapse toward a particular response which may not honor the reference statistics. This problem is known as ensemble collapse (Nejadi, 2014).

Further details about the EnKF algorithm steps are presented in chapter 4.

2.2 Proxy Modeling

Data assimilation and real-time updating of large-scale SAGD reservoirs is a challenging task due to high computational cost and time requirement associated with reservoir simulators. For a conventional reservoir of average size, hundreds or thousands of grid blocks can be used each one with different properties. In forecast step, each realization of the ensemble is evaluated using commercial reservoir simulators at every time step, resulting in high computational cost and time. Various approaches have been discussed recently in literature to reduce the time and number of simulations in the history matching process. One of the approach to reduce the computational cost and time is by reducing the ensemble size by application of various screening/ranking methods (Patel et al., 2015). Another solution to the huge computational cost has been the use of proxy models. Proxy models are referred as mathematically derived models that imitate the output of a simulation model for selected input parameters. In the literature, proxy models are often termed as response surface models or surrogate models (Zubarev 2009). In cases where proxy models can adequately represent relevant output parameters, they can be used as an adequate substitution for full reservoir simulations. Polynomial regression models, ordinary kriging models, artificial neural networks (ANN) and radial basis functions (RBF) are some of the commonly used proxy models for reservoir simulation.

Numerous active applications of these proxy models in uncertainty quantification, history matching, and optimization have been found in the literature (Li &

Friedmann, 2005, Junker et al. 2006, Slotte et al. 2008). Peng & Gupta (2004) presented different experimental design techniques in combination with kriging and polynomial regression models to predict initial hydrocarbon in place. Zangl et al. (2006) used a combination of neural network-based proxy model and genetic algorithms for the production optimization process. He et al. (2011) proposed Trajectory Piecewise Linearization (TPWL) for model reduction and used them with EnKF for history matching of artificial reservoirs. Li & Friedmann (2007) used thin-plate spline interpolation to generate objective function in history matching of two waterflooded reservoirs. Cullick et al. (2006) used a neural network based non-linear proxy model for the history matching process. Costa et al. (2014) demonstrated the application of Artificial Neural Network (ANN) in history matching of a synthetic water flooded reservoir.

Some applications of proxy models in case of the complex thermal recovery process as SAGD are also available in the literature. Akram (2011) considered a polynomial regression model using least square fit for optimization of a SAGD model. Fedutenko et al. (2014) suggested proxy models based on radial basis functions and observed their capability to forecast the outputs of SAGD operations for a given period. Queipo et al.(2002) proposed a proxy methodology for the optimization of well spacing and sub-cooling to maximize the production of a synthetic 2D SAGD reservoir model, using a method based on neural networks and a kriging surface. Vanegas et al. (2008) discussed another strategy for assessing optimal operating conditions in the SAGD process using the design of experiments and polynomial response surface approach. However, these

studies are based on synthetic reservoir models and overlooked the geological uncertainties and complexities of the real SAGD field. It can also be observed that proxy models used for history matching deal with few undiagnosed reservoir parameters due to their inability to incorporate a higher number of parameters. Hence, there is an opportunity of improved method, which can be used for proxy modeling of large-scale reservoirs and can easily be integrated with assisted history matching algorithms such as EnKF to reduce computational cost while maintaining the accuracy of results.

Zubarev (2009) has conducted a comparative study of proxy-modeling algorithms, their prognostic quality, and computational efficacy. He concluded that model complexity, dimension and the selection of the trial runs are the key dependencies for all proxy-modeling methods. Therefore, the ideal proxy selection is unique to the problem for examination and requires a detailed understanding of the advantages and limitations of the proxy model before selection. In almost all cases, some traditional reservoir simulation runs for development or validation purposes are needed. An initial analysis is done to reduce the number of input parameters required for the simulation. The run time and sophistication necessary to generate a proxy model is another significant aspect that differs among different types of models.

In the next section, some of the commonly applied proxy models in reservoir simulation are discussed. A very brief review of these proxy model types is provided. A detailed discussion of the models used in this work will be presented in subsequent chapters.

2.2.1 Polynomial Regression Models

Polynomial regression models is a term generally used to address response surface models (RSM). Because of their flexibility and computational efficiency, these models are widely been used in petroleum industry. A general form of a second order polynomial model can be represented as:

$$y(x) = c_o + \sum_{k=1}^{P} c_k x_k + \sum_{k=1}^{P} c_{kk} x_k^2 + \sum_{j<k} \sum_{k=1}^{P} c_{jk} x_j x_k \quad \ldots\ldots\ldots\ldots 2.3$$

Where P is the parameter space dimensionality, x_k are the input variables, $y(x)$ is the output based on input variables, c_o , c_k , c_{kk} and c_{jk} … are the coefficients for constant, linear, quadratic and cross terms of polynomial expression respectively. In this approach, coefficients are generally calculated using least squares fitting method. This method requires careful selection of training data and calculation of regression coefficients. It has an advantage of quick evaluation but is prone to over fitting (Fedutenko et al., 2013).It also depicts poor prediction accuracy for highly nonlinear multidimensional cases.

2.2.2 Kriging Models

Kriging is a geostatistical technique to quantify the spatial correlation in the data using a function of distance and direction between the two points. It assumes that points are spatially correlated to each other. The extent of correlation is expressed in terms of geostatistical parameters as covariance function and variograms, which are further used to determine the weight of each sample point for the estimation of new values at unsampled locations. This method has high prediction

accuracy compared to polynomial regression models (Fedutenko et al., 2013). However, for an experiment of L observations, it requires inversion of a (L+1) x (L+1) matrix, which is time-consuming and makes this approach difficult for a large value of L.

2.2.3 Thin-plate Spline Models

Thin plate spline is a multidimensional interpolation method for data within a parameter space. It is analogous to bending of a thin sheet of metal. If a particular set of data points is available, this approach creates a weighted combination of thin plate splines which are centered about each data point and provides the interpolation function that passes through the points with minimized bending energy. These models exactly reproduce the input data but require more experiments than the number of uncertainty parameters (Zubarev, 2009). It consists of radial basis functions that define a spatial mapping between two points in space. This method can be widely used as the non-rigid transformation model in image alignment and shape matching (Bookstein, 1989). However, it has been observed that splines already in one dimension can cause severe overshoots. In 2D such effects can be much more critical, which limits the application of this approach for complex multidimensional cases.

2.3 Parameterization techniques

In the case of large reservoir models, it is almost impractical to deal with the full description of a reservoir. History matching is an ill-posed problem which does not have a unique solution. The number of independent parameters required to describe the dynamics of fluid flow within the reservoir is more than the parameters that can be calibrated during assisted history matching process based on limited data and constraints. Thus, there is a need for excellent selection of how many and which parameters have to be calibrated at a particular time. Reduced-order modeling procedures entail the projection of a high-resolution description of a reservoir into a low-dimensional space, hence significantly reducing the number of unknowns and making the problem better posed. This process of dimension reduction is known as random field parameterization. Various approaches for random field parameterization have been published in the literature (Sudret & Kiureghian, 2000).The competency of a random field parameterization method rests on its ability to approximate the original random field accurately with a minimum number of unknowns. Following are some of the parameterization techniques discussed in literature:

2.3.1 Zonation

Zonation is the most basic and commonly used form of parameterization. In this method, the reservoir model is divided into individual zones larger than a single grid block and value of the particular parameter is kept constant within that zone. This method is mainly useful in cases where boundaries of zones are determined before the calibration process so that values of the reservoir properties in each

region are considered to be uncertain. The data misfit with zonation process is larger because of less number of degrees of freedom and inappropriate selection of zone boundaries (Oliver & Chen, 2011). Also, this method suffers from the discontinuity of reservoir properties at the zone boundary due to coarse zonation.

2.3.2 Pilot Point Method

Pilot point method is a parameterization technique in which initial realizations of model parameters are generated based on prior knowledge. Pilot points are selected at certain locations where the parameters are calibrated. The properties at other points are computed using the interpolation techniques and covariance of the model parameters. Initial realizations are conditioned to the hard data along with soft data such as seismic, geostatistical parameters through stochastic simulation. The concern of this approach is the selection of the location of the pilot points. Another practical issue is to specify the number of pilot points to be considered while using this method for parameterization. RamaRao et al., (1995) suggested locating the pilot points at most sensitive points to decrease the overall objective function. Still, the selection of the number of points is a grave concern in this approach. Various factors such as reservoir heterogeneity, the complexity of reservoir model, production mechanism and well pattern should be thoroughly investigated for selection of number of points (Nejadi, 2014). Points should be selected so that they can capture the adjustments made during the optimization process.

2.3.3 Discrete Cosine Transformation

Discrete Cosine Transform (DCT) is a Fourier-based transformation used to represent the spatial distribution of model parameters in terms of coefficients of independent basis functions. The basis functions are defined as cosine functions. This method is extensively used as image compression technique. In place of actual model parameters, the coefficients of retained cosine functions are used. Largest DCT coefficients are selected and maintained to reduce the number of model parameters and hence the size of parameter space. Jafarpour & McLaughlin (2008) have presented a history matching algorithm with a combination of DCT with EnKF, where, the DCT coefficients were involved in the state vector instead of actual model parameters. During History matching the parameter field is unknown; therefore it's hard to determine the related basis functions by ordering the coefficients. However, some prior information about the model parameter field is available but selecting the optimum number of DCT coefficients is still a concern in this approach.

2.4 Summary

A broad literature review on history matching techniques and large applications has been presented in this chapter. The option of using multiple models has benefits compared to unique history matched model. It can be observed from the limitations of different history matching algorithms that there is a requirement of a method which can include both static and dynamic parameters with measurements.

EnKF method seems to fulfill the requirements and has already been extensively reported in the petroleum industry. This approach avoids the computation of adjoint equations and sensitivity coefficients. Another key advantage of using EnKF method for history matching is its independence of any reservoir simulator. The sequential nature of EnKF helps in data assimilation in real time fashion. One of the concerns raised in the literature regarding this approach is the time-consuming simulation runs at each update step. In this work, we tried to overcome this issue by use of proxy models as a substitute of reservoir simulator in history matching process.

Chapter 3:

Proxy Modeling of Geological Reservoirs[1]

3.1 Introduction

Our approach in this work for proxy modeling is based on different techniques including Polynomial Chaos Expansion (PCE) introduced by Wiener (1938) and Artificial Neural Networks (ANN). The main idea is to represent the response of a model through a mathematical relation or network that can imitate the output similar to the commercial simulators. In PCE approach, the model is generated with the help of orthogonal polynomial basis functions. The polynomial chaos expansions are orthogonal with respect to the specific distribution of the input random variables. In other words, the dependency of the output parameter on all significant input parameters is estimated by a polynomial chaos-based proxy. The work of (Ghanem & Spanos, 2003), in which they approximated solutions of stochastic differential equations with truncated Hermite polynomials of Gaussian random variables, re-ignited the interest in the application of PCE for stochastic modeling. This approach offers a highly efficient way to include non-linear effects in stochastic modeling (Zhang & Lu, 2004, Fajraoui et al., 2011).

[1]based on a manuscript "Application of Polynomial Chaos Theory as an accurate and computationally efficient Proxy Model for Heterogeneous SAGD Reservoirs" submitted for publication in Journal of Petroleum Science and Engineering

PCE in simple form can be expressed as:

$Y(\xi) = \sum_{j=1}^{\infty} c_j \varphi_j(\xi)$ 3.1

Where c_j are deterministic coefficients and $\varphi_j(\xi)$ are jth order orthogonal functions of random variable ξ. There are numerous applications of PCE in petroleum engineering problems. (Ghanem & Spanos, 2003) used PCE for uncertainty quantification. Sarma, (2006) used PCE to efficiently quantify uncertainty for closed-loop production optimization. Xiu & Karniadakis (2003) used generalized polynomial chaos for modeling uncertainty in flow simulations. Zhang & Lu, (2004) also used PCE for modeling flow in random porous media. The applications and limitations of the PCE technique have been studied by Augustin et al., (2008). Oladyshkin & Nowak (2012) used PCE for risk assessment via robust design of subsurface flow. Babaei et al., (2015) used PCE for the optimization of water injection rate to maximize oil production from a synthetic 2D reservoir.

One of the main advantages of using PCE over other proxy models is that it converges systematically as the order of the expansion increases. This indicates that PCE does not have problems with overfitting which affects many other proxies. Cameron & Martin (1947) examined the convergence properties of PCE and stated that the convergence rate of PCE is exponential for Gaussian random variables. Another advantage of using PCE is that it is generally applicable to any type of input distribution. The PCE used in this work is based on Hermite polynomials specified for Gaussian distribution of input random variables.

However, for non-Gaussian random variables the convergence rate of PCE might be slow. To improve the convergence rate in such cases, optimal basis functions for PCE needs to be chosen. Not restricted to Gaussian distributions, the input random variables can be of any probability distribution such as uniform, gamma, and beta distributions. Input distributions determine the type of orthogonal polynomials to be used (Xiu & Karniadakis, 2003). For example, Legendre basis functions are recommended for uniform distribution, Hermite for Gaussian distribution and Jacobi basis functions for beta distribution. In reservoir models, the distribution of input random variables can be arbitrary. For such cases, the corresponding orthogonal polynomial expressions need to be constructed. Such a study is beyond the scope of this work.

The precision of the PCE proxy model is dependent on the set of basis polynomials chosen for the expansion and also on the precise calculation of the coefficients of the basis polynomials in the proxy model. Hence, the calculation of coefficients in the PCE proxy is a crucial task. Various methods can be used to determine the coefficients accurately, including the Galerkin projection scheme (Xiu & Karniadakis, 2003) and probabilistic collocation method (PCM) (Webster al., 1996) However, the former method requires access to the equations governing those coefficients. For complex problems, these equations are non-linear partial differential equations, and hence, the procedure may become very complex and computationally demanding (Li et al., 2009). On the other hand, the PCM approach used in this work uses a solution set generated from trial runs ("black box" approach) to compute the PCE coefficients. In this approach, coefficients are

obtained by solving a linear system of equations of size *M*, where *M* is the number of coefficients. However, the PCM approach can be unfavorable for a higher order of PCE as it requires more simulation runs since the number of collocation points exponentially increases with an increase in the dimensionality of the expansion. Isukapalli (1999) suggested the regression-based PCM approach as a modified collocation-based method in which he applied a least square technique to determine the coefficients, hence solving a regression problem in place of solving linear system equations. However, the regression problem will be strictly undetermined and non-unique if the number of simulations is not sufficient. To lessen this issue of dimensionality, another method as suggested by (Li et al, 2011) is to use pure terms of PCE and discard the cross terms. This will reduce the number of coefficients in PCE, but this approach will only work if the true model does not contain any nonlinear cross effects, which is practically not true for reservoir models. This approach also contravenes the primary advantage of PCE, which is guaranteed convergence as the order of PCE increases.

The main drawback of PCE is the exponential increase in the number of PCE terms as the order of PCE increases. This issue becomes stringent in the case of high dimensional problems. Although the different methods discussed above address this issue, still a large number of simulation runs is required to compute PCE coefficients. In this work, we tried to mitigate this issue by using the reduced terms PCE that retains only the relevant terms in PCE based on the spatial distribution of the input parameters. To represent a PCE, we need a strong understanding of the geological formation properties (permeability and porosity),

which will be considered as input parameters for the proxy model. These properties always have high spatial variability. In the case of large reservoir models, it is almost impractical to deal with the full description of a reservoir. Reduced-order modeling procedures entail the projection of a high-resolution description of a reservoir into a low-dimensional space, hence significantly reducing the number of unknowns. This process of dimension reduction is generally known as random field parameterization. Various approaches for random field parameterization have been published in the literature (Sudret & Der Kiureghian, 2000). The competency of a random field parameterization method rests on its ability to approximate the original random field accurately with a minimum number of unknowns. Among the various approaches, Karhunen-Loeve (KL) expansion is an optimal technique for global parameterization. This approach was studied in detail by (Ghanem & Spanos, 1991). (Romary, 2009) used a similar approach to reduce the dimension of the inference problem in the Monte Carlo Markov chains algorithm. The use of KL expansion as an efficient parameterization technique has also been discussed by (Reynolds et al., 1996). Zhang & Lu, (2004) developed a KL expansion-based moment equation (KLME) for uncertainty quantification. KL expansion is a very promising approach for representing stationary and non-stationary processes with clearly known covariance functions. It is a series expansion method for the representation of a random field. The expansion involves a complete set of deterministic functions (Huang et al., 2001). The deterministic functions are the eigenvalues of covariance function that decay steadily. However, when the eigenvalues of the

KL decomposition are not realistically fast decaying, the number of terms in PCE again increases, and hence, there is an increase in the number of simulation runs to compute the coefficients. The proposed work becomes less efficient when the random field cannot be efficiently expressed using KL decomposition.

In this work, we use KL expansion in combination with PCE and ANN proxy model to project the geological parameters (permeability and porosity) into a lower dimension in terms of Gaussian random variables. After parameterization of the random permeability field into random variables, the desired outputs (cumulative oil, steam to oil ratio and oil production rate) will be calculated using the proxy models. The computational requirement of the proxy in comparison to full numerical simulations is emphasized during the course of this study. Hence, we studied different cases by varying the number of random variables, thus changing the number of terms in PCE to get an accurate proxy representation with minimal use of simulators. We also compared the efficacy and accuracy of the proposed approach with ANN and RBF proxy models under the same geological conditions. The results of all the proxy models are compared with numerical simulator results.

As very few applications of proxy models are found for heterogeneous SAGD reservoirs in the literature, hence, in this work we tried to develop and analyze different proxy models for a real heterogeneous SAGD reservoir. This work has demonstrated that PCE as a proxy model outperforms other data-driven proxy model techniques such as ANN and RBF considering the same number of simulations used to train the proxy models. The main advantage of the proposed

approach is that it can easily substitute the need for commercial numerical simulators in order to run different realizations of uncertain input parameters. The proposed method also addressed some of the limitations mentioned above by incorporating the geological uncertainty and complexity of the SAGD reservoir. While the computational time and cost are greatly reduced compared to conventional simulation methods, this work will find prodigious applicability when uncertainty within a SAGD reservoir is a substantial limitation for decision-making processes such as history matching and production optimization tasks.

3.2 Methodologies

3.2.1 Karhunen-Loeve Expansion

In reservoir models, the fluid and rock properties at different grid locations are assumed to have some correlation to each other, and this correlation helps reduce the dimensions of the model without compromising the geological information. If the analytical expression for the correlation function is not known, the numerically derived covariance matrix can be used to perform KL transformation. The KL expansion is a promising approach for representing random fields (permeability in our case) with the help of a covariance matrix. It is a linear relation that de-correlates the random field while preserving the two-point statistics of the field (Bazargan 2014).

The correlation structure of the random field may be described by the covariance function. Since the covariance is bounded, symmetric and positive-definite, it may be decomposed as follows (Ghanem & Spanos, 1991):

$$C(x_1, x_2) = \sum_{i=1}^{\infty} \lambda_i \; f_i(x_1) f_i(x_2) \quad \text{..... 3.2}$$

Where λ_i are the eigenvalues and $f_i(x)$ are the deterministic eigen functions. Now, the random field $Y(x, \theta)$ with mean $\bar{Y}(x)$, can be expressed using KL expansion as:

$$Y(x, \theta) = \bar{Y}(x) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} \; f_i(x) \xi_i(\theta) \quad \text{..... 3.3}$$

Here, x is the spatial variable, $\xi_i(\theta)$ is a set of uncorrelated Gaussian random variables.

KL expansion can be approximated by sorting the eigenvalues λ and the corresponding eigen functions $f(x)$ in a descending order and truncating the expansion after *K* terms. *K* is the number of terms needed to reproduce the *Y* variability with a given accuracy. To truncate KL expansion after *K* terms, the smallest eigenvalues are discarded. The rate of decay in the eigenvalues depends on the correlation strength of the random field (Bazargan 2014). Eqn. (3.3) therefore provides an alternate way for generating the random field realizations. Once the eigenvalues and their corresponding eigen functions are determined, a

realization can be generated with a certain number of values ξi from the standard Gaussian distribution $N(0,1)$.

The discrete form of KL expansion is then given as (Sarma, 2006):

$$[Y]_{R,1} = [f(x)]_{R,K} [\sqrt{\lambda}]_{K,K} [\xi]_{K,1} + [\bar{Y}]_{H,1} \quad \text{3.4}$$

Here, $f(x)$ is the matrix of the eigenvectors corresponding to the K largest eigen functions of the covariance matrix, $\sqrt{\lambda}$ is a diagonal matrix consisting of the K largest square roots of eigenvalues, ξ is a vector of uncorrelated random variables with zero mean and unit variance, and $\bar{Y}$ is the mean value of output **Y**. In practice, $\boldsymbol{K}$ is much less than $\boldsymbol{R}$, where $\boldsymbol{R}$ is the total number of grid blocks in the problem. Thus by using KL transformation, Y is represented by a much smaller set of parameters ξ . The workflow for KL expansion used in this work is presented in Figure 3.

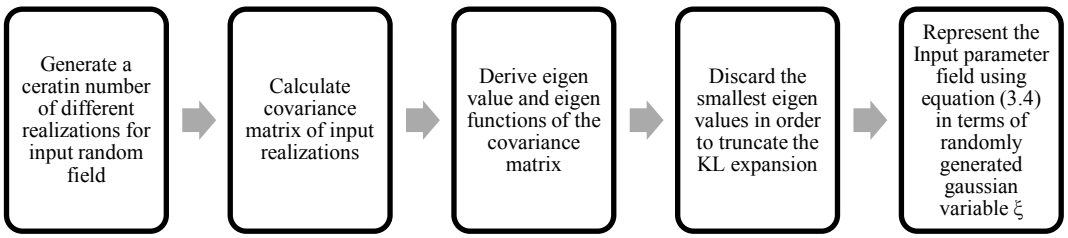


Figure 3: Workflow for dimension reduction using KL expansion

3.2.2 Polynomial Chaos Expansion (PCE)

PCE, introduced by Wiener (1938), can be used to effectively express the dependent (output) parameters. This approach is basically using bi-orthogonal polynomial functions of different orders to predict outputs for several values of input parameters. Outputs obtained from full reservoir simulation runs are used to train the PCE model and to get the best fit coefficients for the *N*th-order polynomials. Each output parameter of interest can be expressed as:

$$Y(\xi) = a_o + \sum_{i_1=1}^{\infty} a_{i_1}\, \Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} a_{i_1 i_2}\, \Gamma_2(\xi_{i_1}, \xi_{i_2}) \ldots\ldots\ldots\ldots\ldots \quad 3.5$$

Where a_o , a_{i_1} and $a_{i_1 i_2 \ldots\ldots i_d}$ are deterministic coefficients, and $\Gamma_d(\xi_{i_1}, \ldots . \xi_{i_d})$ are orthogonal polynomial chaos of order *d* with respect to the random variables $(\xi_{i_1}, \ldots . \xi_{i_d})$. The type of distribution of the input random variable determines the type of orthogonal polynomials to be used. Hermite polynomials form the best orthogonal basis for Gaussian random variables (Ghanem & Spanos, 1991). In this work, we assume that the random variables generated after dimensionality reduction are all standard Gaussian distributions, hence we use Hermite PCE to construct the proxy. In discrete form, PCE can be expressed as Eqn. (3.1), where, ξ is a vector of dimension *K* (same as in KL expansion). The total number of terms *P* in PCE is determined by the random dimensionality *K* and highest PCE order *d*,

$$P = \frac{(K+d)!}{K!d!} \quad \text{..... 3.6}$$

Eqn. (3.6) demonstrates that increasing the dimension of the random field (i.e., increasing the variables in vector ξ or increasing the order of PCE) will exponentially increase the number of terms in PCE. Greater numbers of terms require more simulation runs to determine the coefficients.

In this work, we use PCM to determine the unknown coefficients in PCE. This approach was initially introduced by (Webster et al., 1996). In the PCM approach, the simulator is treated as a black box. In this process, coefficients of PCE are calculated from the outputs by running the commercial simulator with some selected collocation realizations of input parameters. As input parameters have already been represented by KL expansion as the functions of standard random variables in vector ξ, the collocation realizations are constructed by some selected random variables. For a set of output parameters, the collocation realizations must equal the number of terms in PCE. We expressed the discrete form of PCE in Eqn. (3.1), which can be rewritten in the form of linear systems of equations as:

$$Y = C\,Z \quad \text{..... 3.7}$$

Where, Z is the matrix of dimension $P \times P$ and consists of Hermite polynomials evaluated at selected collocation point sets, i.e,

$$Z = \begin{bmatrix} \varphi_o(\xi_1) & \varphi_1(\xi_1) & \dots\dots & \varphi_{P-1}(\xi_1) \\ \varphi_o(\xi_2) & \varphi_1(\xi_2) & \dots\dots & \varphi_{P-1}(\xi_2) \\ & & \cdot & \\ & & \cdot & \\ & & \cdot & \\ \varphi_o(\xi_P) & \varphi_1(\xi_P) & \dots\dots & \varphi_{P-1}(\xi_P) \end{bmatrix} \quad \text{..... 3.8}$$

$$\text{and, } Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_P \end{bmatrix} \quad \text{.....3.9}$$

is the vector containing the P collocation outputs obtained by running the simulators with the corresponding collocation realizations derived on vector ξ . C is the vector of coefficients to be solved, which can be obtained by solving the above linear system of equations (Eqn. (3.7)). It is important to note that the sets of collocation points should be selected such that the matrix Z satisfies the condition of rank (Z)=P (i.e., the matrix should not be a singular matrix).

The performance of the PCM approach strongly depends on the choice of the collocation points. The method used for selecting the collocation points in this work is derived from the Gaussian quadrature technique to numerically solve the integrals. In the Gaussian quadrature technique, we can estimate the integral of a polynomial as a summation by using the roots of next higher order polynomial (Webster et al., 1996). Similarly, in the PCM approach, we use the roots of next higher order polynomial as the collocation points at which we will run the full

simulation. For example, if we use second-order PCE, then the collocation point sets are the combination of three roots (0,-$\sqrt{3}$, $\sqrt{3}$) of the third-order Hermite polynomial $\varphi_3(\xi) = \xi^3 - 3\xi$. We rank these three roots in order of decreasing probability. Because the random variables in vector ξ obey the standard Gaussian distribution, the root 0 is of the highest probability, hence, the first point set is always chosen as (0, 0, 0…0). The other point sets are selected by keeping as many of the variables of high probability as possible. The workflow of the PCE model development using the PCM approach is presented in Figure 4.

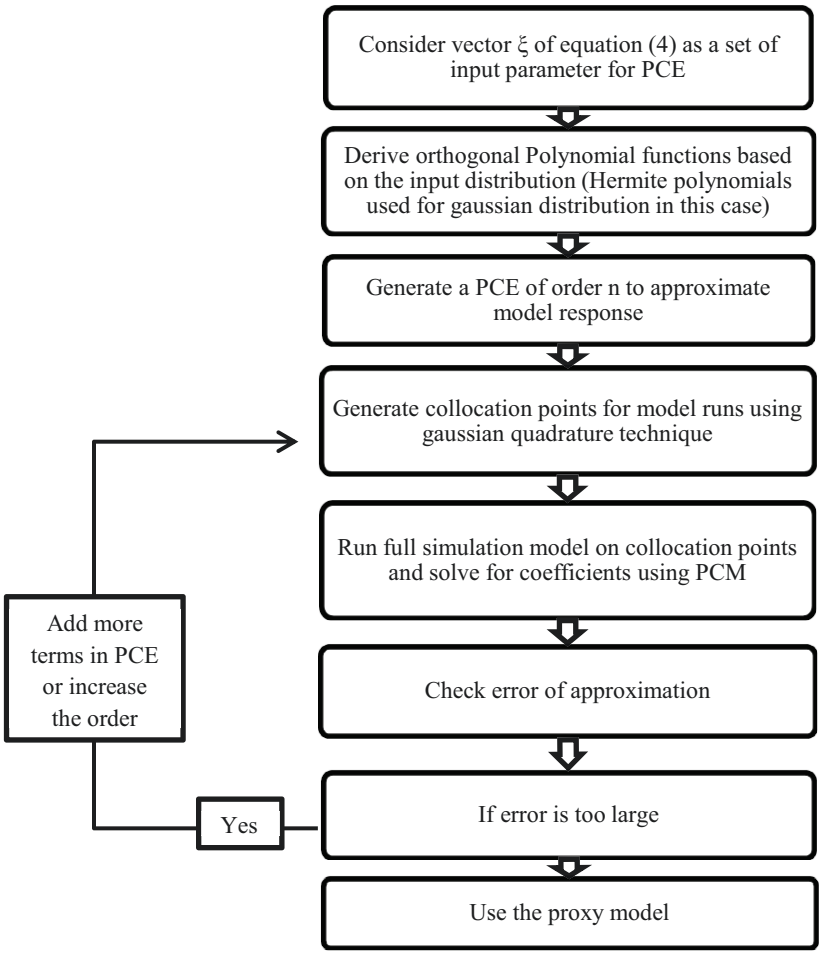


Figure 4: PCE proxy model workflow

3.2.3 Artificial Neural Network (ANN)

Artificial neural networks are similar to the concept of reflecting the reasoning of the human brain. It uses a dataset (training data) to learn the correlations and dependencies between the input and the output parameters. ANN can be a single or a multilayered network in which information runs from input to output through one or more hidden layers. ANN architecture consists of some base elements known as nodes, which are analogous to neurons in a biological system. Nodes of every layer are connected to other nodes in the contiguous layer. The nodes keep the information in the form of weights. The input to any node is modified according to the weight and then passed on to the next node. During the training phase of ANN architecture, the weights are adjusted. Training is completed when the network can predict the given target output (Amirian, 2014). After training and testing, the neural network can be used to compute output values for any given set of input values. In this work, the network created can be utilized as a proxy model to substitute the time consuming, full simulation model runs. A schematic process of an ANN with a single hidden layer is shown in Figure 5.

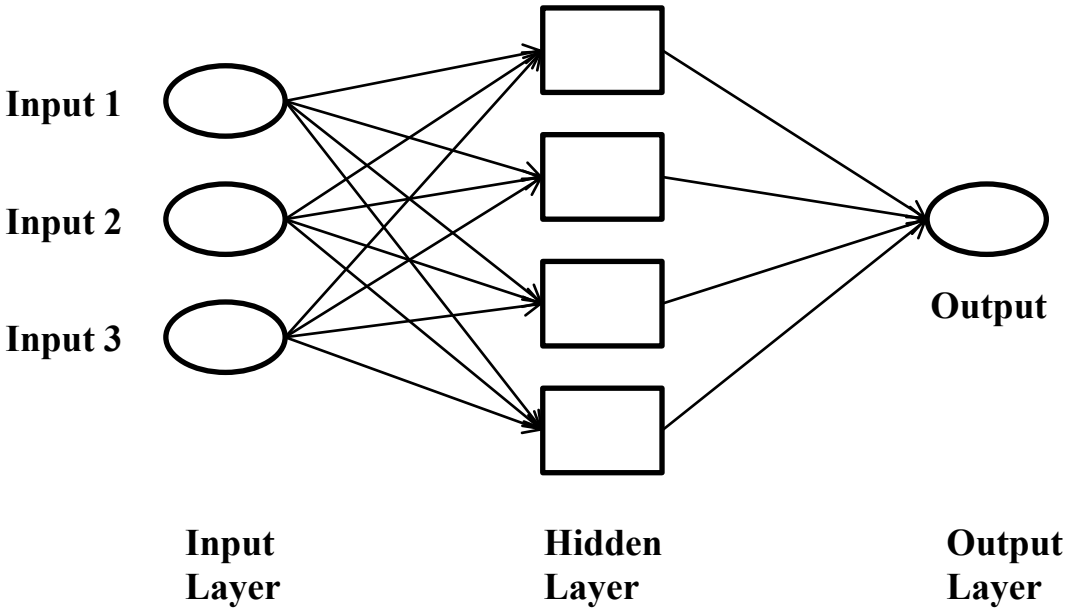


Figure 5: General schematic diagram for an ANN model

To Construct an ANN, it is essential to define the optimum number of hidden layers, the number of nodes per hidden layer and transfer function. The number of hidden layers and nodes influences the efficacy of the network to simulate different degrees of non-linearity (Ma et al., 2015). Neural network configurations are purely based on numerical observations. There is no information stored in the network, which determines the actual physics involved in the real process. If the value of the input exceeds the range that has been used to train the network, the output is not expected to be accurate. It is, therefore a crucial task from user's end to set up and train the neural network in a way that it can be used afterward. During the training of neural networks, the general practice is first to divide the data into two subsets: a training dataset and a validation dataset. The training dataset is used for computing the gradient and updating the network weights and

biases. In the training phase, the errors in the validation set are monitored. The validation error typically decreases during the initial phase of training. However, if the network begins to overfit the data, the error on the validation set usually starts to increase (Math Works 2014). The network is selected at the minimum validation error.

If the network is not satisfactorily accurate, it can be re-initialized and retrained. Every time the network is trained, the network parameters are different and might produce different solutions. A second approach for improving the accuracy of the neural network is to increase the number and size of hidden layers. Increasing the number of neurons enhances the flexibility of the network because the network will have more parameters to optimize. However, a balance need to be maintained between the accuracy and over fitting of data: using too few neurons can have under fitting problems, and the error cannot be minimized, while too many neurons in the hidden layer can lead to over-fitting of the network (Ferreira et al., 2012). The performance of the model can be further improved by using different training algorithms. Some training algorithms, such as Bayesian regularization, produce better generalization capabilities and adequately capture the non-linearity of the process (Amirian, 2014). Finally, if the accuracy of the model is still not acceptable, then additional training data should be used. The workflow for ANN model development is presented in Figure 6.

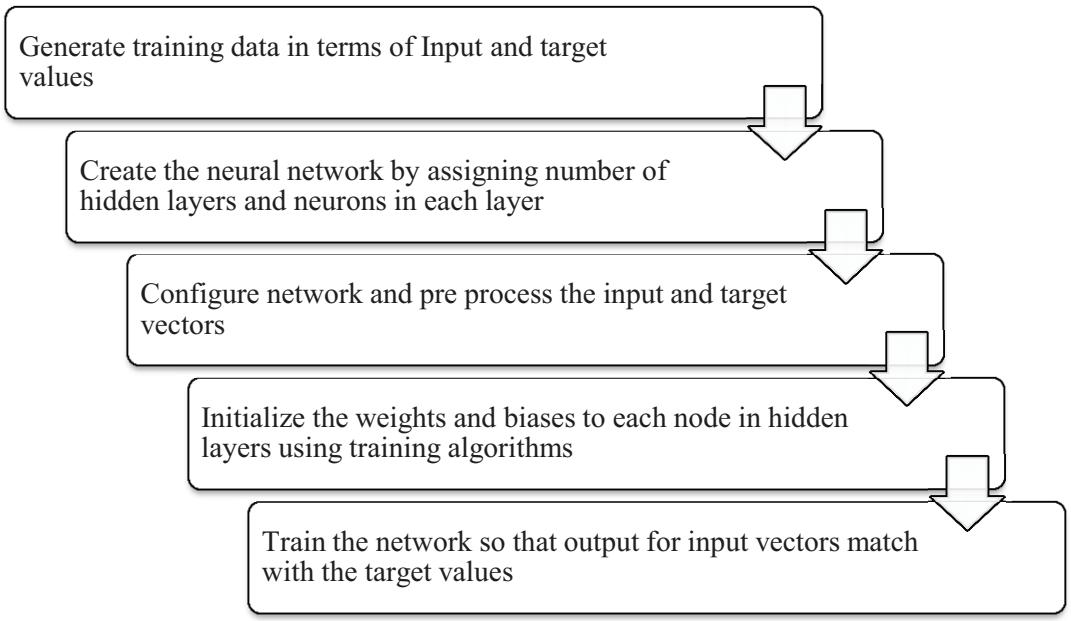


Figure 6: ANN proxy model workflow

3.2.4 Radial Basis Function (RBF)

The RBF network model is another type of neutral network model. As mentioned, ANN models require a proper understanding and large size of training data to effectively use different hidden layers. RBF, on the contrary, is a single layer neural network model, which can be used to efficiently predict the production parameters of reservoir models from a limited number of reservoir simulations.

Broomhead et al., (1988) has proposed a single-layer RBF network model, which has competencies similar to those of the multilayer ANN, but a quicker training period. The RBF network consists of an input layer, a single hidden layer and an output layer of linear weights. The main difference between the RBF network

model and the general feed-forward ANN model is that it uses a radial basis transfer function instead of a sigmoid transfer function.

The following relation describes the input-output mapping of the RBF network:

$$Y(x_k) = w_o + \sum_{i=1}^{M} w_i \phi(x_i, x_k) \quad \text{..... 3.10}$$

Where, $\phi_i(x_i, x_k)$ is the *i*th radial basis function that is calculated on the distance between the input parameter x_k and the center x_i. w_i represents the weights associated with the neurons.

RBF networks take a slightly different approach from the general ANN. In RBF models, the hidden nodes implement a set of radial basis functions (e.g. Gaussian functions). It is comparatively fast in learning and training the dynamics of the model. The weights and biases of each neuron in the hidden layer define the position and width of a radial basis function. Each linear output neuron forms a weighted sum of these radial basis functions. With the correct weight and bias values for each layer and enough hidden neurons, a radial basis network can fit any function and can be used as a substitute for simulations.

3.3 Reservoir Model: A SAGD Field Case Study

To demonstrate the efficacy of the proposed methods and to compare their results with numerical simulation outputs, we implemented all three methods (PCE, ANN, and RBF) to a real SAGD reservoir in northern Alberta. A general

description of the reservoir model is explained in Section 3.3.1. Stepwise implementation of the proposed frameworks is discussed in further subsections.

3.3.1 Description of the Reservoir Model

A 3D heterogeneous SAGD reservoir model was built using the data available from the field. Various types of well logs obtained from the vertical core holes at a particular location were used to build a static model in Petrel. A corner point grid was generated with a total of 20,000 grid blocks: 25 in the *I* direction, 50 in the *J* direction and 16 in the *K* direction. The dimensions of each grid block are 25 m × 2 m × 1.5 m in the *I*, *J* and *K* directions, respectively. Well logs were used to obtain the porosity of the grid blocks containing vertical core holes, and permeability was also calculated (Figure 7). Sequential Gaussian Simulation (SGS) was performed to generate 100 realizations of permeability using the data at the wells as conditioning data. Bitumen viscosity at the initial reservoir temperature (7^{o}C) was 625,000 cp, and at a higher temperature of (216◦C), it was 10 cp. A rock type with appropriate relative permeability curves was used in the model (details of the rock are not provided due to confidentiality). A horizontal well pair 500 m in length with 6 m spacing between the injector at the top and producer at the bottom was modeled. The different constraints of both wells from the field data were used in the simulation model. The permeability values ranged from 1525 md to 7150 md in the realizations. The porosity values ranged from 31.5% to 41.5%, while irreducible water saturation ranged from 0.16 to 0.2 in a 100-member initial ensemble. Realizations were simulated for 1,355 days using

the thermal simulator CMG STARS^{TM} (CMG 2013), and the results are compared with the outputs of the proposed proxy models.

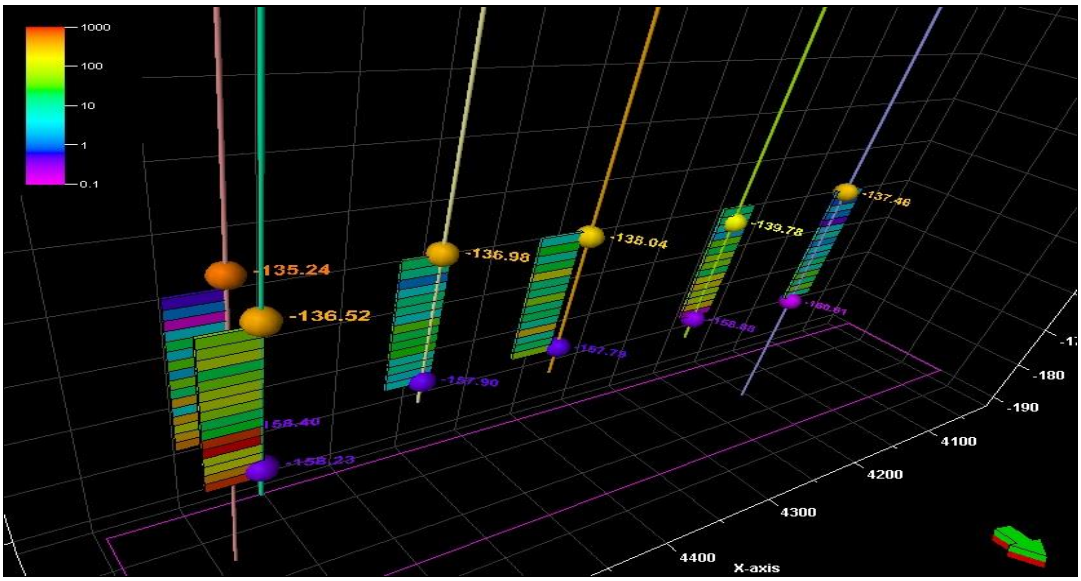


Figure 7: Permeability values of grid blocks containing core holes, which was used as conditional data in sequential Gaussian simulation to create permeability realizations

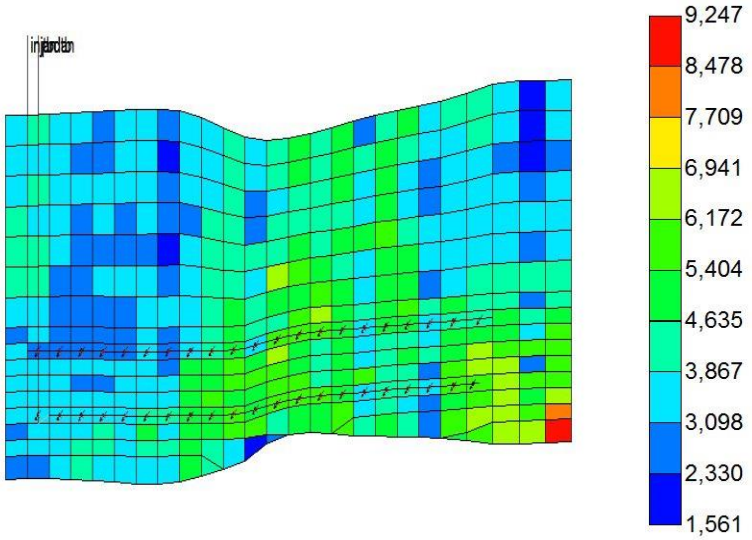


Figure 8: 2D view of the SAGD model

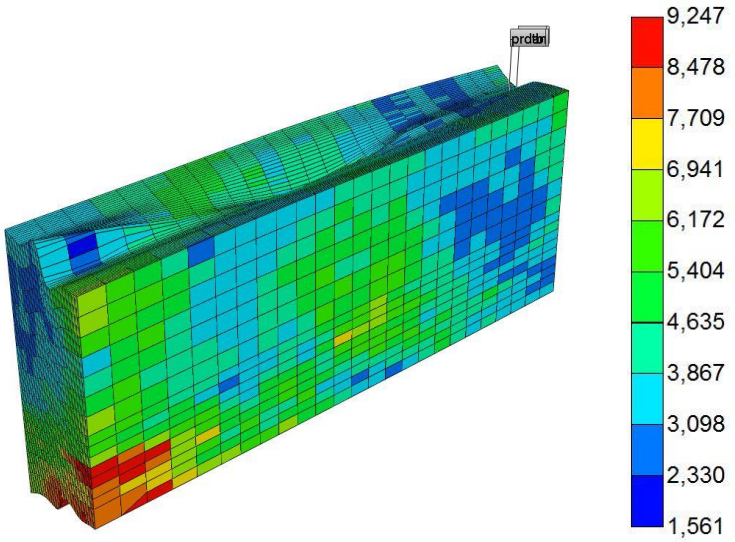


Figure 9: 3D view of the SAGD model

3.3.2 KL - PCE Approach for Proxy Modeling

The case study SAGD model is a heterogeneous model with different values of permeability for different grid blocks. Therefore, if we consider permeability as an input variable, then the number of input variables will be equal to the number of grid blocks, which are 20,000 in this case. It is almost impractical to deal with such a large number, hence, reduced order representation of the permeability field is required. To parameterize the permeability field and to reduce the number of input parameters, we used KL expansion. KL expansion (Eqn. (3.3)) allows the expression of a correlated random field or process in terms of a set of independent random variables while maintaining the covariance structure. A set of 100 random permeability realizations generated using SGS is considered here to obtain the covariance matrix. The KL workflow presented in Figure 3 is followed using these 100 realizations. The covariance structure is expressed in terms of eigenvalues and eigen functions (Eqn. (3.2)). The eigenvalues show a monotonically

decreasing trend as demonstrated in Figure 10. The sum of normalized eigenvalues shows the ratio of energy (variance) kept in the KL terms (Figure 11). Upon plotting the cumulative sum of eigenvalues, we observed that most of the energy is associated with a few initial eigenvalues. To truncate the KL expansion as per Eqn. (3.4), we discarded the small eigenvalues and selected only the initial eigenvalues. Discarding the smaller eigenvalues implies that we are discarding the shorter correlation lengths in the covariance structure. For adequate parameterization, the sum of the truncated normalized eigenvalues should be close to unity (Chang & Zhang, 2009). It is generally not suggested to keep more KL terms as that would increase the number of random variables representing the input field, which eventually increases the number of terms in PCE and requires more simulation runs for the training data. Figure 11 shows that more than 90% of energy is preserved with the first five to six eigenvalues.

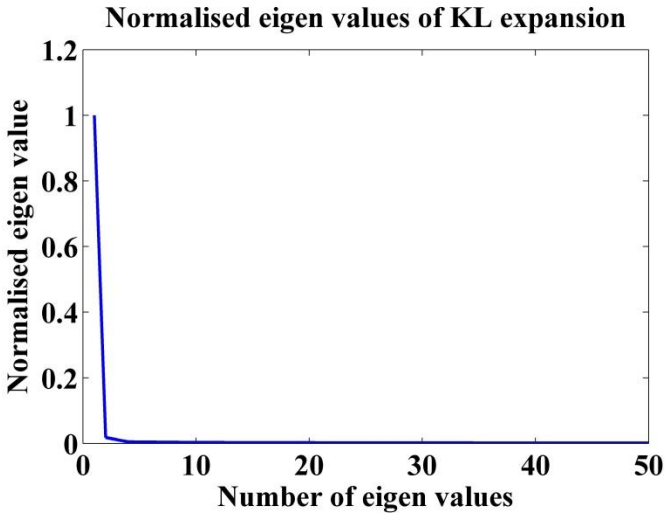


Figure 10: Decay in eigenvalues of KL expansion

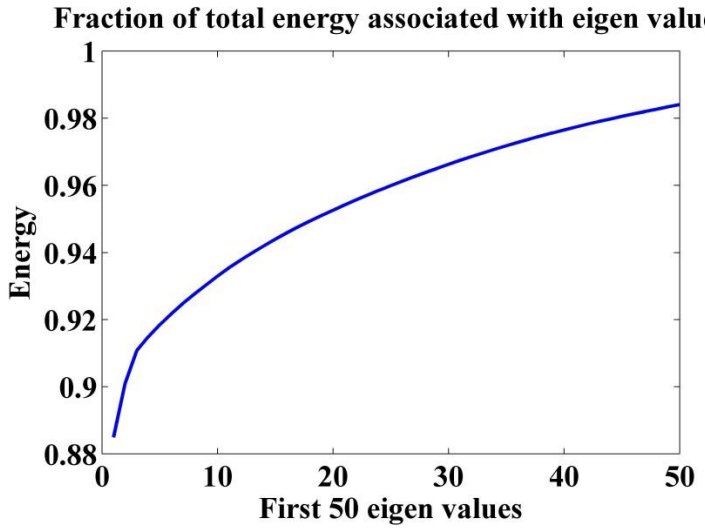


Figure 11: Energy retained in the first 50 eigenvalues of KL expansion

In this work, we consider two cases for the PCE proxy model. In PCE model case 1, we consider only the first three eigenvalues. Hence, 20,000 correlated permeability values were reduced to a vector of three independent random variables by KL expansion (Eqn. (3.4)). For the second degree of PCE approximation, Eqn. (3.6) can be used to calculate the total number of terms, which were 10 in this case. These terms are Hermite polynomial functions of different orders derived from independent Gaussian random variables. To solve Eqn. (3.7) to derive the coefficients, numerical simulation outputs are required at 10 collocation points. More specifically, the PCM algorithm explained in Section 3.2.2 requires 10 sets of collocation points for determining the unknown coefficients. The Gaussian quadrature technique is used to select collocation points (explained in Section 3.2.2). In PCE model case 2, we increased the

number of eigenvalues from three to five. Using Eqn. (3.6), the number of PCE terms becomes 21. This will increase the number of collocation points and simulation runs to 21. At this point, a comparison on computational effort can be made between the two cases. The computational effort to obtain output for each set of collocation points in the proposed approach is the same as solving for a realization in a numerical simulator. PCE case 1 needs only 10 simulation runs to train the proxy model, while 21 different simulation runs are required in case 2 for PCE proxy development. Once we obtained the collocation outputs by running the simulator with the corresponding parameter collocation realizations, we can solve the system of linear equations (Eqn. (3.7)) to obtain the unknown coefficients. For each output parameter of interest, we will get a different set of coefficients. This step of coefficient determination has to be done only once, and values of coefficients for every time interval can be calculated and tabulated for future use.

It is worth mentioning that each random variable vector corresponds to a high-resolution permeability field (Eqn. (3.4)). Hermite polynomials can easily be evaluated for a given set of random variables. Using the coefficients determined above, we can calculate the output results for any number of permeability realizations. As this step only needs a polynomial equation to generate results, it takes a fraction of seconds to give the output, which is a significant advantage over using the numerical simulator again and again for running thousands of realizations. We will later compare the results obtained from both PCE proxy model cases with the outputs of a numerical simulator to demonstrate the efficiency and accuracy of the PCE proxy model.

3.3.3 KL - ANN Approach for Proxy Modeling

In this work, the MATLAB neural network module is used to build the neural network architecture. The workflow presented in Figure 5 is strictly followed to obtain a neural network-based proxy model. To maintain consistency while comparing the efficacy of different proxy models presented in this work, we kept the training data the same for all models. For the ANN proxy, the input data vector is the set of permeability collocation realizations, which were used in PCE approach. The target output vector is obtained by running those realizations on a numerical simulator. The design of a neural network involves the selection of a number of neurons and hidden layers. As mentioned in Section 3.2.3, special attention is required in selecting the number of neurons. (Ferreira et al., 2012) used some thumb rules from the literature for the selection of a number of neurons. They suggested the number of neurons should be between the number of input parameters and output parameters. More precisely, it should be two-thirds the number of input parameters plus the output parameters and should not be more than twice the number of input parameters. In our work, different network configurations were analyzed and ideal network architecture was selected by comparing the prediction error from different configurations.

A three-layer feed forward neural network is used in this work for proxy model development. The first layer consists of neurons representing the input values of permeability in terms of random variables used in Eqn. (3.4). The second (hidden) layer consists of three neurons with sigmoid transfer functions, and the third layer contains one neuron representing the output value of the production parameter (oil

rate, cumulative oil production, and SOR). Pre-processing functions are used to normalize the input data vectors and target output vectors. The network is trained using the results of 10 simulation runs, which were used to develop the PCE model 1. The network uses a back propagation training algorithm that updates the weight and bias values for each neuron. The training algorithm used in this work minimizes a combination of squared errors and weights and then determines the correct combination to produce a network that generalizes well. The process used to train the network is called Bayesian regularization back propagation. During the training process, 80% of the data is used for training and 20% of the data is used for validation, which allows monitoring of the network's general performance and prevents overfitting of the training data. After the setup, training, and validation, the neural network is used to compute the output values for any number of permeability realizations represented by a set of random variables, and the results are compared with other proxy methods as well as with numerical simulation outputs.

3.3.4 KL - RBF Approach for Proxy Modeling

To develop a RBF-based proxy model, we used the same set of training data used for PCE model case 1 and ANN proxy development. The outputs of 10 collocation simulation runs are used to build the corresponding RBF network proxy model. The training data is stored in the form of input data vectors and target outputs (similar to ANN). To design radial basis networks, MATLAB inbuilt functions are used. The function takes matrices of input vectors and targets output vectors, and a constant value for the radial basis layer, and returns a

network with weights and biases such that the outputs match the target outputs for a given set of input data. This function creates, as many radial basis nodes as there are input vectors, resulting in a layer of radial basis nodes in which each node acts as a detector for a different input vector. If there are *K* input vectors, then there will be *K* nodes. The constant value assigned in a function determines the width of an area in the input space to which each node responds. The constant value should be large enough that nodes respond strongly to overlapping regions of the input space. The developed network is further used as a proxy model to simulate the production data for the given set of permeability realizations represented by the vector of random variables, and the results are compared with other proxy models and the numerical simulator.

3.4 Results and Discussions

This section describes the results obtained from the different proxy models. Our primary objective is to substitute the numerical simulator by a proxy model to eliminate the need to run a simulator every time, especially in cases where an output for hundreds or thousands of different input realizations is needed.

To compare the performance and efficacy of different proxy models, two forms of analyses were performed: quantitative and qualitative. A set of 100 realizations, different from those used for training the models, was used for comparison. These realizations were run on CMG STARS^{TM}, and the oil production rate, cumulative oil production, and cumulative steam to oil ratio were calculated. The same set of 100 realizations was tested on different proxy methods, and the results were

compared. All of the production parameters presented in this work are normalized by a target value of respective production parameters for confidentiality purpose.

This section is divided into two subsections. In Section 3.4.1, a quantitative analysis of the different proxy model is discussed, and in Section 3.4.2 a qualitative analysis using performance cross plots is discussed.

3.4.1 Quantitative Analysis

For the quantitative screening, statistical error analysis was used. The output production parameters at the end of 1,200 days were analyzed. The statistical parameters used for the assessment are mean, variance, minimum and maximum data values obtained from 100 experiments. The oil production rate, cumulative oil and cumulative SOR at the end of 1,200 days obtained from different proxy models were compared with results of the numerical simulator.

Table 1 shows the comparison of oil rate from different proxy models with the simulator results. A proxy model will be a good substitute for full simulation if the statistical parameters are the same as those obtained from the simulator. As shown in Table 1, the normalized mean value of the oil rate from the simulator is 0.5387. All the proxies' mean values are close to this value, showing a good estimate of oil rate. Along with mean, the variance is another important parameter which predicts the accuracy of the model. PCE model case 1 and the ANN model shows good agreement in terms of variance. The RBF proxy shows poor performance in terms of the mean as well as variance values. In addition, from Table 1, the minimum and maximum oil rate for different proxy models show their

efficacy to cover the whole range of results obtained from the simulator. PCE model case 1 and PCE model case 2 show similar values; therefore, it is evident that the PCE model correctly represents the original spread in terms of oil rate. Table 2 demonstrates the same statistical parameters for the prediction of cumulative oil production. In the case of mean and variance, all the proxy models except RBF are in agreement with the simulator results, but PCE model case 1 and PCE model case 2 outperforms when minimum and maximum values are considered. This again demonstrates the capability of the PCE model to cover the whole range of output parameters.

Cumulative Steam Oil Ratio (SOR), another important production parameter that depicts the efficiency of the SAGD process, was also studied. Results for SOR from different proxy models are tabulated in Table 3. All of the proxy models are in good agreement with the simulator results. The reason for the good prediction in terms of SOR is due to the small range of SOR output. This is evident from the variance value that is very small (0.0032). Therefore, one cannot observe many variations in the predicted data from the different proxy models. Based on the parameter values shown in Table 3, the PCE model demonstrates a comparatively better prediction than the other proxy models.

Table 1: Normalized Oil Production Rate after 1,200 days

Normalized Oil Production Rate after 1,200 days					
	Simulator	**PCE Case 1**	**PCE Case 2**	**ANN**	**RBF**
Mean	0.5387	0.5323	0.5482	0.5309	0.5797
Variance	0.0117	0.0114	0.0137	0.0115	0.0068
Minimum	0.1708	0.2116	0.2391	0.3124	0.3538
Maximum	0.7147	0.7156	0.7713	0.6627	0.7111

Table 2: Normalized Cumulative Oil Production after 1,200 days

Normalized Cumulative Oil Production after 1,200 days					
	Simulator	**PCE Case 1**	**PCE Case 2**	**ANN**	**RBF**
Mean	0.6233	0.6276	0.6310	0.6140	0.6462
Variance	0.0094	0.0106	0.0112	0.0155	0.0046
Minimum	0.2410	0.2338	0.2289	0.3190	0.4103
Maximum	0.7655	0.7677	0.7553	0.7424	0.7370

Table 3: Normalized Cumulative SOR after 1,200 days

Normalized Cumulative SOR after 1,200 days					
	Simulator	**PCE Case 1**	**PCE Case 2**	**ANN**	**RBF**
Mean	0.2553	0.2550	0.2516	0.2684	0.2480
Variance	0.0032	0.0039	0.0040	0.0055	0.0016
Minimum	0.2026	0.1951	0.1875	0.2029	0.2106
Maximum	0.6098	0.5236	0.5193	0.4737	0.4038

For a detailed quantification of the predictive quality of the proxy models, some other parameters were studied, including the correlation coefficient (Υ), average relative error ($\bar{\delta}$), maximum relative error (δ_{max}), standard deviation of relative error (σ_δ) and root mean square relative error (RMSE). A brief explanation of these parameters and their implications are as follows:

- Correlation coefficient (Υ) is termed as

$$\Upsilon = \frac{\sum_{i=1}^{n}(Y_i^s - \overline{Y^s})(Y_i^p - \overline{Y^p})}{\sqrt{\sum_{i=1}^{n}(Y_i^s - \overline{Y^s})^2}\sqrt{\sum_{i=1}^{n}(Y_i^p - \overline{Y^p})^2}} \quad \text{..... 3.11}$$

Where n is the number of experiments. $\overline{Y^s}$ and $\overline{Y^p}$ are the mean values of the simulated and predicted data, respectively. The higher the proxy quality, the closer the correlation coefficient will be 1.

- Average relative error ($\bar{\delta}$) is defined as

$$\bar{\delta} = \frac{1}{n}\sum_{i=1}^{n} \delta_i \quad \text{..... 3.12}$$

Where, n is the number of experiments, $\delta_i = \left|Y_i^s - Y_i^p\right| / Y_i^s$ is a relative error, Y_i^s is a simulated response from the full numeric model and Y_i^p is a predicted response from a proxy model.

- Maximum relative error is termed as

$$\delta_{max} = \max(\delta_i) \quad \text{..... 3.13}$$

- Standard deviation of the relative error is calculated as

$$\sigma_\delta = \sqrt{\frac{\sum_{i=1}^{n}(\delta_i - \bar{\delta})^2}{n-1}} \quad \text{..... 3.14}$$

- RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \delta_i{}^2}{n}} \quad \text{..... 3.15}$$

A proxy with a small value of error proves to be an efficient and accurate replica of a simulation model. All of the proxy models are compared with numerical simulator outputs. The output of the simulator is taken as a true simulated value and outputs from the proxy models as predicted values. The parameters

mentioned above are determined for oil rate, cumulative oil production, and SOR prediction.

Table 4 shows the values of the above parameters for oil rate prediction. The correlation coefficient is the measure of linearity between the simulated and predicted values. A value close to 1 demonstrates high prediction quality for a proxy model. Table 4 shows that the PCE and ANN models have good agreement with the simulated values with a correlation coefficient close to unity. The RBF model demonstrated the worse performance in this case with a correlation coefficient value of only 0.7592. This cannot be considered a good proxy model because, as a rule of thumb, the correlation coefficient needs to be above 0.85 for the proxy to be considered acceptable. Also, the average relative error in the PCE and ANN models is close to 5% as compared to 13% in the case of the RBF model. However, the PCE models outperform ANN when the maximum relative error is considered, showing its ability to imitate the responses at the desired level. Also, the PCE models have comparatively lower values of RMSE.

Table 4: Quantitative comparison of Oil Rate after 1,200 days

Quantitative comparison of Oil Rate after 1,200 days				
	PCE Case 1	**PCE Case 2**	**ANN**	**RBF**
Correlation Coefficient	0.9216	0.9541	0.9184	0.7592
Average relative error	0.0624	0.0555	0.0579	0.1302
Maximum relative error	0.2517	0.2073	0.8289	1.3066
Std. deviation of rel. error	0.0569	0.0552	0.0944	0.1998
RMSE	0.0842	0.0834	0.1103	0.2375

Similar analyses were performed for cumulative oil and SOR predictions. The results are shown in Table 5 and Table 6, respectively. In these cases, RBF tends to show better results in terms of a correlation coefficient higher than 0.85, but the

PCE models show better performance in predicting the output. It is evident from the data that PCE is superior to other proxy models in almost all cases with a correlation coefficient value close to unity and average relative error always less than 6%.

Table 5: Quantitative comparison of Cumulative Oil after 1,200 days

Quantitative comparison of Cumulative Oil after 1,200 days				
	PCE Case 1	**PCE Case 2**	**ANN**	**RBF**
Correlation Coefficient	0.9847	0.9902	0.9652	0.8745
Average relative error	0.0236	0.0166	0.0577	0.0637
Maximum relative error	0.1715	0.0948	0.3236	0.8145
Std. deviation of rel. error	0.0237	0.0165	0.0609	0.1219
RMSE	0.0334	0.0234	0.0836	0.1370

Table 6: Quantitative comparison of Cumulative SOR after 1,200 days

Quantitative comparison of Cumulative SOR after 1,200 days				
	PCE Case 1	**PCE Case 2**	**ANN**	**RBF**
Correlation Coefficient	0.9428	0.9492	0.8755	0.8617
Average relative error	0.0684	0.0603	0.0817	0.0417
Maximum relative error	0.1769	0.2040	0.2886	0.3606
Std. deviation of rel. error	0.0385	0.0350	0.0894	0.0618
RMSE	0.0784	0.0696	0.1207	0.0742

As per the results are shown in Tables 1 to 6, it has been verified that the PCE proxy model outperforms the other proxy methods and provides predictions with higher accuracy in terms of the highest correlation coefficient, lowest average relative error, the lowest value of the maximum relative error, the lowest value of standard deviation and lowest RMSE.

The two PCE model cases are different in terms of the number of full simulation runs required to develop the proxy model. PCE model case 1 takes only 10 realizations to run on a numerical simulator while PCE model case 2 requires 21

full simulation runs as training data. The performance of PCE model case 2 as evident from Tables 1 to 6 is better than PCE model case 1 in terms of statistical parameters and correlation coefficients. However, the error in prediction for both cases is almost the same range. Based on the results, it can be concluded that increasing the number of terms in PCE can improve the prediction performance, but not drastically in this case. Our primary motive is to develop a proxy model with the minimum computational requirement. In such cases, PCE model case 1 with fewer terms is recommended over PCE model case 2, which unnecessarily increases computational time in terms of more simulation runs.

3.4.2 Qualitative Analysis

For qualitative screening, performance cross plots were used. The cross plot is a graph of the predicted versus measured properties with a 45^{o} reference line to readily determine the proxy models' fitness and accuracy. A perfect proxy model would plot as a straight line with a slope of 45^{o}. Each point on the cross plot represents the output for one realization. We investigated the predicted outputs of 100 realizations from different proxy models against the measured simulation outputs. As we have considered outputs from the simulator as our true measured values, these values are plotted on the *x*-axis of the cross plot, and values obtained from the proxy models are plotted on the *y*-axis of the plot. Figure 12 show the cross plots between the measured oil rate and that predicted by different proxy models. It is evident from Figure 12 that the PCE and ANN models are in good agreement with the simulator results while the RBF model shows more variability

in the oil rate prediction. The low correlation coefficient value and high relative error of the RBF model in Table 4 support its corresponding plot in Figure 12. It can be seen that, for the ANN and RBF proxy models, certain points are very far away from the reference line, showing a mismatch in the measured and predicted value for some realizations. To improve the accuracy of these models, these realizations need to be added in the training data, which in turn increases the computational requirement. However, in the case of the PCE models, all points are near the reference line, which demonstrates that the training data used is sufficient for PCE model development.

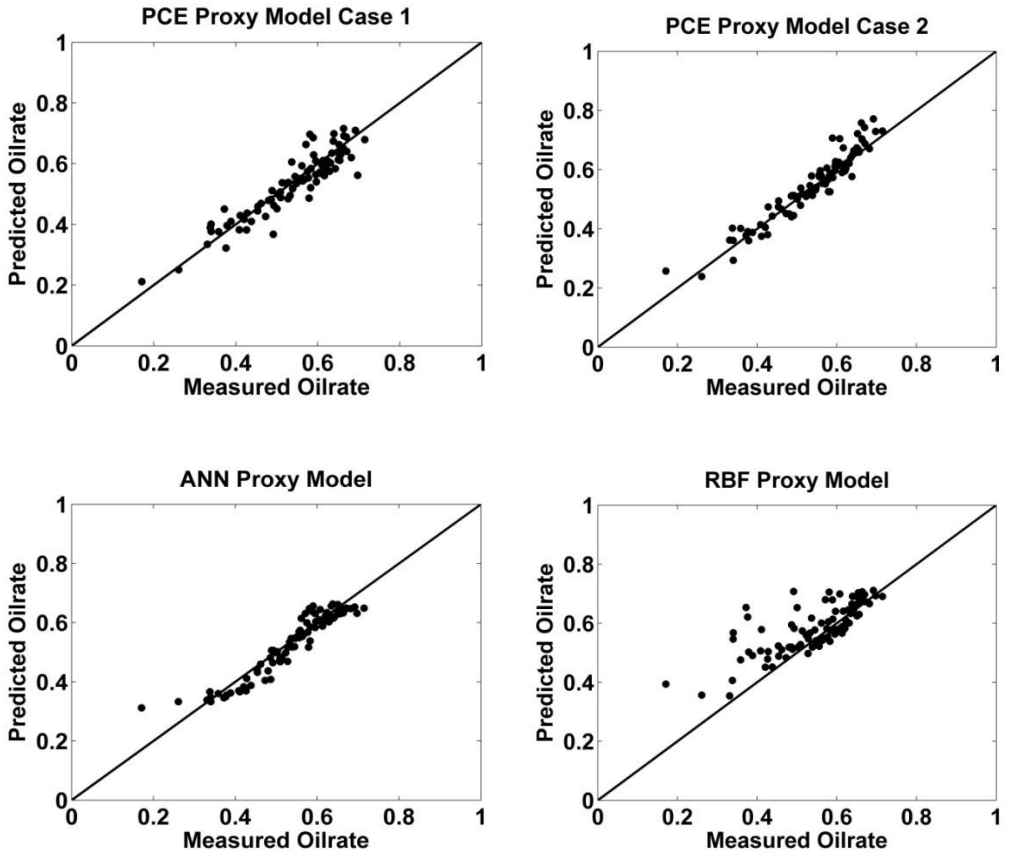


Figure 12: Cross plots for oil rate production at 1,200 days from PCE model case 1, PCE model case 2, ANN and RBF proxy models.

Similar cross plots were generated for cumulative oil prediction and cumulative SOR from different proxy models (i.e., Figure 13 and Figure 14, respectively). Compared to other proxy methods, the PCE model shows the tightest cloud of points around the 45^{o} line with very good clusters at all ranges of output values, indicating the excellent agreement between the measured and predicted data values.

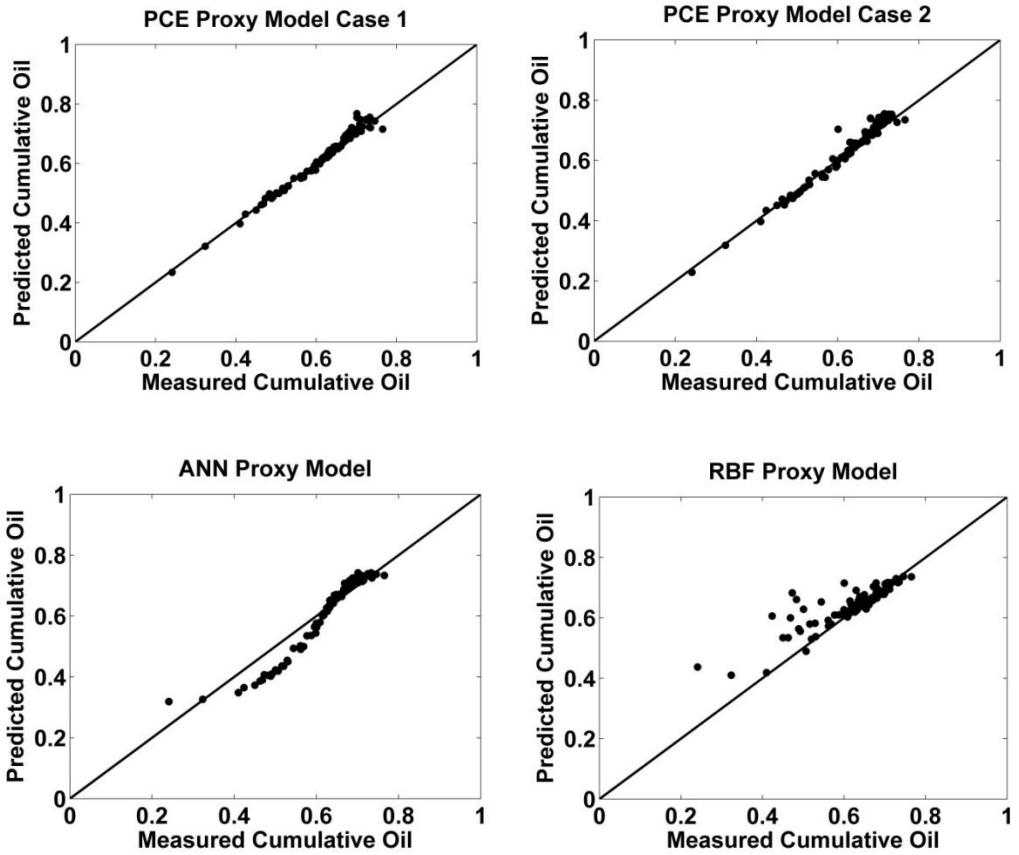


Figure 13: Cross plots for cumulative oil production at 1,200 days from PCE model case 1, PCE model case 2, ANN and RBF proxy model.

The cross plots show that RBF has the worst performance and is unable to predict the nonlinearity of the model. The ANN model shows better agreement compared to RBF. The PCE model outperforms the other proxy models in almost all cases as the data values are very close to the reference line.

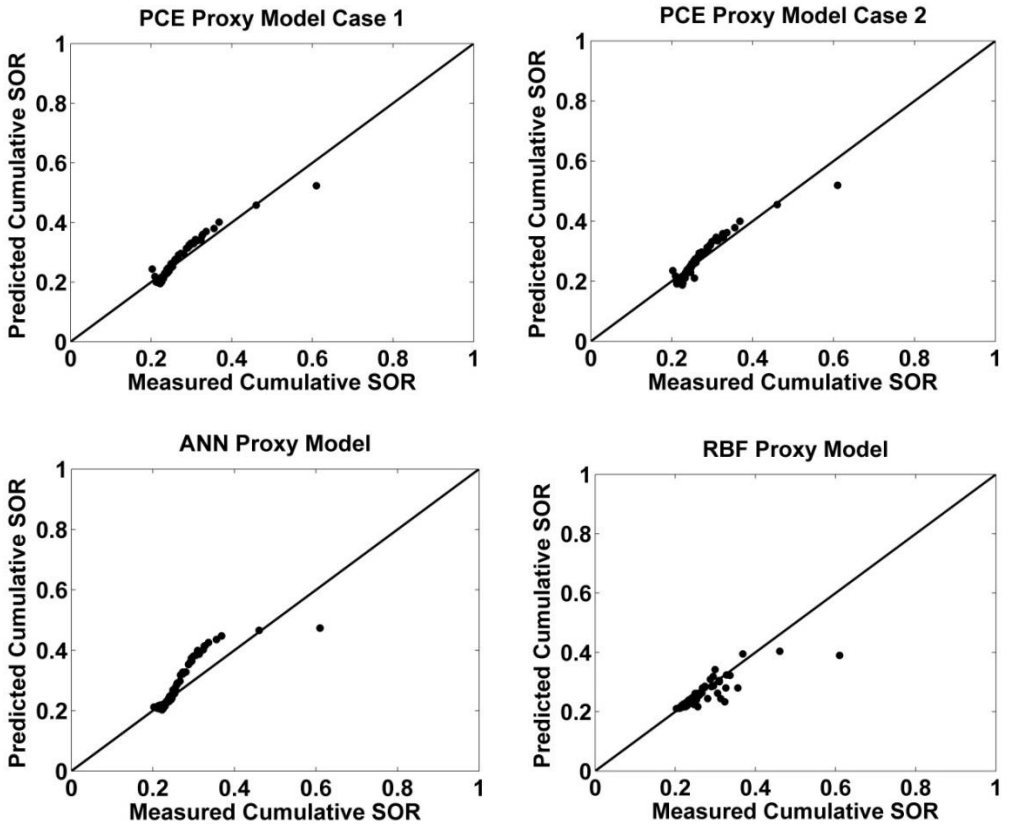


Figure 14: Cross plots for cumulative SOR at 1,200 days from PCE model case 1, PCE model case 2, ANN and RBF proxy models.

In Figure 15, we demonstrated the oil rate trend for the entire simulation period. Oil rate was calculated at a time interval of every 30 days until 1,355 days using the proposed proxy methods and was compared with the numerical simulation output. We obtained the oil rate from 100 different realizations of permeability

using different proxy models and opted to show the median, lowest and highest values of different realizations at each time interval. The black lines in Figure 15 show the median, lowest and highest oil rate value trend obtained from the different proxy models. These trends are superimposed on the spread of oil rate obtained from numerical simulation by running the same 100 realizations. The gray lines in Figure 15 show the output of the numerical simulator. An unbiased and accurate proxy model should cover the entire spread of the simulation results. However, it can clearly be seen that only the PCE proxy method shows good agreement with the oil rate trend obtained from the numerical simulator. ANN and RBF are seen to be a little biased and do not cover the whole range of oil rate predicted by the numerical simulator. One reason for the poor performance of ANN and RBF may be insufficient data on which these models were trained. To improve the accuracy, additional realizations must be added while training the ANN and RBF proxy models. At this point, the study indicates the PCE method is more efficient than ANN and RBF in capturing the whole dynamics of the process with limited training data.

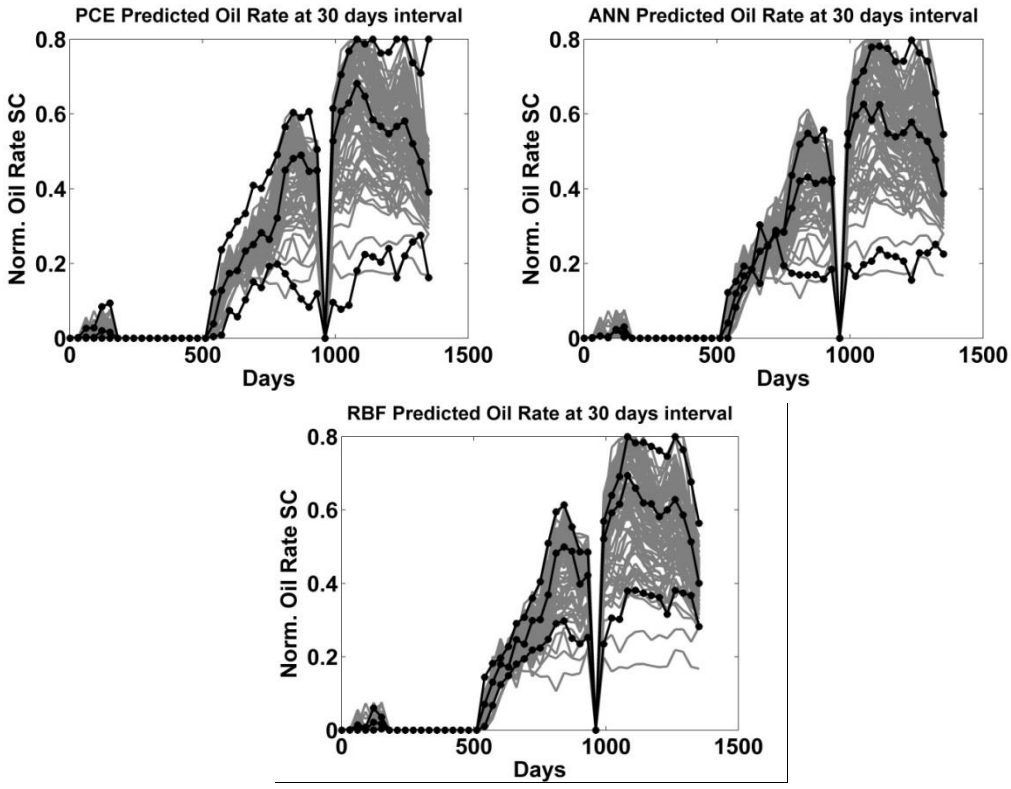


Figure 15: Oil rate trend predicted from different proxy models at a time interval of 30 days. The gray lines are the outputs from the numerical simulator. The black lines show the highest, median and lowest value trends of the PCE, ANN, and RBF proxy model.

3.5 Summary

In this chapter, we proposed a proxy model framework as an alternative to the numerical simulator. A real field SAGD model was used to demonstrate the applicability of the developed approach. The KL expansion was employed to parameterize the permeability field and to reduce the dimensionality of the model. Model parameters were represented by means of standard Gaussian random variables. Further, PCE was developed on these random variables to predict the output. The coefficients of PCE were obtained using PCM. Two other proxy models based on ANN and RBF were also studied in this work. The results of all

the proxy models were compared with those obtained from a numerical simulator. The following are the key findings of this work:

- Our proposed PCE approach, integrated with KL expansion, gives adequate results when compared with the outputs obtained from the numerical simulator. This can be used to run as many realizations as required in a relatively very small amount of time and with very low computational effort.
- The proposed PCE proxy model demonstrates better performance in comparison to the ANN and RBF-based proxy models. The PCE framework is easy to implement even in highly heterogeneous and large reservoir models. The PCE model developed with the help of very few simulation runs is efficient enough to represent the full dynamics and trend of simulation results, thus eliminating the need for time-consuming simulation runs.
- In this case, increasing the number of terms in PCE improves the prediction performance, but not significantly. To develop a proxy model with the minimum computational requirement, PCE with fewer terms is recommended over PCE with more terms. In this case, the latter will unnecessarily increase computational time in terms of more simulation runs.
- ANN-based models can show improved performance if more training data is used.

Chapter 4:

ANN Proxy Model Based EnKF framework for Assisted History matching[2]

4.1 Introduction

In recent years, the research community has shown great interest for Ensemble Kalman filter (EnKF) as a data assimilation technique because of its simplicity and ability to adapt large-scale non-linear systems. As data becomes available, this method sequentially and continuously updates the reservoir model states (saturation, pressure etc.) and parameters (permeability, porosity etc.). It is highly parallelizable and can be readily combined with any reservoir simulator. Evensen (1994) initially introduced this method for oceanic models. Lorentzen et al. (2001) used this method for updating both dynamic variables and model parameters for a two-phase well flow model used in underbalanced drilling. Naevdal et al.(2003) showed the application of this method in history matching with encouraging results. It has also been applied to several synthetic cases for parameter estimation and production optimization (Brouwer et al. 2004, Wang et al. 2007). Haugen et al. (2006) presented a study for a North Sea field using EnKF for assimilating production data.

[2]based on a manuscript "Application of ANN based Proxy Models for Efficient and Fast track Assisted History Matching of SAGD Reservoirs" Published in proceedings of World Heavy Oil Congress

Bianco et al. (2007) applied this method to a saturated oil reservoir and studied the influence of ensemble size on history matching results. Chitralekha et al. (2010) used production data in EnKF to characterize and history match a 3D synthetic steam assisted gravity drainage (SAGD) reservoir. EnKF method for history matching mainly comprises of two steps, the forecast step, and the analysis step. In forecast step, all the realizations in the ensemble are forwarded using numerical simulation from current time step (t_k) to next time step (t_{k+1}). Then uncertainty in production forecast is assessed and if uncertainty is high then analysis step is performed in which unknown model parameters are updated by assimilating production data available from the field at current time step. Again, production forecast of the ensemble with updated model parameters is obtained for the next time step and uncertainty is measured. This process goes on till the uncertainty is reduced up to a level where realizations can be used further to develop field related strategies.

For a conventional reservoir of average size, hundreds or thousands of grid blocks can be used each one with different properties. In forecast step, each realization of the ensemble is evaluated using commercial reservoir simulators at every time step, resulting in high computational cost and time. Various approaches have been discussed recently in the literature to reduce the time and number of simulations in the history matching process. One of the approaches to reduce the computational cost and time is by reducing the ensemble size by application of various screening/ranking methods (Patel et al., 2015). Another solution to the huge computational cost has been the use of proxy models. Proxy models are

referred as mathematically derived models that imitate the output of a simulation model for selected input parameters. In the literature, proxy models are often termed as response surface models or surrogate models (Zubarev, 2009). In cases where proxy models can adequately represent relevant output parameters, they can be used as a substitution for full reservoir simulations. Polynomial regression models, ordinary kriging models, artificial neural networks (ANN) and radial basis functions (RBF) are some of the commonly used proxy models for reservoir simulation

In this work, data-driven proxy models, which can substitute the need of using a reservoir simulator during the assisted history matching process, are developed. Permeability realizations of a large SAGD reservoir, located in northern Alberta, are first parameterized using Karhunen-Loeve (KL) series expansion and represented in the form of uncorrelated random variables. KL expansion represents the initial ensemble in the form of uncorrelated random variables using Eigen decomposition of the covariance function. Artificial Neural Network (ANN) based proxy models are developed using the random variables obtained from the KL expansion as input parameters and predict production parameters as outputs. Established models are further integrated into the EnKF framework to predict production parameters in forecast step while updating the random variables instead of permeability of each grid block in the analysis step of EnKF. Computational requirement of the proxy models during the development as well as deployment as compared to the commercial reservoir simulator is emphasized

in this study. The Primary aim is to perform assisted history matching of SAGD reservoirs in quick, low-cost manner while maintaining the accuracy of results.

4.2 Ensemble Kalman Filter

EnKF is a Monte Carlo method, which is widely used for assisted history matching due to its ability to include available observations sequentially in time. Aanonsen et al., (2009) reviewed the application of EnKF in reservoir engineering for estimation of reservoir parameters. In EnKF procedure an ensemble of model states is used to estimate the covariance matrices which are further used in the model updating process. Model parameters are the reservoir properties such as porosity and permeability. The Initial ensemble is generated based on the prior knowledge of the reservoir derived from various sources as well logs, core, and seismic analysis. In general, simulation techniques such as Sequential Gaussian Simulation (SGS) and Sequential Indicator Simulation (SIS) are used to generate multiple realizations. These realizations are consistent with the initial state of the reservoir. In the next step, all the reservoir models in the ensemble are forwarded typically using the numerical reservoir simulation. Mean and covariance of predicted model states are calculated and used in turn to calculate Kalman gain. Next, in the update (or analysis) step, each geological realization of the ensemble is updated using the Kalman gain.

The primary procedure for EnKF contains two parts. In the forecast step, the forecast model is applied to the ensemble using the commercial reservoir simulator. Each realization is forwarded in time using a transition function(f),

which is based on the solution of the dynamical equations for flow and transport in the reservoir. In mathematical form, it can be expressed as:

$$\begin{bmatrix} s_k^j \\ y_k^j \end{bmatrix} = f \begin{bmatrix} m_{k-1}^j \\ s_{k-1}^j \end{bmatrix} \quad \forall\ j\ \in\ [1\ , N_e] \text{.....} \quad 4.1$$

where s, y and m represent reservoir state parameters, production data, and model parameters respectively while superscript j refers to a particular realization in the initial ensemble and subscript k stands for the current time step, $\mathrm{k}-1$ indicates previous time step. N_e represents the number of independent ensemble members.

The state vector is generated for each realization in the ensemble that contains all the uncertain and dynamic variables which define the state of the system. These variables can be categorized into three parts and can be expressed as:

$$\varphi_k^j = \begin{bmatrix} m_k^j \\ s_k^j \\ y_k^j \end{bmatrix} \quad \forall j \in [1, N_e] \text{.....} \quad 4.2$$

The static parameters (reservoir properties such as permeability and porosity) are represented by m. Initial realizations of m are generated using prior knowledge of the reservoir. Dynamic parameters (such as saturation and pressure) are denoted by s, which are assumed to be known initially. y represents the production data which can be oil rate, water rate, steam to oil ratio, etc. Production data is kept in state vector for the purpose of estimating the correlation between the state vector

and observed data (Chen et al. 2009). State vector of all the realizations is combined in the form of a matrix that can be written as:

$$\alpha_k = [\, \varphi_k^1,\ \varphi_k^2, \dots, \varphi_k^j, \dots, \varphi_k^{N_e}\,] \quad \text{..... 4.3}$$

Ensemble mean of the prediction ensemble is calculated using equation,

$$\overline{\alpha_k} = \left[\overline{\varphi_k^1}, \overline{\varphi_k^2}, \dots, \overline{\varphi_k^j}, \dots, \overline{\varphi_k^{N_e}}\right] = \alpha_k w_k \quad \text{..... 4.4}$$

where w_k is the $N_e \times N_e$ matrix having each of its element as $1/N_e$.

In next step, ensemble perturbation matrix $\Delta\alpha_k$ is calculated by subtracting ensemble mean from state vector of each realization and is represented as,

$$\Delta\alpha_k = \alpha_k - \overline{\alpha_k} \quad \text{..... 4.5}$$

The ensemble covariance matrix can then be expressed as

$$C_{z_k}^p = \frac{\Delta\alpha_k (\Delta\alpha_k)^T}{N_e - 1} \quad \text{..... 4.6}$$

where superscript p represents predicted state of the matrix.

Further, a random Gaussian noise $v_k \sim N(0, R)$, generally known as measurement error is added to original observation y_k^{obs} at time step k,

$$y_k^{j,obs} = y_k^{obs} + v_k^j \quad \forall j \in [1, N_e] \quad \text{..... 4.7}$$

where v_k follows a zero mean Gaussian distribution with measurement noise covariance matrix R and $\mathrm{E}\left[\mathrm{v_k v_k^T}\right] = \mathrm{R}$ assuming that R is known. The addition of random variables as noise is a necessity to maintain variance of updated ensemble (Burgers et al. 1998). In update/analysis step, Kalman gain K_g is calculated using ensemble covariance matrix as follows,

$$K_g = C_{Z_k}^p H^T \left(H C_{Z_k}^p H^T + R\right)^{-1} \quad \text{4.8}$$

H is a measurement operator having only 0 and 1 as its elements. It relates state vectors to theoretical observations. It can be represented as,

$$H = [0 \mid I] \quad \text{4.9}$$

Kalman gain calculated in Eq. (4.8) is finally used to update state vector of each realization in predicted ensemble. Mathematically, this step can be shown as,

$$z_k^{j,a} = z_k^{j,p} + K_g\left(d_k^{j,obs} - H z_k^{j,p}\right) \quad \forall j \in [1, N_e] \quad \text{4.10}$$

In Eq. (4.10) superscript a denotes updated/analyzed state vector. The process is repeated till the uncertainty in production forecast of updated state vector is acceptable. A workflow for history matching using EnKF is shown in Figure 16.

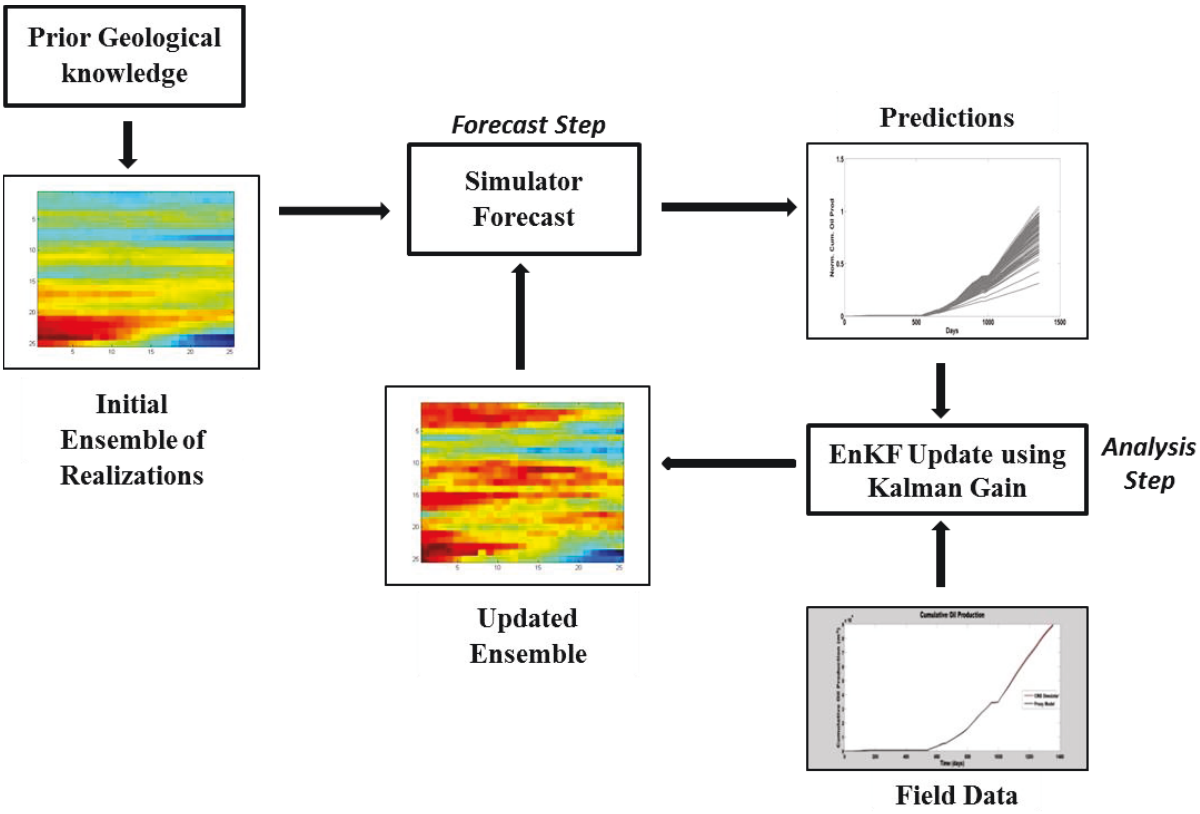


Figure 16: Traditional EnKF Workflow for History Matching

4.3 Integration of ANN based Proxy model in EnKF Framework

After the development of a comparable proxy model based on ANN (discussed in Chapter 3), the next task is to combine it with existing EnKF framework for assisted history matching. Permeability realizations are represented in the form of eigenvalues and corresponding eigen functions along with random variables using KL expansion as described in section 3.2.1. Once the eigenvalues and their corresponding eigen functions are determined after truncating terms in KL expansion, a realization can be generated with a certain number of values ξ_i from the standard Gaussian distribution N(0,1) using Eqn. (3.4). Each vector of the

random variable represents a permeability realization. These random variable vectors are used as input parameters for development of the proxy models. A detailed explanation for the proxy model generation has already been presented in Chapter 3. A brief description of proxy model will be provided in the further section using a real field case study of SAGD reservoir. The proxy models are then used in forecast step of EnKF.

EnKF workflow starts from the construction of a state vector shown in Eqn. (4.2). In KL-ANN-EnKF approach, state vector consists of random variables as model parameters representing the permeability field along with production parameters. In forecast step, reservoir simulator is replaced by the proposed ANN proxy model and production parameters are estimated considering random variables as input parameters. Output spread is analyzed, and state vector is updated using true production data from the field and Kalman gain in analysis step of EnKF loop. The random variables from updated state vector are again used to calculate output parameters and process is repeated until the uncertainty in production forecast of updated state vector is acceptable. Permeability realizations at any time step can be generated by substitution of last updated random variables into Eqn. (3.4). The integrated ANN-EnKF workflow is presented in Figure 17.

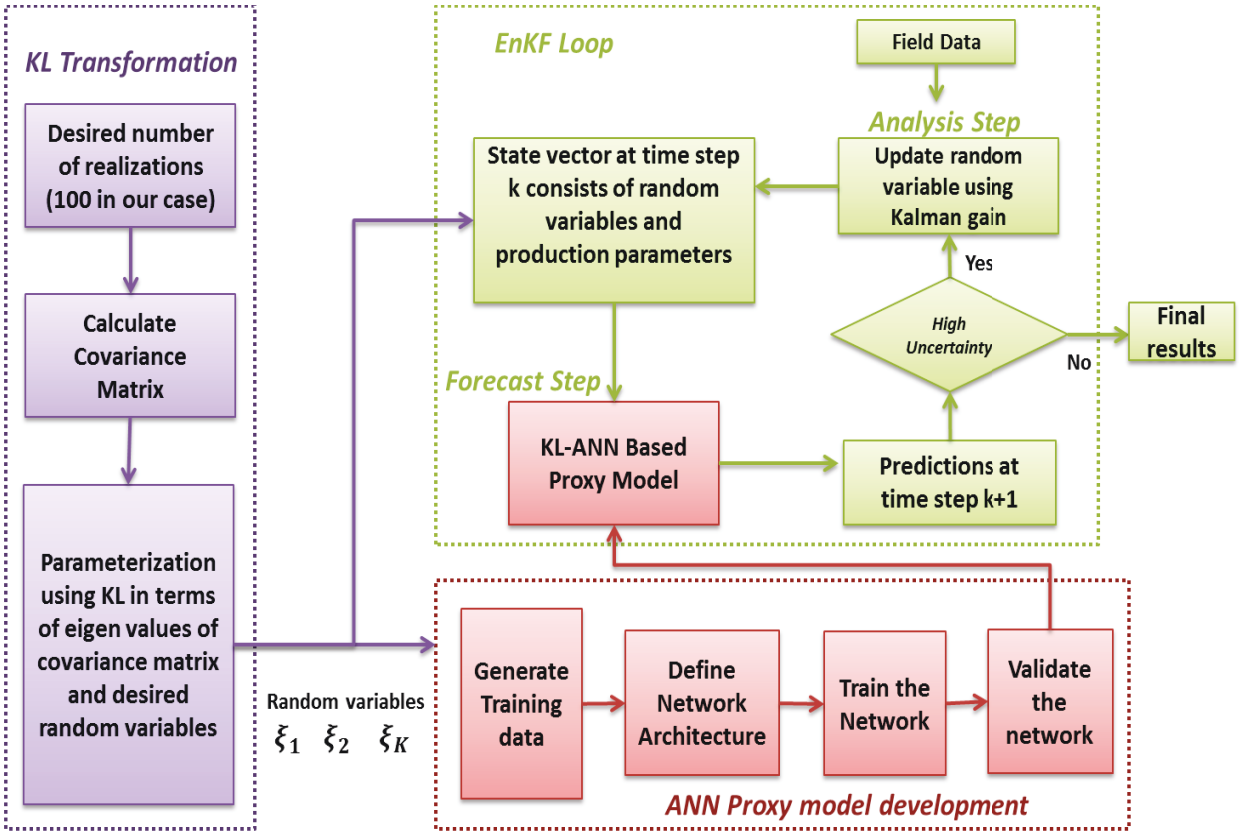


Figure 17: Proposed Integrated ANN-EnKF Workflow

4.4 Application to SAGD Reservoir: Field Case Study

In a view, to demonstrate the efficacy of the proposed method and compare the results with numerical simulation outputs, KL-ANN based proxy model in combination with EnKF framework is implemented on a large-scale SAGD reservoir in northern Alberta. A code is developed using MATLAB® (release R2014a) that integrates CMG STARS™ (CMG 2013a), CMG Results Report™ (CMG 2013b), and ANN based proxy model. A general description of the reservoir model is explained in Section 4.4.1. Stepwise implementation of the proposed frameworks is discussed in further subsections.

4.4.1 Description of the Reservoir Model

A single horizontal well pair, 500 m in length with 6 m spacing between the injector at the top and producer at the bottom, from a field in northern Alberta is considered to build a 3D heterogeneous SAGD reservoir model. Several well logs obtained from the vertical observation wells are imported in Petrel exploration and production software provided by Schlumberger to get information about formation top and depths. A corner point grid is generated with a total of 20,000 grid blocks. The model comprises of 25 grid blocks in the I direction, 50 in the J direction and 16 in the K direction. Each grid block has a dimension of 25 m × 2 m × 1.5 m in the I, J and K directions, respectively. Well logs are used to obtain the porosity of the grid blocks, and permeability is also calculated at those locations. Sequential Gaussian Simulation (SGS) is performed to generate 100 realizations of permeability using the data at the wells as conditioning data. Permeability values range from 1525 md to 7150 md. The porosity values range from 31.5% to 41.5%, while irreducible water saturation ranges from 0.16 to 0.2 within the initially generated 100 realizations. The static model is further imported in CMG BuilderTM (CMG 2013c) to assign dynamic properties. A 2D and 3D view of SAGD reservoir model are shown in Figure 18 and Figure 19 respectively. Initial reservoir temperature is 7^{o}C and bitumen viscosity at the original reservoir temperature is assigned as 625,000 cp. At a higher temperature of 216◦C, bitumen viscosity is 10 cp. A rock type with appropriate relative permeability curves is used in the model (details are not provided due to confidentiality). Injector and producer well-operating constraints are applied as

per the historical field data. Realizations are simulated for 1,355 days using the thermal simulator CMG STARS^{TM} (CMG 2013a).

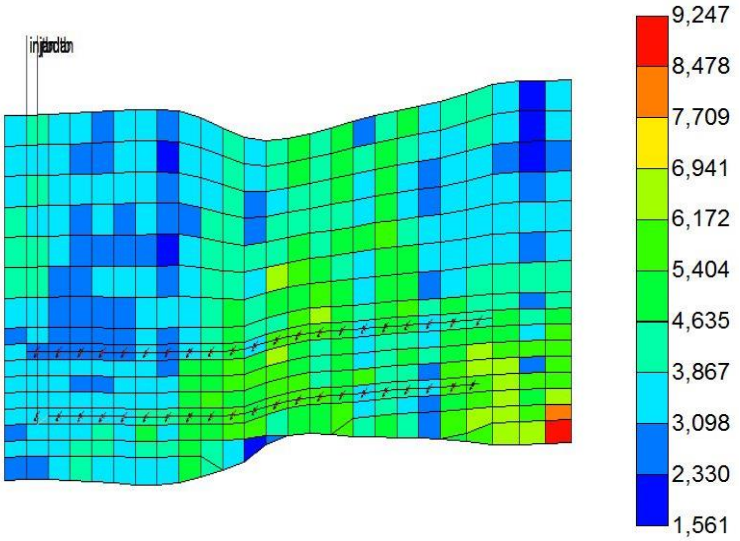


Figure 18: 2D view of SAGD Reservoir Model

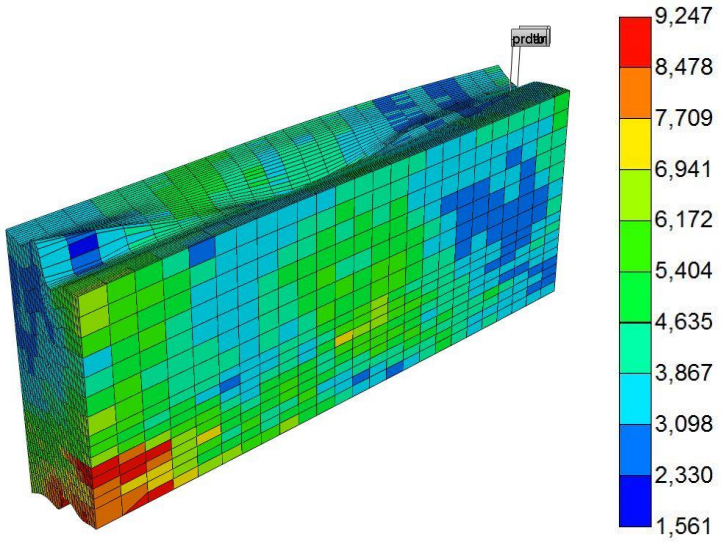


Figure 19: 3D view of SAGD Reservoir Model

4.4.2 EnKF using a reservoir simulator: conventional approach

A base case is prepared using traditional EnKF assisted history matching. Initially, 50 realizations are selected from original 100-member initial ensemble using scenario reduction method (Rahim et al., 2015) proposed for assisted history matching using EnKF. In state vector, the natural logarithm of permeability is used as model parameters while daily oil rate and steam oil ratio as production parameters. State vector φ_k for the base case can be written as,

$$\varphi_k = \begin{bmatrix} \{\ln(\text{permeability})\}_{N_g \times N_e} \\ \{\text{Oil Rate}\}_{1 \times N_e} \\ \{\text{Steam Oil Ratio}\}_{1 \times N_e} \end{bmatrix}_{20002 \times N_e} \quad \text{(4.11)}$$

where N_g denotes the total number of grid blocks (i.e. 20,000) in the reservoir model. N_e refers to the number of sampled realizations (i.e. 50) in the initial ensemble. State parameters are not used in Eqn. (4.11) since their time dependency could cause potential inconsistency (Gu & Oliver, 2006). Since the Gaussian statistics assumption has to be followed, logarithmic permeability is used in the state vector. The state vector is updated at four different time steps i.e. at 601, 760, 840 and 1160 days. After every update, simulations are rerun from the beginning using updated state vector. Production parameters are plotted after each update to compare and analyze the results.

4.4.3 EnKF using KL-ANN based proxy model: Proposed Approach

KL expansion allows the expression of a correlated random field or process in the form of a set of independent random variables while maintaining the covariance

structure. A set of 100 random permeability realizations generated using SGS is considered to obtain the covariance matrix. The KL workflow presented in Chapter 3 is followed using these 100 realizations. The covariance structure is expressed in the form of eigenvalues and eigen functions. The eigenvalues show a monotonically decreasing trend as demonstrated in Figure 10. The sum of normalized eigenvalues displays the ratio of energy (variance) in the KL terms. Upon plotting the cumulative sum of eigenvalues, it is observed that most of the energy is associated with a few initial eigenvalues. In a view to truncate the KL expansion (Eqn. (3.4)), smaller eigenvalues are discarded, and only the first few eigenvalues are selected. It is shown in the previous chapter (Figure 11) that more than 90% of energy is preserved with the first five to six eigenvalues. In this work, the KL expansion is truncated after first three eigenvalues and each permeability realization is represented using only three random variables with normal distribution.

MATLAB® Neural Network Toolbox™ (release R2013a) is used to build the neural network architecture. The workflow presented in Figure 5 is followed to obtain a neural network-based proxy model. A set of 25 permeability realizations in the form of random variables are selected as input training data for ANN model development. The target output vector for training data is obtained by running these realizations on the reservoir simulator. The design of a neural network involves selection of the number of neurons and hidden layers. (Ferreira et al., 2012) used some thumb rules from the literature for the selection of the number of neurons. They suggested the number of neurons should be between the number of

input parameters and output parameters. More precisely, it should be two-thirds the number of input parameters plus the output parameters and should not be more than twice the number of input parameters. In our work, different network configurations are analyzed, and suitable network architecture is selected by comparing the prediction error of various configurations.

A three-layer feed forward neural network is used for the proxy model development. The first layer consists of neurons representing the input values of permeability in the form of random variables used in Eqn. (3.4). The second (hidden) layer consists of three neurons with sigmoid transfer functions, and the third layer contains one neuron representing the output value of the production parameter. Three different models with similar structure are developed for three production parameters as outputs (oil rate, cumulative oil production, and steam oil ratio). Pre-processing functions are used to normalize the input data vectors and target output vectors. The network uses a back propagation training algorithm that updates the weight and bias values for each neuron. The training algorithm utilized in this work minimizes a combination of squared errors and weights and then determines the correct combination to produce a network that generalizes well. The process used to train the network is called Bayesian regularization back propagation. During the training process, 80% of the data is used for training and 20% of the data is used for testing, which allows monitoring of the network's general performance and prevents over-fitting of the training data. After the setup, training, and testing, the neural network is used to compute the output values for any number of permeability realizations represented by a set of random variables.

To show the accuracy of ANN based proxy model, 100 realizations different from those used for training are used for validation purpose. Production parameters are calculated using both reservoir simulator and proxy model. All of the production parameters presented in this work are normalized by a target value of respective output parameters for confidentiality purpose. Table 7 shows the quantitative comparison of oil rate, cumulative oil production, and steam oil ratio after 1200 days regarding mean, variance, minimum and maximum values. It can be seen that all the statistical values calculated using ANN proxy model are almost equal to those calculated using the commercial simulator. The result proves the potential of ANN proxy model as a substitute for reservoir simulator in history matching process.

Table 7: Quantitative comparison of different production parameters obtained using simulator and ANN proxy model after 1,200 days

Quantitative Measures	Normalized Oil Rate		Normalized Cumulative Oil Production		Normalized Steam Oil Ratio	
	Simulator	ANN	Simulator	ANN	Simulator	ANN
Mean	0.539	0.535	0.623	0.618	0.256	0.263
Variance	0.011	0.011	0.009	0.015	0.003	0.005
Min	0.171	0.295	0.241	0.310	0.203	0.202
Max	0.715	0.692	0.766	0.752	0.611	0.493

It is evident from Table 7 that ANN proxy model is in good agreement with the numerical simulator results as the statistical values obtained from ANN model are comparable with those obtained using the simulator. The proposed ANN model is sufficient to calculate required production parameters and can be used further with EnKF in the forecast step as a substitute for the commercial simulator.

Random variables, which represent permeability field through KL expansion, are used as model parameters instead of the natural logarithm of permeability. Modified state vector can be shown as:

$$\varphi_k = \begin{bmatrix} \{Random\ Variables\ (\xi)\}_{3\times N_e} \\ \{Oil\ Rate\}_{1\times N_e} \\ \{Steam\ Oil\ Ratio\}_{1\times N_e} \end{bmatrix}_{5\times N_e} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ 4.12$$

where ξ is a vector of random variables (containing three random variables for each realization) and Ne is the total number of realizations in the initial ensemble (50 in this case). For comparing the results, the state vector is updated at the same time instances as the baseline scenario. After each update, production parameters are plotted and compared. History matching results are shown and discussed in further section.

4.5 Results and Discussions

This section describes the results of assisted history matching using the proposed approach and also its comparison with history matching results of base case obtained using conventional EnKF framework. We compared the performance and efficacy of proposed KL-ANN-EnKF history matching workflow using comprehensive quantitative as well as qualitative analysis.

Quantitative analysis is performed using distinct statistical measures and output parameters. Different production parameters such as daily oil production rate, cumulative oil production, and cumulative steam to oil ratio are considered, actual expected mean values after 1200 days of which are 0.404, 0.562 and 0.268, respectively. Statistical measures used for the assessment are mean, standard deviation, minimum and maximum data values of the ensemble at every update step. Also, two quality measures, R^2 and root-mean-square-error ($RMSE$) are calculated for detailed quantification of the efficiency and accuracy of proposed method. R^2 of a realization i can be calculated as (Chitralekha et al., 2010),

$$R_i^2 = 1 - \frac{\sum_{i_1=t_1}^{t_n} (\hat{y}_{i_1} - \hat{y}_{true})^2}{\sum_{i_1=t_1}^{t_n} (\hat{y}_{true} - \bar{y}_{true})^2} \quad \text{..... 4.13}$$

where R_i^2 is R^2 value for realization i, $\hat{y}_i$ is the simulated value for i^{th} realization, and $\hat{y}_{true}$ is the true field data. $\bar{y}_{true}$ represents the average value of production parameter over the time span from t_1 to t_n. For the complete ensemble of realizations, R^2 can be calculated as mean of all the realizations using equation below:

$$R^2 = \frac{1}{N_e} \sum_{i_1=1}^{N_e} R_{i_1}^2 \quad \text{..... 4.14}$$

The value of R^2 ranges from -∞ to 1, where 1 represents a perfect match. Another parameter used to quantify the accuracy of results is the RMSE value. For a particular realization i, it can be calculated as (Gu & Oliver, 2006),

$$RMSE_i = \sqrt{\frac{1}{t_n}\sum_{i_1=1}^{t_n}(\hat{y}_{i_1} - \hat{y}_{true})^2} \quad \text{..... 4.15}$$

RMSE value for the entire ensemble is calculated by averaging RMSE of all realizations and ranges from 0 to ∞ where 0 denotes a perfect match; equation for which can be written as,

$$RMSE = \frac{1}{N_e}\sum_{i_1=1}^{N_e} RMSE_{i_1} \quad \text{..... 4.16}$$

If model parameters are estimated correctly in the history matching process, then mean of the production parameters of all realizations in the ensemble should be closer to the real value of the respective output parameter after history matching. Also, reduction in standard deviation of the ensemble should be observed which indicates uncertainty reduction in production forecast. However, ensemble variability should be maintained in EnKF to avoid filter divergence and spurious updates of state vectors, widely known as "ensemble collapse" (Chen et al., 2009). Furthermore, when the simulator is replaced with ANN based proxy model in

proposed history matching workflow, quality of history matching results should be consistent concerning average data mismatch. Keeping in mind these prospects, results of standard EnKF and KL-ANN-EnKF are analyzed here.

Table 8: Quantitative analysis of production parameters after 1,200 days at each update step obtained using conventional approach for EnKF

Quantitative Measures	Initial Ensemble	After 1st Update	After 2nd Update	After 3rd Update	After 4th Update
	Normalized Oil Rate SC after 1200 days (True value = 0.404)				
Mean	0.468	0.510	0.530	0.519	0.454
Std Dev	0.131	0.063	0.059	0.046	0.029
Min	0.225	0.403	0.394	0.419	0.405
Max	0.728	0.665	0.654	0.613	0.551
R^2	-0.230	-0.048	-0.153	0.053	0.476
RMSE	0.136	0.129	0.136	0.124	0.093
	Normalized Cum. Oil Production after 1200 days (True value = 0.562)				
Mean	0.549	0.568	0.590	0.586	0.525
Std Dev	0.132	0.055	0.047	0.038	0.024
Min	0.306	0.433	0.453	0.503	0.476
Max	0.734	0.684	0.664	0.668	0.600
R^2	0.842	0.960	0.966	0.976	0.968
RMSE	0.077	0.040	0.037	0.032	0.038
	Normalized Steam Oil Ratio after 1200 days (True value = 0.268)				
Mean	0.295	0.274	0.263	0.264	0.291
Std Dev	0.087	0.026	0.021	0.017	0.012
Min	0.212	0.229	0.233	0.230	0.259
Max	0.490	0.337	0.338	0.305	0.314
R^2	-0.001	-0.198	0.301	0.288	-0.051
RMSE	0.061	0.069	0.052	0.053	0.069

Results of quantitative analysis of the base case and proposed KL-ANN-EnKF approach are shown in Table 8 and Table 9, respectively. Though same initial ensemble of 100 realizations as the base case has been used in recommended workflow, as mentioned in methodology, permeability realizations are represented

using random variables in KL expansion. Therefore, the statistical measures for initial ensemble are different for KL-ANN method then the base case. Since, quantitative measures in Table 9 are shown for realizations using truncated KL; a small variation in output is expected.

Table 9: Quantitative analysis of production parameters after 1200 days at each update step obtained using KL-ANN-EnKF workflow

Quantitative Measures	Initial Ensemble	After 1st Update	After 2nd Update	After 3rd Update	After 4th Update
	Normalized Oil Rate SC after 1200 days (True value = 0.404)				
Mean	0.517	0.568	0.569	0.481	0.467
Std Dev	0.115	0.046	0.027	0.009	0.007
Min	0.210	0.424	0.510	0.462	0.452
Max	0.689	0.678	0.631	0.505	0.485
R^2	-0.228	-0.293	-0.251	0.473	0.510
RMSE	0.137	0.145	0.143	0.094	0.090
	Normalized Cum. Oil Production after 1200 days (True value = 0.562)				
Mean	0.588	0.638	0.638	0.551	0.529
Std Dev	0.101	0.037	0.024	0.011	0.008
Min	0.310	0.534	0.570	0.524	0.510
Max	0.728	0.724	0.689	0.571	0.546
R^2	0.896	0.936	0.943	0.990	0.977
RMSE	0.061	0.052	0.050	0.022	0.033
	Normalized Steam Oil Ratio after 1200 days (True value = 0.268)				
Mean	0.266	0.236	0.237	0.259	0.262
Std Dev	0.078	0.010	0.006	0.002	0.002
Min	0.205	0.206	0.222	0.255	0.258
Max	0.588	0.265	0.251	0.265	0.269
R^2	-0.837	-0.972	-0.968	-0.715	-0.624
RMSE	0.265	0.275	0.275	0.257	0.250

However, by comparing the different measures for all the production parameters after each update step, it can be observed that all values are in the same range as the base case. The normalized oil production rate for the base case and KL-ANN-EnKF after each update step is presented in Figure 20 and Figure 21 respectively. Actual data obtained from the field is shown using black line while the ensemble is represented as gray lines. Smaller ensemble spread for KL-ANN-EnKF workflow after the last update step is evident from Figure 21, the fact which is also endorsed by lower standard deviation of oil rate in Table 9 after last update step, indicating better convergence of model parameters.

In the case of KL-ANN-EnKF, oil rates after the 2^{nd} update are slightly biased from the field data when compared to the base case, most probably due to the mismatch between the prediction of production parameters using ANN proxy model and commercial reservoir simulator. However, it has been taken care of later in the next update step due to sequential data assimilation in EnKF. Though mean of the oil rate for base case is closer to the real value (Table 8), proposed workflow depicts minimum ensemble span with higher minimum and lower maximum values for the oil rate (Table 9) after the last update, suggesting minimal uncertainty in model parameters between both cases. Lower RMSE and improved R^2 values are observed for oil production rate in Table 9 as compared to the base case, indicating fewer data mismatch over the time and hence confirming the capability of the proposed KL-ANN-EnKF workflow.

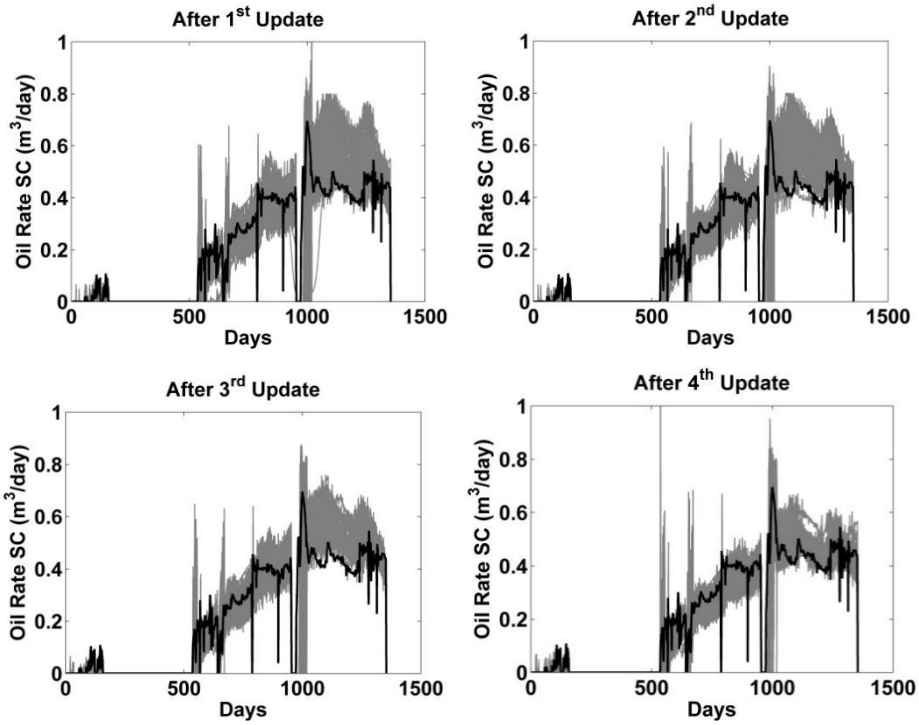


Figure 20: Normalized oil rate after each update of EnKF for the base case in which conventional approach is considered to update the ensemble (gray lines). Black line shows history obtained from field

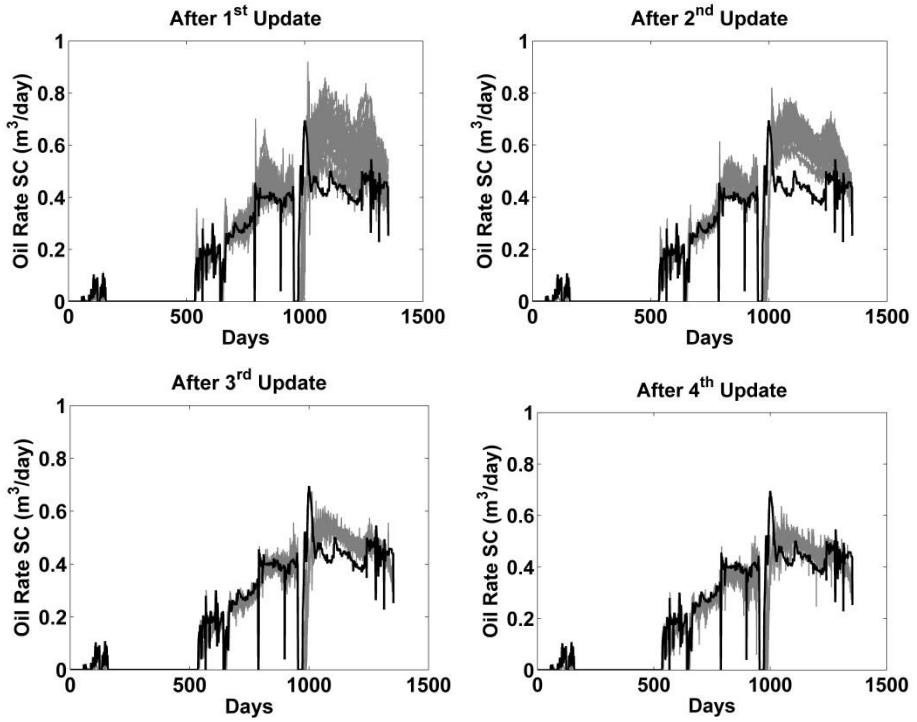


Figure 21: Normalized oil rate at each update step obtained after using KL-ANN-EnKF approach to update the ensemble (gray lines). The black line shows history obtained from the field.

Figure 22 and Figure 23 represent history matching results of normalized cumulative oil production after each update for the base case and KL-ANN-EnKF respectively. It can be seen from the figures that with each update, reduction in ensemble spread (gray lines) in the case of proposed workflow is better than baseline scenario. Quantitative analysis in Table 9 also shows comparable results with significantly smaller standard deviation and better minimum/maximum values of cumulative oil production which reflect reduced ensemble span and hence faster uncertainty reduction in output forecast as compared to the base case after the last update. Also, mean of the cumulative oil after final update (Table 9) is slightly closer to the actual value in case of KL-ANN-EnKF, demonstrating a better convergence of model parameters towards the real value. Furthermore, near to unity R^2 value and almost zero RMSE after final update step display better history matching of cumulative oil production using KL-ANN-EnKF.

Steam oil ratio which exhibits the efficiency of the SAGD process is also plotted in Figure 24 and Figure 25 after each update in history matching, performed using base case and KL-ANN-EnKF, respectively. Similar to other production parameters, a significant reduction of ensemble spread in every update step as compared to base case is evident from the results of KL-ANN-EnKF. Please note that small band spread here does not indicate a loss in ensemble variability, which can be proven by comparing, updated model parameters with reference statistics before history matching as shown in Figure 31 and Figure 32.

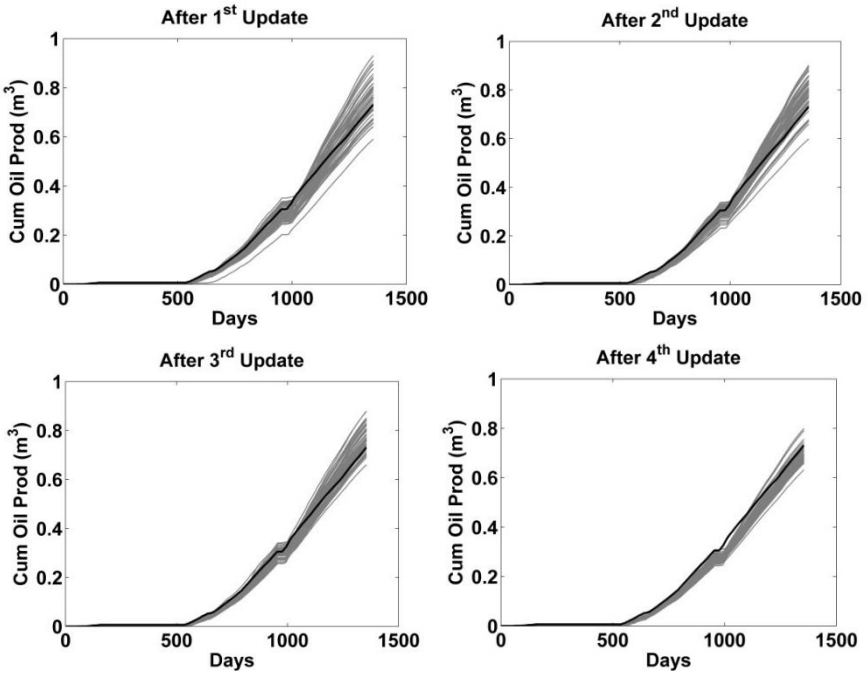


Figure 22: Normalized cumulative oil production after each update of EnKF for base case in which conventional approach is considered to update the ensemble (grey lines). Black line shows history obtained from field

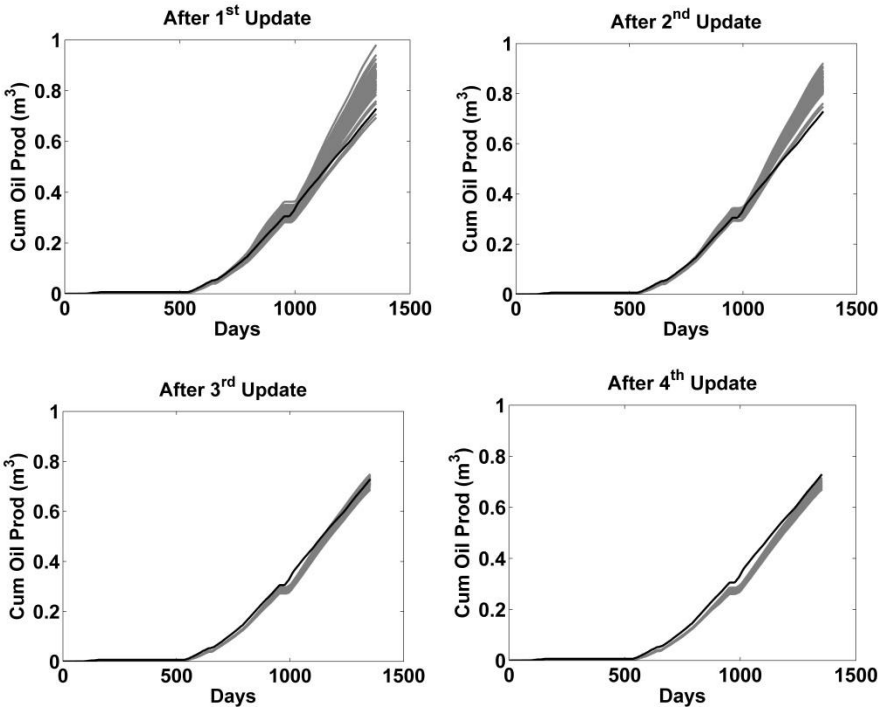


Figure 23: Normalized cumulative oil production at each update step obtained after using KL-ANN-EnKF approach to update the ensemble (grey lines). Black line shows history obtained from field.

Various statistical measures in Table 9 also show promising history matching results using KL-ANN- EnKF as the mean value of steam oil ratio after the last update is almost equal to the real value. The result proves the capability of ANN based proxy model to estimate production parameters accurately in forecast step, which ultimately leads to an adequate update of model parameters in following analysis step. Despite better history matching results using KL-ANN-EnKF, quality measures R^2 is negative while RMSE is higher than the base case (Table 8) for steam oil ratio. This observation can be explained by the fact that steam oil ratio has minuscule standard deviation and hence, even a small deviation from the actual value will considerably contribute to data mismatch, ultimately resulting in reduced quality measures. However, all other statistical measures in quantitative analysis (Table 9) as well as plots in qualitative analysis (Figure 25) express satisfactory history matching results. Overall, ensemble mean for all the three production parameters is in proximity to the respective real value while ensemble spread is reduced significantly with each update in proposed workflow. It is, therefore, evident that ANN proxy model can successfully substitute the use of commercial simulator during the assisted history matching process without compromising the accuracy of history matching results.

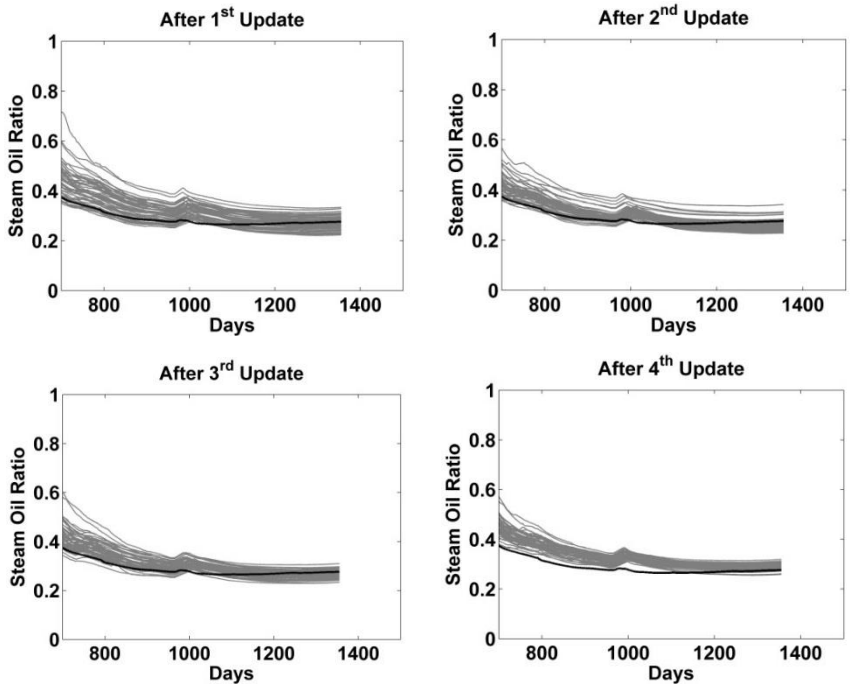


Figure 24: Normalized cumulative steam to oil ratio after each update of EnKF for the base case in which conventional approach is considered to update the ensemble (gray lines). Black line shows history obtained from field

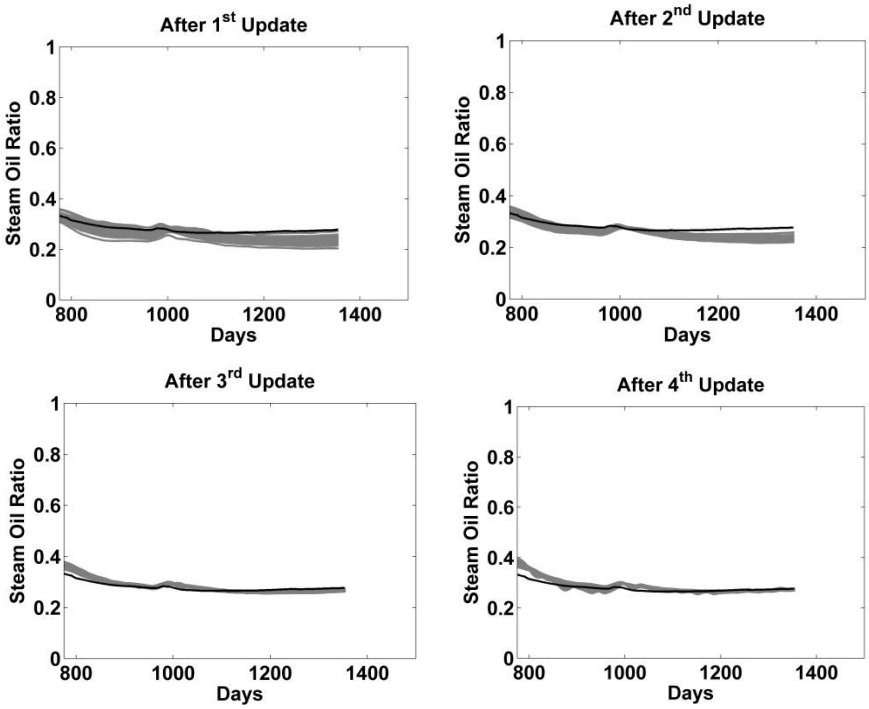


Figure 25: Normalized cumulative steam-oil ratio at each update step obtained after using KL-ANN-EnKF to update the ensemble (gray lines). The black line shows history obtained from the field.

In Figure 26,

Figure 27 and Figure 28, box and whisker plots that demonstrate the change in distribution are plotted for oil production rate, cumulative oil production, and steam oil ratio respectively at 1200 days, before and after history matching. In these plots, whisker ranges from minimum to maximum while box presents 0.25 quantile to 0.75 quantiles of the distribution. Red line and * sign in interquartile range of each box and whisker plot shows median and mean of the distribution respectively while black horizontal line denotes the real value of the respective production parameter. From the box-whisker plots, it can be said that KL expansion can reduce the dimensionality of the history matching problem while maintaining variability in the ensemble since the distribution, as well as mean and median of all production parameters, are similar to those for the base case before history matching. Another important observation is the lower standard deviation for all production parameters after history matching in KL-ANN-EnKF compared to the baseline scenario, which is in agreement with the qualitative and quantitative analysis. Also, concerning mean of output parameters, history matching results is either similar or better compared to the base case. Furthermore, it is observed that difference between mean and median reduced after history matching process using KL-ANN-EnKF, suggesting diminished bias in the ensemble since identical mean and median shows that same numbers of realizations in the ensemble are above and below the average value. This analysis of change in distribution further establishes the effectiveness of proposed integrated KL-ANN-EnKF workflow.

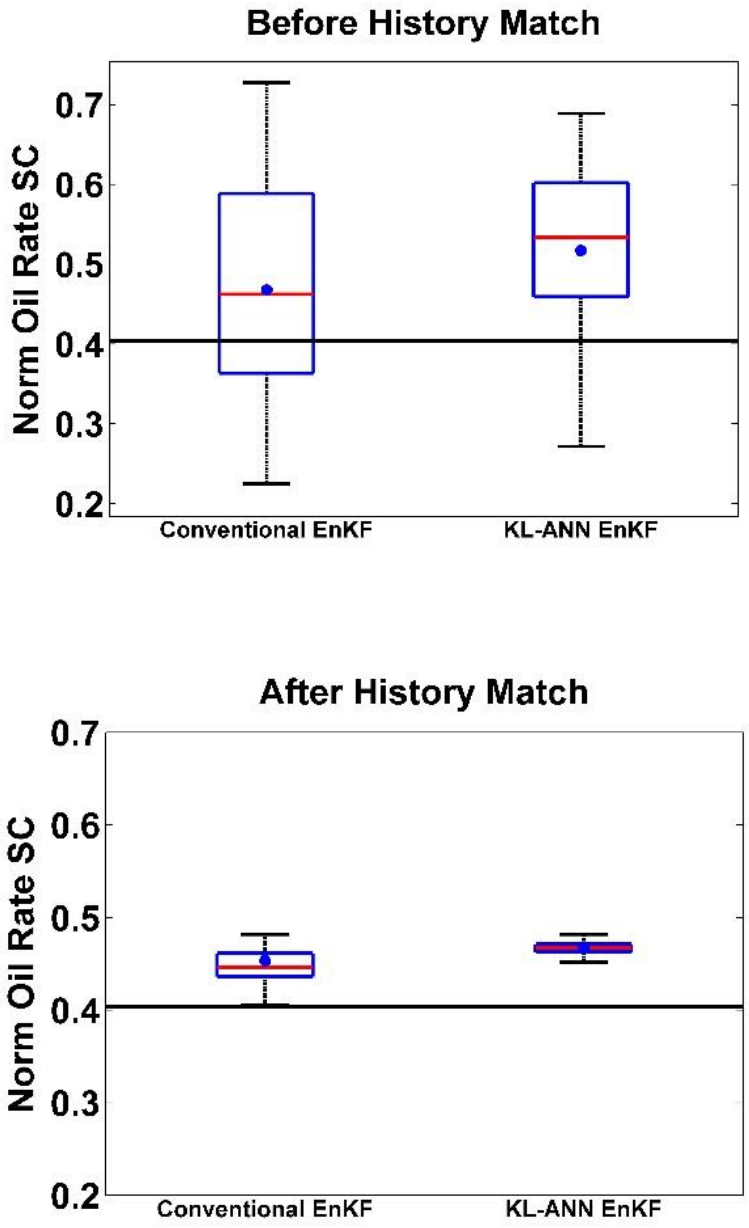


Figure 26: Box and whisker plots representing distributions of normalized oil rate at 1200 days obtained using simulations of all realizations of the initial and updated ensemble using different EnKF approaches for history matching. The red line shows the median of each distribution. The horizontal black line shows the true value.

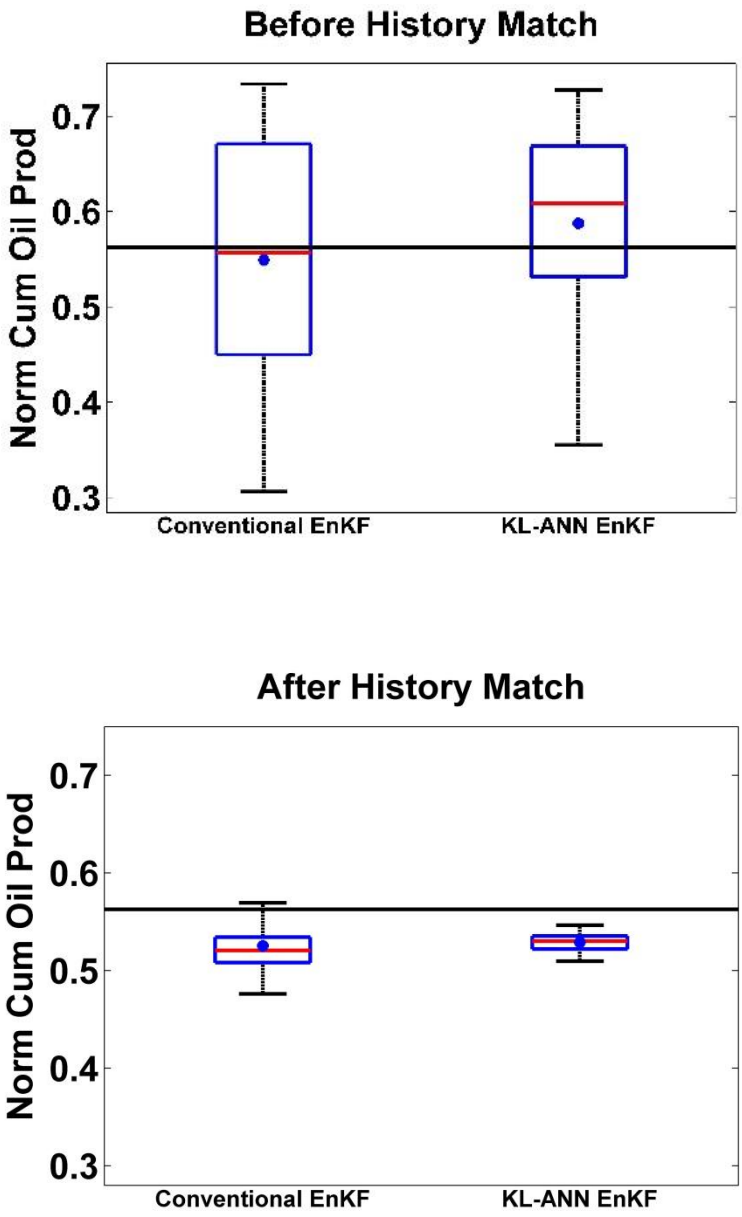


Figure 27: Box and whisker plots representing distributions of normalized cumulative oil production at 1200 days obtained using simulations of all realizations of the initial and updated ensemble using different EnKF approaches for history matching. The red line shows the median of each distribution. The horizontal black line shows the true value.

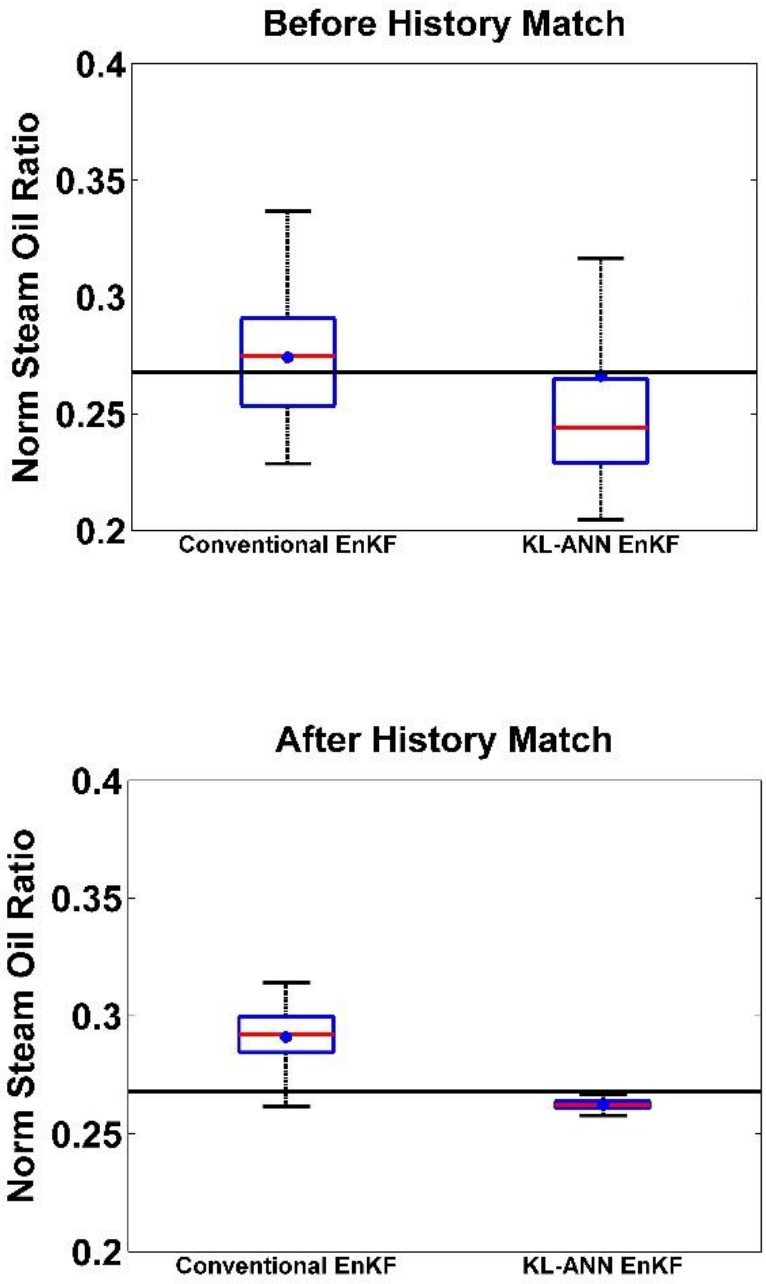


Figure 28: Box and whisker plots representing distributions of normalized steam oil ratio at 1200 days obtained using simulations of all realizations of the initial and updated ensemble using different EnKF approaches for history matching. The red line shows the median of each distribution. The horizontal black line shows the true value.

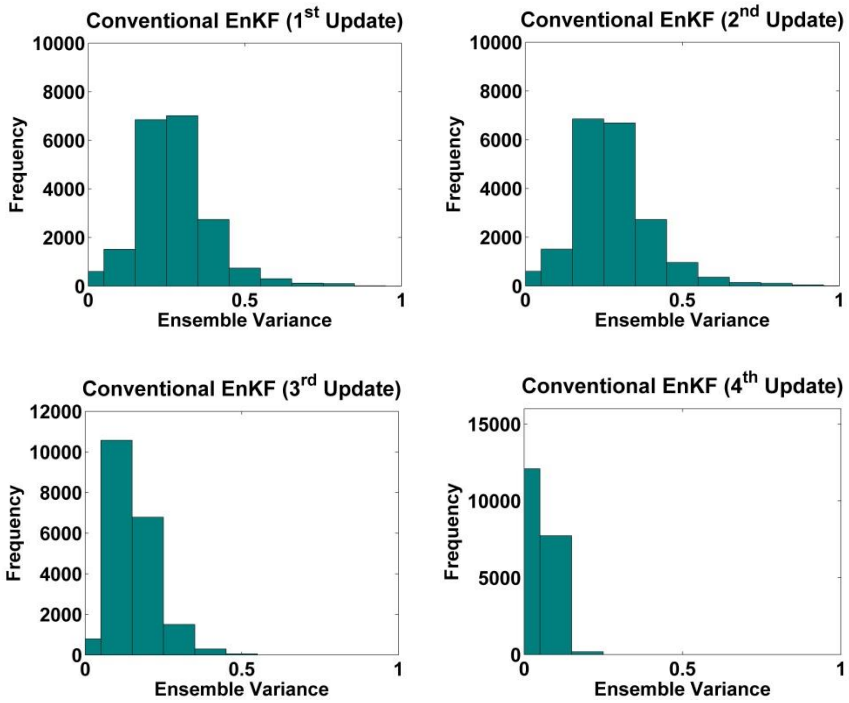


Figure 29: Ensemble variance of model parameters at various assimilation steps during conventional EnKF method.

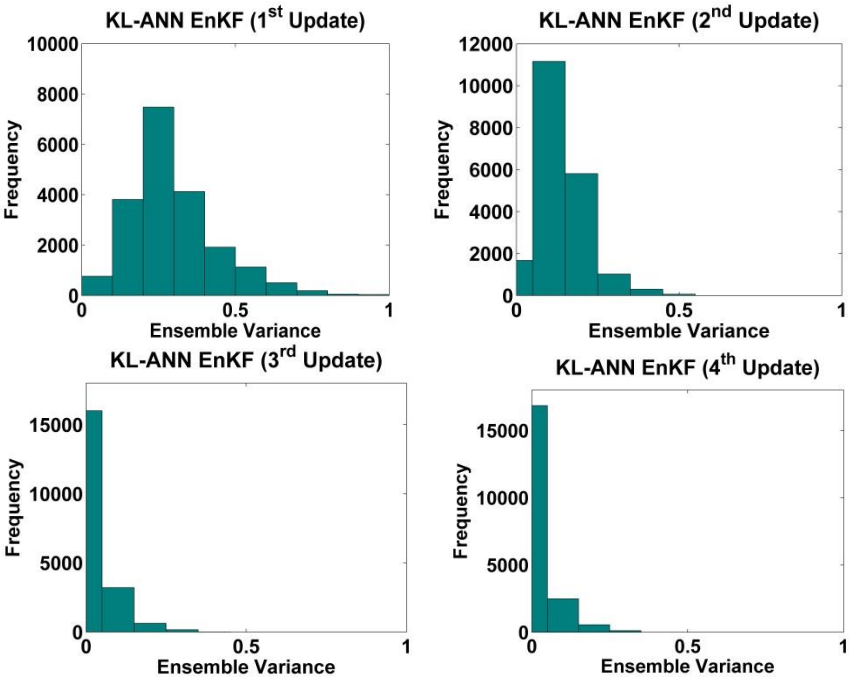


Figure 30: Ensemble variance of model parameters at various assimilation steps during KL-ANN-EnKF method.

It can be observed from Figure 29 and Figure 30 that there is a significant reduction in the ensemble variance of model parameters at successive assimilation steps in conventional as well as KL-ANN based EnKF method. This shows the convergence of ensemble towards the true representation of permeability field with almost zero variance after fourth update step.

It is important to note that in KL-ANN-EnKF, state vectors used in EnKF process are modified. In this approach, individual random variables, which represent permeability field through KL expansion, are used as model parameters in the initial ensemble instead of the natural logarithm of permeability. For obtaining credible history matching results, it is important that parameterization technique is competent enough to represent the initial ensemble. To investigate this fact, the comparison of permeability histogram before history matching is plotted for realizations for the base case and KL-ANN-EnKF is plotted in Figure 31. Irrespective of the different method used for representing the permeability field, the overall distribution is similar in both cases. Also, to ensure the integrity of the history matching process, the occurrence of ensemble collapse should be examined. The variance of different grid blocks in all realizations of the ensemble becomes zero when ensemble collapse takes place. In other words, for all the grid blocks, values of model parameters (i.e. permeability in our case) become equal in different realizations, leading to apparently only one realization in the ensemble and hence loss of ensemble variability. Figure 32 depicts the updated permeability distribution of standard EnKF and KL-ANN-EnKF. It is evident from this figure

that only higher permeability values (which most likely are impractical values for the field) are filtered out.

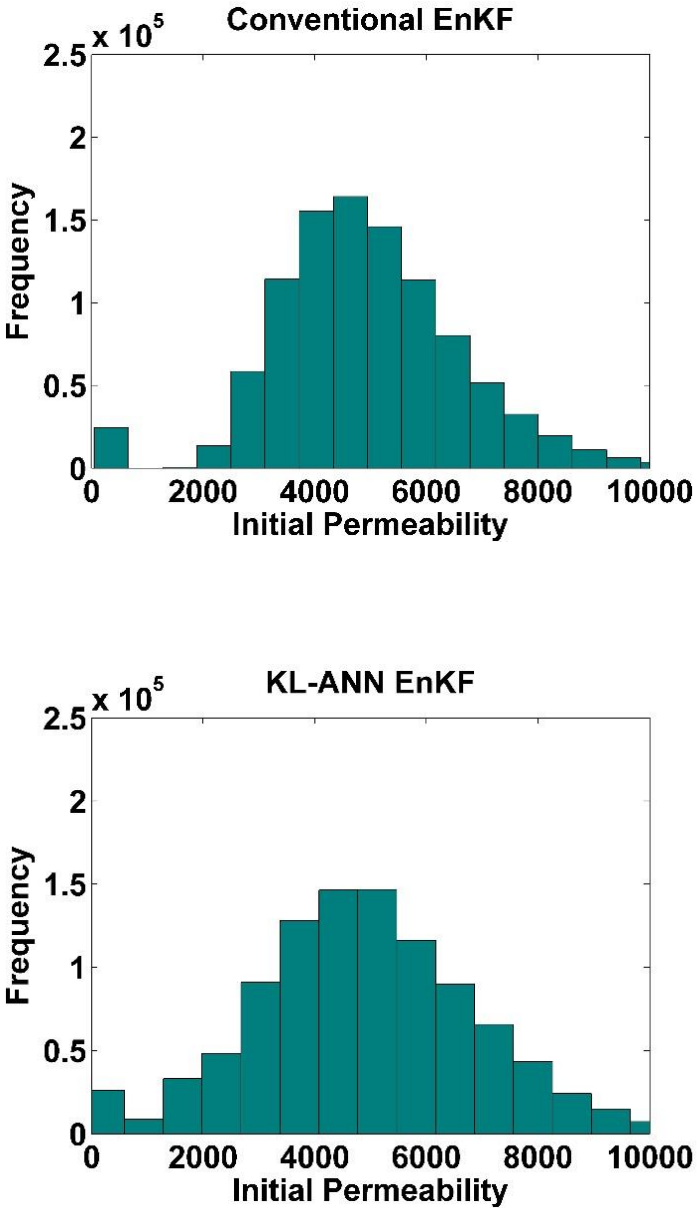


Figure 31: Histogram of Initial Permeability before history match for realizations used during conventional EnKF and KL-ANN-EnKF.

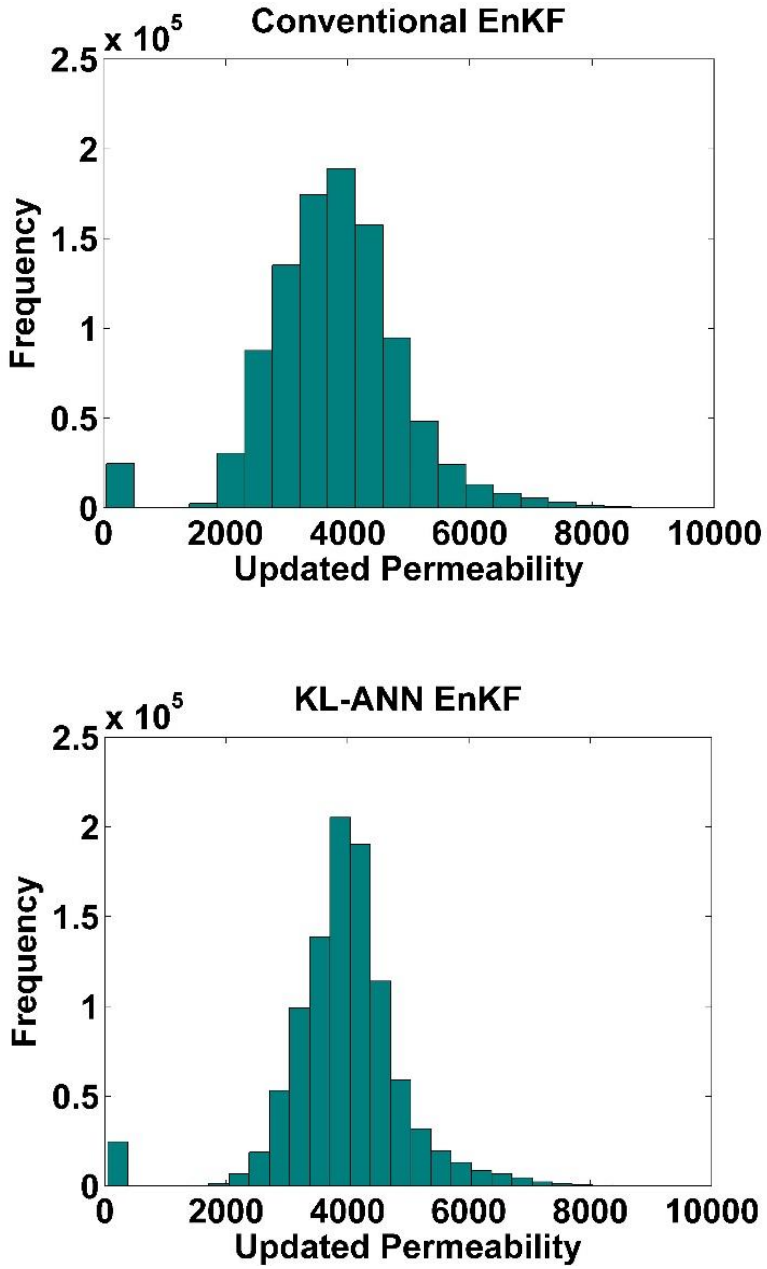


Figure 32: Histogram of Final updated Permeability after history match for realizations obtained during conventional EnKF and KL-ANN-EnKF.

The objective of this work is to reduce the computational time during the assisted history matching process. The commercial simulator is substituted with KL-ANN based proxy model for the forecast step in EnKF. In standard EnKF, for the ensemble of 50 realizations, 50 full simulation runs are needed for each time step. In this work, 4 update steps are considered which required a total of 200 simulation runs. On the other hand, to build KL-ANN proxy model, simulations required are equal to realizations used for training data, which are only 25 in our case. Once these 25 realizations are run using the commercial simulator and output is used to train the ANN model, there is no requirement for commercial simulators in the further process which reduced the computational time and cost significantly. A comparison of computational time in conventional EnKF method with the time required during KL-ANN-EnKF has been shown in Figure 33. Use of KL-ANN based proxy model as forecast model in EnKF algorithm reduces the time requirement up to 90%, yet producing similar history matching results.

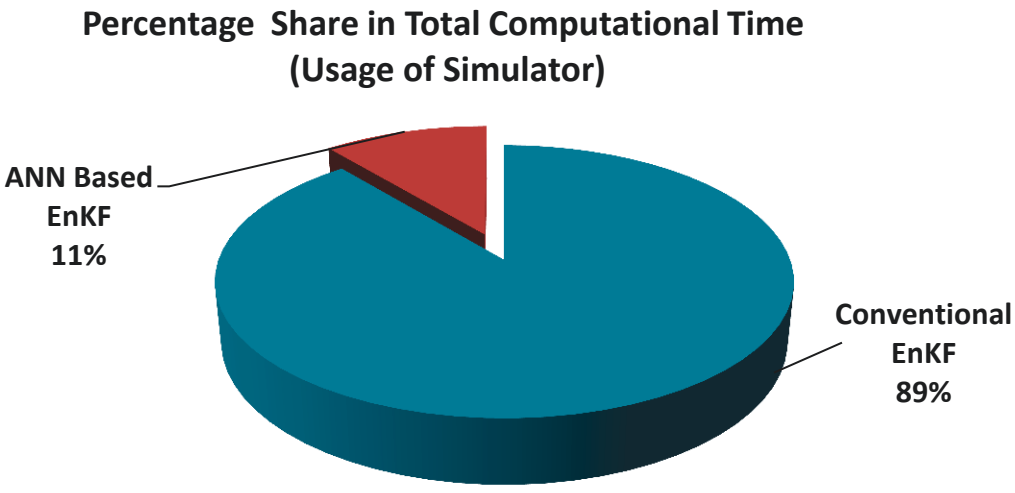


Figure 33: Comparison of computational time requirement for proposed method.

4.6 Summary

In this Chapter, a KL-ANN based proxy model framework as an alternative to the commercial simulator in forecast step of EnKF is proposed. A real field SAGD model is used to demonstrate the applicability of the proposed approach. The KL expansion is employed to parameterize the permeability field and to reduce the dimensionality of the model. Model parameters are represented with the help of standard Gaussian random variables. Further, proxy models are developed on these random variables to predict the output. Results obtained from the proxy models are verified by comparing with the results obtained using the commercial simulator. The following are the main conclusions of this work:

- KL-ANN-EnKF provides adequate results when compared with the outputs obtained from the commercial simulator. Developed model can be used to run as many realizations as required in a relatively very small amount of time and with very low computational effort.
- The proposed framework is easy to implement in large-scale complex reservoirs. Models developed with the help of very few simulation runs are efficient enough to imitate the full dynamics and trend of simulation results, thus eliminating the need for time-consuming simulation runs.
- KL-ANN based proxy model can easily be incorporated in EnKF framework as forecast model and can reduce the time required for assisted history matching, up to 90% for the case presented, while maintaining the accuracy of results.

Chapter 5:

PCE Proxy Model Based EnKF framework for Assisted History matching[3]

5.1 Introduction

After the development of a comparable proxy model based on PCE (discussed in Chapter 3), the next task is to combine it with existing EnKF framework for assisted history matching.

Workflow that shows stepwise procedure to be followed to implement assisted history matching with practical computing cost is shown in Figure 34. Starting point of proposed integrated framework is to represent initial ensemble in terms of eigen values and corresponding eigen functions of covariance matrix along with random variables using Eqn. (3.3). Generally, depending on the pdf of random variables, type of orthogonal polynomial is chosen. In proposed framework, as stated before, Gaussian random variables with zero mean and unit variance are used and as the same random variables are considered as input parameters in PCE, Hermite polynomials are used to construct polynomial chaos. Using Gaussian quadrature technique, sets of random variables(ξ), commonly known as collocation points are obtained.

[3]based on a manuscript "Polynomial-Chaos-Expansion Based Assisted History Matching Workflow for Computationally Efficient Reservoir Characterization: A SAGD Field Case Study," Submitted in SPE Journal of Reservoir Engineering and Evaluation.

Considering each set of random variables as separate input, full physics simulations using commercial simulators are run and output data is collected. System of equations is developed using each input-output pair and coefficients are determined. If the error in approximation using PCE is not significant then the same polynomial model is used in forecast step of EnKF.

The only loop in proposed assisted history matching workflow is of EnKF. It basically starts from construction of a state vector. Since initial ensemble is represented by random variables in KL parameterization, they are used in place of model parameters (i.e. permeability) in state vector along with the production parameters. Then using proxy model instead of commercial simulators in forecast step, production parameters are computed considering the random variables in state vector as an input parameters. Uncertainty in production parameters is analyzed and if it is high then state vector is updated using true production data from the field and Kalman gain in analysis step of EnKF loop. Again random variables from updated state vector are used in KL-PCE based proxy model to calculate production parameters and when uncertainty in production parameters is less enough, realizations can be generated by substituting latest updated random variables into Eqn. (3.3). Best realization from the ensemble then can be used further to outline future field development strategies.

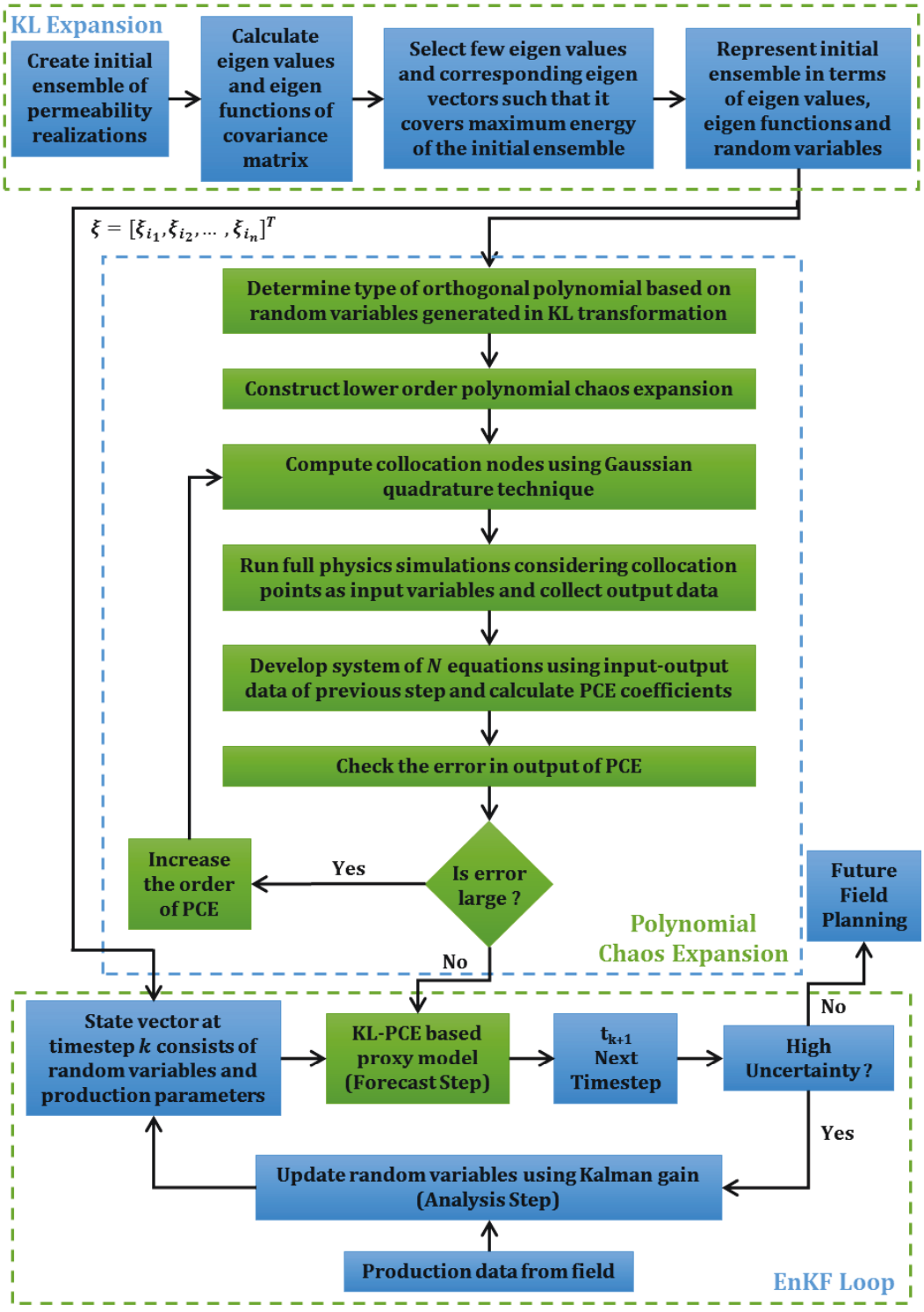


Figure 34: Workflow for integration of KL-PCE based proxy model in existing EnKF framework

5.2 Application to a SAGD Reservoir: Field Case Study

In a view, to demonstrate the efficacy of the proposed method and compare the results with numerical simulation outputs, KL-PCE based proxy model in combination with EnKF framework is implemented on a large-scale SAGD reservoir in northern Alberta. A code is developed using MATLAB® (release R2014a) that integrates CMG STARS™ (CMG 2013a), CMG Results Report™ (CMG 2013b), and PCE based proxy model. A general description of the reservoir model is explained in Section 5.2.1. Stepwise implementation of the proposed frameworks is discussed in further subsections.

5.2.1 Description of the Reservoir Model

A portion of the reservoir in northern Alberta that contains one horizontal well pair and six vertical observation wells was considered to build a reservoir model. Various types of well logs (caliper log, resistivity log, sonic log, gamma ray log, neutron porosity log, self potential log, borehole compensated formation evaluation log) for each observation well were imported in Petrel exploration and production software platform by Schlumberger and well tops as well as top and bottom of formation were determined. Next, top and bottom surface of the formation was defined using well tops and formation depth at each well. A 3D view of reservoir model is shown in Figure 35 grid dimensions of which are 25×50×16 and size of each grid block is 25×2×1.5 m in i (East), j (North) and k (Elevation) directions, respectively. Porosity obtained from well logs was then scaled up and used to calculate permeability for the grid blocks containing vertical observation wells. Using them as conditioning data, 100 permeability realizations

were generated by applying sequential Gaussian simulation. Porosity values range from 0.315 to 0.41 while permeability values range from 1525 md to 7150 md in the 100-member initial ensemble.

In order to include dynamic properties, static model was exported to CMG Builder™ (CMG 2013c). Initial reservoir temperature was defined as 7º C and bitumen viscosity at that temperature was specified as 625,000 cp which decreases to 10 cp at higher temperature of 216º C. Also, rock type with pertinent relative permeability curve was used in the dynamic model which along with viscosity vs. temperature curve is not provided here due to confidentiality purposes. A horizontal well pair with 6 m spacing was included in the reservoir model with injector being at top. Well constraints defined for both horizontal wells were obtained from the field data recorded on daily basis. Dynamic model was simulated using thermal simulator CMG STARS™ (CMG 2013a) and various production parameters were obtained for 1355 days.

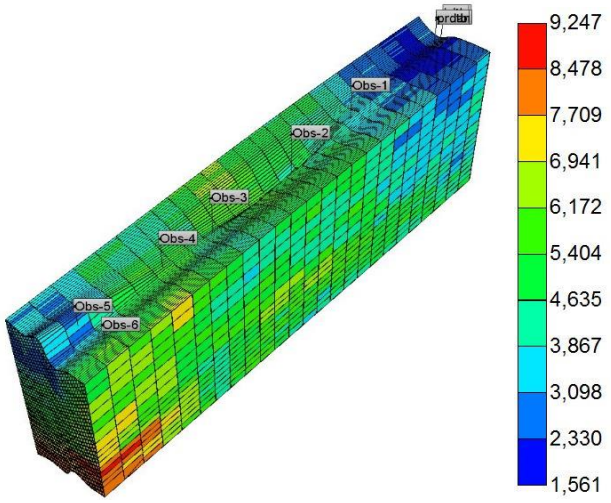


Figure 35: 3D View of reservoir model

5.2.2 EnKF using a reservoir simulator: conventional approach

In order to compare the results of history matching obtained by applying proposed ideas with traditional assisted history matching using EnKF, a base case was prepared in which history matching includes assimilation of only production data. At first, 50 realizations were sampled from original 100-member initial ensemble using scenario reduction method proposed by Patel et al. (2015) to make history matching of the field scale reservoir a practical exercise. In this case, state vector consists natural logarithm of permeability as model parameters while daily oil rate and steam oil ratio as production parameters. State vector Z_k for the base case can be written as,

$$Z_k = \begin{bmatrix} \{\ln(permeability)\}_{N_c \times N_s} \\ \{Oil\ Rate\}_{1 \times N_s} \\ \{Steam\ Oil\ Ratio\}_{1 \times N_s} \end{bmatrix}_{20002 \times N_s} \quad \text{..... (5.1)}$$

where N_c denotes total number of grid blocks (i.e. 20,000) in a reservoir model while N_s refers to number of sampled realizations (i.e. 50) from initial ensemble. Logarithmic permeability in Eq. (5.1) is necessary in order to fulfill Gaussian statistics assumption of EnKF. Furthermore, state parameters are not used in Eq. (5.1) since their time dependency can cause potential inconsistency (Gu and Oliver 2006). State vector was updated at four instances i.e. at 601, 760, 840 and 1160 days. After each update, simulations were rerun from the beginning using updated state vector so that accurate time dependent properties (e.g. pressure, saturation) can be calculated. Also, different production parameters were plotted after each update to compare and analyze the results.

5.2.3 EnKF using KL-PCE based proxy model: Proposed Approach

KL expansion allows the expression of a correlated random field or process in the form of a set of independent random variables while maintaining the covariance structure. A set of 100 random permeability realizations generated using SGS is considered to obtain the covariance matrix. The KL workflow presented in Chapter 3 is followed using these 100 realizations. The covariance structure is expressed in the form of eigenvalues and eigen functions. The eigenvalues show a monotonically decreasing trend as demonstrated in Figure 10. The sum of normalized eigenvalues displays the ratio of energy (variance) in the KL terms. Upon plotting the cumulative sum of eigenvalues, it is observed that most of the energy is associated with a few initial eigenvalues. In a view to truncate the KL expansion (Eqn. (3.4)), smaller eigenvalues are discarded, and only the first few eigenvalues are selected. It is shown in the previous chapter (Figure 11) that more than 90% of energy is preserved with the first five to six eigenvalues. In this work, the KL expansion is truncated after first three eigenvalues and each permeability realization is represented using only three random variables with normal distribution.

After parameterization, based on probability distribution of input random variables, type of orthogonal polynomials to be used in PCE is decided which is Hermite polynomials in our case. In PCE, to start with, 2^{nd} order polynomial was considered. Then collocation nodes were computed using Gaussian quadrature technique. In our case, roots of next higher order orthogonal polynomial i.e. 3^{rd} order Hermite polynomial were obtained and using them collocation nodes were

prepared. Total number of collocation nodes available here are 27. Since total number of terms in 2nd order PCE with 3 random variables is 10 which can be calculated using Eqn. (3.6), only 10 collocation nodes with higher probability were selected among all. Note that each collocation node here is a vector length of which is equal to number of random variables considered in PCE. For each collocation point, permeability field was generated using Eq. (3.4) and full physics simulations were carried out using CMG STARS™ to obtain required production parameters. Finally, all 10 coefficients were determined by solving system of equations shown in Eqn. (3.7) for each PCE constructed to calculate that particular production parameter.

To check the error in output of PCE, 100 realizations different than initial ensemble were generated. Production parameters were obtained using CMG STARS™ as well as 2nd order PCE. Table 10 shows the quantitative comparison of oil rate, cumulative oil production and steam oil ratio (all normalized) after 1200 days in terms of mean, standard deviation, minimum and maximum values. From the table, it can be said that production parameters obtained using PCE are very similar to those calculated using simulator, meaning that higher order polynomial chaos is not required. Hence, 2nd order PCE is sufficient to calculate required production parameters and can be used further in EnKF loop in place of commercial simulator.

Table 10: Quantitative comparison of different production parameters obtained using simulator and 2nd order PCE after 1200 days

Quantitative Measures	Normalized Oil Rate		Normalized Cumulative Oil Production		Normalized Steam Oil Ratio	
	Simulator	PCE	Simulator	PCE	Simulator	PCE
Mean	0.539	0.532	0.623	0.628	0.256	0.255
Std Dev	0.108	0.107	0.097	0.103	0.566	0.062
Min	0.171	0.212	0.241	0.234	0.203	0.195
Max	0.715	0.716	0.766	0.768	0.611	0.524

In EnKF loop of integrated proxy model, as initial ensemble is represented using random variables in KL parameterization, random variables (ξ) are used as model parameters instead of natural logarithm of permeability while production parameters used here are same as base case. State vector for integrated KL-PCE based EnKF workflow can be shown as,

$$Z_k = \begin{bmatrix} \{Random\ Variables\ (\xi)\}_{3\times N_e} \\ \{Oil\ Rate\}_{1\times N_e} \\ \{Steam\ Oil\ Ratio\}_{1\times N_e} \end{bmatrix}_{5\times N_e} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5.2)$$

where ξ is a vector of random variables (containing three random variables for each realization as discussed before) and N_e is the total number of realizations in initial ensemble. In order to compare results with base case, production parameters were calculated using proxy model. State vector was updated at same time instances as base case. After each update in EnKF loop, different production parameters were plotted to evaluate the uncertainty reduction in production forecast.

5.3 Results and Discussions

For qualitative analysis of assisted history matching results of base case, simulations for each ensemble member were rerun after each EnKF update and production parameters as well as surveillance parameters were plotted. Also, detailed quantitative analysis was performed to observe change in production parameters like oil rate, cumulative oil production, and steam oil ratio. Furthermore, box and whisker plots were prepared for all production parameters at 1200 days in order to analyze uncertainty reduction in production forecast after each update when only production parameters are used in state vector as shown in Eqn. (5.1).

If model parameters in history matching are estimated precisely then mean of the production parameters calculated for all realizations in the ensemble should be equal to the true value of the respective production parameter. In addition, reduction in standard deviation of the ensemble ultimately demonstrates reduced uncertainty in production forecast. From Table 8, it can be observed that standard deviation of ensemble is decreasing consistently for all production parameters after each update in history matching of base case. Also, ensemble mean is moving closer to true value of respective production parameter as compare to that of initial ensemble which emphasizes the fact that EnKF can be successfully implemented for assisted history matching of field scale SAGD reservoirs. It is evident from the same table that minimum and maximum values of the ensemble for all production parameters are increasing and decreasing respectively which

denotes the convergence of ensemble towards the true value due to updated model parameters.

Quantitative analysis is performed using distinct statistical measures and output parameters. Different production parameters such as daily oil production rate, cumulative oil production, and cumulative steam to oil ratio are considered, actual expected mean values after 1200 days of which are 0.404, 0.562 and 0.268, respectively. Statistical measures used for the assessment are mean, standard deviation, minimum and maximum data values of the ensemble at every update step. Also, two quality measures, R^2 and root-mean-square-error ($RMSE$) are calculated for detailed quantification of the efficiency and accuracy of proposed method. R^2 of a realization i can be calculated as (Chitralekha et al., 2010),

$$R_i^2 = 1 - \frac{\sum_{i_1=t_1}^{t_n} (\hat{y}_{i_1} - \hat{y}_{true})^2}{\sum_{i_1=t_1}^{t_n} (\hat{y}_{true} - \bar{y}_{true})^2} \quad \text{.....5.3}$$

where R_i^2 is R^2 value for realization i, $\hat{y}_i$ is the simulated value for i^{th} realization, and $\hat{y}_{true}$ is the true field data. $\bar{y}_{true}$ represents the average value of production parameter over the time span from t_1 to t_n. For the complete ensemble of realizations, R^2 can be calculated as mean of all the realizations using equation below:

$$R^2 = \frac{1}{N_e}\sum_{i_1=1}^{N_e} R_{i_1}^2 \text{.....} 5.4$$

The value of R^2 ranges from -∞ to 1, where 1 represents a perfect match. Another parameter used to quantify the accuracy of results is the RMSE value. For a particular realization i, it can be calculated as (Gu & Oliver, 2006),

$$RMSE_i = \sqrt{\frac{1}{t_n}\sum_{i_1=1}^{t_n}(\hat{y}_{i_1} - \hat{y}_{true})^2} \text{ .....} 5.5$$

RMSE value for the entire ensemble is calculated by averaging RMSE of all realizations and ranges from 0 to ∞ where 0 denotes a perfect match; equation for which can be written as,

$$RMSE = \frac{1}{N_e}\sum_{i_1=1}^{N_e} RMSE_{i_1} \text{.....} 5.6$$

From Table 8, it can be noticed that both R^2 and RMSE has improved with each EnKF update for almost all production parameters. In fact, initial ensemble shows negative R^2 value in case of oil rate, which later improved to 0.476 after 4th update. Similarly, for steam oil ratio both quality measures reflects better history matching except the last update in which RMSE is increased and R^2 value becomes negative. Since standard deviation of ensemble is the least after last update, a small deviation of steam oil ratio from true value contributes to the error

considerably, ultimately resulting into poor quality measures although ensemble provides acceptable match with field observations which can be seen in Figure 24.

For qualitative analysis of base case history matching results, oil production rate of the ensemble (grey lines) as well as true data (black line) is plotted in Figure 20. It can be noticed that after each EnKF update, ensemble spread (area consists of grey lines) is shrinking. Since model parameters are converging towards true value, uncertainty in production forecast is decreasing which is reflected in ensemble spread of that particular production parameter. Similar phenomenon can also be observed in case of cumulative oil production which is shown in Figure 22. Steam oil ratio that depicts the efficiency of SAGD process is shown in Figure 24 which demonstrates satisfactory results in assisted history matching.

An attempt is made here to make assisted history matching practically applicable by integrating KL-PCE based proxy model in EnKF framework which reduces the overall computational cost. Simplest way to measure reduction in computational cost is number of simulation runs required to perform history matching. For history matching in base case, each realization in the ensemble is simulated till next time-step. So, for ensemble consists of 50 realizations, 50 full physics simulations are needed for each time-step. Therefore, in total 200 simulation runs are necessary for history matching of base case. On the other hand, to build a KL-PCE based proxy model, simulations required are equal to number of collocation nodes i.e. 10 in our case. So it can be said that computational cost is reduced by 95%. Even if higher order PCE is to be used, reduction in computational cost would be significant. However, it should be confirmed that history matching

results obtained using proposed integrated workflow are consistent. In other words, if a method uses less number of simulation runs and still produces results like traditional EnKF method, then only it will be successful attempt for computing cost reduction. Therefore, to compare the results with base case, similar quantitative and qualitative analysis is carried out for all production parameters obtained in history matching that uses proposed workflow.

Quantitative analysis of initial ensemble as well as history matching results obtained using KL-PCE based EnKF workflow is shown in Table 11. Primarily, it can be observed from quantitative analysis that different measures like mean, standard deviation etc. for initial ensemble in Table 11 are different than those of base case in Table 8. In KL parameterization, each realization of the original 100-member ensemble is represented using few random variables (3 in our case) which are then used to obtain permeability of all grid blocks in a particular realization using truncated KL shown in Eqn. (3.4). Since, quantitative measures in Table 11 are shown for realizations obtained using truncated KL, they are different. However, it can be noticed that they are not much distant from measures of original ensemble meaning that a set of random variables can represent the original ensemble correctly in KL transformation. By comparing the different measures for oil rate after each update with base case, it can be noticed that all values are in the same range as base case. Results for cumulative oil in Table 11 are even better than base case as mean after final update is even closer to the true value. Also, higher R^2 value and lower RMSE indicates better history matching of cumulative oil production when proposed workflow is used. In addition, though

mean and standard deviation for steam oil ratio are almost same as base case, lower RMSE and positive R^2 value indicates appropriate update of state vector. Hence it is logical to use KL-PCE-EnKF workflow as not only comparable but also better results for some production parameters as compare to traditional EnKF can be obtained while consuming only 5% of computational time required originally.

Table 11: Quantitative analysis of production parameters after 1200 days at each update step obtained using KL-PCE-EnKF workflow

Quantitative Measures	Initial Ensemble	After 1st Update	After 2nd Update	After 3rd Update	After 4th Update
	Normalized Oil Rate SC after 1200 days (True value = 0.404)				
Mean	0.511	0.525	0.556	0.500	0.460
Std Dev	0.119	0.116	0.061	0.020	0.017
Min	0.157	0.165	0.392	0.439	0.430
Max	0.726	0.770	0.725	0.542	0.506
R^2	-0.400	-0.362	-0.337	0.161	0.401
RMSE	0.148	0.146	0.147	0.118	0.099
	Normalized Cum. Oil Production after 1200 days (True value = 0.562)				
Mean	0.611	0.625	0.653	0.601	0.554
Std Dev	0.112	0.091	0.048	0.030	0.028
Min	0.266	0.366	0.535	0.474	0.475
Max	0.756	0.748	0.777	0.672	0.672
R^2	0.863	0.886	0.906	0.978	0.987
RMSE	0.070	0.066	0.060	0.030	0.022
	Normalized Steam Oil Ratio after 1200 days (True value = 0.268)				
Mean	0.265	0.255	0.235	0.263	0.295
Std Dev	0.070	0.056	0.025	0.017	0.015
Min	0.194	0.197	0.197	0.237	0.238
Max	0.498	0.438	0.317	0.334	0.334
R^2	0.259	0.518	0.738	0.807	0.586
RMSE	0.053	0.046	0.036	0.031	0.046

To endorse the capability of proposed workflow, qualitative analysis of history matching results after each update is also shown. Figure 36 shows oil rate (grey

lines) obtained after each update when proposed workflow is used for assisted history matching. Plots of oil rate more or less resemble to those of base case in Figure 20 with ensemble spread decreasing and realizations converging towards true data (red line). However, frequent shoot-ups in oil rate can be seen in Figure 36. As PCE tries to accommodate all the changes in daily oil rate, value of deterministic coefficients keeps changing and sometimes combination of all coefficients predicts relatively higher or lower oil rate, mostly when changes in parameter are significant. Also, grey lines in Figure 36 represents whole ensemble but for one realization, there won't be many abrupt changes and hence it will not affect history matching adversely, however, additional constraints in the proxy model or smoothing of output can be considered if needed. Cumulative oil production is plotted in Figure 37 after each update. It can be observed that after final update, ensemble follows the true data and shows satisfactory match. In fact, most of the realizations in case of cumulative oil production are on the same side of true data after 4^{th} update in base case which is not evident in Figure 37 obtained using proposed workflow. In case of steam oil ratio which is shown in Figure 38, results capture the trend over the time and are similar to base case results. Thus, qualitative analysis of results also indicates good potential of the integrated KL-PCE-EnKF workflow.

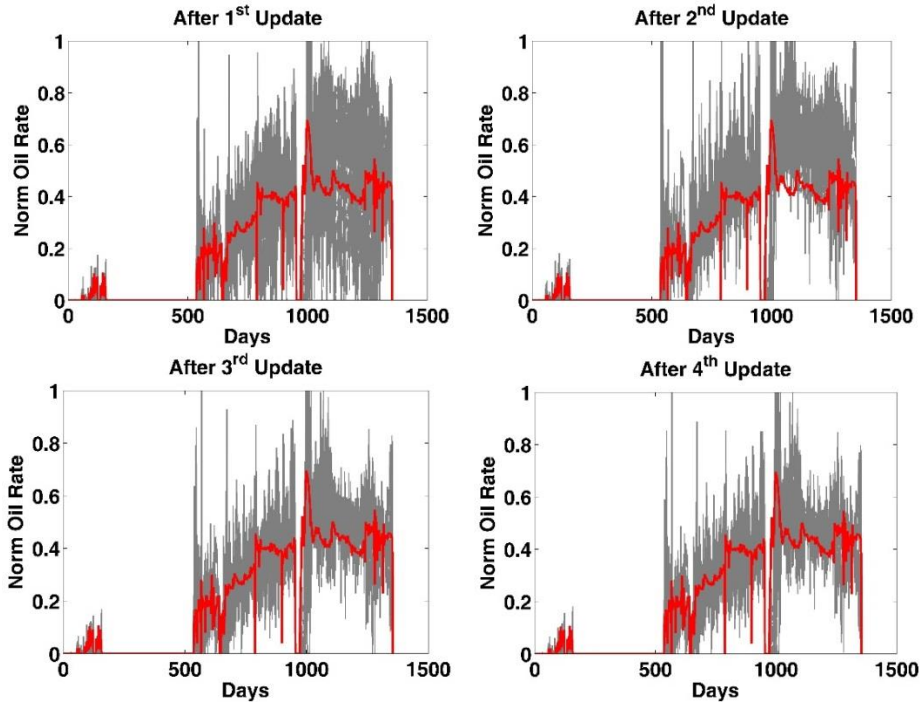


Figure 36: Normalized oil rate at each update step obtained after using KL-PCE-EnKF approach to update the ensemble (gray lines). The Red line shows history obtained from the field.

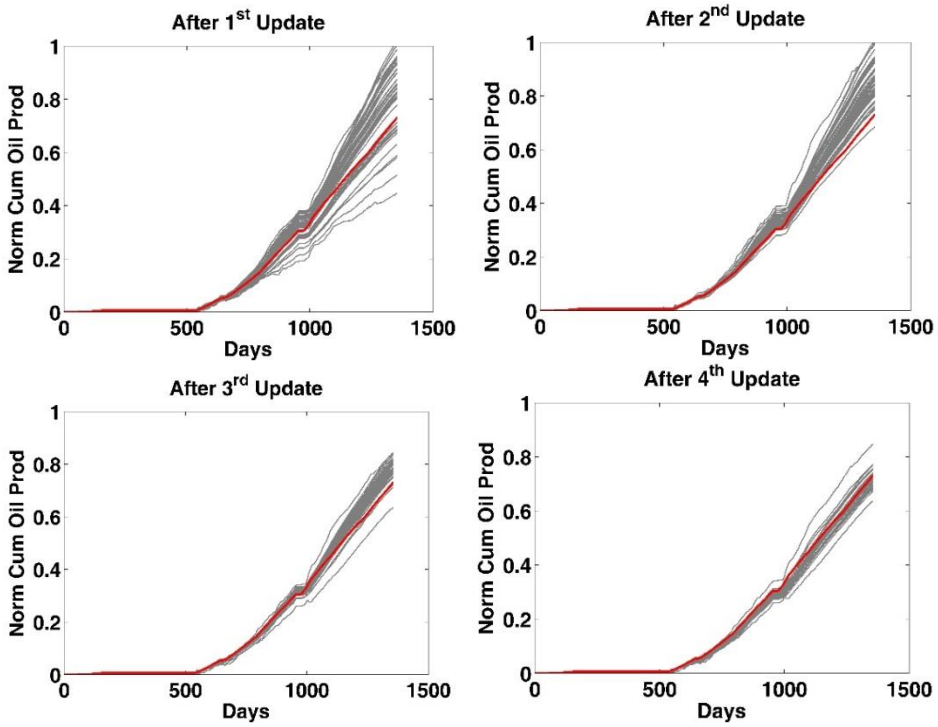


Figure 37: Normalized Cumulative oil production at each update step obtained after using KL-PCE-EnKF approach to update the ensemble (gray lines). The Red line shows history obtained from the field.

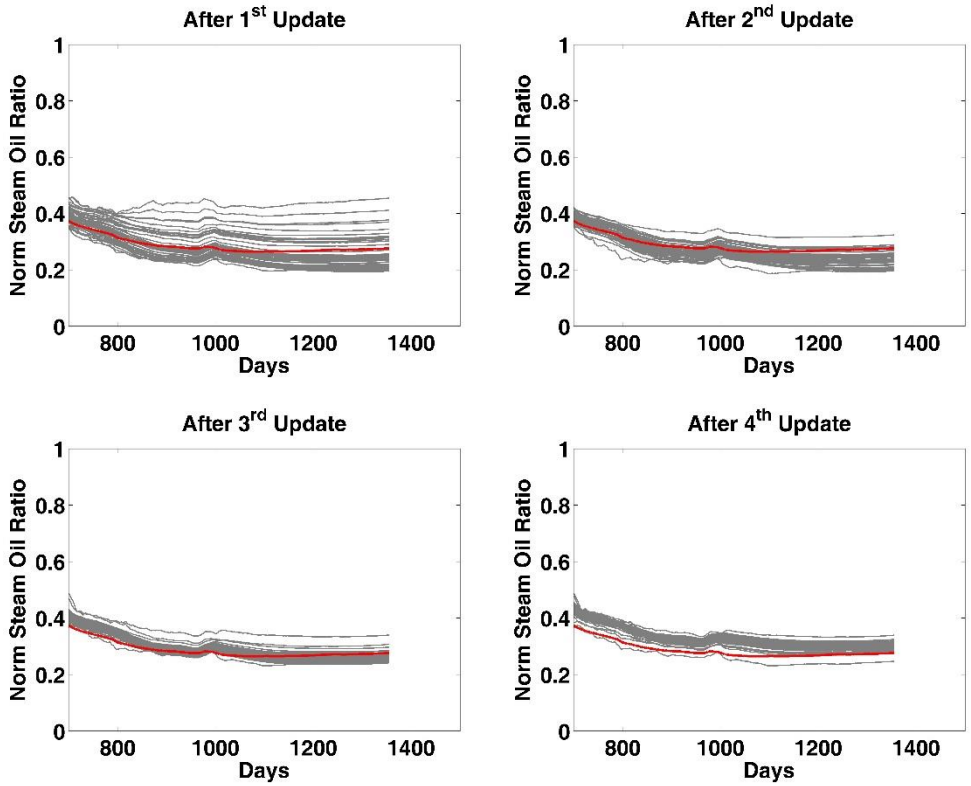


Figure 38: Normalized Steam oil ratio at each update step obtained after using KL-PCE-EnKF approach to update the ensemble (gray lines). The Red line shows history obtained from the field.

In Figure 39, box and whisker plots that demonstrate change in distribution are plotted after each update for all 3 production parameters at 1200 days using conventional EnKF method. In the plot, whisker ranges from minimum to maximum while box presents 0.25 quantile to 0.75 quantile of the distribution. Red line and + sign in interquartile range of each box and whisker plot shows median and mean of the distribution respectively while green horizontal line denotes the true value of the respective production parameter. It is evident from the plots that standard deviation of the distribution is decreasing with each update of EnKF and is minimum after final update for all production parameters. Also, mean in case of oil rate is closest to the true value (i.e. green line) after 4th update.

In case of cumulative oil production and steam oil ratio, mean after last update is not as close to the true value as it was in previous update. However, plots in Figure 39 shows the distribution of both production parameters only at 1200 days while Figure 22 and Figure 24 confirms acceptable closeness of ensemble mean to the true value for cumulative oil production and steam oil ratio over the time. Finally, proximity of the ensemble mean to the true value confirms the overall promising results in history matching of the base case.

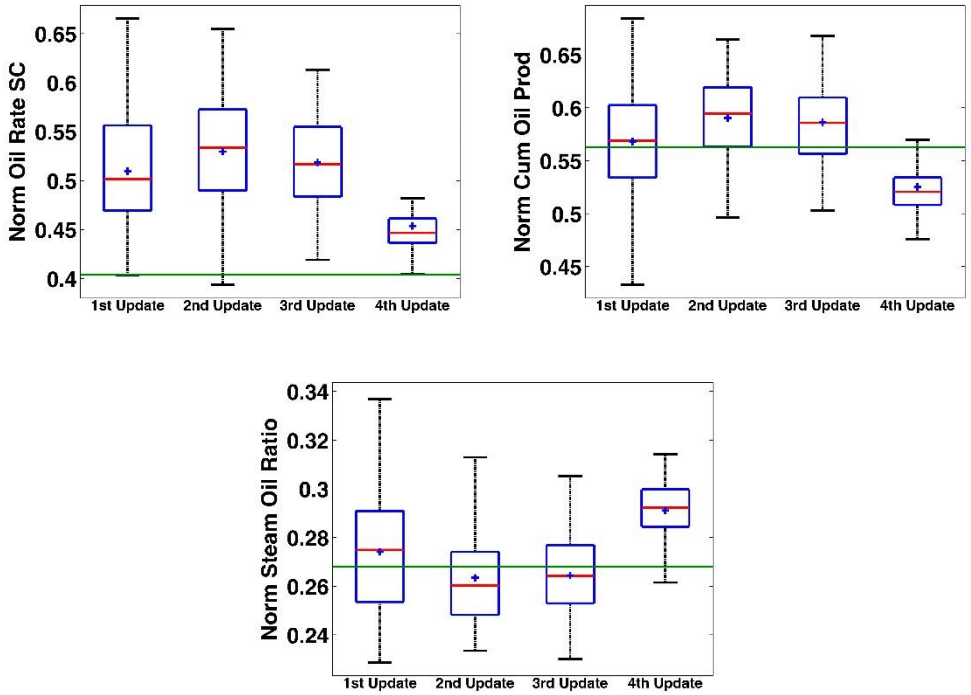


Figure 39: Box and whisker plots representing change in distribution of different production parameters at 1200 days after each update using conventional EnKF. Red line and '+' mark show median and mean of each distribution respectively while continuous green line depicts true value of particular production parameter

Finally, to compare the change in distribution of production parameters with base case at 1200 days, box and whisker plots are shown in Figure 40. Like base case, box spans from 0.25 quantile to 0.75 quantile of the distribution and whisker spans the entire range of distribution. Also, red line, green line and + sign denotes median, true data and mean respectively.

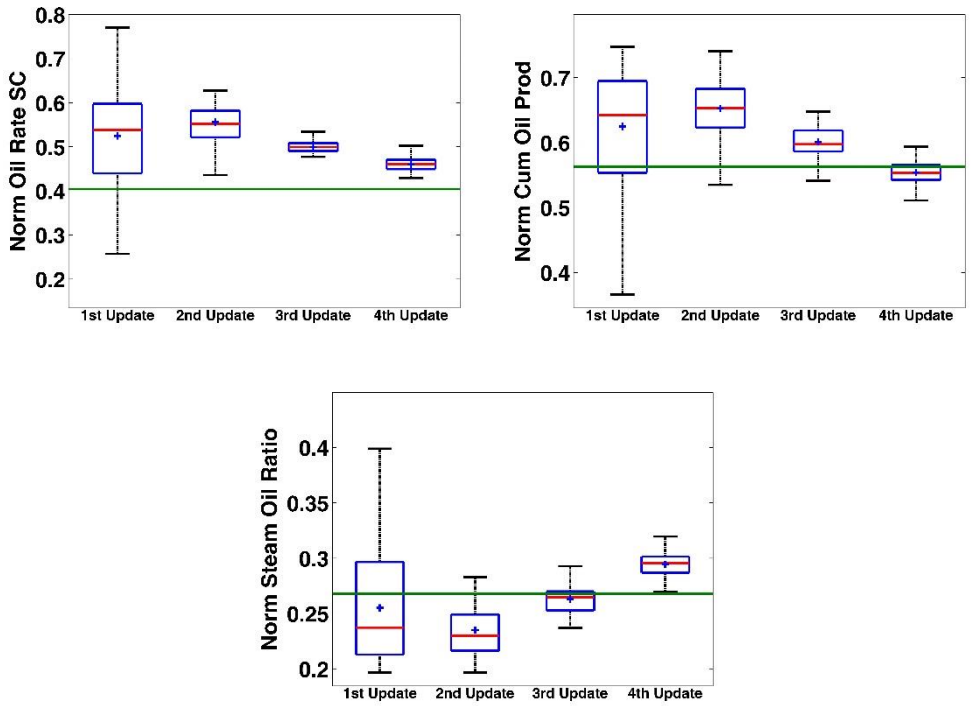


Figure 40: Box and whisker plots representing change in distribution of different production parameters at 1200 days after each update using proposed KL-PCE-EnKF workflow. Red line and '+' mark show median and mean of each distribution respectively while continuous green line depicts true value of particular production parameter

Change in distribution of different production parameters in Figure 40 reflects the perfect history matching of model parameters as standard deviation of distribution is decreasing gradually as oppose to some swift changes observed in base case, mostly in last couple of update steps. Also, mean is advancing towards the true value progressively instead of sudden change for almost all 3 production parameters. In fact, mean of cumulative oil production is nearly same as true value after last update. Furthermore, difference between mean and median is reducing after each update step meaning that biasedness in ensemble is diminishing since identical mean and median shows that same number of realizations in the ensemble is above and below the mean value. This analysis of change in distribution further establishes the effectiveness of proposed integrated KL-PCE-EnKF workflow.

It is also important to show the variance in model parameters after every update step using the proposed workflow. For an efficient application of EnKF approach, the ensemble variance for the model parameters must reach close to zero at the end of final update step. It can be observed from Figure 29 and Figure 41 that there is a significant reduction in the ensemble variance of model parameters at successive assimilation steps in conventional as well as KL-PCE based EnKF method. This shows the convergence of ensemble towards the true representation of permeability field with almost zero variance after fourth update step and hence demonstrates the efficacy of proposed method.

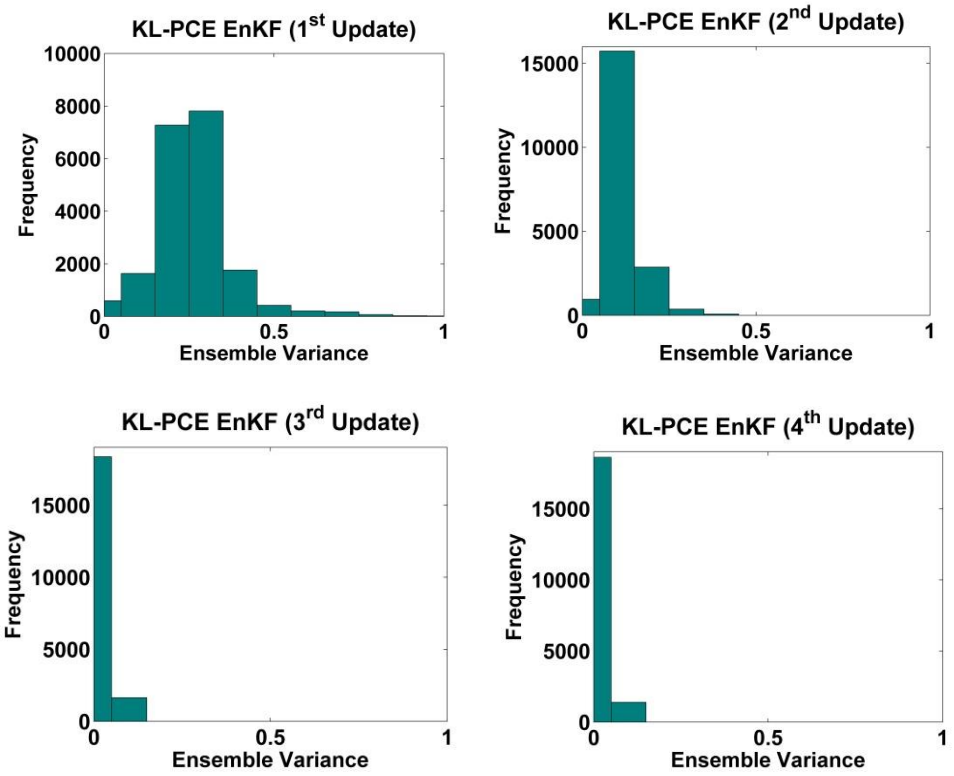


Figure 41: Ensemble variance of model parameters at various assimilation steps during KL-PCE-EnKF method.

It is important to note that in KL-PCE-EnKF, state vectors used in EnKF process are modified. In this approach, individual random variables, which represent permeability field through KL expansion, are used as model parameters in the initial ensemble instead of the natural logarithm of permeability. For obtaining credible history matching results, it is important that parameterization technique is competent enough to represent the initial ensemble. To investigate this fact, the comparison of permeability histogram before history matching is plotted for realizations for the base case and KL-PCE-EnKF is plotted in Figure 42. Irrespective of the different method used for representing the permeability field, the overall distribution is similar in both cases. Also, to ensure the integrity of the

history matching process, the occurrence of ensemble collapse should be examined. The variance of different grid blocks in all realizations of the ensemble becomes zero when ensemble collapse takes place. In other words, for all the grid blocks, values of model parameters (i.e. permeability in our case) become equal in different realizations, leading to apparently only one realization in the ensemble and hence loss of ensemble variability. Figure 43 depicts the updated permeability distribution of conventional EnKF and KL-PCE-EnKF. It is evident from this figure that only higher permeability values (which most likely are impractical values for the field) are filtered out.

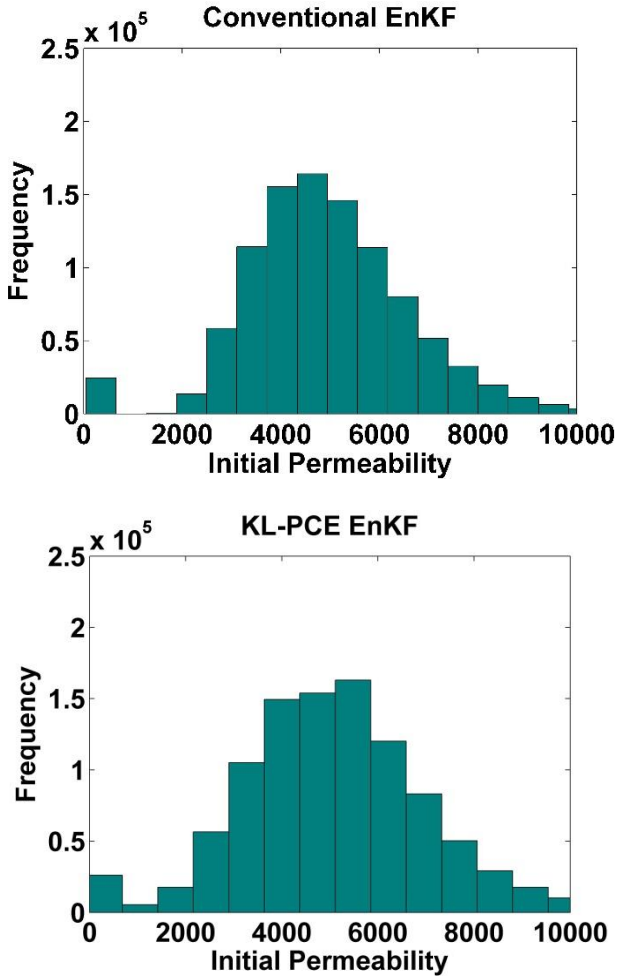
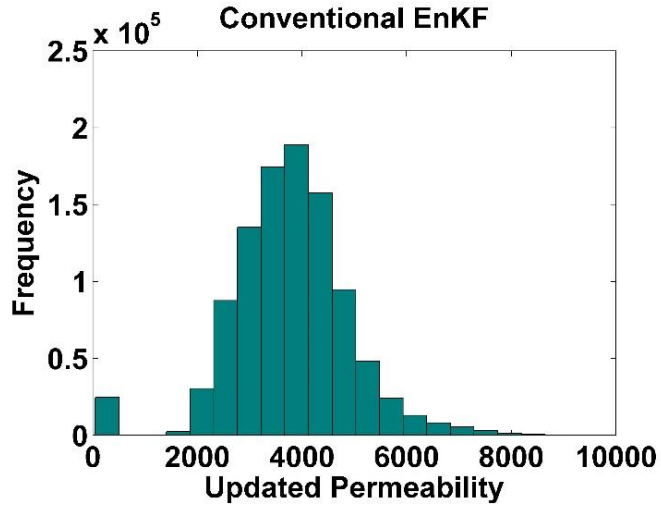


Figure 42: Histogram of Initial Permeability before history match for realizations used during conventional EnKF and KL-PCE-EnKF.



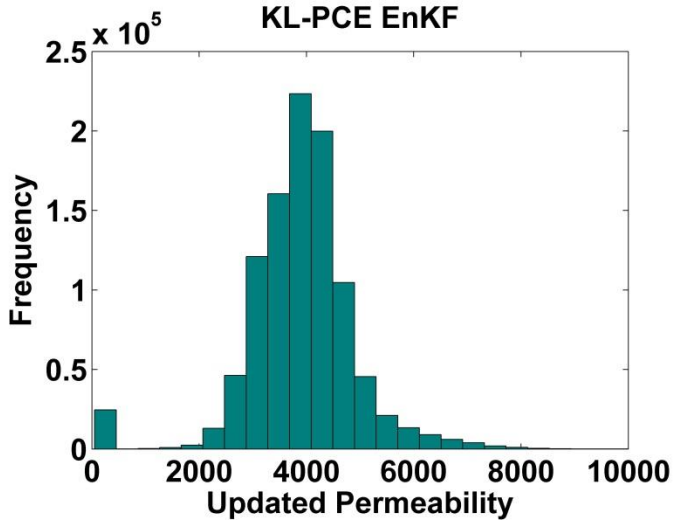


Figure 43: Histogram of Final updated Permeability after history match for realizations obtained during conventional EnKF and KL-PCE-EnKF.

5.4 Summary

In this work, a novel approach is presented to improve the assisted history matching process for SAGD reservoirs. To reduce the computational cost of assisted history matching, integrated KL-PCE-EnKF workflow is proposed. KL

transformation in the workflow parameterizes the initial ensemble using random variables which are then used to build PCE based mathematical model. Reservoir simulator in forecast step of EnKF is replaced by KL-PCE based proxy model and necessary changes are made in the state vector. Proposed approach is verified using a field scale SAGD reservoir model, results of which can be summarized as below:

- KL parameterization can effectively represent the ensemble using set of random variables while PCE is competent enough to calculate daily as well as cumulative production parameters, both of which can be used further in EnKF to estimate model parameters successfully.
- Implementation of proposed KL-PCE-EnKF workflow reduces the computational time required for assisted history matching by 95% for the given field case study without compromising the history matching results.

Chapter 6:

Conclusions

6.1 Conclusions

The following are the key findings of this work:

- Our proposed PCE approach, integrated with KL expansion, gives adequate results when compared with the outputs obtained from the numerical simulator. This can be used to run as many realizations as required in a relatively very small amount of time and with very low computational effort.
- The proposed PCE proxy model demonstrates better performance in comparison to the ANN and RBF-based proxy models when limited training data is used. The PCE framework is easy to implement even in highly heterogeneous and large reservoir models. The PCE model developed with the help of very few simulation runs is efficient enough to represent the full dynamics and trend of simulation results, thus eliminating the need for time-consuming simulation runs.
- In our case, increasing the number of terms in PCE improves the prediction performance, but not significantly. To develop a proxy model with the minimum computational requirement, PCE with fewer terms is recommended over PCE with more terms. In this case, the latter will unnecessarily increase computational time in terms of more simulation runs.

- ANN-based models can show improved performance if more training data is used.
- Data-driven proxy model based EnKF provides adequate results when compared with the outputs obtained from the commercial simulator. Developed model can be used to run as many realizations as required in a relatively very small amount of time and with very low computational effort.
- The proposed framework is easy to implement in large-scale complex reservoirs. Models developed with the help of very few simulation runs are efficient enough to imitate the full dynamics and trend of simulation results, thus eliminating the need for time-consuming simulation runs.
- Data-driven proxy models can easily be incorporated in EnKF framework as forecast model and can reduce the time required for assisted history matching, up to 90% for the case presented, while maintaining the accuracy of results.

6.2 Recommendations for Future Work

For use of proxy models during assisted history matching process, the following future research is recommended:

- In the proposed approach, permeability is used as input parameter for proxy development as well as in state vector during EnKF procedure for assisted history matching. Other parameters that can affect the desired output parameters can be included for proxy model development as well as in state vector of EnKF. For example, parameters like relative permeability, capillary pressure, temperature, well-logging data, and well-testing measurements.

- KL expansion method has been used to parameterize the Gaussian permeability field in this work for SAGD reservoir. However, different types of transformation techniques such as discrete cosine transform, optimization based PCA can be applied to characterize non-Gaussian reservoir models such as 2 facies model, channelized models.
- Application of proxy models with other history matching algorithms can also be investigated.

Bibliography

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

Agarwal, B., Hermansen, H., Sylte, J. E., and Thomas L.K.,1999. Reservoir Characterization of Ekofisk Field : A Giant , Fractured Chalk Reservoir in the Norwegian North Sea -- History Match, *SPE Reservoir Eval. & Eng*. **3** (6): 534-543

Akram, F. 2011. Multimillion-Cell SAGD Models — Opportunity for Detailed Field Analysis. Presented at World heavy Oil Congress, Canada. WHOC 11-534 (1–14).

Amirian, E., Leung, J. Y. W., Zanon, S. D. J., and Dzurman, P. J. 2014. An Integrated Application of Cluster Analysis and Artificial Neural Networks for SAGD Recovery Performance Prediction in Heterogeneous Reservoirs. Presented at SPE Heavy Oil Conference Canada, 10-12 June. SPE-170113-MS

Augustin, F., Gilg, A., Paffrath, M., Rentrop, P., and Wever, U. 2008. Polynomial chaos for the approximation of uncertainties: Chances and limits. *European Journal of Applied Mathematics*, **19**(2): 149–190.

Babaei, M., Alkhatib, A., and Pan, I. 2015. Robust optimization of subsurface flow using polynomial chaos and response surface surrogates. *Computational Geosciences*, **19**(5): 979–998.

Bazargan, H., 2014. *An efficient polynomial chaos-based proxy model for history matching and uncertainty quantification of complex geological structures*. Ph.D. Thesis, Heriot-Watt University, Edinburgh, UK.

Begum, N. 2009. *Reservoir Parameter Estimation for Reservoir Simulation using Ensemble Kalman Filter (EnKF)*. MSc Dissertation, Norges teknisk-naturvitenskapelige universitet, Trondheim.

Bianco, A., Cominelli, A., Dovera, L., Naevdal, G., and Valles, B. 2007. History Matching and Production Forecast Uncertainty by Means of the Ensemble Kalman Filter: A Real Field Application. Presented at SPE Europec/EAGE Annual Conference & Exhibition London, United kingdom, 11-14 June. SPE- 107161

Bookstein, F. L. 1989. Principal Warps: Thin-Plates Splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(6): 567–585.

Broomhead, D. S. and Lowe, D. 1988. Multivariable Functional Interpolation and

Adaptive Networks. *Complex Systems*, **2**, 321– 355.

Brouwer, D. R., Naevdal, G., Jansen, J. D., Vefring, E.H., and Kruijsdijk, C. P. J. W. Van. 2004. Improved Reservoir Management Through Optimal Control and Continuous Model Updating. Presented at SPE Annual Technical Conference & Exhibition Houston, Texas, 26-29 September. SPE- 90149.

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

Cameron, R. H., and Martin, W. T. 1947. The Orthogonal Development of Non-Linear Functionals in Series of Fourier-Hermite Functionals *Annals of Mathematics*, **48**(2): 385–392.

Cancelliere, M., Verga, F., and Viberti, D. 2011. Benefits and Limitations of Assisted History Matching. Presented at SPE Offshore Europe Oil & Gas Conference & Exhibition Aberdeen, United kingdom, 6-8 September. SPE-146278

Carlson, M. R. 2003. *Practical Reservoir Simulation*. Oklahoma: Penn Well Corporation.

Chang, H., and Zhang, D. 2009. A Comparative Study of Stochastic Collocation Methods for Flow in Spatially Correlated Random Fields. *Commun. Comput. Phys.*, **6**(3): 509–535.

Chen, Y., Oliver, D. S., and Zhang, D. 2009. Data assimilation for nonlinear problems by ensemble Kalman filters with reparameterization. *Journal of Petroleum Science and Engineering*, **66**(2009):1–14.

Chitralekha, S. B., Trivedi, J. J., and Shah, S. L. 2010. Application of the Ensemble Kalman Filter for Characterization and History Matching of Unconventional Oil Reservoirs. Presented at Canadian Unconventional Resources & International Petroleum Conference,Calgary, Canada, 19-21 October. SPE-137480.

Christie, M., MacBeth, C., and Subbey, S. 2002. Multiple history-matched models for Teal South. *The Leading Edge*, **21**(03): 286–289.

Computer Modelling Group (CMG). 2013a. *STARS User's Guide, Version 2013*. Calgary, Canada: Computer Modelling Group Ltd.

Computer Modelling Group (CMG). 2013b. *Results Report User's Guide, Version 2013*. Calgary, Canada: Computer Modelling Group Ltd.

Computer Modelling Group (CMG). 2013c. *Builder User's Guide, Version 2013*. Calgary, Canada: Computer Modelling Group Ltd.

Costa, L. A. N., Maschio, C., and Schiozer, D.J. 2014. Application of artificial neural networks in a history matching process. *Journal of Petroleum Science and Engineering* **123** (2014): 30–45.

Cullick, A. S., Johnson, D., and Shi, G. 2006. Improved and More Rapid History Matching With a Nonlinear Proxy and Global Optimization. Presented at SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 24-27 September. SPE-101933.

Dake, L. P. 2007. *Fundamentals of Reservoir Engineering*. Elsevier, Netherlands

Ertekin, T., Abou-Kassem, J. H., and King, G. R. 2001. *Basic Applied Reservoir Simulation*. Richardson: Society of Petroleum Engineers.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* **99** (5): 10143-10162.

Fajraoui, N., Ramasomanana, F., Younes, A., Mara, T. A., Ackerer, P., and Guadagnini, A. 2011. Use of global sensitivity analysis and polynomial chaos expansion for interpretation of nonreactive transport experiments in laboratory-scale porous media. *Water Resources Research*, **47**(2): 1–14.

Fedutenko, E., Yang, C., Card, C., and Nghiem, L.X. 2014. Time-Dependent Neural Network Based Proxy Modeling of SAGD Process. Presented at SPE Heavy Oil Conference Canada, 10-12 June. SPE-170085-MS

Ferreira, I., Gammiero, A., and LLamedo, M. 2012. Design of a Neural Network Model for Predicting Well Performance After Water Shutoff Treatments Using Polymer Gels. Presented at SPE Latin America and Caribbean Petroleum Engineering Conference, Mexico City, Mexico, 16-18 April. SPE-153908

Ghanem, R. G., and Spanos, P. D. 1991. *Stochastic Finite Elements: A Spectral Approach*. New York: Springer- Verlag.

Ghanem, R. G., & Spanos, P. D. 2003. *Stochastic Finite Elements: A Spectral Approach*. Courier Devor Publications.

Gomez, S., Gosselin, O., and Barker, J. W. 2001. Gradient-Based History Matching With a Global Optimization Method. *SPE Journal*, (February), 3–6.

Gu, Y., and Oliver, D. S. 2006. The Ensemble Kalman Filter for Continuous Updating of Reservoir Simulation Models. *Journal of Energy Resources Technology*, **128**(1): 79-87.

Haugen, V., Natvik, L.J., Evensen, G., Berg, A., Flornes, K., and Naevdal, G. 2006. History Matching Using the Ensemble Kalman Filter on a North Sea Field Case. Presented at SPE Annual Technical Conference and Exhibition,

San Antonio, Texas, 24-27 September. SPE-102430.

He, J., Sarma, P., and Durlofsky, L.J. 2011. Use of reduced-order models for improved data assimilation within an EnKF context. Presented at SPE Reservoir Simulation Symposium, The Woodlands*, Texas,* 21-23 February.

Holland, J. 1975. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. *The MIT Press,* Cambridge, England

Huang, S. P., Quek, S. T., and Phoon, K. K. 2001. Convergence study of the truncated Karhunen-Loeve expansion for simulation of stochastic processes. *International Journal for Numerical Methods in Engineering*, **52** (9): 1029–1043.

Isukapalli, S. S. 1999. *Uncertainty Analysis of Transport-Transformation Models*. New Brunswick, New Jersey.

Jafarpour, B., and McLaughlin, D. B. 2008. History matching with an ensemble Kalman filter and discrete cosine parameterization. *Computational Geosciences*, **12** (2): 227–244.

Junker, H. J., Plas, L., Dose, T., and Little, A. 2006. Modern Approach to Estimation of Uncertainty of Predictions With Dynamic Reservoir Simulation-A Case Study of a German Rotliegend Gasfield. Presented at SPE Annual Technical Conference, San Antonio, Texas, 24-27 September. SPE-103340

Kaleta, M. P. 2011. *Model-reduced gradient-based history matching*. MSc Dissertation, Delft University of Technology, Netherlands.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. 1983. Optimization by Simulated Annealing. *Science*, **220** (4598): 671–680.

Li, B., and Friedmann, F. 2005. Novel Multiple Resolutions Design of Experiment/Response Surface Methodology for Uncertainty Analysis of Reservoir Simulation Forecasts. Presented at SPE Reservoir Simulation Symposium, Houston, Texas, 31 January-2 February. SPE-92853-MS.

Li, B., and Friedmann, F. 2007. Semiautomatic Multiple Resolution Design for History Matching. *SPE Journal*, **12** (4): 24–27. SPE 102277

Li, H., Chang, H., and Zhang, D. 2009. Stochastic Collocation Methods for Efficient and Accurate Quantification of Uncertainty in Multiphase Reservoir Simulations. Presented at SPE Reservoir Simulation Symposium, The Woodlands, Texas, 2-4 February. SPE 118964

Li, H., Sarma, P., and Zhang, D. 2011. A Comparative Study of the Probabilistic-Collocation and Experimental-Design Methods for Petroleum-Reservoir

Uncertainty Quantification. *SPE Journal*, **16** (2): 429–439.

Liang, B. 2007. *An Ensemble Kalman Filter Module For Automatic History Matching.* Ph.D Thesis, The University of Texas at Austin.

Lorentzen, R. J., Fjelde, K. K., Froyen, J., Lage, A. C. V. M., Naevdal, G., and Vefring, E. H. 2001. Underbalanced and Low-head Drilling Operations: Real Time Interpretation of Measured Data and Operational Support. Presented at SPE Annual Technical Conference, New Orleans, Louisiana, 30 September-3 October. SPE-71384-MS

Ma, Z., Liu, Y., Leung, J. Y., and Zanon, S. 2015. Practical Data Mining and Artificial Neural Network Modeling for SAGD Production Analysis. Presented at SPE Canada Heavy Oil Technical Conference, Calgary, Canada, 9-11 June. SPE-174460-MS

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** (02): 83–92.

MathWorks. 2014. *MATLAB® Release R2014a.* The MathWorks Inc., Natick, Massachusetts.

MathWorks. 2013. *Neural Network Toolbox™ Release R2013a.* The MathWorks Inc., Natick Massachusetts

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. 1953. Equation of state calculations by fast computing machines. *Journal Chemical Physics*, **21** (6): 1087–1092.

Mitchell, H. L., and Houtekamer, P. L. 2000. An adaptive ensemble Kalman filter. *Monthly Weather Review*, **128** (2): 416–433.

Naevdal, G., Johnsen, L. M., Aanonsen, S. I., and Vefring E.H. 2003. Reservoir Monitoring and Continuous Model Updating Using Ensemble Kalman Filter. Presented at SPE Annual Technical Conference, Denver, Colorado, 5-8 October. SPE-84372.

Nejadi, S. 2014. *Re-Sampling the Ensemble Kalman Filter for Improved History Matching and Characterizations of Non-Gaussian and Non-Linear Reservoir Models*. Ph.D Thesis, University of Alberta, Canada.

Oladyshkin, S., and Nowak, W. 2012. Polynomial Response Surfaces for Probabilistic Risk Assessment and Risk Control via Robust Design, Dr. Yuzhou Luo (Ed.), InTech, 317–344.

Oliver, D. S., and Chen, Y. 2011. Recent progress on reservoir history matching: A review. *Computational Geosciences*, **15** (1): 185–221.

Oliver, D. S., Reynolds, A. C., and Liu, N. 2008. *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. Cambridge: Cambridge University Press.

Patel, R. G., Trivedi J.J., Rahim, S., and Li, Z. 2015. Initial Sampling of ensemble for steam-assisted-gravity-drainage-reservoir history matching. *Journal of Canadian Petroleum Technology* **54** (6): 424-441. SPE-178927-PA.

Peng, C.Y. and Gupta, R. 2004. Experimental Design and Analysis Methods in Multiple Deterministic Modeling for Quantifying Hydrocarbon In-Place Probability Distribution Curve. Presented at SPE Asia Pacific Conference on Integrated Modeling and for Asset Management, Kuala Lumpur, Malaysia, 29-30 March. SPE-87002.

Queipo, N. V., Goicochea, J. V., and Pintos, S. 2002. Surrogate modeling-based optimization of SAGD processes. *Journal of Petroleum Science and Engineering*, **35** (1-2): 83–93.

Rahim, S., Li, Z., and Trivedi, J. 2015. Reservoir Geological Uncertainty Reduction: an Optimization-Based Method Using Multiple Static Measures. *Mathematical Geosciences*, **47** (4): 373–396.

RamaRao, B. S., LaVenue, A. M., Marsily, G. D., and Marietta, M. G. 1995. Pilot Point Methodology for Automated Calibration of an Ensemble of conditionally Simulated Transmissivity Fields: 1. Theory and Computational Experiments. *Water Resources Research*, **31** (3): 475–493.

Reynolds, A. C., He, N., Chu, L., and Oliver, D. S. 1996. Reparameterization Techniques for Generating Reservoir Descriptions Conditioned to Variograms and Well-Test Pressure Data. *SPE Journal*, **1** (4): 413-426

Romary, T. 2009. Integrating production data under uncertainty by parallel interacting Markov chains on a reduced dimensional space. *Computational Geosciences*, **13** (1): 103–122.

Romeu, R. K. 2010. History Matching and Forecasting. *JPT*, (April), 80–88.

Sambridge, M. 1999. Geophysical inversion with a neighborhood algorithm - I. Searching a parameter space. *Geophysical Journal International*, **138** (2): 479–494.

Sarma, P. 2006. *Efficient Closed-Loop Optimal Control of Petroleum Reservoirs under Uncertainty*, Ph.D Thesis, Stanford University.

Slotte, P. A., and Smørgrav, E. 2008. Response Surface Methodology Approach for History Matching and Uncertainty Assessment of Reservoir Simulation Models. Presented at SPE Europec/EAGE Annual Conference & Exhibition Rome, Italy, 9-12 June. SPE 113390.

Sudret, B., and Der Kiureghian, A. 2000. *Stochastic Finite Element Methods and Reliability A state-of-the-Art Report*. Department of Civil & Environmental Engineering, University of California, Berkeley.

Vanegas Prada, J. W., and Cunha, L. B. 2008. Assessment of Optimal Operating Conditions in a SAGD Project by Design of Experiments and Response Surface Methodology. *Petroleum Science and Technology*, **26** (17): 2095–2107.

Wang, C., Li, G., and Reynolds, A. C. 2007. Production Optimization in Closed-Loop Reservoir Management. Presented at SPE Annual Technical Conference and Exhibition, Anaheim, California, 11-14 November. SPE 109805

Webster, M., Tatang, M. A., and McRae, G. J. 1996. Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model. Report Massachusetts: MIT Joint Program on the Science and Policy of Global Change.

Wiener, N. 1938. The Homogenous Chaos. *American Journal of Mathematics*, **60** (4): 897–936.

Williams, G. J. J., Mansfield, M., MacDonald, D. G., and Bush, M. D. 2004. Top-down reservoir modeling. Presented at SPE Annual Technical Conference and Exhibition, Houston, Texas, 26-29 September. SPE 89974

Xiu, D., and Karniadakis, G. E. 2003. Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics*, **187** (1): 137–167.

Zangl, G., Graf, T., and Al-Kinani, A. 2006. Proxy modeling in production optimization. Presented at SPE Europec/EAGE Annual Conference and Exhibition, Vienna, Austria, 12-15 June. SPE- 100131-MS

Zhan, W., Reynolds, A. C., and Oliver, D. S. 1999. Conditioning Geostatistical Models to Two-Phase Production Data. *SPE Journal*, **4** (2): 27–30.

Zhang, D., and Lu, Z. 2004. An efficient, high-order perturbation approach for flow in random porous media via Karhunen-Loeve and polynomial expansions. *Journal of Computational Physics*, **194** (2): 773–794.

Zubarev, D. I. 2009. Pros and Cons of Applying Proxy-Models as a Substitute for Full Reservoir Simulations. Presented at SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 4-7 October. SPE-124815