Permeability Estimation Using a Non-Parametric Bayesian Network

14th February 2015

Richard Tymotheus Purba
Title : Permeability Estimation using a Non-Parametric Bayesian Network

Author(s) : Richard Tymotheus Purba

Date : 14 February 2015

Professor(s): Prof. Dr. Ir. Jan-Dirk Jansen; Prof. Dr. Ir. Arnold Heemink

Supervisor(s): Aurelius A. Zilko, MSc

Postal Address: Section for Petroleum Engineering
Department of Geoscience & Engineering
Delft University of Technology
P.O. Box 5028
The Netherlands

Telephone : (31) 15 2781328 (secretary)
Telefax : (31) 15 2781189

Copyright ©2015 Section for Petroleum Engineering

All rights reserved.
No parts of this publication may be reproduced,
Stored in a retrieval system, or transmitted,
In any form or by any means, electronic,
Mechanical, photocopying, recording, or otherwise,
Without the prior written permission of the
Section for Petroleum Engineering
Abstract

This thesis addresses the problem of permeability estimation of an oil reservoir using oil production data, a process known in the oil industry as computer-assisted history matching or data assimilation. Zilko (2012) compared the well-known Ensemble Kalman Filter method for data assimilation with the Non-Parametric Bayesian Beliefs Network (NPBN) approach to recover a permeability field in a 2D reservoir model. Due to computational limitations of the NPBN method, this earlier study estimated the permeability field in separate regions, resulting in an image with discontinuities between the regions and avoid these discontinuities by using interpolation. We test the application of NPBN in the Egg Model, an ensemble of 100 channelized, 3D reservoir models often used for history matching studies. Because the NPBN methods requires many more model realizations than the available 100, we use the Stanford Geostatistics Modeling Software (SGeMS) and its algorithms, including the SNESIM algorithm, to generate additional realizations. Thereafter, we perform a twin experiment, where the MATLAB Reservoir Simulation Toolbox (MRST) of SINTEF is used for two-phase fluid flow simulation. We test several interpolation methods for the permeability field: Cubic, Spline, and Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) techniques. To measure the performance of the NPBN and interpolation in recovering the permeability field, we use the root mean square error between the “true” and the recovered permeability fields. We also compare the change in liquid production/injection profiles. Visual observations of the results of the data assimilation do not show entirely impressive result for the Egg model with the number of wells and Gaussian noise that are used in this project. The NPBN method produces a permeability field which is not identical to the truth, and the forecasted production profiles of the estimated models also do not match the truth. They are different by the factor of, approximately, two. The addition of new measurement at later times does not improve the permeability field. As the correlation of the measurement and the state variables is very high at the first time step, the addition of more wells in the system is thought to be more significant than performing more
measurement over time. The interpolation using the Spline method and the PCHIP method give very little difference to recover the remaining, unassimilated permeability field. However, at the peripheral area, where the extrapolation is conducted, the PCHIP method gives a better approximation, as indicated by a relatively lower absolute difference.

**Keywords:** Non-Parametric Bayesian Network, Data Assimilation, Reservoir Simulation, Parameter Estimation, Computer Assisted History Matching, Interpolation
In dedication to my mother, Ir. Orleans Ginting, for her hard work to facilitate education for the children and for her unwavering emotional support over the making of this thesis.
Acknowledgement

I have always captivated to search the answer for the question “*Quid est Veritas.*” Partly, that is why I took seriously the problem of history-matching in my thesis: how to know the truth of the reservoir. Moreover, I keep thinking how to make a scientific solution in a commercial framework. Although the latter has not been successfully realized, I would like to express my gratitude to some persons during the making of this thesis.

I would like to thank Aurelius Zilko for his innumerable support during the making of this thesis. I am probably not the best student in mathematics, but your resilience and patience help me to understand the magic and wonder of the subject. Perhaps, it was my stubbornness and perseverance that made me really fond of your in-depth critics and suggestions. *Terima kasih Mas!* My sincere gratitude goes to other thesis committee, Prof. Jan-Dirk Jansen and Prof. Arnold Heemink, and my previous supervisor Dr. Anca Hanea. I am grateful to be given an opportunity to expand my curious mind. Taking this novel research is one of my rarest scientific research opportunity in my life, so far. I thank also Pascal Schmidt, my study advisor, for his understanding with my circumstances. I thank John Stals for his RAS support as well.

For the computer, I thank David Ginting to give me the privilege of using his exceptional laptop. *Bujur, meuah-juah* David. I would like to thank to the developer of mGstat, SGeMS, and MRST open source software that allows a rigorous reservoir modeling and simulation research. I thank Dan Ababei, developer of UNINET, to use the software for this research and explanation on the interaction between MATLAB and UNINET.

The struggle for the making of this thesis has been a great experience to know myself. Having a part-time job and doing a novel research simultaneously was not a healthy combination. I thank my mother for her emotional support during the depression and great tiresome. I thank my sister Sophia Purba for her motivational suggestion during the time of need. Paul the Apostle inspires me during the hardships, where he wrote “*Three times I pleaded with the Lord to take it [suffering] away from me. But he said to me, ‘My grace is sufficient for you, for my power is made perfect in weakness.’ Therefore I will boast all the more gladly about my weaknesses, so that Christ’s power may rest on me. That is why, for Christ’s sake, I delight in weaknesses, in insults, in hardships, in persecutions, in difficulties. For when I am weak, then I am strong.*” I finish the thesis, thereby my Master program, with the faith, hope and love that is living within me.
# Table of Contents

1  Introduction .................................................................................................................. 7  
   1.1  Inverse Problems and Non-Parametric Bayesian Networks ............................. 8  
   1.2  The Purpose of the Research .............................................................................. 10  
   1.3  Organization of Thesis Report ............................................................................ 10  
2  The NPBN and The Experimental Setup .................................................................... 11  
   2.1  The NPBN ............................................................................................................ 11  
      2.1.1  State-space Models and Bayesian Network ............................................. 11  
      2.1.2  Copula and NPBN ...................................................................................... 12  
   2.2  The Generation of Initial Model ........................................................................ 12  
      2.2.1  Twin Experiment and Measurement Generation ..................................... 14  
   2.3  The State Vector and Initialization .................................................................... 15  
   2.4  Egg Model Partitioning for Data Assimilation ................................................. 17  
3  Interpolation ............................................................................................................... 20  
   3.1  Cubic Interpolation .............................................................................................. 21  
   3.2  PCHIP Interpolation ........................................................................................... 21  
   3.3  The Spline Interpolation Technique .................................................................. 24  
   3.4  Interpolation Result for Bimodal Distribution .................................................. 26  
4  Investigation of the Result ......................................................................................... 28  
   4.1  Comparison and Analysis of the Result .............................................................. 28  
      4.1.1  Visual Analysis ............................................................................................. 28  
      4.1.2  Root Mean Square Error and Absolute Difference ............................... 31  
   4.2  Forecasting Well Profiles .................................................................................... 34  
5  Conclusion and Recommendation ............................................................................. 45  
   5.1  Conclusion .......................................................................................................... 45  
   5.2  Recommendation ................................................................................................. 46
1 Introduction

Parameters estimation of a petroleum reservoir for an oil company is important to understand the reservoir as an asset. Gaining more knowledge allows the company to make a better strategy for decision-making process and other critical things. Being a porous medium, Darcy’s Law describes a petroleum reservoir parameter relation. One of the parameters that is important is the permeability, which is a measure of the ability of the porous medium to transmit fluid.

Computer modeling and simulation help the company to model the parameters relationship. The repeatability of these parameters relationship, hence the computer model, gives the power of prediction of the reservoir. Reservoir simulation deals with this prediction, which is generally called forward simulation. However, a model which reflects the reality has been a great challenge for reservoir simulation. Figure 1–1 shows the description of the problem.

![Figure 1–1 Production Profile of a reservoir](image)

Figure 1–1 Production Profile of a reservoir, where q(t) is the production rate of the oil as a function of time. The blue, solid line is the solution from the computer model. The star is the field measurement. The solid–dot, red, vertical line is the current time. The dashed, black line is the forecasting of the computer model.

The response of the reservoir simulation is the blue line. The red, solid–dot, vertical line is time at now. At this point, we want to know what would be the behavior of the oil rate at some time in the future. The verification between the simulation of the model and the reality is usually done to evaluate the accuracy of the model is in reservoir simulation. It is shown from Figure
1–1 that the Field Measurement is not in agreement with the prediction of the model (dashed black line), which is usually the case in the industry.

1.1 Inverse Problems and Non–Parametric Bayesian Networks

In a reservoir model, there usually exists difference between the computer model and the observed data due to the lack of knowledge of the reservoir. The challenge to match this difference is usually called *history–matching* problem.

To summarize the entire process in a petroleum industry to approximate the history–matching problem is described in the following workflow (adapted from Caers [2011]):

![Schematic workflow for solution of History Matching problem](image)

Each of the box represents every element for modeling, simulating and verifying the reservoir model. The dark–yellow boxes are the elements for geological modeling. There are various processes within this procedure, which are outside the scope of this thesis. The blue boxes deal with the numerical simulation of the reservoir model. The red box is the observation. The mismatch is the difference between the reservoir model result and the field measurement. Hence, we modify the flow model parameter or the geological parameter.

Inverse problem, within the Bayesian framework, is the study to approach the Likelihood of a parameter given the measured data. Data assimilation is a subject where we perform inverse problem automatically as a solution for the history–matching problem. We study a relatively new method to approach the Likelihood.

Non–Parametric Bayesian Network (NPBN) is a method for data assimilation. Gheorge (2010) introduced the NPBN method, as an alternative
to the state of the art method of Ensemble Kalman Filter (EnKF). She used a 2-dimensional reservoir and analyzed the performance of the NPBN using Gaussian noise as the measurement noise. The method appears promising because the computational speed is not expensive, which takes seconds to complete the conditionalization with the size of 1000 sample, and it produces good result as well. However, she discovered a limitation in the number of the variables which can be involved in the model. With 900 samples, the NPBN appears to be able to handle around 100 variables. Hence, it means that the NPBN can recover only part of the whole reservoir field.

Zilko (2012) extended this research, with the same 2-dimensional reservoir with five wells. In his research, he included saturation in the model. To tackle the dimensionality problem, he performed the NPBN method only on several parts of the reservoir, and he combined the result by using interpolation. He used the Cubic interpolation as the interpolation method. Moreover, the measurements had been generated with non-Gaussian noise. In order to measure the performance of NPBN, he used the RMSE and the forecasting results of the estimated model. Zilko (2012) showed that the NPBN appears as a promising technique for data assimilation.

Previous projects on making comparison between NPBN and EnKF show novel performance result of recovering heterogeneous permeability field, whereas the EnKF estimation has very smooth transition between very different permeabilities. Moreover, the spatial dependence between the permeability variables has different result for both of the two methods. The NPBN analysis reveals positive and negative correlation between variables. In comparison, EnKF finds, predominantly, positively correlated variables. These comparison motivates this project to study NPBN in advance.

This study extends the work of Zilko (2012). First, a more realistic model of a 3-dimensional model is used. The field is a channelized reservoir and it has more wells. Initial realizations of this type of reservoir are made. The Gaussian distribution is chosen for the measurement noise. To recover the unassimilated area, a number of interpolation techniques are applied. This project
uses the Cubic Interpolation and two more interpolation techniques: the Spline method and the PCHIP method.

1.2 The Purpose of the Research

The aim of this project is the evaluation of NPBN as a data assimilation method for a more realistic, 3-dimensional reservoir, namely the Egg model. We estimate the permeability as the spatial parameter and study different interpolation methods, namely Cubic, Spline, and PCHIP interpolation method, for recovering the remaining part of the permeability after the data assimilation. The verification of the hypothesis will be done using visual analysis (geologically satisfactory image) and the quality of history matching using forecasting (liquid production/injection profile).

1.3 Organization of Thesis Report

The report starts, in Chapter 1, with an introduction of the history matching problem and the use of NPBN method in data assimilation. For a more theoretical background of this thesis, interested readers are referred to Gheorge (2010) and Zilko (2012). In Chapter 2, we provide the theoretical background of the NPBN method. In the same chapter, we define the state vector and partition the permeability field. The realizations of the channelized Egg model is explained, including populating the permeability variable. The twin experiment is conducted: one realization is picked from the ensemble as the truth and treat the rest for data assimilation. Generation of measurements is obtained from the chosen truth. In Chapter 3, the remaining, unassimilated permeability field is recovered with interpolation. We start with Cubic interpolation then move on to the Spline and PCHIP interpolation method. Chapter 4 investigates the result of the interpolation, data assimilation, and the updated state. The thesis is closed by conclusion and recommendation in Chapter 5.
2 The NPBN and The Experimental Setup

2.1 The NPBN

2.1.1 State-space Models and Bayesian Network

Any model which has observation process $Y_t$ and a state process $K_t$ at current time $t$ is a state-space model. The objective is generally to estimate a hidden state which we would like to estimate. The evolution of the state vector occurs in time and we want to infer the hidden state. This is done recursively using the Bayes' rule: we need a prior ($P(K_1)$), a state-transition function ($P(K_t|K_{t-1})$), and an observation function/likelihood $P(Y_t|K_t)$.

A Bayesian Networks (BN) contains the directed acyclic graph (DAG) and a set of probability functions implied by the graph. A dynamic BN, which is collection of static BN linked in discrete slices of time, allow a revealing representation for a state-space. The structure of a DAG is defined by two sets: the set of nodes, which represents univariate random variables, and the set of arcs, which represents the flow of influence between the variables.

Consider a BN containing $n$ nodes, random variable $K_1$ to $K_n$, taken in that order. We are interested to calculate a particular value in the joint distribution $P(K_1 = k_1, K_2 = k_2, K_3 = k_3, ..., K_n = k_n)$, or more compactly as $P(k_1, k_2, k_3, ..., k_n)$, where we can factorize the joint distribution (Pearl, 2000):

$$P(k_1, k_2, k_3, ..., k_n) = \prod_j P(k_j|k_1, ..., k_{j-1}); j \text{ from } 1 \text{ to } n$$

The term $k_1, ..., k_{j-1}$ at the right-hand side of the equation above, which is usually called parents, is the direct predecessors of a node $k_j$. Hence, the influence flows from parents to that node. The arcs in the BN is quantified by the terms in the right-hand side of the above formula.

---

1 Generally, the state process is represented as $X_t$, whereas in this project we use $K_t$ to represent permeability as the estimated state, which will be explained later.
2.1.2 Copula and NPBN

As the name suggests, the NPBN method utilizes a BN in order to model the dependency between the variables. Given the structure of the DAG, the NPBN models the dependency implied by the structure by means of a set of (conditional) copula. Copula is the joint distribution in \(d\)-dimensional unit cube. Sklar (1959) defines the following relation

\[ F_{XY}(x, y) = C(F_X(x), F_Y(y)) \]

where \(X\) and \(Y\) represent the random variables, joined by the copula \(C\), \(F_{XY}\) is the cumulative joint distribution function of \(X\) and \(Y\), and \(F_X\) and \(F_Y\) are the marginal of each corresponding random variable distribution. Sklar (1959) introduced this concept in order to decompose \(n\)-dimensional distribution function \(F\) into two parts: the marginal distribution functions \(F_i\), and the copula \(C\).

One copula that is of interest in this project is the Normal Copula. The bivariate Normal Copula is defined as follows

\[ C_{\rho}(u, v) = \Phi_{\rho}(\Phi^{-1}(u), \Phi^{-1}(v)) \quad u, v \in [0,1]^2 \]

where \(\Phi_{\rho}\) is the bivariate normal cumulative distribution function (cdf) with correlation \(\rho\), and \(\Phi^{-1}\) represents the inverse of the standard univariate normal distribution. This copula is of interest in this project since it allows rapid updating (Gheorge, 2010, Chap. 3).

In the NPBN, the one dimensional marginal distribution is taken from the empirical distribution of each variable and the NPBN model assumes that the normal copula underlies the dependence between variables (Gheorge, 2010, pg. 40).

2.2 The Generation of Initial Model

Jansen et al., (2013[a]) developed, as part of the PhD dissertation of Maarten Zanvliet and Gijs van Essen, a 3-dimensional reservoir model with a dimension of \(60\times60\times7\) cells, and the size of \(8\times8\times4\) meter. There are of 25,200 grid cells in total, which is a rectangular shape reservoir. The shape is reduced into 18,533 grid cells because there are inactive cells, leaving an egg-shaped reservoir model. This model is a geological model with channels, the high permeability facies, and background mud, the low permeability facies.
It has eight injection wells and four production wells, where perforation is made at every layer.

The channelized feature in the Egg model is the peculiar geological shape in the studied model. Previously made in computer assisted drawing tools to reproduce the channel, this project attempts to create independent and identically distributed samples where each of them has channel feature. The project utilizes the SNESIM algorithm (Strebelle, 2000; Remy, Boucher, and Wu, 2009) that uses image for the spatial architecture, which is usually called Training Image. However, since training image for SNESIM algorithm demands stationarity in the image, the TIGENERATOR algorithm is utilized to create the training image and to satisfy such condition. For complete explanation of this algorithm, the reader is referred to Remy, et. al (2009), while the complete parameters are given in Appendix B.

The author uses MATLAB platform to work for the entire project. At the first stage, to create realizations, the author uses mGstat to interact between MATLAB and Stanford Geostatistical Modeling Software (SGeMS). After the spatial distribution per each sample is made, the SGSIM algorithm is applied to populate the facies, where the mean and covariance at any location in the reservoir is estimated by using Simple Kriging. The entire algorithm can be found at Jensen, Lake, Corbett, and Goggin (2000). Some of the final results of the ensemble is shown in Figure 2-1.

Gheorge (2010) observes that the NPBN appears to be able to work with approximately 100 variables at a time, with a size of 900 samples. In this project, we design a set of 1000 samples.

The whole procedures produce the mean of sand-permeability spatial distribution is $5 \times 10^{-13} \text{m}^2$ (5 Darcy) and its variance is $1.042 \times 10^{-26} \text{m}^2$. The distribution of the shale-permeability has the mean of $3 \times 10^{-15} \text{m}^2$ (3 mDarcy) and the variance of $2.62 \times 10^{-31} \text{m}^2$. Clearly, there are two different spatial distributions in the same space domain of a geological model, a huge discrepancy between one and another.

2 (http://mgstat.sourceforge.net/, which is developed by Prof. Thomas Mejer Hansen of Technical University of Denmark)
Figure 2–1 Six of one thousand realizations of SGSIM realizations of Egg Model. The shape of each model is not square, since we have eliminated the inactive cells. The illustration above is vertically exaggerated.

2.2.1 Twin Experiment and Measurement Generation

In reality, the true reservoir is unknown. In this project the Twin Experiment is applied. One of 1000 realizations is picked to be the truth, reducing the size of the ensemble for analysis to 999. In this case, the first model is chosen as the truth. Figure 2–2 shows the truth of the model and the wells thereby in the log-scale\(^1\) and its spatial distribution. It is observed that well ‘PROD1’, ‘INJECT3’, ‘INJECT6’, and ‘INJECT8’ are the wells with more permeable area (the precise location of the well is the cell below the first letter of the wells’ indicator in the figure), which infers that the profile of the liquid production/injection will be much larger than the rest.

---

\(^1\) Hereafter, the permeability field and spatial distribution will be served in logarithmic, thus the term permeability and log-permeability are used interchangeably
For the measurement, synthetic observations are conducted for the total flow rate of the whole, twelve wells. The Gaussian noise is used for this experiment, with 5% of variance. The design of the measurement network is made every month for 15 months after the first month. This is because in the first 16 months of the profile undergoes significant changes, more than any other time domain. This is shown in Figure 2–3. The bottom hole pressure value of the injection wells is fixed at 405 Bar. For the bottom hole pressure at the producing wells is set to be 395 Bar.

2.3 The State Vector and Initialization

First of all, our variable of interest, namely, the permeability, which is denoted by \( k(t) \), is included in the state vector:

\[
X(t) = (k(t))
\]

The vector for measurement at each time \( t \) is as follows:

\[
Z(t) = \begin{pmatrix} q_{INJ}(t) \\ bhp(t) \end{pmatrix}
\]

where \( q_{INJ}(t) \) is the total injection rate at time \( t \); \( bhp(t) \) is the bottom-hole pressure of the well at time \( t \).

In this data assimilation technique, because the measurement are important part to the data assimilation, we augment the measurement vector to the state vector as well. Therefore, we have:

\[
X(t) = \begin{pmatrix} k(t) \\ q_{INJ}(t) \\ bhp(t) \end{pmatrix}
\]
We recollect that permeability is in the order about of either $-15$ or $-13$. Because the number is very small order, the natural logarithmic scale is used for permeability. This should not be a problem if we are operating in Field Unit, using millidarcy (mD) as an alternative. To this end, we have the subsequent augmented state vector:

$$X(t) = \begin{pmatrix} \log(k(t)) \\ q_{INJ}(t) \\ bhP(t) \end{pmatrix} \in \mathbb{R}^{18545},$$

where the log-permeability tally with the number of grid cell (18533) and the rate corresponds to the number of the wells.

The pattern of the directionality of the arcs itself is inspired from previous research. Gheorge (2010, pg. 53–56) explains how the arcs are determined and the Bayesian Network is constructed. In this project, the bottom hole pressure values of the injection wells are above the allowable reservoir pressure.
Due to this circumstance, the modification of the measurement vector is necessary. We change the bottom hole pressure variable into the total injection rate at the injection wells. Although it may not be real, this decision is made for the sake of simplicity. As a result, the final state vector is of the form

$$X(t) = \begin{pmatrix} \log(k(t)) \\ q_{INJ}(t) \\ q_{PROD}(t) \end{pmatrix} \in \mathbb{R}^{18545}$$

The flow of influence between these three variables is illustrated in the following figure.

![Diagram showing the flow of influence between $q_{INJ}(t)$, $\log(k(t))$, and $q_{PROD}(t)$]

Figure 2–4 The directionality of arcs in this project; $\log(k(t))$ is log–permeability; $q_{INJ}(t)$ is the rate of the injections; $q_{PROD}(t)$ is the rate of the producers.

### 2.4 Egg Model Partitioning for Data Assimilation

Assimilation of new measurement results in changes in the state vector. Because of the dimensionality limitation, gridblocks\(^4\) of size 9×9 are chosen. Each gridblock has 81 permeability variables. With the twelve measurements, the total state vector is 93, which is below 100 variables limit.

Figure 2–5 illustrates the partitioning of the Egg model to conduct the data assimilation\(^5\). None of this gridblocks are overlapping. In addition, one gridblock and another is separated with a difference of, at least, two grid cells; this is to prevent obtaining very different permeability values from one assimilated gridblock to the next gridblock in case of no separation. The location of some of the gridblocks corresponds to the location of the wells; so we can see that some of the gridblocks have association with the well name. The gridblock’s

---

\(^4\) The term gridblock is used to define the partitioning of the (Egg) Model into smaller group of grid cells in a rectangular shape as an object of NPBN data assimilation process.

\(^5\) The creation of the gridblocks is done by algorithm in Appendix A.
location nearby the well is chosen because intuitively there should be stronger
correlation of the nearby cells with the corresponding well. After these gridblocks
are acquired, nineteen more gridblocks are chosen to recover the rest of the
area, as shown by the yellow boxes in Figure 2–5. The size ranges from the
largest of 9×9 to the smallest of 2×2, with the same two grid cell difference
one to another. In total, the entire gridblocks cover for 57.69% of the total
reservoir field. It is also pretty hard to cover optimally for every layer, since
the shape and boundary for each layer of the Egg model is different.

In the experiment, each augmented state vector is translated to an SAE
file, a comma separated file. That consists of variables column-wise and their
corresponding samples row-wise. To perform
the calculation, we use the
UNINET\(^\text{6}\) software. The SAE file, which is a comma-separated file which
contains the state and measurement variables, is read by UNINET with specific
commands from MATLAB. To provide an illustration, the network of NPBN
looks like the left-hand side figure below with a lesser number of variables,
while in our case, the network is presented in the right-hand side for one
gridblock.

Figure 2–5 Illustration of the discretization with the gridblocks on which the data assimilation
is conducted. The gridblocks with the well name are the gridblocks nearby the well. The
gridblocks with the yellow boxes are the remaining gridblocks.

---

\(^\text{6}\) http://www.lighttwist.net/wp/uninet
Figure 2–6 Saturated NPBN for observable case (left), taken from Gheorge (2010) and the saturated NPBN of this project (right) with 93 variables, which is not visible to observe.

The NPBN as shown in Figure 2–4 is called Saturated Bayesian Network where it means that each node is connected in the Bayesian Network with all other nodes. Figure 2–6 shows the difference of the saturated NPBN with lesser variables (left) and numerous variables (right), where we cannot see clearly the structure.

Previous project uses UNINET software to do the conditionalization. Due to the numerous variables and a significant number of gridblocks, we use MATLAB–UNINET interaction to automate the process. However, there is problem in the bridging command between MATLAB and UNINET during the making of this project. For this reason, we write a MATLAB script with the exact same algorithm for the saturated NPBN.
3 Interpolation

The NPBN method recovers only part of the reservoir field by the means of gridblocks. Figure 2–5 shows that the choice of gridblocks necessitates us to recover the unassimilated area of the reservoir field. Zilko (2012) tackles this problem by means of interpolation; where the interpolation is conducted on the mean of the ensembles. The algorithm of the NPBN does not preserve the sample ordering before conditionalization. Therefore, due to the Egg model’s partitioning, it is impossible to perform interpolation at the sample level, because, for instance, sample number 1 from gridblock INJECT3 after conditioning does not necessarily correspond to the first sample from adjacent gridblock PROD1, as the location is shown in Figure 2–5.

In this project, we consider three 1-D interpolation techniques: the Cubic interpolation, the PCHIP interpolation, and the Spline interpolation. The interpolation of the permeability is done in the log-scale.

Figure 3–1 illustrates how the interpolation works. For Cubic interpolation, it searches four nearest (informed) cells to the upper–lower cells (yellow, solid arrow) and changes the search to left–right cells (yellow, dashed arrow) if there is not any data for the former. For the Spline and PCHIP interpolation, it searches the nearest gridblocks and use the entire, corresponding row (column) from those two gridblocks as the informed cells to estimate the query cell(s) of the same row (column).

Figure 3–1 describes four type of interpolation problems in the Egg model. In case number 1 and number 2, we interpolate either horizontally (1) or vertically (2). Some cells are of case number 3 where the cell’s location is neither in between two gridblocks. For this area, we use previously interpolated cells as the informed cells for the data. From Zilko (2012, pg. 71), we perform horizontal interpolation to recover these cells. For problem number 4, we perform extrapolation.

The extrapolation procedure finds up to five informed data cells in order to estimate up to two uninformed cells. The extrapolated area may produce unrealistic value of permeability. It can produces value down from as low as –
4.6 \times 10^{-7}~\text{Darcy} to as high as \(6.9 \times 10^{10}~\text{Darcy}\). This is of course unrealistic. To tackle this problem, we introduce the lower and upper bound. The lower bound is chosen to be \(-39.2\) (0.1 mDarcy) and the upper bound is \(-25.3\) (100 Darcy).

Figure 3–1 Illustration of how interpolation is performed in the Egg model

3.1 Cubic Interpolation

Zilko (2012) uses cubic interpolation as the interpolation technique to recover the entire permeability field. He extensively analyses this method on a 2-D reservoir. As a starter, this project uses the same interpolation to see the performance at a different type of reservoir.

In this chapter, for the purpose of understanding the result, we will use the truth from Chapter 2.2.1.

Figure 3–2 describes the interpolation result of the egg model using cubic interpolation. The picture is served only for the top layer. This is adequate to observe the interpolation result as it gives us how likely it resembles the architecture of the truth.

3.2 PCHIP Interpolation

Another interpolation method is the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) method.
If we consider Figure 3–3, let us define $h_k$ as the length of the $k$–th subinterval: $h_k = x_{k+1} - x_k$, where $x_k$ is the location of data at subinterval $k$ and $x_{k+1}$ is the subsequent location. Subsequently, we define the first divided difference, $\delta_k$, for the permeability value, $y_k$, which is given by $\delta_k = \frac{y_{k+1} - y_k}{h_k}$ where $y_{k+1}$ is the permeability value at the next subinterval. Let the slope at one point is $d_k = P'(x_k)$, where $P$ will be defined very shortly. We would like to estimate the permeability value of $x$, which is between $x_k$ and $x_{k+1}$. Let $s = x - x_k$ and $h = h_k$. The cubic function for $P$ is as follows.

$$P(x) = \frac{3hs^2 - 2s^3}{h^3}y_{k+1} + \frac{h^3 - 3hs^2 + 2s^3}{h^3}y_k + \frac{s^2(s-h)}{h^2}d_{k+1} + \frac{s(s-h)^2}{h^2}d_k$$

$P$ is a function of $s$, hence a function of $x$, which satisfies four interpolation conditions:

- $P(x_k) = y_k$,
- $P(x_{k+1}) = y_{k+1}$,
- $P'(x_k) = d_k$,
- $P'(x_{k+1}) = d_{k+1}$,
We are left to the question: “how do we determine \( d_k \) and \( d_{k+1} \)?” Any function satisfies these derivatives is called Hermite or Osculatory interpolants. The method of PCHIP defines \( d_k \) and \( d_{k+1} \) using harmonic mean of the two discrete slopes:

\[
\frac{1}{d_k} = \frac{1}{2}\left(\frac{1}{\delta_{k-1}} + \frac{1}{\delta_k}\right), \quad \text{Eq (4.1)}
\]

with similar equation for \( d_{k+1} \).

Figure 3–3 Illustration for pchip interpolation result: the red circles are the data and the black line is the result. \( x \) is the grid location, and \( y \) is the permeability value

We use the \( \delta_k \) formula for each of the data points. We then calculate \( d_k \). PCHIP requires that if \( \delta_k \) and \( \delta_{k-1} \) have opposite sign (i.e., different gradient) or if either one of them is zero, then \( x_k \) is discrete local minimum or maximum (refer to Moler, 2004, Chap. 3 pg. 9); otherwise the harmonic mean, as in Equation 4.1, will yield undefined values. In this case, the first data point has zero \( \delta_{k-1} \). Another consideration is that \( \delta_k \) at data point 2 and 3 have different subinterval \( h_k \). Different subinterval will be evaluated with weighted harmonic average defined as:

\[
\frac{w_1 + w_2}{d_k} = \frac{w_1}{\delta_{k-1}} + \frac{w_2}{\delta_k},
\]

where \( w_1 = 2h_k + h_{k-1} \), and \( w_2 = h_k + 2h_{k-1} \).

The PCHIP interpolation is regarded as shape-preserving interpolation which means that the global shape of the data is preserved.

The result of this method, by using the truth, is as follows:
3.3 The Spline Interpolation Technique

We shall continue to another interpolation technique: the Spline method. The spline technique uses four criteria.

1. We require that each curve segment passes through the data. If the segment is \( f(x) \), then \( f(x_i) = y_i \) and \( f(x_{i+1}) = y_{i+1} \). This enforces us to have the \( C^0 \) continuity, where segment from one data is continuous to the next data. The following picture describes that in the left hand side, the function does not pass the data, hence not \( C^0 \) continuous, whereas the right hand side, the function does, hence \( C^0 \) continuous.

Figure 3–5 Criteria 1 of Spline: \( C^0 \) continuity
2. The technique demands that the slopes of \( f(x) \) at which the segments cross are the same. In other words, \( f'_i(x_{i+1}) = f'_{i+1}(x_{i+1}) \). This guarantees the \( C^1 \) continuity. The picture below shows that although the function passes the data, we demand that the gradient from previous segment is the same to the next segment.

![Figure 3-6 Criteria 2 of Spline: continuity on the slope of the point where the segments are met](image)

3. Spline method necessitates that each curve of the segments have the same curvature, i.e., \( f''_i(x_{i+1}) = f''_{i+1}(x_{i+1}) \). In other words, the second derivatives of the two segments must be equal. This is called the \( C^2 \) continuity. The illustration below states that Spline interpolation requires that the radius of the curvature at previous segment is equal to the next segment. The left hand side of the picture below shows it does not have the same radius of curvature.

![Figure 3-7 Criteria 3 of Spline: continuity on the curvature of each segment.](image)

4. At both ends of the Spline data, we employ the ‘boundary condition’, which regulate the behavior of the end points of the interpolation curve. The algorithm in MATLAB uses the natural spline, where the boundary condition is chosen to be \( f'_0(x_0) = f'_n(x_n) \) where the index 0 corresponds to the first data and the index \( n \) is referring to the last data.
The following figure gives the result of interpolation of the truth.

![Figure 3-8 The interpolation result of Spline method. The whole recovery (Left) and the interpolation result (Right, the gray–black area is the assimilated one)](image)

### 3.4 Interpolation Result for Bimodal Distribution

Of all the three methods, the analysis reveals that cubic and Spline interpolation return more extreme values for the interpolated and, in particular, the extrapolated area than PCHIP. This is shown in Figure 3-9, where the frequency of the upper and lower bounds for the interpolated and extrapolated cells shows a high value for these two methods. It is likely that such phenomenon is found in the extrapolation method, where there exists high oscillation for the result of the estimated cell(s). The frequency of the extreme values for these methods has approximately similar result. However, because it is widely known that the cubic interpolation method can lead to the Runge’s phenomenon, this project works with Spline and PCHIP as the methods to do the interpolation and extrapolation.
Figure 3–9 The CDF and histogram of interpolated/extrapolated values for Cubic method (Left), Spline method (Middle), and PCHIP method (Right).

Being a channel based reservoir, the Egg model has bimodal distribution in the permeability field. Figure 3–10 shows the spatial distribution for a sample at initial model after interpolation and extrapolation for three interpolation methods.

Figure 3–10 Spatial Distribution of sample number 2 after interpolation/extrapolation with Cubic (Left), Spline (Middle) and PCHIP (Right) interpolation technique.

Unsurprisingly, due to the continuous and smooth property of all three methods, these results show non-bimodal distribution. The interpolation results method will provide a misinformed prior spatial distribution for every sample before the (next) data assimilation takes place. This corroborates the decision to perform interpolation in the Mean of the ensembles.
4 Investigation of the Result

4.1 Comparison and Analysis of the Result

In the previous chapter, the comparison is performed between three methods of interpolation, the Spline and PCHIP interpolation. In this sub-chapter, we are analyzing the Mean of the Ensemble and observe the performance of NPBN method coupled with the two interpolation method. We use visual observation, RMSE, absolute difference, forecasting, and forecasting evolution.

4.1.1 Visual Analysis

The result of various time steps with NPBN data assimilation with the two interpolation techniques are presented in Figure 4–1.
Figure 4–1 The evolution of the Mean of the Ensemble for NPBN data Assimilation from time step 0 (initial) to 15. The left column is the data assimilation with Spline Interpolation. The right column is the data assimilation with PCHIP Interpolation. The ‘truth’ is shown at the first row, left picture. The Initial, which is the same for both cases, is shown at the first row, right picture.

At the first time step of data assimilation, one can notice the existence of a channel in the estimated model, indicated by the pronouncement of more of higher permeability grid cells nearby well PROD1, INJECT6 and INJECT8. More importantly, the data assimilation recognizes the lower permeability in INJECT7 and INJECT3 at this time step.

The NPBN shows a visible learning process, as it is shown in Figure 4–1. As time evolves, one notices that the method learns to resemble the truth, albeit slowly. For example, we focus on the sand area between INJECT5 and PROD4. In the truth, this sand channel is a quite narrow channel. At time step 3, the channel has a much bulkier size. When the new measurements are assimilated, the size of this channel decreases, hence trying to imitate the truth. However, we do not see evolution of the size at the latter stages, in particular from time step 9 to the final time step.
The method hardly learns much about the sand-permeability area development on the left of INJECT6, where the truth shows that the gridblock of INJECT6 is next to a greater area of shale-permeability (South-West area). In addition to that, the development of the South-West channel area (the V-shape channel) as shown in the truth does not show a good appearance in the estimated model.

At the West area, we observe that truth has a small area of sand-permeability. The estimated model does not give similar response. At the North-West area, we notice that the area is dominated by shale-permeability in the truth. The shale-permeability at the estimated model develops very little during the data assimilation.

The truth has a longest channel, spanning from East to South-East, right-bottom corner. The pronunciation of the sand-permeability in the estimated model suggests good learning process at this particular area, although the size is wider. If we look globally at the final result of the data assimilation, the result shows resemblance of the shape of the channel, though it is not impressive.

4.1.2 Root Mean Square Error and Absolute Difference

The measure of performance of Root Mean Square (RMSE) has been used in the previous research (Gheorge, 2010, and Zilko, 2012). The result of RMSE in this project shows similar behavior from Zilko (2012), as shown in Figure 4-2, where we notice that the RMSE is increasing at the entire time span (except at time step 1 for the PCHIP method), at the interpolation area and especially at the extrapolation area. The increasing of RMSE in the whole area is predominantly influenced by large RMSE of the extrapolation area. This is because it is very hard to recover the permeability at the extrapolation area with the current interpolation technique. However, the assimilated area, depicting by the black line, shows a decreasing behavior of the RMSE at time step one of assimilation and almost constant afterwards.
Figure 4–2 The RMSE of Spline method and PCHIP method at entire area, interpolated area, extrapolated area, and assimilated area. We pronounce the Legend in the lower part of the figure.

Next, we calculate absolute difference between the truth and the estimated model with both types of the interpolation. Figure 4–3 gives the evolution of the absolute difference at selected time steps. Channel shape, at time step 0, is observed, where it has smaller absolute values, relative to the rest of the area. As time evolves, the absolute difference shows an improvement at some area, where the light blue area, at time step 0, is turning to darker blue area, at final time step. On the contrary, some area, which is previously a darker blue area at time step 0, is turning to lighter blue. At the interpolation area, the absolute difference between two methods of interpolation does not show a significant difference. However, at the peripheral of the Egg model, the smaller absolute difference reveals that, generally, the PCHIP method is visually better than Spline method.
Figure 4-3 The absolute difference between truth and the Mean of the ensemble with Spline (Left) and PCHIP (Right) interpolation method. We provide the result for time step 0, time step 1, time step 9 and time step 15.
The histogram of the absolute difference and its corresponding CDF for Spline Method (Left) and PCHIP method (Right).

The absolute difference of the result coming from the PCHIP interpolation method produces lesser high absolute difference, especially in the extrapolation area. We compare the Spline interpolation and the estimated model with PCHIP interpolation for the last time step. This is shown in Figure 4–4, where PCHIP shows relatively less frequency of absolute difference values in the values more than 8. In other words, Figure 4–4 suggest that we have more higher absolute difference by using spline interpolation method. Hence, PCHIP interpolation method gives smaller discrepancy towards the truth at the peripheral area.

4.2 Forecasting Well Profiles

Forecasting the reservoir model is forward simulating the model through time and to view the difference between the simulated rate of the estimated model and the truth. The estimated models that we are taking are from the last time step of the data assimilation of the Mean of the Ensemble for both of the Spline interpolation and PCHIP interpolation. We shall view the difference of production profile of these two model.

We shall analyze two of the four main wells to view their production profile (we present the rest of the production profiles of the remaining wells in Appendix C). PROD1 is the production well, which is shown in the black box of Figure 4–5. The responses of the PROD1 in spline and PCHIP method give similar profile. What is interesting is that the response of the truth is much lower than the response of the Mean of the ensemble. We shall explain this behavior, why this is the case.
The calculation of the production of oil rate in the well is given by:

\[ q_o = -J(p_R - p_{bh}) \]  

(Eq. 4.1)

\( q_o \) is the rate of the oil production (negative term indicates a production well; positive is injection well; for water injection, we use subscript \( w \)), \( p_R \) is the reservoir pressure, \( p_{bh} \) is the bottom hole pressure at the well, and \( J \) is the productivity index (or Injectivity index for Injection Well).

If we have radial type of well, \( J \) is as follows:

\[ J = \frac{2\pi k_o h}{\mu B \ln(r_e/r_w)} \]

where \( k_o \) is the phase permeability of oil, \( h \) is the thickness of the reservoir, \( \mu \) is the viscosity of the oil, \( B \) is the oil formation volume factor, \( r_e \) is the radius of encroachment, and \( r_w \) is the radius of the well. Interested reader can refer these terminologies to Dake (1983), Lie (2014) or Jansen (2013[b]). In our case, \( B \) is equal to one, assuming that we do not have significant changes of the oil volume from sand-face/reservoir to the surface. The well is in prescribed condition, which is the bottom hole pressure (\( p_{bh} \) is constant). The production of the well is related to the value of the saturation
at the cell of the well. We shall see into detail the State (saturation or pressure) nearby of well PROD1 for the truth, the Mean with Spline interpolation method and the Mean with PCHIP interpolation method. We shall focus Figure 4–5 into the permeability area near the black box.

![Image](image_url)

Figure 4–6 The permeability nearby the black box in Figure 4–5: The left figure is the Spline model, in the middle is the PCHIP model and the right figure is the truth.

The sand-permeability area in the truth near well PROD1 has the shape of a channel, thus allowing more pronounced shape of saturation change according to the shape of the higher permeability channel. If we compare these three permeability field, nearby zone of the cell of well PROD1 at the Mean of the ensemble for both of interpolation method has much wider sand-permeability area. Even though the permeability range at the estimated model is lower (there are much of orange-to-red cells), the water breakthrough is much faster for the estimated models. To prove this, Figure 4–7 describes the dynamical simulation of the water saturation in the estimated model and the truth. We picked 3 time step: around 1\textsuperscript{st} year, 2\textsuperscript{nd} year and 3\textsuperscript{rd} year, where at the second year, both of the estimated model has the water breakthrough and at 3\textsuperscript{rd} year the truth has the water breakthrough. Observation reveals that, at Layer 4, the water breakthrough occurs in the estimated model since there exist a permeability streak between INJECT3 and PROD1. The shape of the saturation spread implies the shape of the permeability: the estimated models do not reveal the shape of a channel as we see in the truth. This suggests that the development of the permeability with NPBN is limited after gaining more observations.
Figure 4–7 The Water saturation change of three model for the fourth layer: the first column is \textit{Mean} of the ensemble with Spline Interpolation, the second column is the Mean of the ensemble with PCHIP interpolation, and the third column is the truth. The time at the first row is after 1\textsuperscript{st} year, the second row is after 2\textsuperscript{nd} year and the last row is after 3\textsuperscript{rd} year. The yellow–dashed box is the location of well PROD1.

Looking at Figure 4–8, we understand that the water reaches PROD1 at the Layer 4 and Layer 5 of the perforations. This map is provided at which water is breaking through at PROD1, which is about 1.6 year. The water reaches the PROD1 at Layer 4 and 5. Figure 4–9 helps us to describe the situation: the permeability estimation at Layer 4 and 5 develops a ‘permeability streak’ (focus at black box in the figure) where the sand–permeability develops very continuous at this area. This suggest that the development of the permeability field is not entirely the same at all layers. Further investigation shows that the shape of global permeability is comparably the same but with different features, as shown in Figure 4–10. The NPBN method shows an unimpressive approximation at every layer. Moreover, it is imperfect to reproduce an identical shape of permeability field for every layer, as what we have in the truth.
Figure 4–8 The slice of 3D water saturation map at which water is breaking through at PROD1.

Figure 4–9 The corresponding slice 3D permeability field near PROD1 and INJECT3. The black box shows relatively higher permeability value at layer 4 and 5.

Figure 4–10 The permeability field difference at, for example, Layer 5 and Layer 6.

Figure 4–11 gives the pressure map of the Egg model at some time step. The cell at which Well PROD1 is located in the picture shows a black cell, precisely in its location. The Mean of the ensemble, for both of the interpolation method, returns almost the same reservoir pressure at the cell where PROD1
is located. However, we can clearly see that the truth shows very low pressure area in the left bottom of the Egg model (as shown with darker area).

Figure 4–11 The reservoir pressure of Egg Model for the first layer: the first column is Mean of the ensemble with Spline Interpolation, the second column is the Mean of the ensemble with PCHIP interpolation, and the third column is the truth. The time at the first row is after 1\textsuperscript{st} year, the second row is after 2\textsuperscript{nd} year and the last row is after 3\textsuperscript{rd} year. The yellow–dashed box is well PROD1

The parabolic shape of the pressure behavior/shape nearby well PROD1 in the truth of Figure 4–11 is resulted due to the very high anisotropy (variation of values in every direction) of the permeability, a phenomenon that we find with relatively smaller pronouncement in both of the Mean of the Ensemble. After the first year, the truth (top, right column) shows very fast pressure depletion. Equation 4.1 states that this fast depletion is most likely caused by different shape of the permeability field compared to the Mean of the ensemble for both of the interpolation methods.

The NPBN seems hardly learn the state after gaining more observation. Observation at time step one from Figure 4–1 shows that the learning process is obvious at this time. However, addition more observation, especially from
time step ten, demonstrates little improvement to the state. We shall analyze this from the correlation matrix of the gridblock at which PROD1 is located.

The correlation matrix has a very short range as shown in Figure 4–13; the channelized reservoir has a characteristic of a short distance variation for permeability. Initially, the correlation matrix indicates a strong, positive correlation at the nearby grid cells, as indicated in Figure 4–13, left figure. Notice that the variables are divided into two parts: from 1 to 81 is the permeability state. As moving further from a particular cell of the permeability state, the correlations are still positive, although we find some of the light-green and light blue area, indicating a negative correlation. These cells gradually gain more correlation as the time progresses from time step one. At time step fifteen, the data assimilation makes the strongest correlation at top-left of the gridblock and the bottom-right. The correlation range of the state variable at gridblock PROD1 is from $-0.1617$ to $0.5565$ at time step one and from $-0.3794$ to $0.5691$ at time step fifteen.

Variables from 82 to 93 is the measurement state. The arrow shows that the correlation of the permeability state and the measurements are very high at the beginning. The black box is the measurement of INJECT3 and red box is the measurement of PROD1, where we have a negative correlation because we assign negative total flow rate for the production wells. NPBN method has a very high correlation at the beginning of data assimilation, resulting a significant drop of variance for the both type of the variables. Interestingly, the correlation
between the measurement variables and the state variables, after time step one, shows very little magnitude near zero. This insignificant correlation values implies that any addition of new measurement yields no significant improvement to the state variables.

![Figure 4-13 The correlation matrix of 9x9 gridblock PROD1. The correlation scale is from -1 (blue) and +1 (red). The left figure is the correlation matrix at time step 1 and the right is at time step 15.](image)

The evolution of the profile of the liquid production might help this observation. We provide, in Figure 4-14, the evolution of the profile of the liquid production for some time step of data assimilation. Careful assessment shows that the profile after time step 9 does not improve the profile, as to match it to the truth (black line). We notice that the initial guess is very close to the truth for the oil production, whereas the water breakthrough is not. The profiles, both for the oil production and water production, at time step 2 are perhaps better than the final time step. The evolution of the profile shows a very little change, insinuating that adding more measurement for well PROD1 does not give any significant changes to the state.

Figure 4-15 corroborates this phenomenon, where we provide the water saturation state after time step 9 during at year 1.6 and layer 1. The state of the water saturation, which implies the permeability state, of each time step shows very little improvement from time step 10.
Figure 4–14 The evolution of the profile of the oil production (Left) and water production (Right) of well PROD1.

Figure 4–15 The evolution of the water saturation state at Layer 1 at year 1.6 of estimated model with Spline method. The ‘t’ indicates the time step of data assimilation.

If we look Figure 4–16, the water profile is very different to the truth. After year four, the water injection is increasing steadily. The pressure map nearby the location of this well, as shown in Figure 4–11, shows a relatively higher pressure in comparison to the truth. At the first year, the truth describes a very fast pressure drop nearby well PROD1, especially in South–West area. Increasing of pressure from INJECT3 at the truth gives very little significance at the subsequent years. Therefore, we do not see an increasing gradient of oil production of PROD1 at the truth around year four. This is not the case.
for the Mean of ensemble, where the pressure drop is not fast and extensive in the South-West area. Increasing of pressure by INJECT3 influences the pressure around well PROD1, making the oil production of PROD1 larger than the truth. Similarly, the water injection of INJECT3 is much larger than the truth.

Figure 4–16 The water injection profile for well INJECT3 (top). The bottom row is the Mean with Spline interpolation Method, the Mean with PCHIP interpolation method, and the truth, respectively, with the black box indicating the location of well INJECT3.

The correlation of INJECT3 reveals similar behavior in comparison to PROD1, as shown in Figure 4–17. The correlation is, at the final time step, positively high around the top–right corner (recall Figure 4–12). Initially, we find that the range of correlation of the state variable starts from $-0.1862$ to $0.5287$, whereas at the final time step the range is from $-0.3612$ to $0.5839$.

We observe that all main wells reveal the same behavior of the profile of the liquid, either injection or production wells. Generally, at the first nine time steps, the evolution of the profiles of these main wells shows a larger profile.
as the new data is assimilated, making the new profile more distant from the truth. The addition of new measurement after time step nine shows a learning behavior towards the truth. However, it has very little influence to match the profile to the truth.

Figure 4–17 The correlation matrix of 9x9 gridblock INJECT3. The correlation scale is from -1 (blue) and +1 (red). The left figure is the correlation matrix at time step 1, the middle is at time step 9 and the right is at time step 15.

Figure 4–18 The evolution of the profile of the water injection of Well INJECT3.

However, we observe different behavior for the well profile that are not the main ones. The evolution of forecasting profiles for these wells shows a consistent evolution towards the truth. This contrasts the inconsistent behavior of the profiles of the main wells, as we have observed before.
5 Conclusion and Recommendation

5.1 Conclusion

1. The visual observation of the NPPBN method does not show very convincing results for the Egg model with the number of wells and Gaussian noise that are used in this project.

2. Using the 3D geological model, the NPBN method produces a permeability field which is not identical at every layer, as what we have in the truth.

3. The forecasted well profiles of the estimated models do not match the truth. They are different by the factor of, approximately, two.

4. The evolution of the liquid profiles for every time step of data assimilation shows a learning behavior. For the wells inside the channel, the learning behavior is fluctuating, where after time step nine they show a slow learning behavior towards the truth. For the wells in the shale background, the learning behavior is consistent towards the truth.

5. Additional measurements, especially after time step 9, does not improve the permeability field. As the correlation of the measurement and the state variables is very high at the first time step, adding more wells in the system is thought to be more significant than assimilating more measurement over time.

6. The high correlation at the beginning of the time step reveals a high variance reduction for both of types of variables after time step one.

7. The interpolation using the Spline method and the PCHIP method give very little difference to recover the remaining, unassimilated permeability field. However, in the peripheral area, where the extrapolation is conducted, the PCHIP method gives a better approximation, as indicated by relatively a lower absolute difference.

8. The use of RMSE is quite inconsequential at this project, given the geospatial nature of a bimodal distribution, namely an enormous difference.
between permeabilities. However, the assimilated area gives a decreasing profile of RMSE.

5.2 Recommendation

1. Addition of more wells or the design that every gridblock contains a well is recommended. This effort will increase the information quality to recover better permeability field. The addition of the well in the channel area (or sand-permeability area) is more important since it gives more information to the surrounding gridblocks, located in the shale-permeability area. This is deducted from the fact that high correlation exists at the first time step between any gridblock which contains a well and the related well.

2. The observation of bad measurements in Zilko (2012) is due to the high variance of these measurements. It is interested to study non-Gaussian or high variance measurement noise and its effects to the permeability recovery of the Egg model.

3. The search of a good, quantitative measure of performance in for NPBN method is still on going. Image analysis technique could be useful to compare the differences between permeability fields.

4. In this project, all gridblocks covers around 57% of the entire field. It is of interest to use different shape for the assimilated area. Using triangle, we perhaps recover much larger permeability field, especially with the irregular shape of the Egg model. It must satisfy the two cells difference between a gridblock and another gridblock as well.
Bibliography


A. MRST and Its Grid Structure

MATLAB Reservoir Simulation Toolbox is an open-source toolboxes for MATLAB, which, according to Lie (2014), “contains a comprehensive set of routines and data structures for reading, representing, processing, and visualizing structured and unstructured grids, with particular emphasis on the corner-point format used within the petroleum industry.” As a part of the research program from SINTEF, Norway, MRST has been developed for basic flow and transport solvers for incompressible, single/two phase, on basic unstructured grids. One can download the entire module at http://www.sintef.no/Projectweb/MRST.

The grid structure in MRST is distinguished in comparison to other commercial reservoir simulator packages. For example, in Eclipse, every cell is assigned with three-number: the x coordinate, y coordinate, and z coordinate. In MRST, it uses cell number from 1 to the maximum of the matrix. Lie (2014, pg. 70) states that the Grid structure in structure G which contains three fields: cells, faces and nodes. The faces structure is consisted of 3 array, where the first is redundant because we could calculate from the second and third array (see Lie (2014, pg. 71) in the part of faces). We reconstruct the information regarding the structure with the following command:

```
rldecode(1:G.cells.num, diff(G.cells.facePos), 2).
```

To make it tangible, we use the following picture (redrawn from http://www.sintef.com/Projectweb/MRST/Tutorials/Grid-factory-tutorial/).
A 1 MRST grid cell structure

The red circle is the numbering of the cell; the green square is the face number; and the blue diamond shape is the node number. We show that we do not have cell number 2 next to cell number 1. So, in order to detect cell next to cell number 1, we have to use the same information that is shared to grid cell number 1. The author uses Tag id, as shown in the right part of the picture. Tag id indicates the wind direction of the adjacent cells. This is useful to create the algorithm which searches the nearby cells; therefore it is possible to make grid blocks assimilation automatically.

For example, we use the 9 by 9 blocks

The letter W indicates the well location. We know for sure the cell number for the well location, as structure W (structure which gives information about the wells) from MRST gives. We use the ridecode to give the direction of cell nearby searching. First, the author uses the ‘backbone’ of the grid blocks,
which is the yellow column in the picture. As done with it; each cells of the yellow backbone (and the well cell thereof) will propagate to East and West direction simultaneously, searching the cell numbers. The way it is shown is because the numbering of the cells is from Left to Right that makes it easier if we propagate towards Left or Right. But it is extremely time-consuming to search adjacent cell from North to South (or otherwise) direction. That is because we have to call the rlidecode for entire domain (not just the cell of interest) and do the logical operation with numerous size of matrix.

B. Algorithm Parameters for SGeMS

SGeMS software uses _py_ (python) script to conduct the entire command. The parameters of each of the algorithm are served in the following figures.

```xml
<parameters>  <algorithm name="snesim_std"/>  <Min value="1"/>  <Constraint_Marginal_ADVANCED value="0.5"/>  <resimulation criterion value="1"/>  <resimulation_iteration_nb value="1"/>  <NB_Multigrids_ADVANCED value="3"/>  <Debug_Level value="0"/>  <Subgrid_choice value="0"/>  <expand_isotropic value="1"/>  <expand_anisotropic value="0"/>  <aniso_factor value="0"/>  <Region_Indicator Prop value="snesim_std_real"/>  <Active_Region_Code value=""/>  <Use_Previous_Simulation value="0"/>  <Use_region value="0"/>  <GridSelector_Sim value="SIM"/>  <Property_Name_Sim value="snesim_std"/>  <NB_realizations value="10"/>  <Seed value="211175"/>  <PropertySelector_Training grid="II" region="property"/>  <Properties_sim value=""/>  <Marginal_CDF value="0.73 0.27"/>  <Max_Cond value="60"/>  <Search_Ellipsoid value="80 80 800 0.0"/>  <Hard_Data_grid=""/>  <ProbField_properties count="0"/>  <ProbField_properties count="0"/>  <ProbField_properties count="0"/>  <ProbField_properties value=""/>  <Use_affinity value="0"/>  <Use_Rotation value="0"/> </parameters>
```

B 1 The Parameters for SNESIM algorithm
B 2 The Parameters for SGSIM algorithm for Sand region

```xml
<parameters>
  <algorithm name="sgsim" />
  <Grid_Name value="Egg" region="Sand" />
  <Property_Name value="Permeability" />
  <Nb_realizations value="1" />
  <Seed value="14012379" />
  <Kriging_type value="Simple Kriging (SK)" />
  <Trend value="0 0 0 0 0 0" />
  <Local_Mean_Property value="Facies" />
  <Assign_Hard_Data value="0" />
  <Hard_Data_Grid value="region" />
  <MaxConditioning_Data value="12" />
  <Search_Ellipsoid value="0 0 0 0 0 0" />
  <AdvancedSearch use_advanced_search="0" />
</parameters>
```

B 3 The Parameters for SGSIM algorithm for Shale region

```xml
<parameters>
  <algorithm name="sgsim" />
  <Grid_Name value="Egg" region="Shale" />
  <Property_Name value="Permeability" />
  <Nb_realizations value="1" />
  <Seed value="14012379" />
  <Kriging_type value="Simple Kriging (SK)" />
  <Assign_Hard_Data value="1" />
  <Hard_Data_Grid value="region" />
  <MaxConditioning_Data value="12" />
  <Search_Ellipsoid value="0 0 0 0 0 0" />
  <AdvancedSearch use_advanced_search="0" />
</parameters>
```

B 4 The transformation parameters for the random variables of the Sand region

```xml
<parameters>
  <algorithm name="trans" />
  <Grid value="Egg" region="Sand" />
  <Props_count value="1" />
  <Out_suffix value="transf" />
  <Is_cond value="0" />
  <ref_type_source value="uniform" />
  <ref_min_source value="-3.9" />
  <ref_max_source value="3.9" />
  <ref_type_target value="Uniform" />
  <ref_min_target value="1e-13" />
  <ref_max_target value="5e-13" />
</parameters>
```
B 5 The transformation parameters for the random variables of the Shale region

```xml
<parameters>
  <algorithm name="trans" />
  <grid value="2D" region="Shale" />
  <props count="1" value="Permeability__real0" />
  <out_suffix value="transf" />
  <is cond value="0" />
  <ref_type_source value="Uniform" />
  <ref_min_source value="3.9" />
  <ref_max_source value="3.9" />
  <ref_type_target value="Uniform" />
  <ref_min_target value="le-15" />
  <ref_max_target value="se-15" />
</parameters>
```

B 6 The TIGENERATOR parameter script, the right part area (the lines start with "<!--") are only comments/remarks for the corresponding lines

```xml
<parameters>
  <algorithm name="tigenerator" />
  <grid value="sim_grid" />
  <prop name="eggreal" />
  <nb_realizations value="1" />
  <seed_rand value="1439833" />
  <nb_geobodies value="1" />
  <geobodySelector index="1" prop="0.2" type="sinform">
    <cdf_type value="Uniform" />
    <const min="70" max="100" />
  </sinform>
  <sinLen cdf_type="Constant" />
  <sinLen const mean="1000" />
  <sinEm cdf_type="Constant" />
  <sinEm const mean="2" />
  <sinWid cdf_type="Constant" />
  <sinWid const mean="1" />
  <sinHtk cdf_type="Constant" />
  <sinHtk const mean="1" />
  <sinAmp cdf_type="Uniform" />
  <sinAmp const min="3" max="8" />
  <sinwav cdf_type="Uniform" />
  <sinwav const min="20" max="30" />
  <interaction erosion="" />
  <overlap min="0.9" max="1" />
  <flag no_self_ovl="0" />
</parameters>
```

<geobodySelector> <!-- this is important to end comments/remarks -->
C. The rest of the well profiles

C 1 The Oil Profile (Left) and Water Profile (Right) of well PROD3

C 2 The Oil Profile (Left) of well PROD2. The Water Production is zero.

C 3 The Oil Profile (Left) of well PROD4. The Water Production is zero.
The Injected Water profile of Injection Well INJECT1 (Left) and well INJECT2 (Right)

The Injected Water profile of Injection Well INJECT3 (Left) and well INJECT4 (Right)

The Injected Water profile of Injection Well INJECT5 (Left) and well INJECT6 (Right)
The Injected Water profile of Injection Well INJECT7 (Left) and INJECT8 (Right)