NTNU
Det skapende universitet

History Matching: Effekten av tilgjengelig informasjon

Håvard Johnsen Reitan

Petroleum Engineering
Innlevert: Juni 2012
Hovedveileder: Jon Kleppe, IPT

Norges teknisk-naturvitenskapelige universitet
Institutt for petroleumsteknologi og anvendt geofysikk

Acknowledgement

This Masters thesis was carried out during the Spring of 2012 at Norwegian University of Science and Technology, Department of Petroleum Engineering and Applied Geophysics.

I would like to thank Professor Jon Kleppe for his help with this thesis. Also a great thanks to fellow students for valuable discussions and inputs to the thesis.

In addition I would like to thank Professor Bjarne Foss, Richard Rwechungura and Eka Suwartadi for help with the EnKF methodology and the code used. Also thanks to IRIS for developing the EnKF Matlab Code used.

Trondheim, 5th June 2012

Håvard J. Reitan

Abstract

A good reservoir model, which correctly represents both the static parameters and flow properties, is essential to optimize production from any reservoir. Much time and effort is put into the description of the reservoir and major financial decisions rest on forecasts from the model. The use of multiple models for forecasts is increasing in popularity and history matching methods Ensemble Kalman Filter are suitable for this.

The Ensemble Kalman Filter was proposed by Evensen in 1994 as a data assimilation method for large-scale nonlinear systems, and have been reviewed and tested several times in the petroleum industry since then. The filter uses an ensemble of state vectors to represent multiple realization of the same system and the ensemble covariance is used to represent both error statistics and the reservoir response to state and model parameters.

In this Thesis, the EnKF has been used to history matching the PUNQ S-3 synthetic case, in order to see the effects of additional information. This was done through two different method, the first one using boundaries to constrain model parameters. In the second case the evolution of the same history match was presented through different timesteps. The models were evaluated through their forecast ability and model parameter match.

Although flow performance in terms of forecasted rates improved for all models, the model parameters did not improve. An evolution towards a wrong parameter solution was seen, even though this solution made a good forecast for future production. The prior geological knowledge was not well enough defined in the updating step, thus the conditioned models evolved further from the truth than

the initial model.

Sammendrag

En god reservoarmodell, som både representerer de statiske parametrene og strømningsegenskaper, er avgjørende for å optimalisere produksjonen fra hvilket som helst reservoar. Mye tid og krefter blir brukt til å beskrive reservoaret så godt som mulig og store økonomiske beslutninger hviler på prognosene fra denne modellen. Prognoser utført ved bruk av flere realisasjoner basert på samme modell blir stadig mer populære for å fange usikkerhet. Historietilpasningsmetoder som Ensemble Kalman Filter er godt egnet for dette.

EnKF ble foreslått av Evensen i 1994 som en data-assimilasjon metode innen oceanografi, og har blitt utviklet og testet flere ganger innen petroleumsindustrien siden da. Filteret bruker et ensemble av vektorer for å beskrive reservoarparameterne og hver av disse vektorene beskriver en realisasjon av reservoaret. Kovariansen mellom disse vektorene brukes til å representere både spredning og reservoarets respons til parameterverdier.

I denne oppgaven har EnKF blitt brukt til historietilpasning av PUNQ S3, en syntetisk reservoarmodell, for å se effekten av tilgjengelig informasjon. Dette ble gjort gjennom to ulike simulering, hvor den første ble gjennomført med grenseverdier for å begrense de statiske parameterne. I det andre tilfellet ble utviklingen av en historietilpasning presentert gjennom ulike tidsskritt. Modellene ble evaluert på bakgrunn av sin prognose for fremtidig produksjon, samt sine avvik i parameterverdi sammenliknet med de sanne parameterne.

Selv om prognosene ble forbedret for samtlige modeller, ble det ikke observert noen forbedring i reservoarparameterne. En utvikling mot en falsk løsning ble observert. Denne løsningen hadde feil parameterverdier, men gav en prognose for

fremtidig produksjon som var veldig lik sannheten. Den geologiske kunnskapen ble ikke anvendt i oppdateringen, noe som førte til at de oppdaterte modellene var lengre unna sannheten enn det opprinnelige utgangspunktet.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Reservoir management is an important aspect in order to optimize hydrocarbon production and thus field value. Optimizing of production can be done innumerable ways and could include drilling of new wells or selection of an increased recovery method. Common for most projects is that they begin with the reservoir model. The value of such a model, built on geophysical and geological data, is proportional with its ability to predict the field production. Inaccuracies in such a model lead to a deteriorating forecast ability with time. To keep the predictive ability of the model, it is updated using production data and in some cases data from geophysical 4D surveys.

The updating sequence of a reservoir model is called history matching or conditioning procedure. Going two decades back in time, this was almost solely done manually. With increased computational power, semi-automated methods have become more prominent. Ensemble Kalman filter has been chosen for this Thesis due to the easy implementation and its continuous update property. It can also update both model and state parameters. While being semi-automated, the conditioning method still relies on prior knowledge that needs to be defined. The aim of this study was to see how a difference in information given would change both the conditioning process and the future forecasts.

The first chapter of this thesis gives an introduction to fundament statistics. Both Chapter 4 and 5 later build on this basic statistics. In Chapter 3, a general intro-

duction to reservoir modeling and forecasts is given. Also included is a short summary of errors present in both reservoir simulation and history matching. Chapter 4 lists some of the challenges met in history matching and explains various objective functions used. A link between objective functions and the prior statistics is given. Both the Kalman Filter and the Ensemble Kalman Filter is derived in Chapter 5. A short review of field application is also given, but as this method still increases in popularity, a full review is not given. The EnKF was applied in two different settings on the same synthetic case. Background information of the reservoir model and the software used is given in Chapter 6, while results are given in Chapter 7 and 8. As both cases have similar results, an overall discussion is given in Chapter 9 and the most important observations is summarized in the following conclusion. In the end, references, nomenclature and appendices are included.

Chapter 2

Statistical fundament

Some elements of the Ensemble Kalman Filter is based on a Bayesian update procedure. This chapter will serve as a review of the basic statistical fundament for continuous variables used in the derivation of EnKF, which is covered in Chapter 5. First a definition of the probability density distribution is introduced; follow by an explanation of the most simple statistical moments. This provides a basic fundament before introducing the Monte Carlo simulations, on which the Ensemble Kalman Filter methodology is based.

2.1 Probability density functions

The probability density function can be derived from the cumulative density function. Given a random continuous variable x, we can assign F(X) as a cumulative density function for this variable. F(X) is a distribution function which denotes the probability that x will take a value less than or equal to X. This relationship can be expressed by

$$F(X) = P(x \leq X) = \int_{-\infty}^{X} f(u)du \tag{2.1.1}$$

where f, if it exist, is the probability density function of the random variable x. If f is continuous at X, one could write

$$f(X) = \frac{d}{dX}F(X) \tag{2.1.2}$$

The probability density function is defined as non-negative everywhere and the total area under its curve is equal to one. The probability of x taking a single given value is zero, as the upper and lower limit would coincide.

$$f(X) \geq 0 \qquad forall X, \tag{2.1.3}$$

$$\int_{-\infty}^{\infty} f(X)dX = 1 \tag{2.1.4}$$

$$P(x = a) = \int_{a}^{a} f(X)dX = 0 \tag{2.1.5}$$

The area under the pdf curve can in other words be described as the likelihood of a continuous random variable, x, to take a value within an interval assigned.

$$P[a \leq x \leq b] = \int_{a}^{b} f(X)dX \tag{2.1.6}$$

One of the most common probability density functions is the Gaussian distribution, also called the normal distribution, which is completely described by its mean and variance. The Gaussian distribution can be written in the form

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\ exp(-\frac{(x-\mu_x)^2}{2\sigma^2}) \tag{2.1.7}$$

The Gaussian distribution is bell shaped and symmetric, as seen in Figure 2.1.1

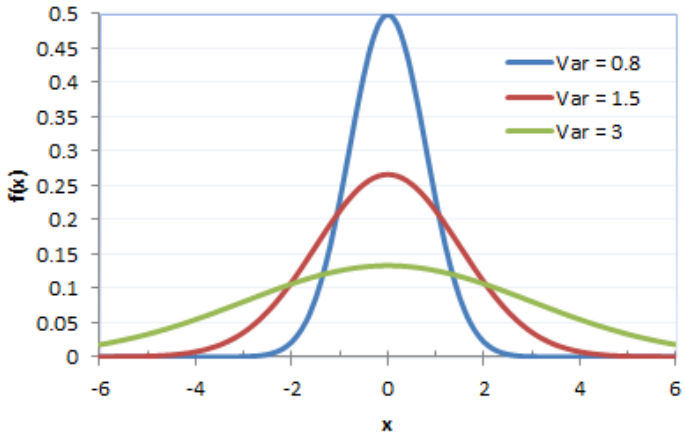

Figure 2.1.1: Three gaussian probability distributions with zero mean and variances of 0.8, 1.5 and 3.

2.2 Statistical moments

Moment in statistics is loosely defined as a quantitative measure of the shape of a set of point. For the first two moments, these will describe the location and the width of the set. Below follows a statistical definition of the expected value and variance in addition to equations to calculate this from sample sets.

Expected value

The expected value, denoted E[X], of a set of random variables $\boldsymbol{x}$ is the weighted average of all possible values the variable can take. The weights are described by the probability density of the variable.

$$E[X] = \int x f_X(x) \tag{2.2.1}$$

For a sample of data containing N independent realizations of a variable x, the expected value can be written as a sample mean by

$$E[X] \simeq \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2.2.2}$$

When N approaches infinity, the difference between the sample mean and the expected value approaches zero.

Variance

The variance is a measure of the average squared distance between a set of data point and their mean value.

$$VAR(X) = E[(X-\mu)^2] \tag{2.2.3}$$
$$= E[X^2] - (E[X])^2 \tag{2.2.4}$$

The sample variance can be approximated using a similar expression as the expected value

$$VAR(\boldsymbol{x}) \simeq \overline{(x-\bar{x})^2} \tag{2.2.5}$$
$$= \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2 \tag{2.2.6}$$

The normalization by dividing the sum by N-1 is a method to create an unbiased sample variance. Since one of the degrees of freedom is used to calculate the sample mean,$\bar{x}$, the degrees of freedom is reduced to N-1.

Covariance

The relationship between two random variables can be described using the covariance as a measure of their linear dependency. If both variable entries are greater, or smaller, than their respective means, their contribution to the covariance takes

a positive value. Covariance is dependent on the scale used, thus not a good measurement for linear dependency across variable sets.

$$COV[X,Y] = E[(X - E[X])(Y - E[Y])] \tag{2.2.7}$$
$$\simeq \overline{(x - \bar{x})(y - \bar{y})} \tag{2.2.8}$$
$$= \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) \tag{2.2.9}$$

These basic statistical terms forms the basis for analyzing data in the Ensemble Kalman Filter.

2.3 Monte Carlo simulations

Computing an integral, A, over a high dimensional space with uncertainty input variables can be challenging computational. Considering such an integral

$$A = \int_R h(\boldsymbol{x})d\boldsymbol{x} \tag{2.3.1}$$

Given the vector $\boldsymbol{x}$ defined in the domain R with d dimensions. Assuming that uniform samples are drawn to evaluate the integral numerically, say a minimum of s samples in each direction are required to represent a grid on which the function is to be evaluated. The computational workload, W, to calculate the function values for the samples, are proportional to:

$$W \propto s^d \tag{2.3.2}$$

In addition to calculating the function values, it's also needed to store and keep track of the grid used.

To cope with this workload, alternative sampling methods exist. Monte Carlo (MC) methods are such a class of computational algorithms using random sam-

ples to compute the result. The use of random samples makes this a stochastic alternative to the deterministic grid method described above. The methods are especially applicable for simulations with significant uncertainty in inputs and systems with a high degree of freedom.

Using a MC method, N independent and identically distributed (i.i.d) samples of x are drawn from the domain R. An approximating to A can be done with these samples by:

$$\hat{A}_N = \frac{1}{N}[h(\boldsymbol{x}_1) + ... + h(\boldsymbol{x}_N)] \tag{2.3.3}$$

Given a set of independent random variables with a common and finite mean and variance, the mean of the samples will converge towards the expected value when the sample size approaches infinite.

$$\lim_{N\to\infty} \hat{A}_N = A \tag{2.3.4}$$

The accuracy of this sampling is given by the central limit theorem:

$$\sqrt{N}(\hat{A_N} - A) \xrightarrow{R} N(0, \sigma^2), \qquad as\ n \to \infty \tag{2.3.5}$$

$$\sigma^2 = Var[h(\boldsymbol{x})] \tag{2.3.6}$$

Seen from Equations 2.3.5 and 2.3.6, the accuracy of the sampling estimate does not depend of the dimensionality of the space sampled, but only on the finite variance of the variable. The variance is normal increasing with higher dimensionality, so for a high dimension problem it could be very large, thus requiring a larger sample size for the sampling to be accurate. Errors with the use of this method is proportional to the square root of the sample size, N.

Several versions of sampling schemes exist, with MCMC (MarkovChain Monte-Carlo) being one that is frequently used in reservoir simulation. One of the benefits of using such a method in reservoir simulation is through the evolution of the parameters. Considering a single predefined best estimate realization of a reser-

voir model. By letting the parameters change with time through the use of a simulator, a future state vector,$\boldsymbol{\psi}$ can be defined though the used of a non-linear function operator, $\boldsymbol{A}$, as seen in Equation 2.3.7. Similarly one can use an ensemble of models,$\boldsymbol{\varphi}^j$ based on the same mean as the single estimate and let these models evolve with the use of the same simulator.

$$\boldsymbol{\psi}_{k+1} = \boldsymbol{A}(\boldsymbol{\psi}_k) \tag{2.3.7}$$

$$\boldsymbol{\varphi}^j_{k+1} = \boldsymbol{A}(\boldsymbol{\varphi}^j_k) \tag{2.3.8}$$

As stated in Equation 2.3.9, the mean of the future ensemble is not necessarily the same as the future best estimate vector. Based on the stochastic nature of the Monte Carlo simulations, the mean of the future ensemble will neither have to be the same every time.

$$\boldsymbol{\psi}_{k+1} \neq \overline{\boldsymbol{\varphi}^j_{k+1}} \tag{2.3.9}$$

Chapter 3

Reservoir modelling and forecast

Any uncertainty dynamic model used for predictions needs to be updated in order to get a more accurate forecast. In the petroleum industry these forecasts are based on the reservoir model, which is introduced in the first section of this chapter. Further a short analysis of error present in modelling is given and a short discussion of forecast accuracy is presented towards the end.

3.1 Reservoir modelling

In short terms, the reservoir model is a computer model based on the geological model. The purpose of the reservoir model is to help facilitate numerical calculations in an effective matter. To achieve this, a huge detailed geological model must go through a reduction in parameters to reduce simulation time, usually done by dividing the geological model into a grid system where each grid block is completely homogeneous. Each grid block contains quantitative parameters from the static geological model in addition to dynamic parameters. This makes the reservoir model a good base for consistent analysis of the hydrocarbon presents, thus also giving a good basis for further economical assessment. It is also used for production optimization through well placement and tertiary recovery techniques.

Dynamic parameters are parameters that change during production history. These

could be well rates, grid block pressure or water saturation. Some of the dynamic parameters, like grid block pressure, can vary significantly during years, or even months, of production. Matching these parameters are very challenging, as one during production does not have data specific to one grid cell, but rather data only from the production and injection wells which is a result of actions from many grid-blocks. Static parameters, like mineralogy and net to gross ratio, do not change during production history. These kinds of parameters should thus be easier to model, as they would have the same value during the entire history matching period.

With all these parameters that needs to be defined, data is acquired from several different sources in the petroleum industry. Each type of data has a certain spatial extent associated to it. A large-scale geophysical survey, like a seismic survey, could provide us with rock-types, fault placement and in some cases porosity estimates. These data does not reveal much of the heterogeneity within a cubic meter of rocks. Through well tests one can get estimates of the product of permeability and thickness of the reservoir, but as with geophysical surveys these data are a sum of actions across a large volume of the reservoir, thus does not explain the actual parameters in one region or in one grid cell. Further data collection through well logs and core samples could give data like net to gross ratio, electric resistivity, hydrocarbon content and porosity. Applying these data to a reservoir model needs to be done with consideration to both spatial variance and measurement errors. With all this data incorporated in the model, a final quality check needs to be done in order to see whether the model is geological plausible. This requires a geological interpretation, and could be performed by seeing if e.g. porosity and permeability values are consistent with rock type. Mezghani and Roggero[1] presented an approach where both the fine-scale geostatistical model and the course-scale reservoir flow-model is updated with dynamic data, in order to have a consistency between the two models.

An assumption of spatial relationship in the parameters is one of the cornerstones of multivariate geostatistics. By knowing the property value in several locations, an estimation of the maining nearby locations are available. This could be done by interpolation techniques like kriging. The spatial dependence of such a parameter

is described by a variogram.

Using commercial reservoir simulators, one can use the reservoir model to predict future production. The accuracy of the prediction is influenced by several uncertain factors. Possible sources of error include, among others, the reservoir model, simulator and flow physics. This prediction is often used as a basis for major reservoir development decisions and huge decisions are based on the expected recovery from the simulations. Development in offshore oil field is especially sensitive for these simulations, due to low economic return and limited flexibility in form of low amount of information[2].

3.2 Uncertainty and sources of error

Within regular mechanics in deterministic systems, uncertainty can be seen as a subjective lack of knowledge. This means that an increased amount of information, through further analysis or surveys, might reduce the uncertainty associated with the system. The uncertainty in production forecast can be traced back to two main categories of errors[3], measurement errors and modeling errors.

Measurement error is generally well observed, as measurement tools have a known uncertainty and can often be calibrated. In addition to a standard deviation from the true value due to accuracy, measurement errors can also be present as a continuous drift from the true value due to wear or operating conditions.

The introduction of continuous well data, like down-hole pressure and multi-phase flow rates, make a lot of information available for matching[4]. The accuracy of flow rates is generally not good, and often separator rates are considered a better estimate. Due to this discrepancy between well measurements and field measurements, back-allocation of produced volumes needs to be distributed either uniformly across producing wells based by their performance, or by a weighting scheme based on experience with the field in question if certain well flow rates are more uncertainty than others. This could lead to a bias in production data used

for history matching, which can increase the uncertainty.

Carter[5] list two types of modeling error present; simplified assumptions in the reservoir model and numerical schemes used in the simulator.

To summarize one could attribute this error to averaging of data during upscaling. Since modeling error in general is hard to quantify, the simulator predictions must be presented with an estimate of their accuracy and limits of their applicability. DeVolder et al.[6] states that simulator results must come with error bars as is expected of experimental results. In production forecasts the results could, as we have seen from the posterior, be given as a spread of values in a probability distribution. Achieving this is not straight forward with a single reservoir model. In order to quantify this uncertainty, the modeler must display the error based on his prior information and assumption. The receiver should in this case be aware of the modeler's view on uncertainty. Some approaches to this are listed by Floris et al.[7].

The errors associated with modeling could also be present in the wellbore or near wellbore area. Perforations, skin effect or blockings of other kinds could restrict the flow, thus resulting in inaccurate production data[8]. Separating these factors in the modeling is difficult based on the production data only.

3.3 Predictive capacities

Through the process of history matching a reservoir model to field production data, a process covered in Chapter 4, it is often assumed that the model will have a forward modeling ability that give results corresponding to the future production. Carter et al.[3] showed that their reservoir model had poor predictive capacity, even though it had identical physics and identical spatial and temporal resolution as the true model. This contradicts the explanation of model errors based on improving resolution or inclusion of more physics. The reservoir in question was a relatively simply model, a cross-section with alternating layers of good and poor sand and a vertical fault. Three parameters were adjusted during the conditioning phase

with monthly data over three years, and the model forecasted production for the following seven years. The conditioning was done with a genetic algorithm.

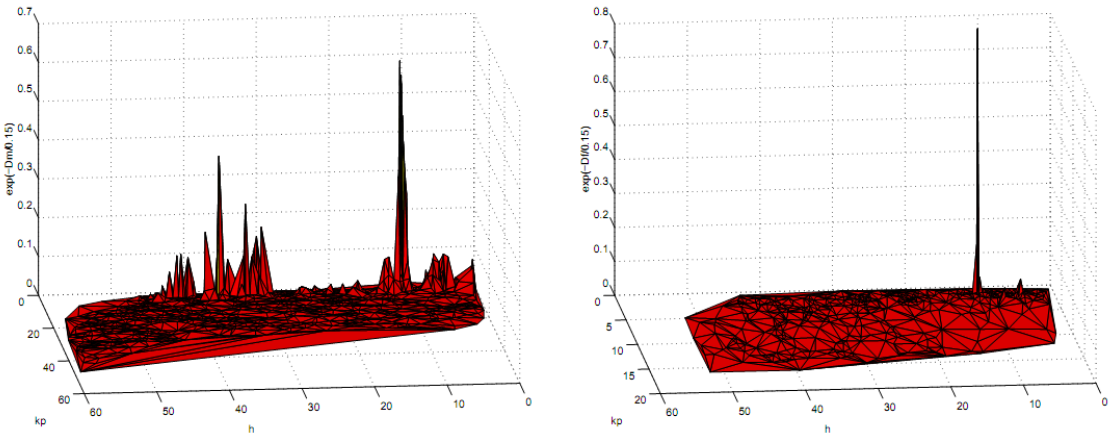

Figure 3.3.1: Left: Objective function of model fit versus fault height and the poor-sand permeability.
Right: Objective function of model forecast versus fault height and the poor-sand permeability. No modeling error. Figure by Carter et al.[3].

As seen from the left plot in Figure 3.3.1 several local minima exists that would be considered a match. Only one (the true case) could represent the prediction period. They were able to obtain a good fit to production rates and a bad fit to parameter, and vice versa. Tavassoli et al.[9] show a relationship between forecast accuracy and the minimized objective function for the same case, still without modeling errors.

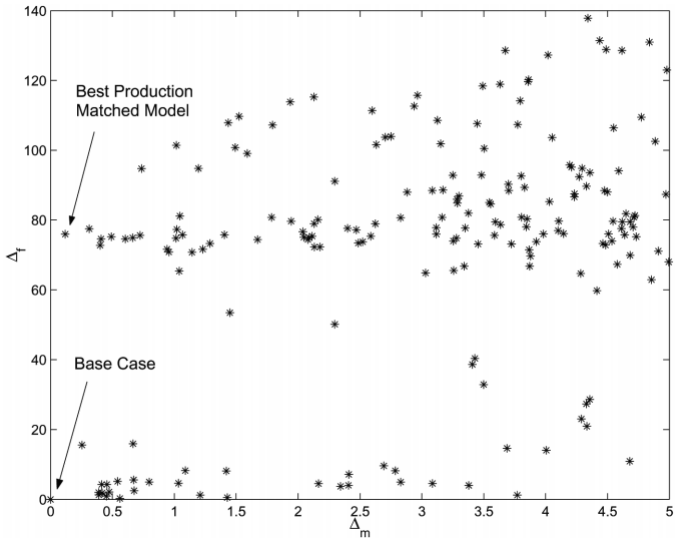

Figure 3.3.2: $\Delta_{forecast}$ vs $\Delta_{modelfit}$. Figure by Tavassoli et al.[9]

The risk associated with the uncertainty in production forecast can be assessed by a sensitivity analysis of the key variables, but due to the non-linear nature of the oil field such an analysis doesn't have to be conclusive. Tavassoli et al.[9] state that bad parameter fits could still produce a good forecast, while being far away from the true case. Since the true model represents the complex geological structure of the reservoir, one cannot guarantee a true representation. This is also seen in Figure 3.3.2. An updated reservoir model will have an associated parameter uncertainty, but this model uncertainty is not directly correlated with production uncertainty[10].

Chapter 4

History Matching

The reservoir model is the basis for many decisions in the oil industry and inaccuracy could significantly change the estimated value of projects. First a general introduction to history matching and it's value is presented. This is followed by some of the challenges involved in estimating the input parameters based on the system output. Evaluation of the match of data is performed through some objective function. Three widely used functions is presented, and a short statistical analysis on the update is given.

4.1 History Matching in general

History matching is defined as an adjustment, or conditioning, of the reservoir model using the data history in order to make it reproduce the observed behavior[11]. The purpose of such updating is mainly in order to make better predictions of the future production by the use of a numerical simulator.

Production rates, pressure and GOR are some among several observable parameters that could be used as input variables in a history match. This could be done manually or as a semiautomatic procedure.

Manual history matching is often done in two steps[12], a pressure match and a saturation match. Since the model is updated by manually changing reservoir

parameters, few parameters are changed in each updating step. The heterogeneous model with thousands of grid blocks makes the engineers experience an important factor for an improved model.

Parameters like absolute and relative permeability is often changed by multiplying the base parameter from the latest model, with a constant for the entire region updated. Since small changes is used in order to control the evolution of the model, this could be time consuming and rely on the reservoir engineers experience not only with the model, but also with the field itself. The engineers needs to use his own assigned probabilities in order to decide which parameters that is going to be updated and what restrictions these parameter are subject of. The flexibility regarding which parameters to update is great, but for a large number of variables this is not an optimal technique.

With large reservoir models the time required to run the simulator might be significant in order to match data. Manual history matching then becomes a time consuming task with the resulting update being a single case reservoir model, hopefully better than the last version. In addition, this will not lead to a better assessment of uncertainty

4.2 Value of Information

The sole purpose of history matching is the introduction of new information to a reservoir model. Available field data is processed and the reservoir model description is optimized with respect to these data. The amount of work and time spent on history matching is significant, and is also reflected in the amount of papers published. However, conditioning the reservoir model has by itself no meaning unless the added information to the model could be used to change a decision. Such a decision might be the choice of well locations or enhanced oil recovery methods. If no such decisions exist or is planned, the only advantage of history matching is the reduced uncertainty of the forecasts, which by itself has no value.

Value of information is an expression often used in quantification, in terms of money, of value gained from gathering new information. In short terms, value of

information (VOI) can be thought of as the value of a project or decision with the information in question versus the same project or decision without the additional information. By calculating the Net Present Value (NPV) for a project with and without added information, we could conclude whether added information actually has a value. The cost of information must be included in the calculations. Thus for information not given for free, it must bring additional monetary value higher than the cost in order to be justifiable.

The uncertainty of new information is an aspect to be considered. As history matching is a complex inverse problem, several solutions exist. We know that a history matched model isn't a perfect representation of the reality, thus the information given by the history match is imperfect. Real life information is often imperfect; examples are seismic surveys, well-logs and core samples. Imperfect information can be described with a probability less than 1 of being true or give an estimate with expected deviation (continuous variables), while perfect information will have a probability of 1. This doesn't necessary mean that imperfect information bring less value to a project than perfect information. The VOI is dependent on a change in decision, so both imperfect information and perfect information can give then same calculation of NPV.

Some might argue that a history matching model is better because of its improved predictive capacities. Forecast inaccuracy put aside, reduction of uncertainty itself is virtually worthless. If the cumulative hydrocarbon production doesn't increase, revenues will stay the same, because oil in place (OIP) doesn't change with the assessed probabilities. An assumption in this statement is that the history match doesn't increase the production rates early in production history without increasing the recovery factor. An uncertainty reduction might add flexibility to companies where a accurate prediction of revenues is necessary in order to support other projects, but the increase accuracy comes with the price of new information.

An assumption made in the sections above, is that the history match actually increases production predictability. As mentioned by Carter et al.[3] this is not necessarily the case, and this topic will be covered in Chapter 7 and 8 as well.

4.3 Assisted History Matching

Assisted history matching routines uses optimization algorithms to mathematically describe the observed mismatch and update the reservoir model. Such routines can be divided in two groups, deterministic and stochastic models. The deterministic models require calculations of Jacobian or Hessian matrices in order to approach the true solution. Stochastic models focuses on minimizing a objective function without the use of gradients. This is generally more computationally demanding because of the model parameter space being searched can be very large, but can generally be applied to non-linear cases where gradient calculations are very hard. Major challenges for both methods are the huge parameter sets, the non-linearity of reservoir simulation and several local minimas in the objective function.

Even though these methods are called assisted or automatic, subjectivity from the reservoir engineer still plays a major role in how these methods are applied. Subjectivity is present in boundary condition as well as the uncertainty analysis of data used. This means that any gut feeling used in a manual history match, needs to be quantified and placed into the model.

4.4 Inverse Theory

By combining a model description of a physical system with the physical theories governing the behavior of the system, we can make predictions of the system outcome[13]. This prediction of a problem is called a forward problem and in a deterministic setting, this forward problem has a unique solution. For instance, oil production can't be both high and low at the same time, it has a certain value.

The physical theories can often be summarized by mathematical equations, dividing the system model into smaller processes. Processes that are assumed less important for the application of the model, is in many cases excluded[14]. This is a source for modeling error, as described in Section 3.2.

During a history matching procedure, we already have the output from the real-life system and we want to optimize the model description. This is called an

inverse problem, and in contrast to the forward problem, it does not necessarily have a unique solution. Albert Tarantola[13] postulated the concept of the state of information over a parameter set. In the textbook cited, the author states that the most general way to describe a state of information is by defining a probability density over a parameter space.

In the case where model parameters are directly observable in nature, data acquired in the field makes the basis for the probability density of the observable parameter. For some parameters, like production data or neutron density in a wireline log, information of the physical correlation between the observed parameter and the model parameter is necessary.

The parameter space is defined by the physical nature of the parameter and based on geological or reservoir engineering knowledge, additional boundaries in the parameter space can be made. An example is the porosity definition; based on the physics we know that it can't have a negative value. In a certain field, our experience might tell us that the porosity value is not likely to be higher than 30% based on cementation environment, thus creating an upper boundary value. Equivalently, apriori information can constitute a lower boundary for a parameters.

Central in Tarantolas text is that observable parameters, apriori information of model parameters and physical correlation can all be described by a probability density. The inverse problem can then be solved as a combination of all this information. This general solution to inverse problems can be used to solve both linear and non-linear problems. The solution is not *a* specific solution, but rather a sample from the posterior probability distribution. Understanding the results presented in a posterior probability distribution is further explored in the following sections.

4.5 Objective function

Assessment of the quality of a history matching procedure is normally done by minimizing some objective function. Oliver and Chen[15] lists three different purposes for the objective function;

1. minimization of mismatch to the observed data
2. computation of the a posteriori solution
3. obtaining samples from the posterior pdf of the model parameters

With the first one being the one most people associate with, the use of Markov-Chain Monte-Carlo (MCMC) simulation makes the latter points highly relevant. The objective function is also used as a quality indicator of the history match. This represents a challenge, since the goal of history matching often is a more accurate prediction, not just the matching of previous observed behavior.

Oliver and Chen[15] specifies the goal of history matching to find some variable m such that the squared norm of the data is minimized;

$$J(\boldsymbol{m}) = \frac{1}{2}||g(\boldsymbol{m}) - \boldsymbol{d}_{obs}||_D^2 \tag{4.5.1}$$

where $g(\boldsymbol{m})$ is the model output. This is a representation of the widely used sum of squared error method, but it can also be normalized with the inverse of the noise variance. Such an equation is typically used in history match methods where the number of variables is small. Another method is proposed in the literature based on the latter, with the inclusion of the model covariance as regularization;

$$\begin{aligned} J(\boldsymbol{m}) = \frac{1}{2}(g(\boldsymbol{m}) - \boldsymbol{d}_{obs})^T \boldsymbol{C}_D^{-1}(g(\boldsymbol{m}) - \boldsymbol{d}_{obs}) \\ + \frac{1}{2}(\boldsymbol{m} - \boldsymbol{m}_{pr})^T \boldsymbol{C}_M^{-1}(\boldsymbol{m} - \boldsymbol{m}_{pr}) \end{aligned} \tag{4.5.2}$$

where $\boldsymbol{C}_D^{-1}$ is the noise covariance and $\boldsymbol{C}_M^{-1}$ is the model parameter covariance. When the number of parameters in $\boldsymbol{C}_M$ becomes large, inversion of the matrix becomes impractical due to computational requirements. The solution of this object function often has multiple minima, which could cause problems for gradient-based history matching methods. Equation 4.5.2 could be rewritten as the generalized least square:

$$J(\boldsymbol{m}) = \frac{1}{2}(1-\beta)(g(\boldsymbol{m}) - \boldsymbol{d}_{obs})^T \boldsymbol{C}_D^{-1}(g(\boldsymbol{m}) - \boldsymbol{d}_{obs}) + \frac{1}{2}\beta(\boldsymbol{m} - \boldsymbol{m}_{pr})^T \boldsymbol{C}_M^{-1}(\boldsymbol{m} - \boldsymbol{m}_{pr}) \tag{4.5.3}$$

where β is a subjective weighting factor.

4.6 Bayesian likelihood and posterior pdf

Equation 4.5.2 is an often used objective function, due to its Bayesian interpretation. This is especially the case with MCMC simulations. Floris et al.[7] and Oliver and Chen[15] respectively describes functions for the likelihood and posterior probability distribution based on Equation 4.5.2.

From Bayes theorem one can write[14] :

$$\begin{aligned} f(\phi, \psi) &= f(\phi|\psi)f(\psi) \\ &= f(\psi|\phi)f(\phi) \end{aligned} \tag{4.6.1}$$

$$f(\psi|\phi) = \frac{f(\phi|\psi)f(\psi)}{f(\phi)} \tag{4.6.2}$$

where the conditional probability distribution of ψ given ϕ, is given in terms of the likelihood, probability of ϕ given ψ, and the marginal probability distributions of ψ and ϕ. In Bayesian probability theory and history matching the marginal distribution of ψ is called the prior, the marginal distribution of ϕ is called the preposterior, while ψ given ϕ is called the posterior. In reference with reservoir modeling, the prior is typically the distribution of our existing models, while the posterior is our output after conditioning the data to the likelihood. Equation 4.6.2 can be rewritten without the normalizing denominator, by saying the posterior is

proportional to the product of likelihood and prior:

$$f(\psi|\phi) \propto f(\phi|\psi)f(\psi) \tag{4.6.3}$$

The likelihood is a statistical approach to estimate the probability of observation, given the model parameters. By assuming the measurement and model errors are Gaussian, the likelihood can be written as

$$f(\boldsymbol{d}_{obs}|\boldsymbol{m}) = c \times exp\left[-\frac{1}{2}(g(\boldsymbol{m}) - \boldsymbol{d}_{obs})^T \boldsymbol{C}_D^{-1}(g(\boldsymbol{m}) - \boldsymbol{d}_{obs})\right] \tag{4.6.4}$$

and the prior as

$$f(\boldsymbol{m}) = exp\left[-\frac{1}{2}(\boldsymbol{m} - \boldsymbol{m}_{pr})^T \boldsymbol{C}_M^{-1}(\boldsymbol{m} - \boldsymbol{m}_{pr})\right] \tag{4.6.5}$$

Combining these two as in equation 4.6.3 gives the posterior

$$\begin{aligned} f(\boldsymbol{m}|\boldsymbol{d}_{obs}) &= f(\boldsymbol{d}_{obs}|\boldsymbol{m}) \times f(\boldsymbol{m}) && (4.6.6)\\ &= c \times exp\left[-\frac{1}{2}(g(\boldsymbol{m}) - \boldsymbol{d}_{obs})^T \boldsymbol{C}_D^{-1}(g(\boldsymbol{m}) - \boldsymbol{d}_{obs})\right] && (4.6.7)\\ &\quad \times exp\left[-\frac{1}{2}(\boldsymbol{m} - \boldsymbol{m}_{pr})^T \boldsymbol{C}_M^{-1}(\boldsymbol{m} - \boldsymbol{m}_{pr})\right] \\ &= c \times exp\left[-\frac{1}{2}(g(\boldsymbol{m}) - \boldsymbol{d}_{obs})^T \boldsymbol{C}_D^{-1}(g(\boldsymbol{m}) - \boldsymbol{d}_{obs})\right. \\ &\quad \left. - \frac{1}{2}(\boldsymbol{m} - \boldsymbol{m}_{pr})^T \boldsymbol{C}_M^{-1}(\boldsymbol{m} - \boldsymbol{m}_{pr})\right] && (4.6.8) \end{aligned}$$

where c is the normalization constant caused by $p(\boldsymbol{d_{obs}})$ in the denominator. As seen the posterior probability distribution will also be Gaussian given that the prior and the likelihood is.

The posterior distribution is recommend by Floris et al.[7] for conditioning reservoir models, in order to reduce ill-posed mathematics compared to using only the likelihood function.

Chapter 5

Ensemble Kalman Filter

This chapter explains the mathematical background of the Kalman Filter and the derivation and application of the Ensemble Kalman Filter. Section 5.1 gives a short induction to the filter and an overview of some of terminology used later in this chapter. Section 5.2 describe the updating step of the original Kalman Filter, then the forecasting step. Sections 5.3, 5.4 and 5.5 cover the Ensemble Kalman Filter and some of the differences from the deterministic predecessor. Iterative Kalman Filter and handling of constraints in the filter update is covered in Section 5.5 and 5.6. In the end Section 5.7 summarizes some the published application of EnKF in the petroleum industry.

5.1 Background and terminology

The original Kalman Filter is a mathematical method named after Rudolf E. Kalman[16], who introduced the method in 1960 to estimate the state of a system with noise or other inaccuracies in the existing model and the measurements. The method described combines the predicted state with the observations, with the use of their respective uncertainties. It can be thought of as a simple weighting function, given most weight to the state with least uncertainty. An assumption in the theoretical formulation of the Kalman filter is that the underlying system is lin-

ear and it's uncertainties through error terms and measurements can be described using Gaussian probability distributions.

The Kalman Filter consists of two main steps, the forecast and the analysis. In the forecast step, the model observed is allowed to evolve through time. In the original Kalman Filter this was done with a linear matrix operator, but newer method like Extended Kalman Filter allows for non-linear operators. The second step is also called the updating step, and is where the model is conditioned to measurements available. One of the major advantages, especially in reservoir engineering, is the sequential assimilation in the various types of Kalman Filter. This allows the user to continue working with the model after the update is applied, without the need to rerun the model from start.

For the following sections, some terminology needs to be defined. The reservoir is described with model parameters and state parameters.*Model parameters*, also called static parameters, are parameters that normally does not change with time. Some model parameters are porosity and permeability. These parameters could change with the production of the reservoir in extreme cases, primarily due to pressure changes, but for all practical purposes they remain fairly constant. When these parameters are updated using a data assimilation method, they are not updated due to change, but rather due to the inherent uncertainty associated with the initial estimation. *State parameters* are in contrast to model parameters time-dependent and changes with the production of the reservoir. Due to their changing nature, they are also called dynamic parameters. In reservoir engineering state parameters could include gas-oil-ratio, pressures and saturation of fluids. The uncertainty of state parameters is often directly influence by the model parameters in addition to other sources.

The data we observe in a field is often derived directly from state parameters and indirectly from the model parameters, for instance fluid flow rates and well pressure. The data observed is the reservoir response to the static and dynamic parameters present in the system, and the covariance matrix of the state vector is an approximation to explain this correlation.

The Ensemble Kalman Filter derived later has several properties which make it a suitable method for many history matching projects. The simple conceptual formulation creates a lower threshold for engineers, together with the ease of implementation compared to many other methods, especially gradient based methods. It could in many cases reduce the computational requirements through the sequential update.

A full review of the Ensemble Kalman Filter can be found in Evensen[14].

5.2 Deriving the Kalman Filter

Considering a stochastic system governed by a linear differential equation of the form

$$\boldsymbol{\psi}_k = \boldsymbol{A}_k \boldsymbol{\psi}_{k-1} + q_{k-1} \tag{5.2.1}$$

where $\boldsymbol{A}$ is the linear model operator, q is the unknown model error and ψ is the system state vector. When advancing a numerical model through time, the evolution of the state vector and the covariance matrix will be according to

$$\boldsymbol{\psi}_k^f = \boldsymbol{A}_k \boldsymbol{\psi}_{k-1}^a \tag{5.2.2}$$

$$\boldsymbol{C}_{\psi\psi}^f(t_k) = \boldsymbol{A}\boldsymbol{C}_{\psi\psi}^a(t_{k-1})\boldsymbol{A}^T + \boldsymbol{C}_{qq}(t_{k-1}) \tag{5.2.3}$$

with

$$\boldsymbol{C}_{qq}(t_k) = q_k q_k^T \tag{5.2.4}$$

This summarizes the equations used in the forecasting step of the Kalman filter.

If we ignore the time-dependency of the variables, we can write the following equation for $t = t_k$. First we define the state vector for a system by

$$\boldsymbol{\psi} = \begin{bmatrix} m \\ f(m) \\ g(m) \end{bmatrix} \tag{5.2.5}$$

Here the state vector is a combination of the model parameter, state parameters and observations. The vector is sufficiently defined to describe the unforced motion of the system, given an understanding of the physics and the control functions acting on the system. Since the true parameters to the system is not known, any representation of the system though ψ is an approximation with an associated uncertainty. Thus we can write

$$\boldsymbol{\psi}^f = \boldsymbol{\psi}^t + \boldsymbol{p}^f \tag{5.2.6}$$

where $\boldsymbol{\psi}^f$ is a predicted state, $\boldsymbol{\psi}^t$ is a the true state and $\boldsymbol{p}^f$ is the error in the model. The data associated with the system is also uncertain and can be written as

$$\boldsymbol{d} = \boldsymbol{M}\boldsymbol{\psi}^t + \boldsymbol{\epsilon} \tag{5.2.7}$$

where $\boldsymbol{d}$ is the observation, $\boldsymbol{\epsilon}$ is the associated measurement error and $\boldsymbol{M}$ is a forward operator that relates the system state vector to the observation. For finite-difference calculation of reservoir performance, a simulator is used to calculated the data, $g(m)$, stored in the state vector. $\boldsymbol{M}$ in the form $[0|I]$, is just an operator matrix that selects the relevant rows in ψ. With the exception of $\boldsymbol{M}$, all of the right-side parameters in Equations 5.2.7 and 5.2.8 are unknowns. By making some assumptions about the error terms, we can try to estimate the true state:

$$\begin{aligned} \overline{\boldsymbol{p}^f} &= 0 & \overline{\boldsymbol{p}^f(\boldsymbol{p}^f)^T} &= \boldsymbol{C}^f_{\psi\psi} \\ \overline{\boldsymbol{\epsilon}} &= 0 & \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T} &= \boldsymbol{C}_{\epsilon\epsilon} \\ \overline{\boldsymbol{\epsilon}\boldsymbol{p}^f} &= 0 \end{aligned} \tag{5.2.8}$$

The measurement error covariance matrix, $\boldsymbol{C}_{\epsilon\epsilon}$, is a diagonal matrix if the errors are uncorrelated. In the further derivation this is assumed. In the updated estimate of the true state, a linear relationship between the predicted state,$\boldsymbol{\psi}^f$, and the observations is used. By minimizing the analyzed state error variance, $\boldsymbol{C}^a_{\psi\psi}$, we obtain the following set of equations for the analyzed state and its error covariance:

$$\boldsymbol{\psi}^a = \boldsymbol{\psi}^f + \boldsymbol{K}(\boldsymbol{d} - \boldsymbol{M}\boldsymbol{\psi}^f) \tag{5.2.9}$$

$$\boldsymbol{K} = \boldsymbol{C}^f_{\psi\psi}\boldsymbol{M}^T(\boldsymbol{M}\boldsymbol{C}^f_{\psi\psi}\boldsymbol{M}^T + \boldsymbol{C}_{\epsilon\epsilon})^{-1} \tag{5.2.10}$$

$$\boldsymbol{C}^a_{\psi\psi} = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{M})\boldsymbol{C}^f_{\psi\psi} \tag{5.2.11}$$

The minimization of $\boldsymbol{C}^a_{\psi\psi}$ implies a maximation of the posterior pdf for the state vector within the Bayesian context. $\boldsymbol{K}$ in Equation 5.2.10 is named the Kalman gain, and is the weighting scheme used to update the state vector. When $\boldsymbol{K}$ approached one, the term $\boldsymbol{C}^f_{\psi\psi}\boldsymbol{M}^T$ being significantly larger than $\boldsymbol{C}_{\epsilon\epsilon}$, all weight is based on the measurement. Otherwise, when $\boldsymbol{K}$ approaches zero, not trust is put in the measurement, and thus no update is performed.

A more thoroughly derivation of the Kalman Filter can be found in Evensen [14].

This constitutes the basis for the Kalman Filter. The filter is sequential, so new data can be applied once they are available without the need to run the model from time zero. If no data is available, the system is forecasted to the next step, $t = t_{k+1}$. Non-linear dynamics can be solved using the Extended Kalman Filter (EKF) where the forward operator, $\boldsymbol{A}$, is replaced with a non-linear function. In reservoir engineering, state dimensions are large and calculations of nonlinear state derivatives can be computationally demanding[17]. Even for the regular Kalman

Filter, maintaining the covariance matrix through time is not computationally feasible for high-dimensional systems. This makes the EKF not suitable for history matching of reservoirs.

5.3 EnKF

In 1994 Evensen[18] introduced the Ensemble Kalman filter (EnKF). EnKF is related a class of particle filters, which is a numerical method for implementing a recursive Bayesian filter through Monte-Carlo simulation. This separates the EnKF from the KF and EKF by being a stochastic method rather than the former deterministic methods. Particle filters can solve state estimation problems for multi-modal and non-Gaussian distributions, but the EnKF make the assumption that all probability distribution is in fact Gaussian. This makes it much more computational efficient than many other particle filters.

5.4 Analysis scheme

The error covariance matrices from the original Kalman filter are approximated by the ensemble covariance around the ensemble mean, $\overline{\boldsymbol{\psi}}$

$$\boldsymbol{C}^f_{\psi\psi} \simeq (\boldsymbol{C}^e_{\psi\psi})^f = \overline{(\boldsymbol{\psi}^f - \overline{\boldsymbol{\psi}^f})(\boldsymbol{\psi}^f - \overline{\boldsymbol{\psi}^f})^T} \tag{5.4.1}$$

$$\boldsymbol{C}^a_{\psi\psi} \simeq (\boldsymbol{C}^e_{\psi\psi})^a = \overline{(\boldsymbol{\psi}^a - \overline{\boldsymbol{\psi}^f})(\boldsymbol{\psi}^a - \overline{\boldsymbol{\psi}^a})^T} \tag{5.4.2}$$

In 1998 Burgers et al.[19] showed that the observations need to be treated as random variables having a distribution with the mean equal to the observation and a covariance equal to $\boldsymbol{C}_{\epsilon\epsilon}$. The ensemble of observations was described as:

$$\boldsymbol{d}_j = \boldsymbol{d} + \boldsymbol{\epsilon}_j \tag{5.4.3}$$

$$\boldsymbol{C}^e_{\epsilon\epsilon} = \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T} \tag{5.4.4}$$

where j counts from 1 to number of ensemble members, N_e. Evensen [14] states that the approximation to the error covariance is justified since the true observation error covariance is poorly known and errors introduced by this perturbation can be made less than the uncertainty in the true $\boldsymbol{C}_{\epsilon\epsilon}$ by choosing a large enough ensemble size. It is later commented, among others by Evensen [20] 2004, that this may be an additional source of error. The use of a square root scheme has been proposed [21].
By rewriting Equation 5.2.9 with Equation 5.2.10, 5.4.3 and 5.4.4 :

$$\boldsymbol{\psi}^a_j = \boldsymbol{\psi}^f_j + (\boldsymbol{C}^e_{\psi\psi})^f \boldsymbol{M}^T (\boldsymbol{M}(\boldsymbol{C}^e_{\psi\psi})^f \boldsymbol{M}^T + \boldsymbol{C}^\epsilon_{\epsilon\epsilon})^{-1} (\boldsymbol{d}_j - \boldsymbol{M}\boldsymbol{\psi}^f_j) \ , \tag{5.4.5}$$

$$\overline{\boldsymbol{\psi}^a} = \overline{\boldsymbol{\psi}^f} + (\boldsymbol{C}^e_{\psi\psi})^f \boldsymbol{M}^T (\boldsymbol{M}(\boldsymbol{C}^e_{\psi\psi})^f \boldsymbol{M}^T + \boldsymbol{C}^\epsilon_{\epsilon\epsilon})^{-1} (\overline{\boldsymbol{d}} - \boldsymbol{M}\overline{\boldsymbol{\psi}^f}) \tag{5.4.6}$$

By subtracting Equation 5.4.5 from 5.4.6, the difference $\boldsymbol{\psi}^a_j - \overline{\boldsymbol{\psi}^a}$ can be used to derive an expression for $(\boldsymbol{C}^e_{\epsilon\epsilon})^a$:

$$\boldsymbol{\psi}^a_j - \overline{\boldsymbol{\psi}^a} = (\boldsymbol{I} - \boldsymbol{K}_e\boldsymbol{M})(\boldsymbol{\psi}^f_j - \overline{\boldsymbol{\psi}^f}) + \boldsymbol{K}_e(\boldsymbol{d}_j - \overline{\boldsymbol{d}}) \tag{5.4.7}$$

$$\boldsymbol{K}_e = (\boldsymbol{C}^e_{\psi\psi})^f \boldsymbol{M}^T (\boldsymbol{M}(\boldsymbol{C}^e_{\psi\psi})^f \boldsymbol{M}^T + \boldsymbol{C}^\epsilon_{\epsilon\epsilon})^{-1} \tag{5.4.8}$$

$$(\boldsymbol{C}^e_{\psi\psi})^a = (\boldsymbol{I} - \boldsymbol{K}_e\boldsymbol{M})(\boldsymbol{C}^e_{\psi\psi})^f \tag{5.4.9}$$

The covariance matrix, $\boldsymbol{C}_{\boldsymbol{\psi}\boldsymbol{\psi}}$, is the essence in the updating of the state vector. As stated in Section 4.5 one of the difficulties regarding Kalman Filter and other deterministic algorithms deriving from it, is the calculation, inversion and storage of this matrix.

$$\boldsymbol{C}^e_{\psi\psi} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (y^f_i - \bar{y}^f)(y^f_i - \bar{y}^f)^T \tag{5.4.10}$$

The covariance matrix seen in equation 5.4.10 has the dimension $N_y \times N_y$, with N_y being the length of the state vector. In the synthetic case evaluated in this thesis, the number of variable in the state vector was more than 15000. For a large oil-field, this number could be significantly higher, thus increasing the computational requirements. This is bypassed in the EnKF by storing only the square root of the matrix[22]. This square root matrix will include the ensemble members deviation from the mean as seen in Equation 5.4.11.

$$\boldsymbol{L}_e = \frac{1}{\sqrt{N_e - 1}}(y_i^f - \bar{y}^f) \tag{5.4.11}$$

$$\boldsymbol{C}_{\psi\psi}^e = \boldsymbol{L}_e \boldsymbol{L}_e^T \tag{5.4.12}$$

Replacing the covariance matrix with the square root matrix in Equation 5.4.11 into Equation 5.2.10 to get a new expression for the Kalman gain.

$$\boldsymbol{K} = \frac{\boldsymbol{C}_{\psi\psi}^f \boldsymbol{M}^T}{(\boldsymbol{M}\boldsymbol{C}_{\psi\psi}^f \boldsymbol{M}^T + \boldsymbol{C}_{\epsilon\epsilon})} \tag{5.4.13}$$

$$\boldsymbol{K}_e = \frac{\boldsymbol{L}_e(\boldsymbol{M}\boldsymbol{L_e})^T}{(\boldsymbol{M}\boldsymbol{L_e})(\boldsymbol{M}\boldsymbol{L_e})^T + \boldsymbol{C}_{\epsilon\epsilon}} \tag{5.4.14}$$

$$\tag{5.4.15}$$

The significant benefit of this replacement is the dimensions of the matrices used in calculating the update. The product $\boldsymbol{M}\boldsymbol{L_e}$ is the last N_d rows of $\boldsymbol{L_e}$ and has the dimensions $N_d \times N_e$, where N_d is the number of measurements and N_e is still the ensemble size.

An additional assumption made is that measurement errors are uncorrelated in time. The advantage of EnKF over deterministic methods, like the KF or EKF, is that the Kalman Gain, $\boldsymbol{K}_e$, is approximated from the ensemble of members, thus decreasing the relative computational workload.

The EnKF is a sequential updating filter, which could be applied every time new

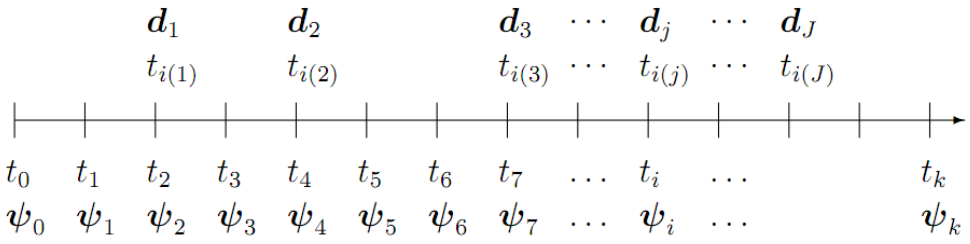

Figure 5.4.1: Model state and observation discretized in time. Observations only available at discret subset of times. Figure by Evensen[14].

data is available as seen in Figure 5.4.1. In contrast to the Kalman Filter and the Extended Kalman Filter, only the ensemble state vectors are updated by the arrival of new measurements. In the KF and EKF the covariance matrix is also updated, but as seen in the derivation above the EnKF uses the updated state vectors for this calculation. Given an infinite ensemble size, the EnKF will give the same results as the KF for a linear case. Still, this EnKF scheme is an approximation by its sampling and processing of non-Gaussian priors for $\boldsymbol{\psi}$. It doesn't solve the Bayesian update for non-Gaussian pdfs, but since the updates that are added to the prior are linear, the updated ensemble will inherit some of non-Gaussian properties.

A finite sized ensemble will always provide an approximating to the error covariance matrix. Based on the Berry-Esseen theorem[23], the errors of the sampling will decrease proportional to $\frac{1}{\sqrt{N_e}}$, as seen in Section 2.3. Evensen[24] states that small mis-specification of the initial ensemble will not influence the results much over time, and recommends an approach of adding perturbations to the best-guess estimate and verifying that the ensemble is stable over the characteristic time scale. The initial ensemble has also been research by, among others, Oliver and Chen[25]. A reduction in ensemble size was attempted with the use of eigenvalues from the initial ensemble and added orthogonality between the realizations, with a biased result for a non-linear case.

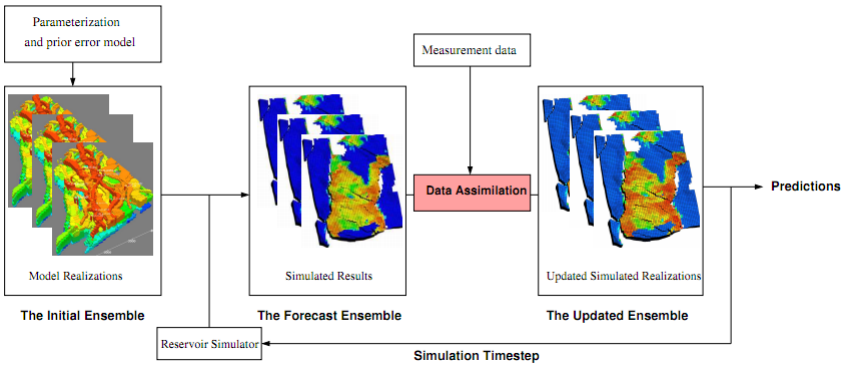

Figure 5.4.2: The general EnKF workflow for petroleum application. Figure by Seiler et al.[26].

As seen by the workflow of EnKF in Figure 5.4.2 the ensemble changes with time, but the ensemble size stays the same. The reservoir simulator is used as the forward operator, both for predictions and forecasting of the state vector.

5.5 Iterative EnKF

Even though the Ensemble Kalman Filter consists of a linear update, it still captures some non-linear behavior. When the non-linearities becomes significant or with a large update of the state vector, the regular EnFK could fail updating the model[27]. Often seen, the state variables are assigned value after the update which is outside their physical domain. Modeling of a water front through S_w could result in such a fail, because of the bimodal probability distribution.

An iterative approach could be incorporated in order to cope with these challenges. In an update consisting of a significant change in both model parameter and state parameters, the state parameters values will depend on the model parameters, thus there is a need to update the model parameters first. Seen below is a representation of the iterative workflow.

$$\psi_{k-1}^{a} \xrightarrow{forecast} \psi_k^f \xrightarrow[parameters]{update\,model} \psi_k^{a'} \tag{5.5.1}$$

$$\psi_k^{a'} = \psi_0^{a'(k)} \begin{cases} model\,parameters\,from\,t = t_k \\ state\,parameters\,from\,IC \end{cases} \tag{5.5.2}$$

$$\psi_0^{a'(k)} \xrightarrow{forecast} \psi_k^{f'(k)} \xrightarrow[parameters]{update\,state} \psi_k^{a} \tag{5.5.3}$$

In the first step the model is forecasted in the same way as the regular EnKF, but in the analysis only the model parameters are updated. A criterion could also be used, revoking the iterative function only if the update is significant enough. The reservoir model is then rerun from the beginning with updated model parameters and initial dynamic conditions. The EnKF loses its sequential properties, but the dynamic conditions are conditioned to the static parameters. Initial dynamic conditions is often described through a form of equilibrium, thus they would be less uncertain than any updated conditions.

It can also be shown that this update is a correct statistic approach. From Gu and Oliver [27]

$$\begin{aligned} f(m_k, f_k|d_{obs,k}) &\propto f(d_{obs,k}|m_k, f_k) f(m_k, f_k|d_{obs,k-1}) \\ &\propto f(d_{obs,k}|m_k, f_k) f(f_k|m_k, d_{obs,k-1}) \\ &\quad \times f(m_k|d_{obs,k-1}) \\ &\propto f(d_{obs,k}|m_k) f(m_k|d_{obs,k-1}) \end{aligned} \tag{5.5.4}$$

5.6 Constraint handling

In Chapter 7 simulations using a constraint on some parameters is performed. Given an assumption that these parameters earlier were distributed with a Gaussian function, a disruption in this distribution can cause some changes in the EnKF. In Figure 5.6.1 below, such a disruption is presented through an lower limit on the

porosity by assuming a resampling from the Gaussian prior if the porosity value is below the limit.

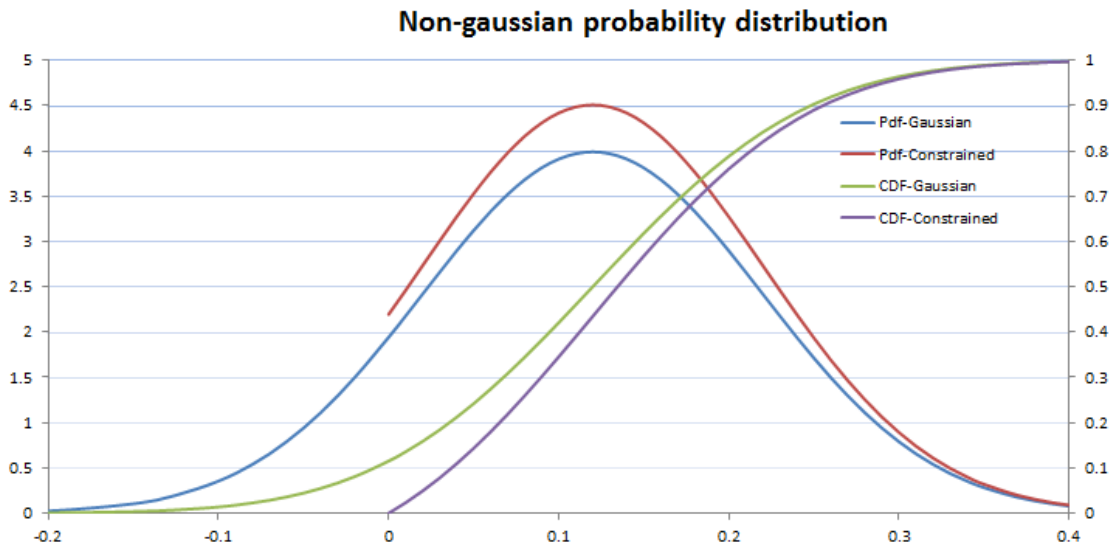

Figure 5.6.1: Constrained cumulative and probability distributions created from a prior Gaussian distribution with mean=0.12 and a standard deviation=0.1, by using a lower limit of 0 to avoid non-physical values.

We often come across variables in the nature which has a non-Gaussian probability density function. In reservoir simulation two such parameters could be permeability and water saturation. Permeability has a log-normal distribution, while the water saturation has a bimodal distribution. These distributions cannot be as easily explained as the Gaussian distribution which only needs the mean and the variance. When handling a Gaussian distribution, the mean, mode and the median is in the same place. As this rarely is the case with non-Gaussian distributions, the mean is often not the most interesting parameter. When working with a multimodal distribution, knowing the modes is beneficial. Adding constraint to a previously Gaussian distribution without changing the explaining parameters; mean and variance, will cause the distribution to gain non-Gaussian elements.

5.7 EnKF in practice

Aanonsen et al.[28] summarize some of the studies done within the petroleum industry since the first application was done in 2001. During the subsequent years, history matching of permeability and porosity was done on several synthetic cases, before applying it to field cases.

Bianco et al.[29] claimed to be the first to publish the use of EnKF to evaluate uncertainty in the production forecast. An on-shore oil field in West Africa was history matched using three ensemble sizes, with 50, 100 and 135 ensemble members. A normalized least square objective function was used to analyze the history matching results, using the GOR and WCT. Considerable improvement was seen for all ensemble sizes, and the forecast uncertainty was assessed by predicting future production using the update ensemble.

Haugen et al.[30] used the EnKF on a North Sea reservoir model with 45000 active grid cells and five years of production history. Updated permeability and porosity field match production better than the manually updated reservoir model. Evensen et al.[31] applied the EnKF to a larger North Sea case, with over 80000 grid blocks, to assimilate oil rate, gas rate and water cut data. The objective match improved significantly, but was not able to reproduce all data from this reservoir. Both processes used an ensemble of 100 members. In 2005 Skjervheim et al.[32] applied a combination of the EnKF and the Ensemble Kalman Smoother (EnKS) on a reservoir model using 4D seismic data.

Seiler et al.[26] applied the EnKF to the Omega field in the North Sea, a field with high spatial and temporal variations. The parameters evaluated were porosity, permeability, relative permeability and fault multipliers. These parameters were history matched against 6 years of monthly production data, with 20 ensemble members. The posterior distribution has a much smaller uncertainty span than the priors. The initial ensemble did not predict a correct water breakthrough and the updated ensemble was not able to correct this bias. Results showed a decent prediction capacity during the following months of the history match, with a deterioration as time progressed. Still, the results show that the last updated

ensemble has a prediction closer to the measured value than the former ensemble. Seiler et al.[26] states that the predictions are not directly comparable to the measurement, due to workovers and choke manipulation at the platform.

A combination of the EnKF and the RML has been proposed in order to maximize the likelihood. This method is an iterative version of the EnKF, in order to enforce constraints and correctly sample the posterior pdf. Because of its iterative nature, it is computationally more expensive than the EnKF, but has a better predictive capacity in non-linear applications. Gu and Oliver[27], Chen et al.[33] recommend combination of EnKF and EnRML, where the EnRML is applied only when state variable corrections are large.

Chapter 6

Simulation

This chapter serves to give a short overview of the software and the model used for the simulations and history matching, in addition to the field geology and development.

6.1 Software

The application of the Ensemble Kalman Filter is done through a protected MATLAB-code provided by the International Research Institute of Stavanger, IRIS. MATLAB is a numerical computing environment and programming language. The entire ensemble is managed, stored and updated using MATLAB, while the reservoir forecasts is performed by Schlumberger's ECLIPSE 100. ECLIPSE 100 is a fully-implicit, three phase black oil simulator using a finite volume method to solve material and energy equations. The software have been run on a Unix platform. Some of the results have been displayed and exported using S3Graf.

6.2 The PUNQ S3 Synthetic case

6.2.1 Geological description

The PUNQ-S reservoir models is based on a reservoir engineering study by Elf Exploration & Production. The geological model was made up from regional geology which was known from adjacent fields and wildcat wells. The depositional history is from a deltaic, coastal plain environment. The reservoir is divided into 5 layers, where the layer thickness is in the order of 5meter for all layers.

Layers 1, 3 and 5 are deposited similarly and consist of fluvial channel fills encased in floodplain mudstone. They all have linear streaks of high-porous sands with an azimuth somewhere between 110 and 170 degrees southeast, which is perpendicular to the paleocurrent. Estimates of the streak width range are respectively 800m, 1000m and 2000m and the porosity is estimated to be higher than 20%. The sand streaks are enclosed by low-porosity shales, with porosity less than 5%. The streak spacing is estimated to be in the range 2km to 5km for layers 1 and 3 and in the range 4km to 10km for layer 5.

Layers 2 and 4 facies can be described respectively as marine or lagoonal shale and mouthbar or lagoonal deltas, with a lower porosity than the other layers. In layers 2, irregular spots of porosity higher than 5% can be found, but the shaly sediments is predominantly made up with porosity less than 5%. In layer 4, the delta depositions can be modeled with an intermediate porosity while embedded by a low porosity lagoonal clay deposit. The intermediate porosity is estimated to be in the order of 15% and shaped like a lobe, similar to an ellipsoid, with the longest axis having an azimuth between 110 and 170 degrees southeast.

As the PUNQ-S3 was a base for comparison of conditioning and uncertainty in reservoir models, a truth case was created. A geostatistical model generated the porosity and permeability field based on Gaussian Random Field, with geostatistical value like mean and variograms consistent to the original geological model.

6.2.2 Reservoir Model

Based on the real reservoir, a simulator grid model was constructed for the Production UNcertainty Quantification project. The PUNQ-S3 model consists in total of 2660cells, divided in a 19x28x5 grid where 1761cells is defined as active. This active definition creates a shoebox shaped reservoir, where the simulation is not affected by the cells outside the actively defined reservoir. The extent of the reservoir is approximately 3.5x5.0km

The reservoir is bounded to the east and south by an impermeable fault. To the north and west it's supported by a strong aquifer. Due to this pressure support, no injection wells have been drilling. A gas cap is present in the upper, dome shaped region. Initially six production wells were drilled in the reservoir, but a later part of the PUNQ-S3 project was to capture the effect of adding five additional wells. These five wells are not included in the presented work.

The reservoir is modeled for use with Schlumberger's ECLIPSE 100. Cornerpoint geometry has been used to describe the spatial extent. Aquifer and PVT data stems from the original model. The relative permeability is set using a power law function and no capillary pressure is assumed.

6.2.3 Production Schedule

Data from the first eight years of operation was provided in the synthetic PUNQ-S3 case, inspired by the reference reservoir. The first year was devoted to extensive well testing, divided into four three-month sessions where each had its own production rate. This was followed by a three year shut-in period. This formed the basis for the first part of the history matching comparison. After this four years of actual field production followed. The first two weeks of each year was devoted to a shut-in in order to collect pressure data.

The production in Eclipse was controlled using the keyword WCONHIST and

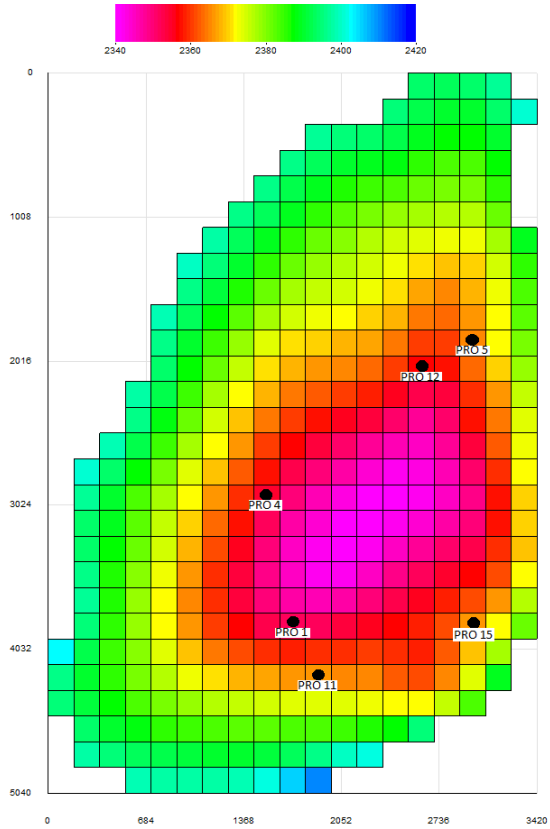

Figure 6.2.1: The PUNQ S3 model seen from above. Grid color refer to depth.

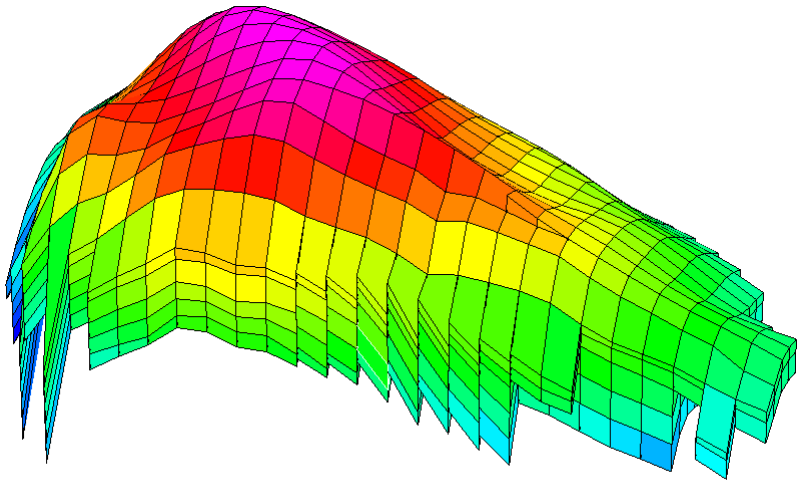

Figure 6.2.2: A 3D representation of the reservoir model. Relative X:Y:Z-scale = 1:1:25

matching with oil rate through ORAT. After the conditioning period producing at given oil rates, the oil rate constraint was changed to a maximum of $150SM3/day$ per well or production at 120bar BHP. An upper limit on gas oil ratio was applied from 1st of July 1975, at 200, with oil as the control phase to adjust.

	Well 1	Well 4	Well 5	Well 11	Well 12	Well 15
3			x	x		
4	x	x	x	x	x	x
5	x	x			x	

Table 6.2.1: Completion by layers

Seen from Table 6.2.1 all wells are completed in layer 4 and some in layer 3 and 5. Wells 1 and 11 are located to the southwest of the top of reservoir, while wells 5 and 12 are located to the northeast. Wells 4 and 15 are located respectively to the west and southeast.

6.3 History matching the PUNQ S3 model

Production values from a true solution were calculated using the true porosities and permeability to provide measurement data for the conditioning. Dynamic variables were initially defined through the equilibrium keyword, EQUIL, in ECLIPSE. The measurement data consisted of bottom hole pressure (BHP), gas to oil ratio(GOR) and water cut (WCT). Since oil production data was used to control the simulations in Eclipse, they were not useable for a conditioning process.

The EnKF was used to update seven variables; three static parameters and four dynamic parameters. The static parameters to be used were porosity, horizontal permeability and vertical permeability. As the equilibrium was initially defined with a certain depth, the initial dynamic condition were considered reliable, but as their evolution is dependent on static parameter a correct initial estimate does not necessarily mean they will be correct after a certain time with production. Thus BHP, GOR and water and gas saturation were updated in every step.

Chapter 7

Problem 1: Effects of boundaries

In the first problem examined, the effects of setting boundaries on the static parameters were investigated. Preliminary results from the history matching process on PUNQ-S3 using EnKF showed some very high values for the porosity. An attempt to reduce the extreme porosity values is presented through two constrained cases with a comparison to an unrestricted case. Section 7.1 introduces the problem background and the boundaries chosen, while the results and a discussion is presented in Section 7.2.

7.1 Background

The initial porosities are determined from geological facies, but since the placement of the porous streaks is not known the initial porosity distribution is very uncertain. In addition, the oil reserves could also be very uncertain, as the pore volume is proportional to the porosity. Conditioning the porosity to the production data could then give a more accurate estimate of the reserves in the reservoir and could thus tell something about future income given a certain recovery factor. In a setting like the real reservoir the PUNQ-S3 model is built upon, a history matching of the first 8 years with respect to oil in place might affect the need of five additional wells. Too account for flow performance, permeability is also updated. The permeability

is normally correlated with porosity, but no such connection is explicitly defined in the code. Understanding the flow patterns in the reservoir is essential to increased recovery.

Physics define some boundaries for the porosities. We know that as the porosity is defined by the ratio of pore volume to the bulk volume, it will take values in the range $[0, 1]$. The same lower limit applies to permeability, so lower limits for both porosity and permeability is incorporate in the code for both simulations to avoid non-physical values.

In addition the geological model sets some boundaries for the estimate parameters that are hard to define in a linear updating sequence like the EnKF. As the porosities are defined for each layer which represent the same depositional time, some upper limits could be set for the porosity layers as well. Based on the knowledge provided by the geological description, upper limit are not explicitly defined so an estimate have been used. In the layers 1, 3 and 5 the porosity ranges from higher than 20% in the high-porous streak too less than 5% in the enclosing shales. In the EnKF code it is possible to define ranges for every single cell if the streaks are known, but as they are not, upper boundaries are defined according to layers. The assigned boundaries are based on the geological description and are summarized in Table 7.1.1.

Initial simulations show that very weak constraints limiting values above 35% for layer 1,3 and 5, and respectively 25% and 15% for layer 2 and 4, did not show a significant change in static parameters. Several attempt where made until the boundaries listed in Table 7.1.1 was decided upon.

All simulations were conditioned to 8 years of production history, and the same measurements was used in all instances.

Preliminary results showed that the Constrained case did not perform as well as the unbounded case when the forecast was analyzed. As explained later in the following section the update step was not complete, so the analyzed state was not fully conditioned to the available data. An iterative version of the EnKF was used with the same input data as the constrained case.

Layer	PoroUB [%]	PermxUB [mD]	PermzUB [mD]
1	0.30	1000	550
2	0.15	250	100
3	0.30	1000	550
4	0.23	550	200
5	0.30	1000	550

Table 7.1.1: Boundaries in EnKF run

This problem has been examined in two parts after the history match. For the first part, a prediction of an ensemble with 100 members is performed to see the forecasting ability of the updated reservoir models. Here the field performance has been the primary point of interest. The second part of problem is a more thorough look at the parameterization of static properties. In this case the mean static parameters have been used to compare the different cases. As the simulations were run from time zero in part two of the problem, dynamic parameters were defined with the same equilibrium conditions as the true case.

7.2 Results

ENSEMBLE FORECAST From Figure 7.2.1, 7.2.2 and 7.2.3, showing the forecast of the Full case, Constained case and the Iterated case, one can observe that neither of the cases completely span the true solution for production data at all times. Looking at the uncertainty of the production forecast, one of the main arguments of using ensemble method is to capture the true production in the forecast. The span of the oil and gas forecasts is in this case very narrow and is not spanning the true data all the time. The increase in the forecast span with time is expected because of the evolution of the model error. When looking at the three forecasts, it is easily observed that the Constrained and Full case perform similarly compared to the Iterated case. One of the observable differences is that the ensemble outliers are slightly further out in the Full case compared to the Constrained case. Because of this, comparison of the Full case and the Iterated case is prioritized in the following paragraphs.

Looking at the oil and gas produced, the Full case performs slightly better than the Iterated case. Both cases span the gas production at all times, with the Full case having the true production closer to the middle of the ensemble at the end of observed time. When it comes to oil production, both cases underpredict the production early in the production history, but approach the true data towards the end. The Full case encloses around the true production during the last years, while the Iterated case never spans the true solution.

The total discrepancy in the gas and oil production is not very big, the deviation in cumulative oil production between the mean of the full case and the true data was in 1979 approximately 2.4%. This was the highest deviation observed for oil production. Although this deviation was small, the ensemble was still expected to span the true solution from the beginning of the forecast.

Contrary to the oil and gas production, the water production shows a more significant deviation. A clear difference between the Iterated case and the Full case is seen. The Full case never spans the true data, and consequently overestimates the water produced. The trends of the water production is similar to the true data, but clearly offset. The Iterated Case does not span the true data for the first 4 years either, but is fairly close. In the following 4.5 years, its ensemble encloses the true data and towards the end of the observation time the true data is located close to the middle of the ensemble. Even though the ensemble spans the true solution after some time, this behavior is opposite of what is expected. Since the updated case is continuously conditioned to data during the first 8 years, it is expected to have a decent forecast the first following time period with gradually declining accuracy. Seen from the ensemble production plots, the ensemble has a very narrow span during the first years, as all the members are updated to the same measurement. Adding of measurement noise and increasing the estimate of measurement inaccuracy could have helped to avoid this.

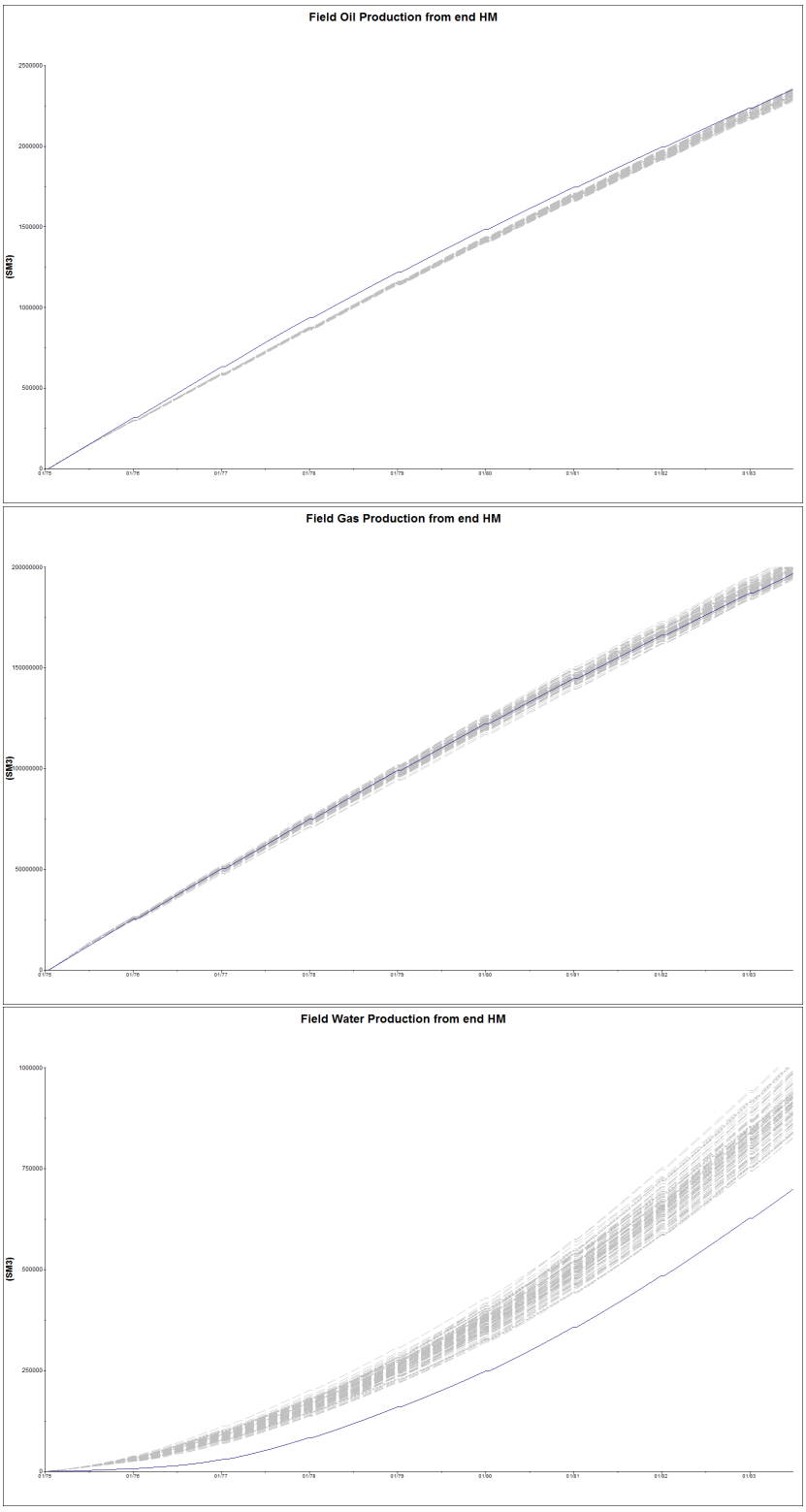

Figure 7.2.1: Production forecasts from end of history match for ensemble members in the Constrained case

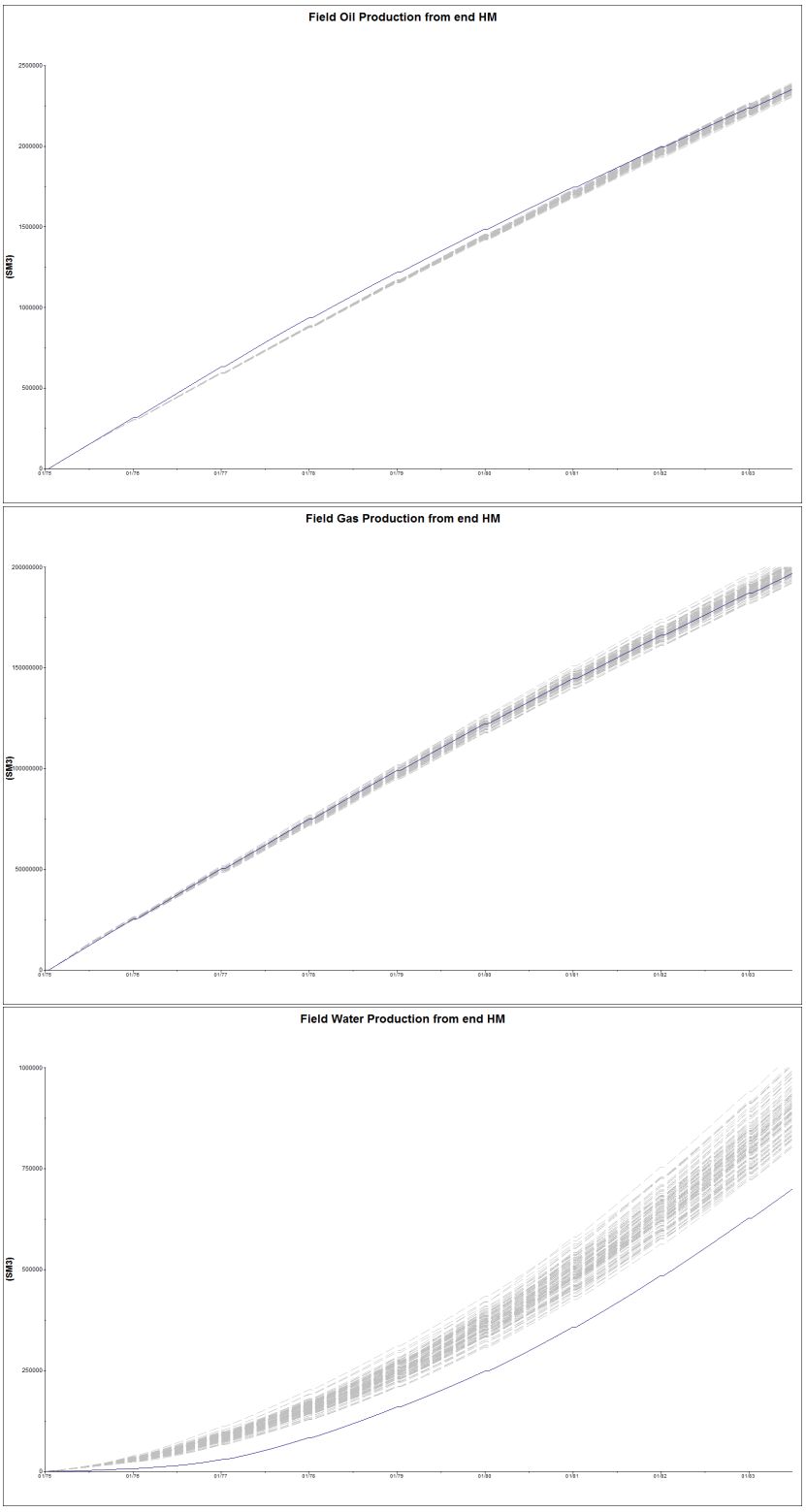

Figure 7.2.2: Production forecasts from end of history match for ensemble members in the Full case.

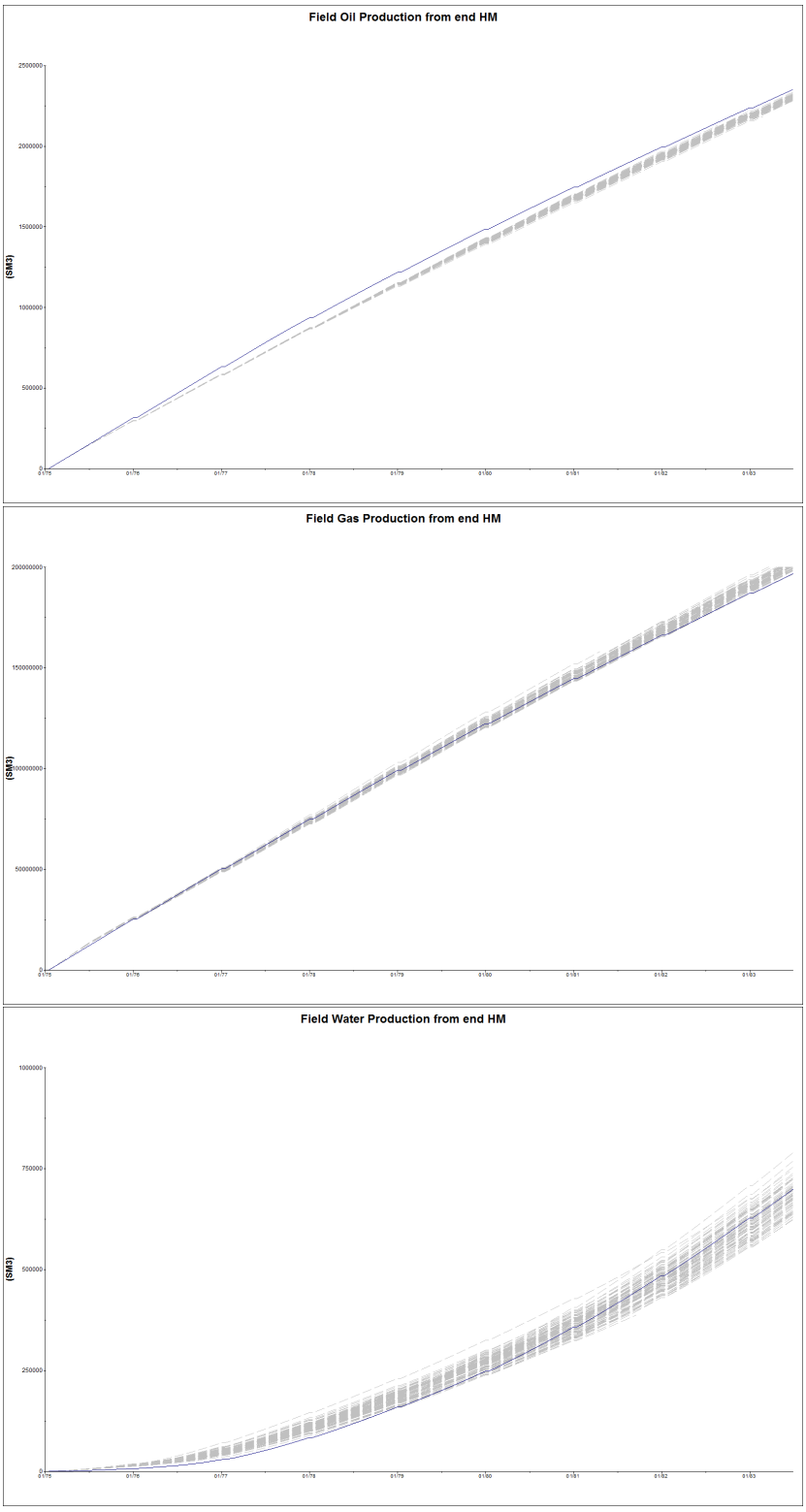

Figure 7.2.3: Production forecasts from end of history match for ensemble members in the Iterated case

Given these field data, it can be seen that none of the three versions is sufficient to describe the truth. While the Full case outperforms the Iterated case when it comes to oil and gas production, the significant discrepancy in water production for the Full case makes the Iterated case a better forecast for water production. The production forecast based on the mean of the ensembles is presented later in the section.

The forecasted production data is one of the primary concerns of the history matching process, in addition to a correct parameterization of the static parameters. From the ensemble forecasts, we see that the measurement span does not enclose the true data everywhere. It is not given that the span of the forecast will capture the true data, even though the span of the reservoir parameters might include the true solution.

STATIC PARAMETER FIELDS A graphical representation of the conditioned porosity and permeability for layer 1 is shown in Figure 7.2.4 and 7.2.5. Plots of the remaining four layers in the reservoir can be found in the Appendix. All grid cells with a dark blue color is cells with a porosity or permeability value above the scale.

The first observation when taking a quick look at these plots is that none of the history matched cases has the same significant geological pattern as the true solution. As the updating step is only a linear interpolation using the covariances as a weighting factor, no geological conditioning is included. The consequence of this is that each model is allowed to evolve without any concern about the underlying porosity and permeability distribution, thus not creating the streaks of high porous sand which the true solution contains. This has to be explicitly defined in the updating algorithm, but is left for further work.

From Figure 7.2.5 we can see the the same lack of geological pattern also exists in the permeability plots. Based on layer 1 we can see that the zones with low permeability are hard to capture using the EnKF. As an example, the lower left third of layer 1 should have permeability close to 0mD, but instead the updated

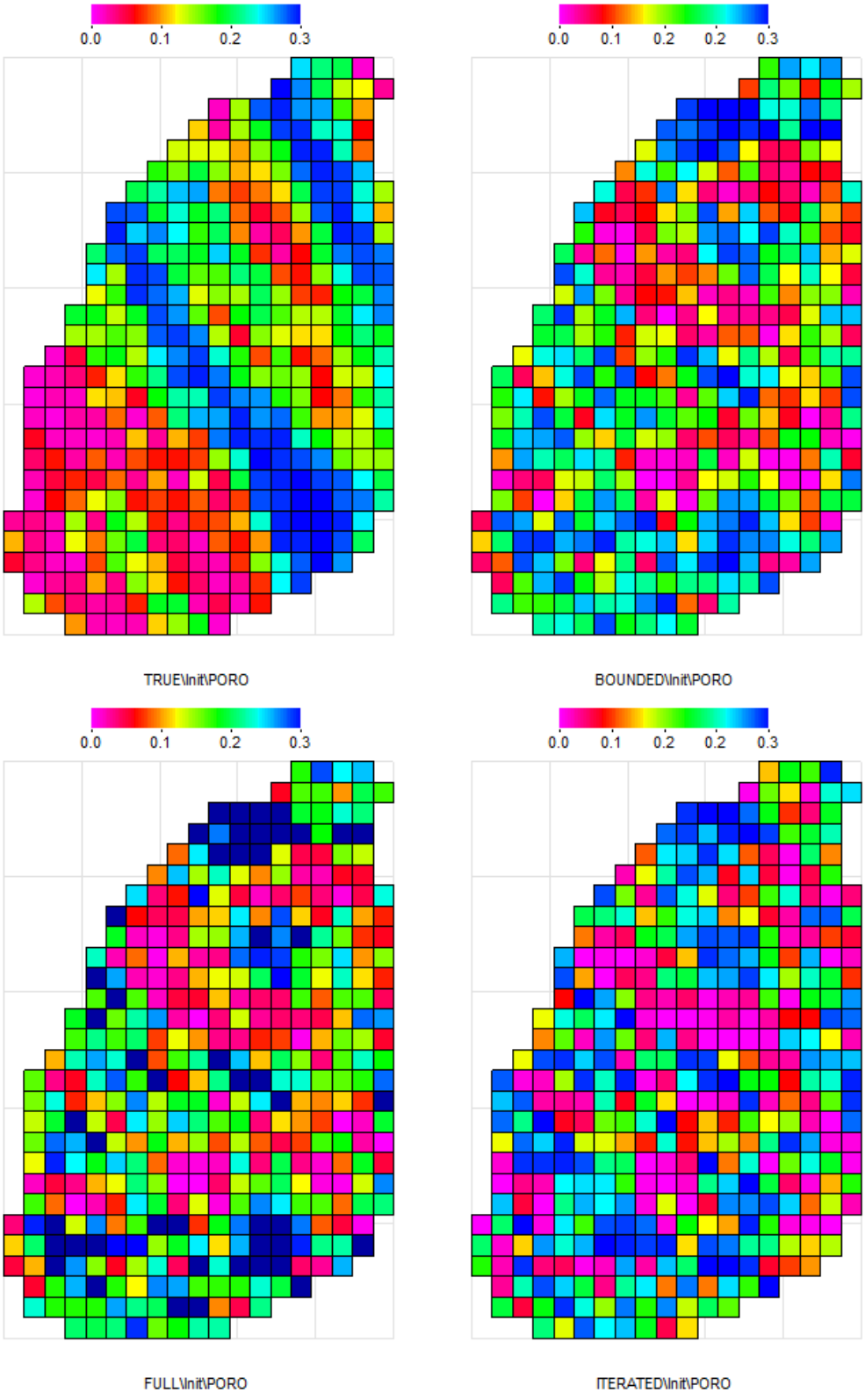

Figure 7.2.4: Porosity in layer 1.

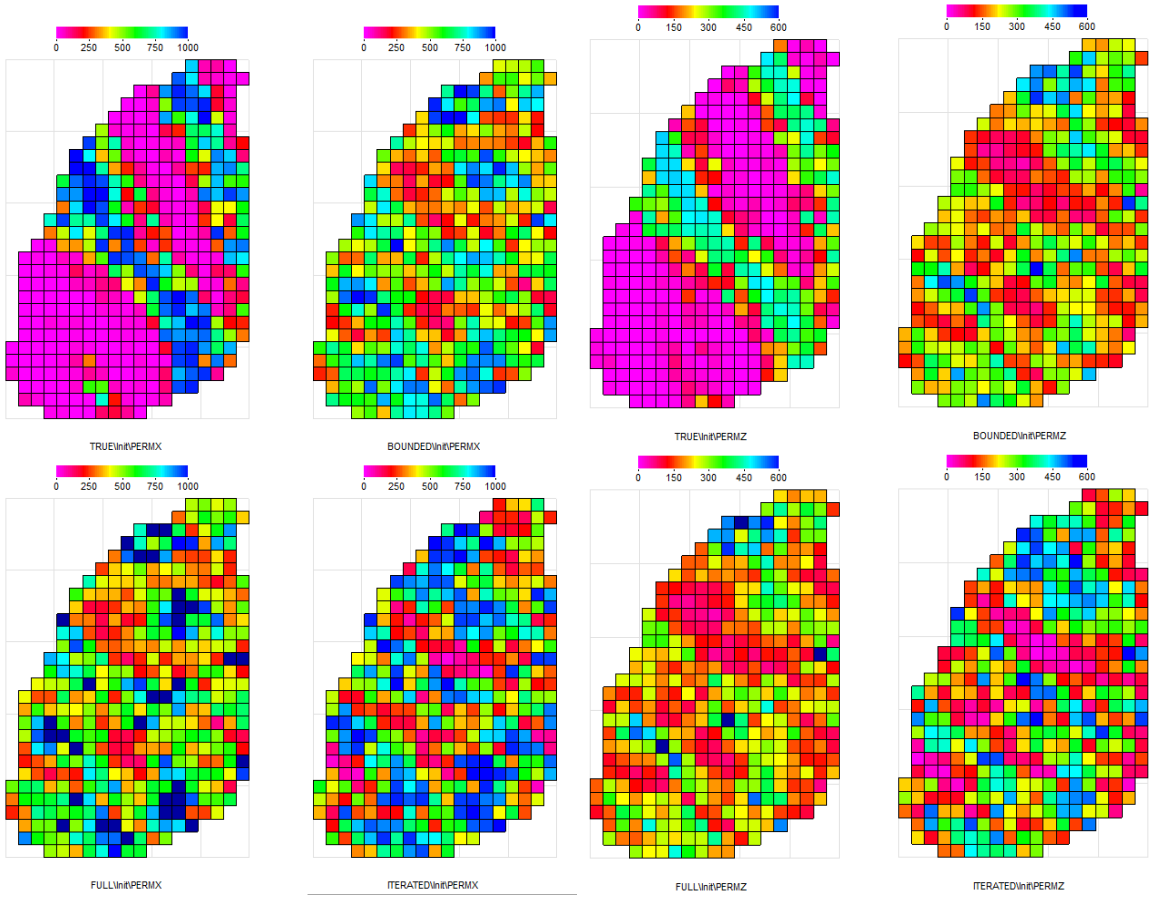

Figure 7.2.5: Permeability in x-direction and z-direction for layer 1

cases have spots of permeability with high value, even above 1000mD for the Full case. The Iterated case is partly able to capture the streaks of shale enclosing the porous streak in the middle of the layer, but the contrast is not stunning.

PERFORMANCE OF MEAN CASES As stated in Section 6.2.3, the production constraint was set to a maximum of 150SM3/day. Seen from Figure 7.2.6, the three examined cases was not able to produce at this rate during the entire production phase. This results in some challenges when interpreting other field performance numbers, but as the deviation in oil production is much smaller than the other observed deviations and the cases should only be considered against each

other, this deviation was accepted.

Figure 7.2.7 and 7.2.9 show the total field gas and water production for the three cases examined. The mean of the static parameters in the ensemble is used, and the simulation is run from 1967 with the same initial conditions as the true case. The figures show the production from 1975, and some of the same trends as seen in ensemble analysis is present here as well. The gas production for all three cases is clearly offset from the true production, but the distance does not increase significant throughout the production time. From Figure 7.2.8 we see that the gas deviation has grown large during the conditioning phase and continue to increase during most of the first year of observation. As seen from the ensemble figures the Iterated case has the highest deviation from the true case, but all cases is in the same range. When looking at the cumulative production the difference between the three cases in gas production is not large compared the offset from the true data.

When looking at the water production forecast, the differences between the models become more evident. All three cases are separated, but they all show some of the same trends as the true case. This is a consequence of all models being based on the same initial ensemble, they follow the same updating algorithm principles and they are subject to the same measurements. Seen from Figure 7.2.9 the Constrained case and the Full case does not have the same starting point as the true case after the conditioning. This corresponds with the gas production offset as seen earlier. An interesting observation is that the Constrained case produces a less accurate water production forecast compared to the Full case. This separation between these two parameters increases steadily with time, in addition to an increased deviation from the true data. Seen from Figure 7.2.10 the similar trends is very pronounced. Note that Figure 7.2.10 has 1974 as a starting point, thus capturing the last year before the production forecast phase. The deviation from the true case starts before the forecasting phase, while the oil production in this time period was the same. Both the Constrained case and the Full case show a water breakthrough very early in Well 5 compared to the true case, and in all three inspected cases the breakthrough in Well 12 is significantly earlier than the true case. Another observation made from the well data, is that the true case has a much earlier water breakthrough

in Well 4 than the conditioned cases. Figure A.4 in the Appendix show the water cut from these wells.

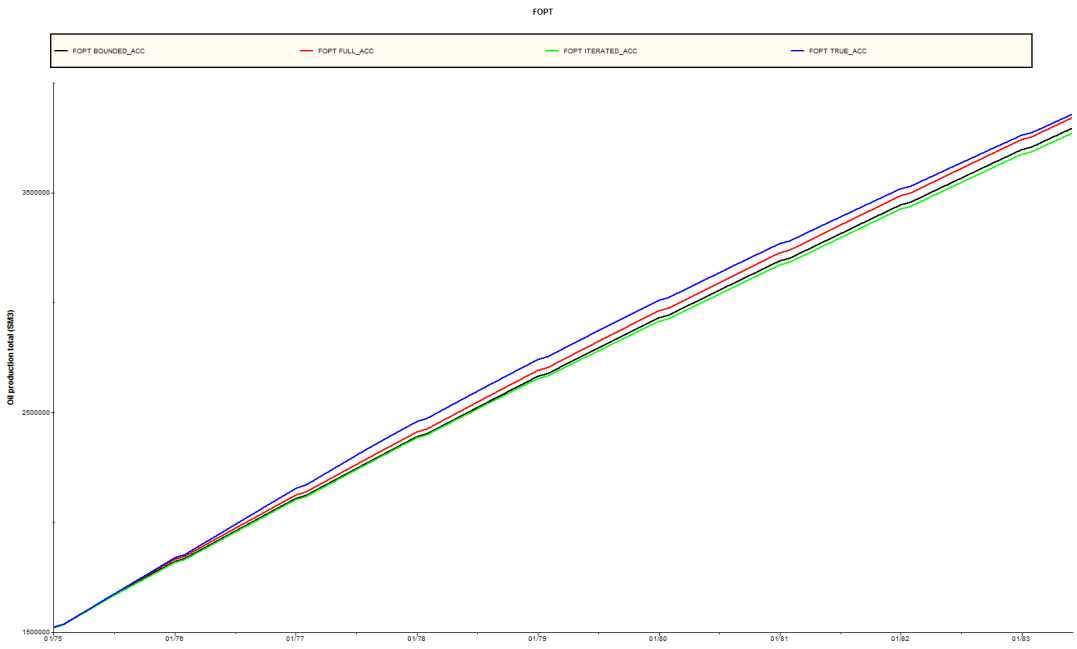

Figure 7.2.6: Field oil production from 1975.

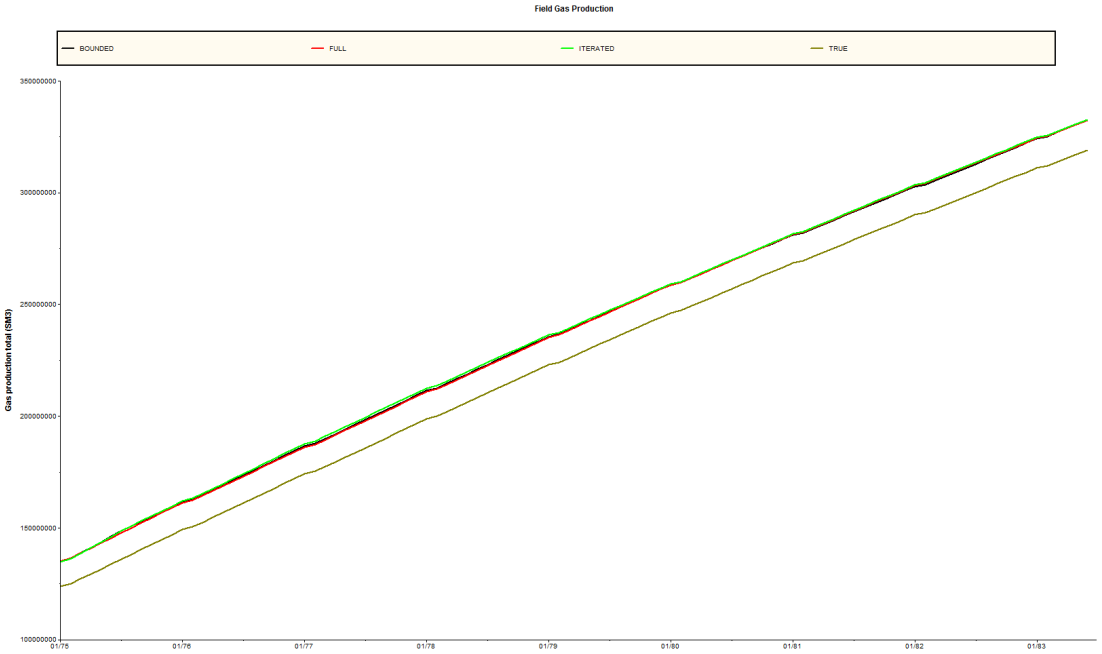

Figure 7.2.7: Field gas production from 1975.

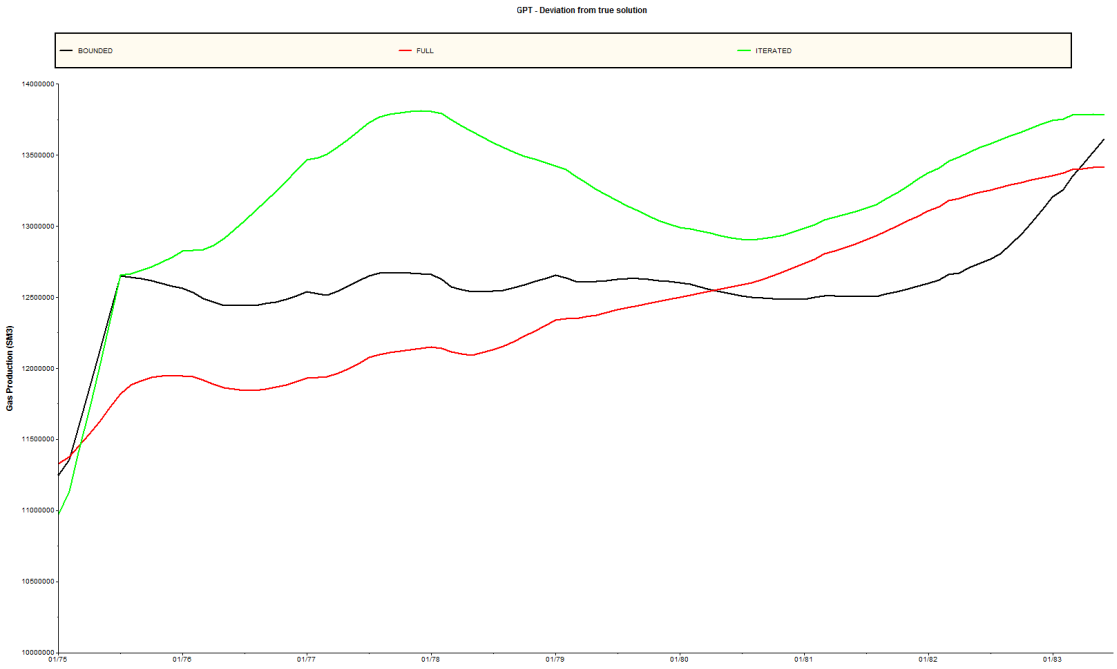

Figure 7.2.8: Deviation in gas production from 1975 compared to the true solution.

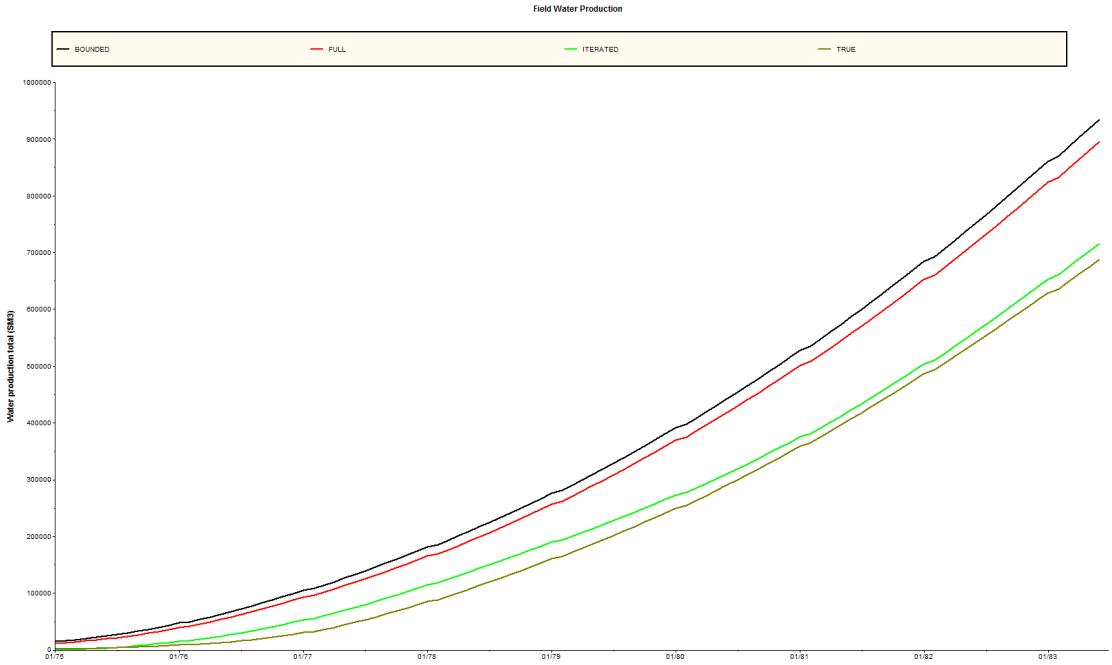

Figure 7.2.9: Field water production from 1975.

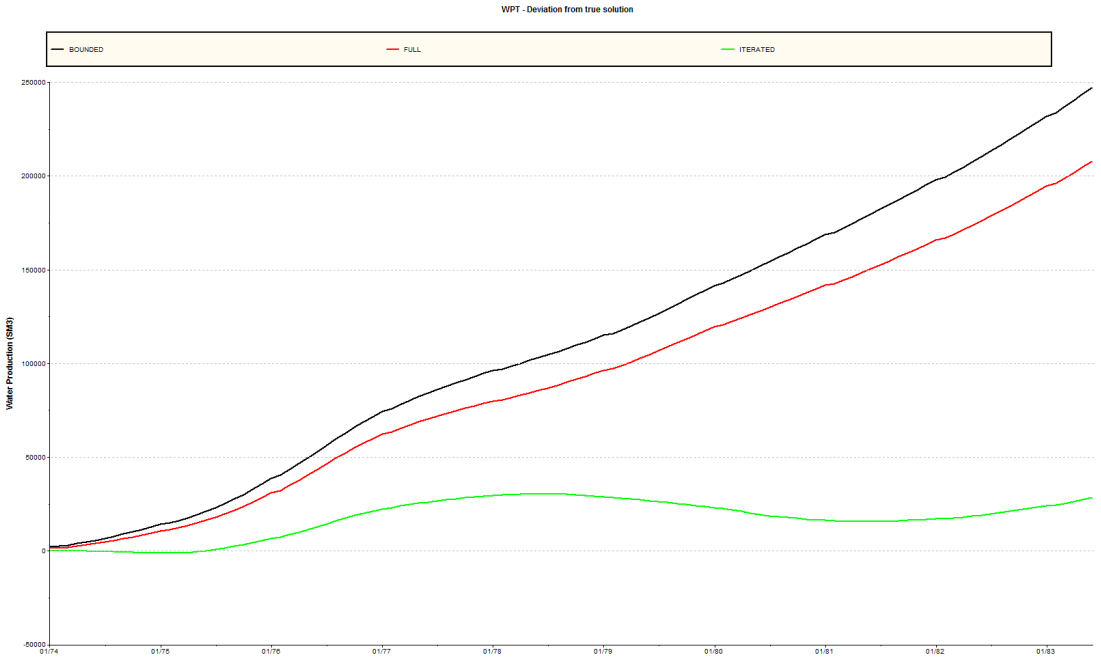

Figure 7.2.10: Deviation in water production from 1974.

The production forecasting ability of the mean cases seen in the conjunction with the parameter fit reveals an interesting result. The Constrained case has the most accurate parameter fit , but at the same time the worst forecasting ability out of the three. Similarly the Iterated case has the best forecast with the worst parameter fit. Due to the non-linear nature of a reservoir simulator, a single answer to explain these results is hard. The Iterated is subject to the same additional information as the Constrained case, while the Full case does not have any additional information and served as a basecase. This means more prior information in the constrained cases, but also a parameter probability distribution that is less Gaussian. For the Iterated case, the uncertainties in the dynamic parameters follow the static parameter, so with the iterative workflow, a theoretical more correct update of the dynamic parameters is performed though the update of dynamic parameters based on the constrained model parameters. This could help to explain why the Iterated case outperform the other two, since the Constrained case also had the same information given, but these have different updating workflows. It was surprising to see that the Full case outperformed the Constrained case in field performance. One reason for this might be the non-gaussian probabilty distributions involved in the update. As the Kalman Gain is calculating without any regards to the constraints involved, the update is applied and *then* corrected to stay within the boundaries.

PARAMETERIZATION MATCH. An attempt is made to capture the parameterization match compared to the true solution and is seen in the table below. The numbers presented are the mean of the squared deviation from the true solution in all active cells. The simple Constrained case outperforms the others when it comes to static parameter match. Surprisingly the Iterated case with the same boundaries as the Constrained case show a high deviation from the truth compared to the Full case without boundaries. Presented in Table 7.2.2 is the deviation based on layers.

As seen on the porosity values from all cases, the deviation from the true case is significantly higher in layer 1. This deviation could be partially explained from the completion data. None of the wells are completed in layer 1 or 2, thus no

direct measurements are made from these layers. This means that the correlation between measured data and grid cell values in layer 1 is hard to define. Layer 2 is subject to heavy constraints and the deviations cannot grow very large, resulting in the lowest deviation values across all cases.

Case	Porosity	Perm-X	Perm-Z
Constrained	0.0097	3.875	3.511
Full	0.0104	3.924	3.536
Iterated	0.0119	4.032	3.852

Table 7.2.1: Sum of squared deviation compared to true solution

Constrained Case			
Layer	Porosity	Perm X	Perm Z
1	0.0198	6.619	6.433
2	0.0022	2.413	2.075
3	0.0101	2.190	2.197
4	0.0059	4.914	3.570
5	0.0101	2.950	2.942
Full Case			
Layer	Porosity	Perm X	Perm Z
1	0.0222	6.686	6.400
2	0.0021	2.399	2.074
3	0.0099	2.342	2.254
4	0.0063	4.973	3.559
5	0.0104	2.940	3.075
Iterated Case			
Layer	Porosity	Perm X	Perm Z
1	0.0216	6.589	6.395
2	0.0029	2.727	2.454
3	0.0124	2.747	2.945
4	0.0068	4.984	3.685
5	0.0159	2.829	3.521

Table 7.2.2: Squared deviation for each case compared to the true solution

The deviation in all three cases follow the same trends, with the highest porosity deviation in the layers with highest porosity; layer 1, 3 and 5. It was expected to see a high deviation in layers 1, partially due to dynamics from grid cells that are

hard to observe and partially due to the large range of possible porosity values. Layers 3 and 5 had a significant lower deviation than layers 1, which was expected since the observability should be higher with, respectively, two and three out of six wells completed in these layer. The permeability deviation was higher in layer 5 than layer 3, both in x-direction and z-direction. Based on the well completion, the observability of layer 3 with two wells could be lower than layer 5 with three wells, when they are subject to similar geological deposition and simulation constraints. This is not the case neither in the Constrained or Full case, where the difference between them are insignificant. In the Iterated case both the porosity and some of the permeability values are higher than the two other cases, but the horizontal permeability in layer 1 and 5 is more accurate. We can also see that the permeability deviation in layer 3 and 5 is more consistent with the observability mentioned above.

As stated in the introduction of this problem, a conditioned porosity field could affect the initial reserves. Table 7.2.3 summarizes the initial hydrocarbon volumes. As seen, the three cases investigate is actually further away from the true case than their common initial model. Even though the hydrocarbon estimate becomes less accurate after the history matching, the flow performance of all three history matched model increase significantly when it comes to replicating the performance of the true case.

	Constrained	Full	Iterated	Initial	True
OIP[SM3]	1.856E+07	1.879E+07	1.845E+07	1.805E+07	1.737E+07
Percent of true	106.8 %	108.2 %	106.2 %	103.9 %	100 %
GIP[SM3]	1.719E+09	1.743E+09	1.703E+09	1.696E+09	1.650E+09
Percent of true	104.1 %	105.6 %	103.2 %	102.7 %	100 %

Table 7.2.3: Hydrocarbons in place at the beginning of production in 1967.

Restrictions in the Constrained case and the Iterated clearly affect the porosity estimates, creating an observable separation from the Full case. The difference between the two cases with boundaries also show reveal some information of their updating workflow. As stated earlier, in the Iterated the dynamic parameters are conditioned to the static for each timestep. The resulted in an overall smaller update of porosity than in the Constrained case.

Chapter 8

Problem 2: Effects of additional conditioning time

One of the goals for this thesis was to see how additional time spent on conditioning a reservoir model would affect the both the parameters estimation and the forecast ability. To start off, a short description of why this is important is given. Afterward, the simulation results are presented and discussed.

8.1 Background

The purpose of conditioning the reservoir is to add more information to the model. This increase of information could be used to reduce the uncertainty in future predication or add more information to a decision being made in the field development. For the case evaluated in this thesis, such a decision could be the drilling of the five new production wells. The time spent on history matching is assumed to be correlated with the amount of information provided. It is expected that an increase in time spent on history matching will provide additional information that in turn will improve the model in question.

Some moments during the life of a field producing is especially important with regards to history matching. As the well is producing oil, and gas, the flow perfor-

mance of the reservoir with regards to the permeability tensor is given continuous new information. This production will also give information about some of the dynamic conditions, like the gas to oil ratio. In this PUNQ-S3 case, it seems the water breakthrough is a critical phase giving much information. Given an understanding of the primary streamlines of the flow pattern in the reservoir, this water breakthrough could provide information about both static parameters like porosity and the dynamic water saturation. One of the cases evaluated was conditioned past the first water breakthrough in the true model, seen in Figure 8.2.7, thus should include some information that the other two cases do not have.

Multiple simulations is performed on the PUNQ S3 case using the same updating algorithm, initial condition and initial ensemble as presented in Problem 1. No upper boundaries are specified for the static parameters. Results are presented with case names Half, Semi and Full according to the time spent on history matching. They are each separated with 2 years of production following a common schedule. This means the information give in the time between the Half case and the Semi case is quantitatively equal to the information given in the time between the Semi case and the Full case.

As in Chapter 7, result from the ensemble predictions are presented first, followed by the parameterization match and the predictions from the mean of the respective ensembles.

8.2 Results

ENSEMBLE FORECASTING As in Problem 1, the conditioned models were not able to produce at the given oil rate for the entire production period. This difference is present both in the ensemble forecast and the forecast of the mean state vector. Looking at the oil rates in Figure 8.2.1, 8.2.2 and 8.2.3 we can see an convergence towards the true oil production as the models are conditioned to more time. Since the wells are given an oil rate limit, the bottom hole flowing pressure is adjusted to get the target rate. In the cases analyzed the drawdown was increased to a maximum, but the target flow rate was not reached. This is telling

us that there is a difference in the flow patterns between the true model, which was able to reach the target rate, and the conditioned cases. Without drilling injection well, this is also an estimate for the maximal possible production. This becomes even clearer in the analysis on the ensemble means.

From the same figures one can see that the gas rate is close to the truth is all three cases, but slightly overestimated for the Half and Semi case. For the Half case the ensemble spans the true solution from 1977 with one outlier. Towards the end of production time, a total of three ensemble member have a lower estimate than he truth while the main part of the ensemble still overestimates. In comparison, the Semi case has seven members below the true production, and the Full case has one third of the members below. In both the oil and the gas production, the ensemble approaches the cumulative true production towards the end of production. This means that the estimate production rates are too high in first period of production, and too low towards the end.

As seen in Chapter 7, the biggest deviation from the truth is found in the water produced. This is also where the biggest changes are seen between the three cases. It easy to observe from looking at a single case, that the ensemble span increases with time, much as expected. From the Half case we observe that the cumulative water production is continuously increasing compared to the true case. As more time is given for the conditioning of the models, the trend is not only a more accurate forecast in terms of being closer to the truth case, but also a narrower ensemble span. Having the ensemble close to the truth is in every setting beneficial, but a decrease in the span without having the truth enclosed is adverse. Both the Half and Semi case has some outliers creating a large span. In a forecast analysis, these could be considered outliers with reduced predictive capacity. A more uniform distribution would be considered more reliable, as seen in the Full case.

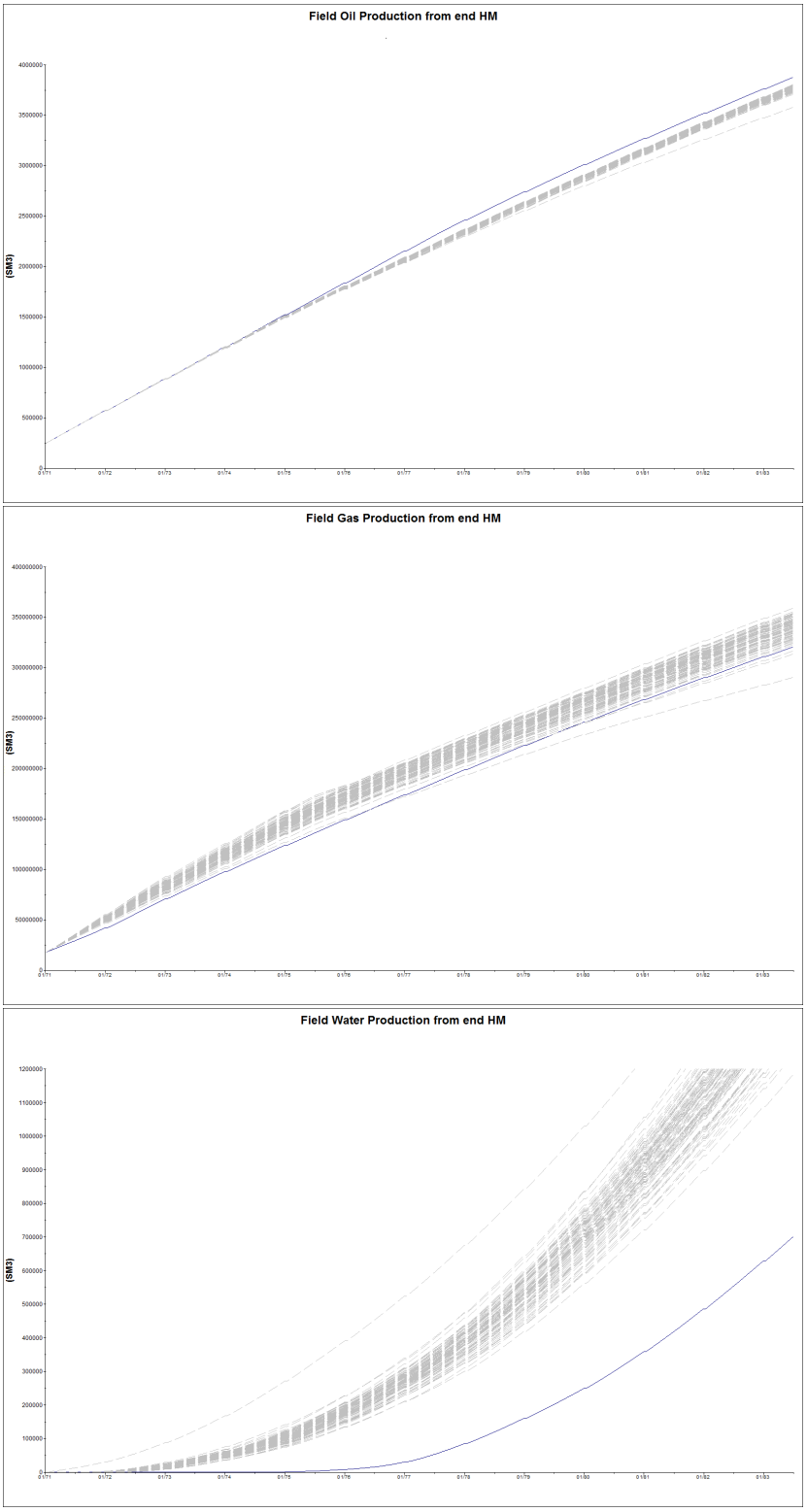

Figure 8.2.1: Production forecasts from end of history match for ensemble members in the Half case.

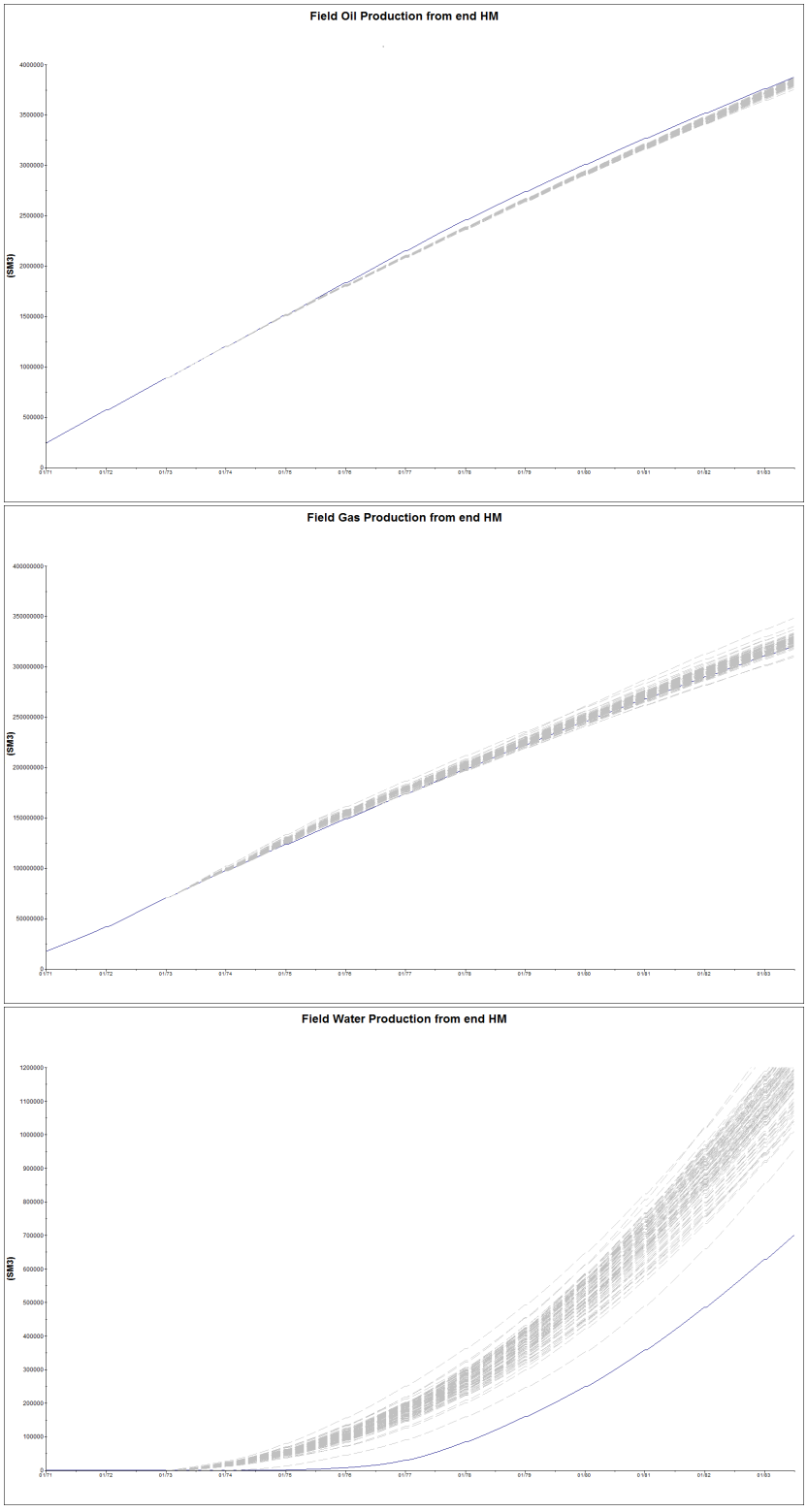

Figure 8.2.2: Production forecasts from end of history match for ensemble members in the Semi case.

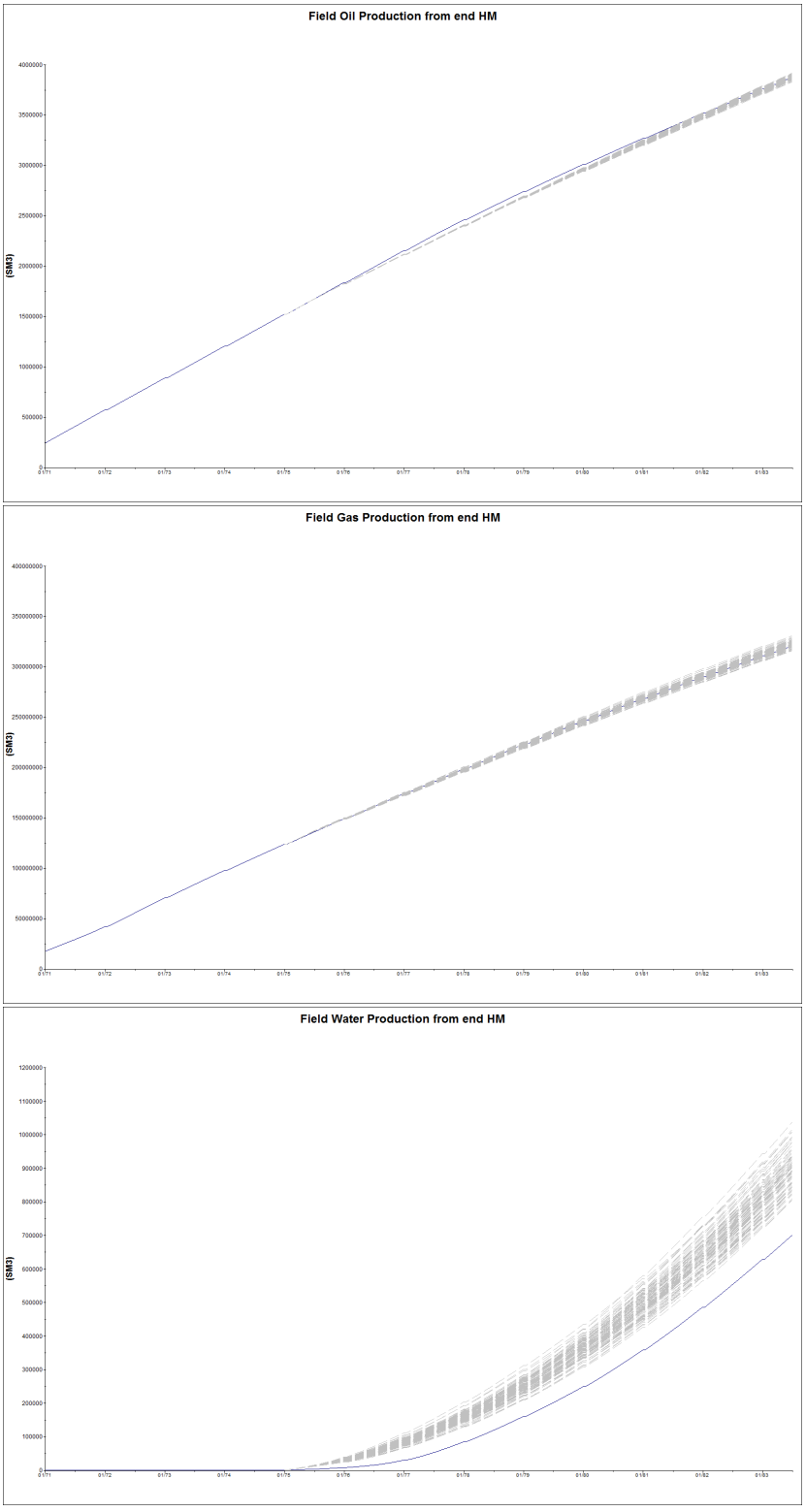

Figure 8.2.3: Production forecasts from end of history match for ensemble members in the Full case.

MEAN ENSEMBLE PROGNOSE. Much of the same characteristics is seen both in the ensemble forecast and the forecast produced by the mean of the ensemble, which was run from time zero with the same initial conditions as the true case. This approach using the mean parameters results in a dynamic view of the parameterization match through the production profiles. One of the interesting aspects of this comparison is to see how the magnitude of the additional conditioning time presents itself in the forecast.

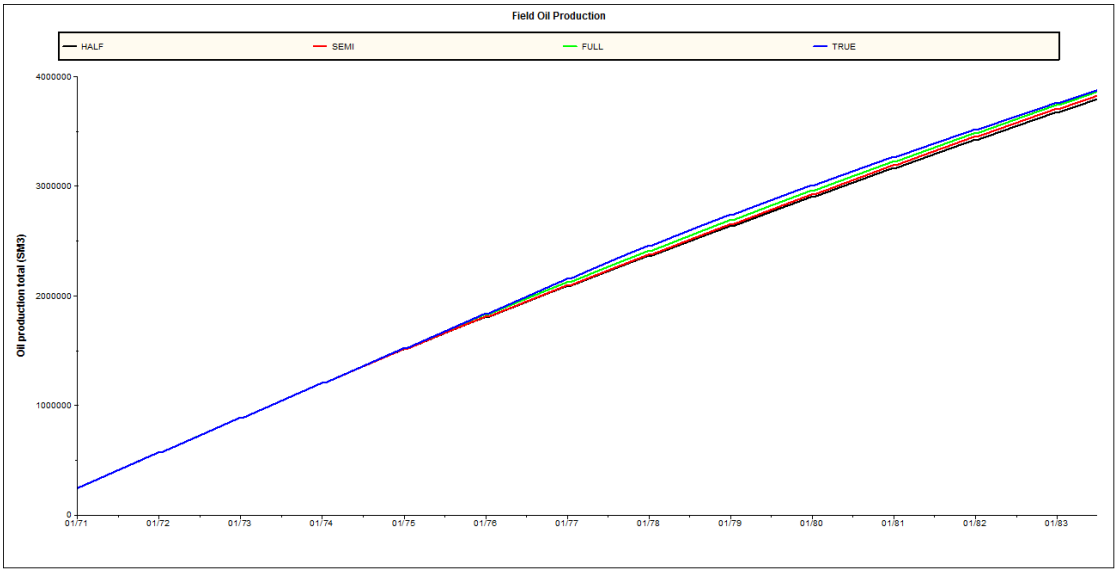

Figure 8.2.4: Cumulative oil production from 1967.

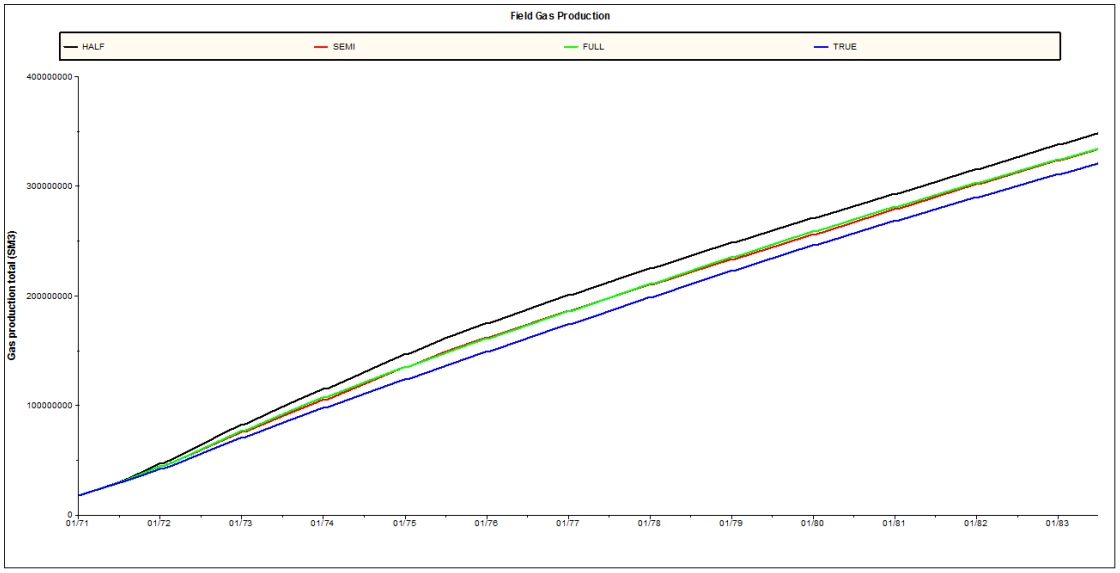

Figure 8.2.5: Cumulative gas production from 1967.

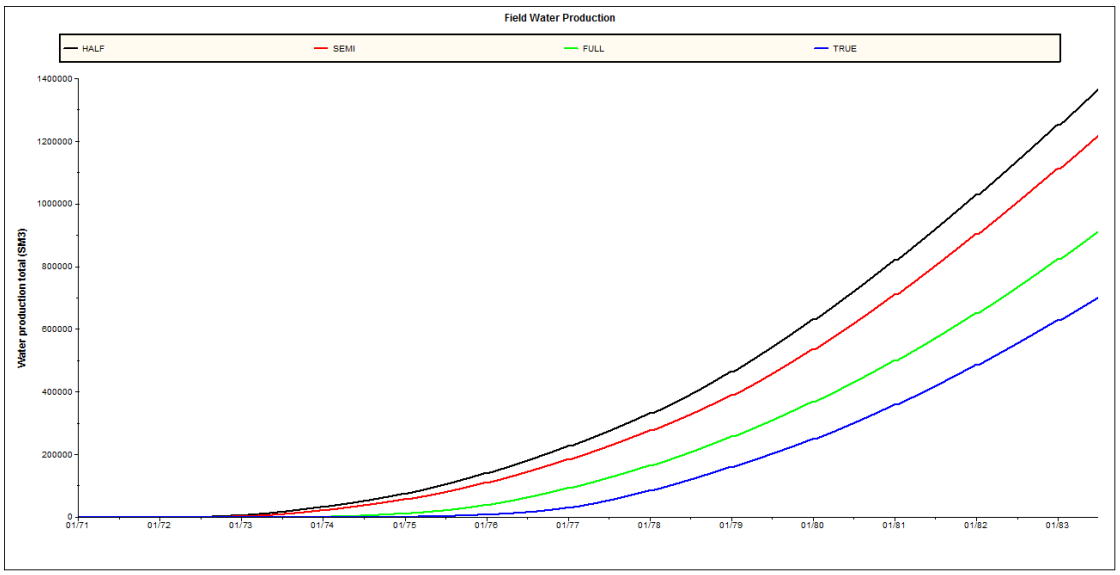

Figure 8.2.6: Cumulative water production from 1967.

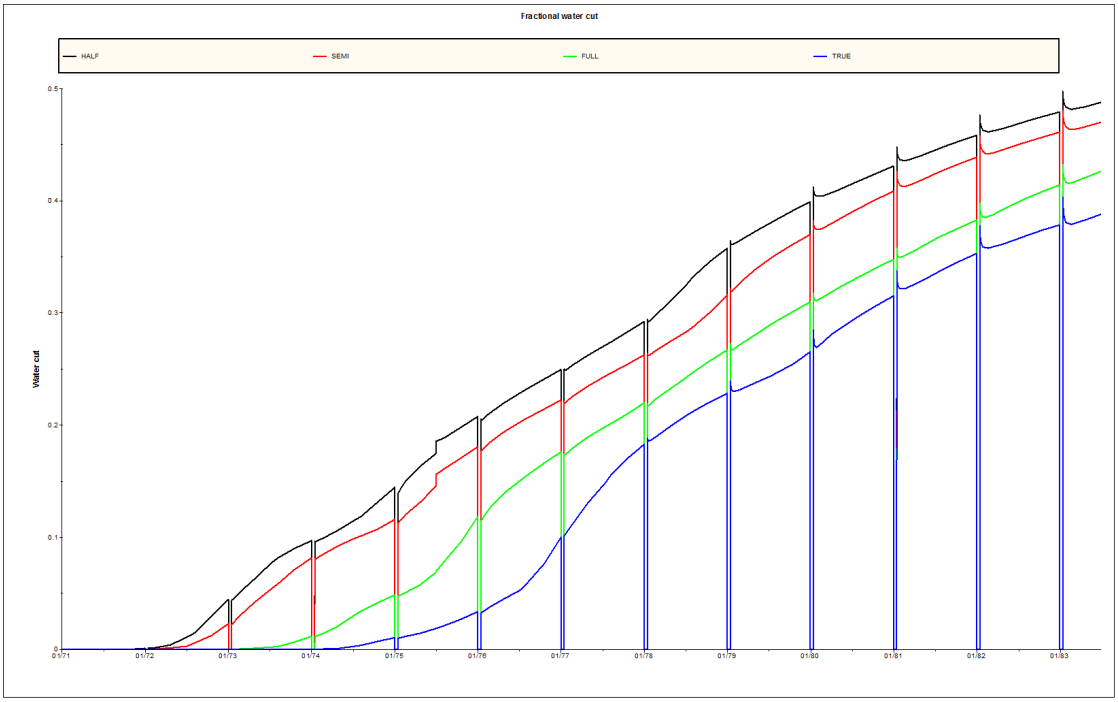

Figure 8.2.7: Watercut for the four cases starting from 1971.

Seen in Figure 8.2.5 the difference in gas production between the Semi and Full cases is negligible. These two cases place themself in between the Half case and the True case. From Figure 8.2.4 we can see that although the production from the different cases is quite similar, the Semi case is closer to the Half case than the Full case. All three cases fail to meet the production schedule, which the true case is able to. In addition, the Half and Semi case is influenced by the GOR limit, causing an oil rate cutback in Well 15. The difference in gas production is steady after 1975, which means the static parameters fail to reproduce the reservoir behavior even in the conditioned period. This was also seen in the analysis of constrained cases in Chapter 7. Since the gas production is dependent on oil production, a closer look on the relative gas production compared to oil is presented in Figure 8.2.8. As seen in the total gas production, the performance of the Semi case and the Full case is similar and both these cases differ significantly from the true case.

As stated earlier, the prediction of a water production could in many fields be a very important parameter. On production facilities offshore, significant water

production could give challenges both in the separation process and in the purification and removal of the water. An increase in water production accuracy corresponding to time spent on conditioning the model is seen in Figure 8.2.6. The early production phase evaluated in this thesis, is clearly important for the water forecast. The introduction of a water breakthrough in a well clearly gives important information in the EnKF algorithm. One of the reasons for this could be the bimodal probability distribution of the water saturations. Often saturation values behind the water front take very high values, in comparison with the very low value ahead of the front. Using the EnKF algorithm it can be hard to model this due to the Gaussian probability distribution assumption.

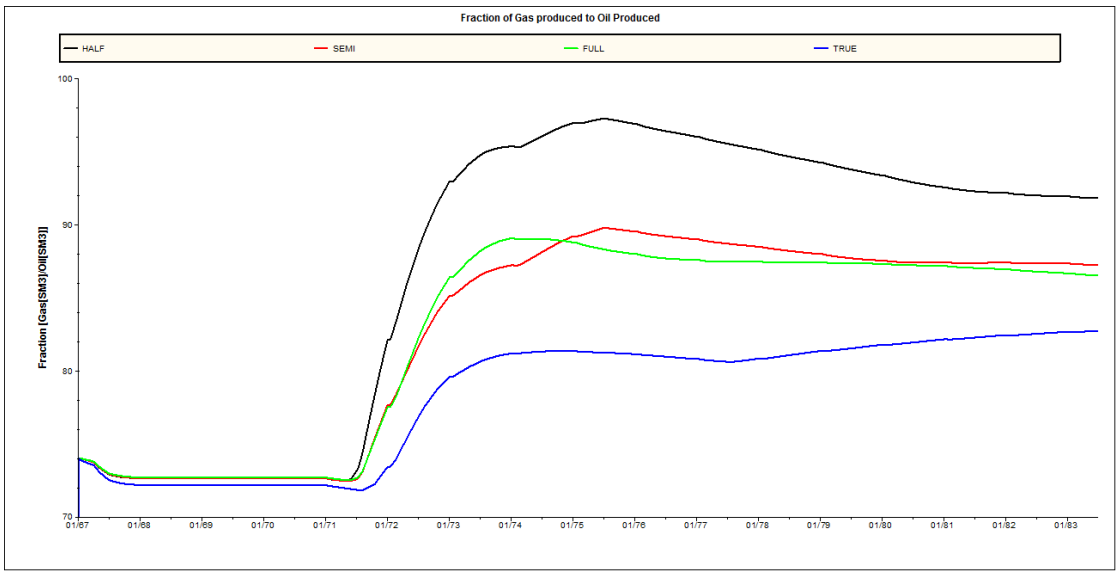

Figure 8.2.8: Fraction of gas produced to oil produced from 1967.

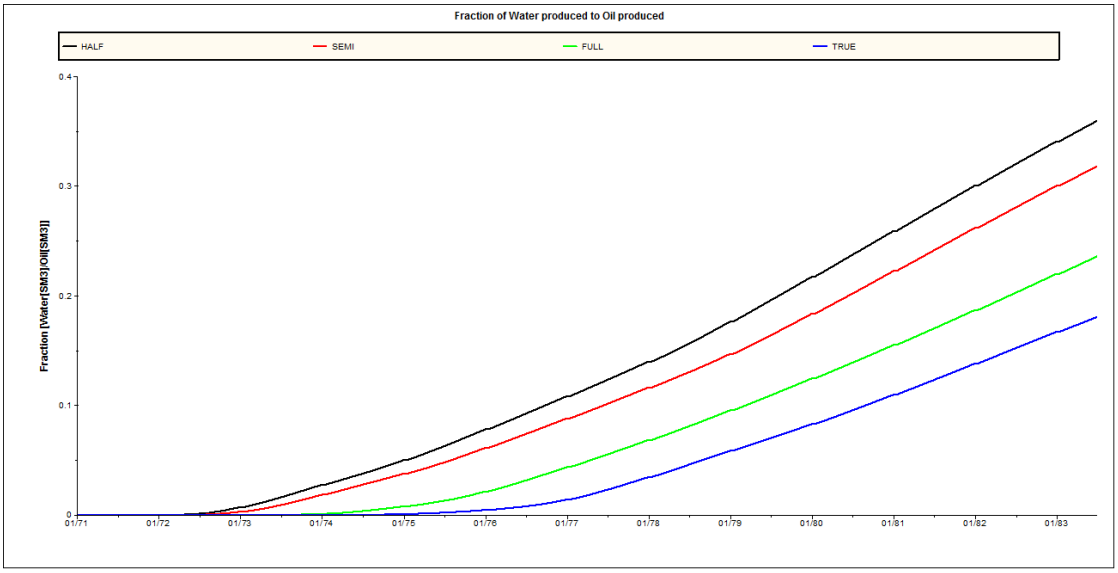

Figure 8.2.9: Fraction of water produced to oil produced from 1971.

Common through the analysis of production data from these three cases, the performance of the Full case is closest to the truth with the Half case being the worst predictor. The Semi case places itself between the two others, being close to the Full case in terms of oil and gas production and closer to the Half case in terms of water production.

PARAMETERIZATION MATCH Production profiles are very dependent on flow pattern, thus largely controlled by the permeability. As in the problem reviewed in Chapter 7, the porosity was also conditioned. Since the cumulative production of both oil and gas was quite close to the truth for all cases examined, having an estimate on remaining oil in the reservoir would influence the remaining value of reservoir. The initial oil in place is presented in Table 8.2.1. As seen, and further explored below, there is an increase in the deviation from truth case as more time is spent conditioning the model. This could lead to an overestimation of the field value, the affecting the NPV of a project.

	HALF	SEMI	FULL	Initial	True
OIP[SM3]	1.82E+07	1.84E+07	1.88E+07	1.81E+07	1.74E+07
Percent of true	104.8 %	105.8 %	108.2 %	103.9 %	100 %
GIP[SM3]	1.66E+09	1.69E+09	1.74E+09	1.70E+09	1.65E+09
Percent of true	100.9 %	102.7 %	105.6 %	102.7 %	100 %

Table 8.2.1: Hydrocarbons in place at the beginning of production in 1967.

In comparison with Probem 1, where that same trend of increase in initial oil in place also was seen on the gas in place, the table above show another evolution of the gas in place for Problem 2. Starting in time from the inital model there was first an increase of oil in place, but a decrease in gas in place. In the period between the Half case and the Full case the initial gas in place increases significantly. As seen in the statistics presented later, this could be attributed to an increase of porosity in layer 1 where the gas cap was present. Another explaination could be an update in gas saturation.

Tables D.1 through D.4 presents the deviation of the three cases with respect to each other and with respect to the true case. Both the sum of squared error and the sum of absolute error is presented. One of the reasons behind two sets of deviation parameters is to estimate whether outliers are the main reason for the squared deviation or if it is a uniform difference between the respective cases.

It was expected that the deviation from the truth would be highest in the Half case and reduced in the Semi and Full case. Seen from Table D.3 and D.4 the opposite is presented, with the Full case having the highest deviation from the truth both in term of absolute and squared deviation. By looking at the same deviation parameters distributed by layers, the same trend is seen in every layer.

As the information given in the time between the cases is quantitatively equal, the deviation between the cases would show if the two different time periods had different effects on the parameter update. Without any knowledge of the true production schedule it was expect that the deviation between the Half case and the Semi case would be lower than the deviation between the Half case and the Full case, showing a convergence towards a solution. As shown in Table D.2 there is a much bigger change in parameters during the time between the Semi and the Full

case than the previous phase. More information is given in the second phase, even though this results in a higher deviation from the truth. During this phase, two wells experienced water breakthrough which would help explain why the update is much higher in this phase.

A closer look at the porosity and permeability field are presented in Appendix C. As seen in the previous chapter, the field is not very smooth thus not capturing the continuity of the geological facies.

Chapter 9

Discussion

For the evaluated problems with a simple, but clearly multifacies geology, we can observe from the porosity and permeability plots that the EnKF algorithm is not able to distinctively separate the facies. No information about the geology was explicitly defined in the updating algorithm, and it is challenging to incorporate such information in the code used since the updating step is very dependent on the ensemble covariance matrix. In the EnKF, this matrix is derived from the ensemble thus not a good geostatistical parameter restrictor. An approach defining upper and lower boundaries for every cell in the reservoir could be performed, but this is time consuming and requires a very detailed knowledge of the reservoir. Other options can be history matching by facies, where an attempt is done to update the facies distribution, where every facies has parameter values associated, rather than a direct parameter update. A big loop model, where the updates are applied to a geological model before returned to the reservoir simulator could also be an option. This requires a good communication platform between the simulator and the geological modeling software.

One of the most interesting observation from both Chapter 7 and 8 was the accuracy of the parameter match versus the forecasting ability. Especially given that the cases evolved from the same ensemble, it was surprising to see the case with worse parameter fit having the more accurate forecast. Through the analysis performed in Problem 2 it was shown that the deviation from the truth case increases

with time. Since the conditioned model and the true model are run on the same simulator using the same spatial resolution, model errors in the simulation are not very likely to influence this result. Some plausible explanations arise:

- the initial model does not sufficiently span the true solution
- the reservoir response to the initial ensemble it not observable
- the main contributor for the deviation in static parameters has, in practice, a too small influence on the production data

Based on the data presented, it seems that the models converge towards a solution other than the truth, but with similar production data. This leads us back to the inverse theory, with multiple models giving the same response. Locating the reason for the deterioration of the parameter match is left for further work.

For Problem 2 where the evaluated cases were different time steps of the same updated model, the change in parameter values was easily evaluated. An increase in the change of the parameters where seen in the last two years, which was the time of the water breakthrough. Based on this we can state that a change in produced fluids can bring significant amounts of information to a reservoir conditioning process. In the Full case, we can see a significant increase in forecast accuracy compared to the Semi case resulting from this information.

In Problem 1 the ensemble covariance matrix, essential in the update of ensemble members, was challenged in a constrained setting. As the updated step of the EnKF is linear, constraining the model parameter after the update is applied will lead to a too large update in state variables. Given a non-linear system, or a non-Gaussian probability distribution, calculating a change in state variables equivalent to the constrained change in model parameters is difficult.

The evolution of dynamic parameters is, in addition to being non-linear, uncertain due to their correlation with static parameters. In a reservoir simulation setting the dynamic conditions could be correctly represented by the initial condition, but by history matching of the static parameters the dynamic parameters will no longer be correct for the new static parameters. One of the advantages of the Kalman Filters is the update of both model and state parameters, but the EnKF will in non-linear cases have trouble updating the dynamic parameters to

the newly updated model parameters, even without constraints. By rerunning the model from time zero with the same initial condition, as done in the iterative EnKF, the dynamic parameters are conditioned to the new static parameters. This workflow eliminates the sequential property of the EnKF, thus increasing the computer workload significantly. A criterion to invoke the iterative module could be incorporated in an EnKF algorithm to avoid unnecessary reruns from time zero.

The effect of information have been examined in both problems reviewed. In Problem 1, most apriori information was present in the two constrained cases. The Iterated case outperforms both the other two models, but with the least accurate parameter fit. The static parameter match in the Iterated case should be examined closer in further work, to see whether this case was able to capture more of the geological distribution. When comparing the Full case to the Constrained case, the latter have a smaller discrepancy in the parameter match, but are significantly outperformed in the forecast analysis. While both cases seem to converge towards a wrong model parameter solution, the flow performance of the Full case is closest to the truth out of these two. Although the Iterated case gave the most accurate forecast, it was given most information and most computational time. For Problem 2 the same tendencies in forecast accuracy versus information given are shown. As expected the Full case outperformed the other two, but the parameter estimation was a surprise. The value of this increase in forecast accuracy could be used in a further analysis to estimate the value of additional time used on conditioning the model.

Chapter 10

Conclusion

The Ensemble Kalman Filter have been derived in this thesis and applied to a synthetic reservoir model. The history matching effects of increased information have been examined in two different setting on the same reservoir using the EnKF. Results show a successful history match with regards to flow performance, but the EnKF method failed to correctly update the model parameters. An increase in amount of information given was seen to correlate with increased flow performance, but also with a reduced parameter match.

The reviewed problems show that the updated models converge towards a solution with wrong model parameters. The system in question is unobservable with multiple solutions creating correct flow performance, but the underlying geological description was not captured in the update.

The quality of information is seen to be high during water breakthrough for the flow performance of the field in this thesis. In essence, changes in production style create more information than monotonous steady production of a single fluid.

Based on the results presented, an economical calculation of the consequences of such an history match would be beneficial to see the value of the history match.

Further Work

For further work, finding the reason for the wrong convergence of the conditioned models is important. As a correct convergence will make the simulations more meaningfull, it might also change the conclusions of this paper.

Using EnKF in a big loop model communicating with a geomodel is also something that should be looked into. Incorporating the EnKF in the geomodel could, in addition to a better constraint on apriori knowledge, facilitate other sources of information, like 4D geophysical surveys.

Putting all the information gathering and history matching in a concrete setting where a decision about further field development needs to be taken, would be interesting in order to see how a history matched model might steer a decision. This will also give a value of the history match.

Nomenclature

Abbreviations

BHP	=	Bottom Hole Pressure
EKF	=	Extended Kalman Filter
EnKF	=	Ensemble Kalman Filter
EnRML	=	Ensemble Randomized Maximum Likelihood
GIP	=	Gas In Place
GOR	=	Gas to Oil Ratio
KF	=	Kalman Filter
MCMC	=	Markov-Chain Monte-Carlo
OIP	=	Oil In Place
pdf	=	Probability Density Funciton
VOI	=	Value of Information
WCT	=	Water Cut

Symbols

A	=	forward modeling function
$\boldsymbol{C}^a_{\psi\psi}$	=	analysed covariance matrix of ensemble
$\boldsymbol{C}^f_{\psi\psi}$	=	forecasted covariance matrix of ensemble
$\boldsymbol{C}_{\epsilon\epsilon}$	=	covariance matrix of measurement error
$\boldsymbol{C}_{qq}$	=	covariance matrix for model errors
d_{obs}	=	observed measurement data
f	=	dynamic condition function
g	=	reservoir response function
J	=	objective function
K	=	Kalman Gain
L_e	=	square root matrix of ensemble covariance
m	=	model parameters
M	=	linear operator
N_y	=	length of state vector
N_e	=	number of ensemble members
N_d	=	number of measurements
$\boldsymbol{\psi}$	=	state vector

Bibliography

[1] M. Mezghani and F. Roggero. Combining gradual deformation and upscaling techniques for direct conditioning of fine scale reservoir models to dynamic data. 2001.

[2] DJ Schiozer, EL Ligero, C. Maschio, and FVA Risso. Risk assessment of petroleum fields—use of numerical simulation and proxy models. *Petroleum Science and Technology*, 26(10-11):1247–1266, 2008.

[3] JN Carter, PJ Ballester, Z. Tavassoli, and PR King. Our calibrated model has poor predictive value: An example from the petroleum industry. *Reliability Engineering & System Safety*, 91(10):1373–1381, 2006.

[4] B. Lenoach and G. Bowen. Optimization and scaling of reservoir models. 2005.

[5] JN Carter. Using bayesian statistics to capture the effects of modelling errors in inverse problems. *Mathematical geology*, 36(2):187–216, 2004.

[6] B. DeVolder, J. Glimm, JW Grove, Y. Kang, Y. Lee, K. Pao, DH Sharp, and K. Ye. Uncertainty quantification for multiscale simulations. *Journal of Fluids Engineering*, 124: 29, 2002.

[7] FJT Floris, MD Bush, M. Cuypers, F. Roggero, and A.R. Syversveen. Methods for quantifying the uncertainty of production forecasts: a comparative study. *Petroleum Geoscience*, 7(S):S87–S96, 2001.

[8] Per Arne Slotte, 2011. SPE Presentation in Trondheim, 25.Oct 2011.

[9] Z. Tavassoli, J. Carter, and P. King. Errors in history matching. *SPE Journal*, 9(3):352–361, 2004.

[10] CFM Bos et al. Production forecasting with uncertainty quantification. *Final report of EC project, NITG-TNO report NITG*, pages 99–255, 2000.

[11] Schlumberger Oilfield Glossay, 2011. URL `http://www.glossary.oilfield.slb.com/Display.cfm?Term=historymatching`.

[12] R. Rwechungura, M. Dadashpour, and J. Kleppe. Advanced history matching techniques reviewed. 2011.

[13] Albert Tarantola. *Inverse Problem Theory: and Methods for Model Parameter Estimation*. SIAM, 2005.

[14] G. Evensen. *Data assimilation: the ensemble Kalman filter*. Springer, 2009.

[15] D.S. Oliver and Y. Chen. Recent progress on reservoir history matching: a review. *Computational Geosciences*, pages 1–37, 2010.

[16] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(Series D):35–45, 1960.

[17] J. Prakash, S.C. Patwardhan, and S.L. Shah. Constrained nonlinear state estimation using ensemble kalman filters. *Industrial & Engineering Chemistry Research*, 49(5):2242–2253, 2010.

[18] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *JOURNAL OF GEOPHYSICAL RESEARCH-ALL SERIES-*, 99:10–10, 1994.

[19] G. Burgers, P.J. van Leeuwen, and G. Evensen. Analysis scheme in the ensemble kalman filter. 1998.

[20] G. Evensen. Sampling strategies and square root analysis schemes for the enkf. *Ocean Dynamics*, 54(6):539–560, 2004.

[21] M.K. Tippett, J.L. Anderson, C.H. Bishop, T.M. Hamill, and J.S. Whitaker. Ensemble square root filters*. 2010.

[22] D.S. Oliver, A.C. Reynolds, and N. Liu. *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge Univ Pr, 2008.

[23] E. Kreyzig. *Advanced Engineering Maths*. John Wiley, 1967.

[24] G. Evensen. The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367, 2003.

[25] D.S. Oliver and Y. Chen. Improved initial sampling for the ensemble kalman filter. *Computational Geosciences*, 13(1):13–27, 2009.

[26] A. Seiler, G. Evensen, J.A. Skjervheim, J. Hove, and J.G. Vabø. Advanced reservoir management workflow using an enkf based assisted history matching method. 2009.

[27] Y. Gu and D. Oliver. An iterative ensemble kalman filter for multiphase fluid flow data assimilation. *SPE Journal*, 12(4):438–446, 2007.

[28] S. Aanonsen, G. Naevdal, D. Oliver, A. Reynolds, and B. Valles. The ensemble kalman filter in reservoir engineering–a review. *SPE Journal*, 14(3):393–412, 2009.

[29] A. Bianco, A. Cominelli, L. Dovera, G. Naevdal, and B. Valles. History matching and production forecast uncertainty by means of the ensemble kalman filter: a real field application. 2007.

[30] V.E. Haugen, G. Naevdal, L.J. Natvik, G. Evensen, A. Berg, and K. Flornes. History matching using the ensemble kalman filter on a north sea field case. *SPE Journal*, 13(4): 382–391, 2008.

[31] G. Evensen, J. Hove, H. Meisingset, E. Reiso, K.S. Seim, and Ø. Espelid. Using the enkf for assisted history matching of a north sea reservoir model. 2007.

[32] J.A. Skjervheim, G. Evensen, S. Aanonsen, B.O. Ruud, and T.A. Johansen. Incorporating 4d seismic data in reservoir simulation models using ensemble kalman filter. 2005.

[33] Y. Chen, D. Oliver, and D. Zhang. Efficient ensemble-based closed-loop production optimization. *SPE Journal*, 14(4):634–645, 2009.

[34] N. Darman, T. Bui, and J. Moreno. Eliminating subjectivity in history match evaluation through systematic data mining and application of history match quality index. In *SPE Paper 132072 ; SPE International Oil and Gas Conference and Exhibition in China*, 2010.

[35] E. Ligero, C. Maschio, and D. Schiozer. Quantifying the impact of grid size, upscaling, and streamline simulation in the risk analysis applied to petroleum field development. 2003.

[36] D. Uldrich, S. Matar, and H. Miller. Using statistics to evaluate a history match. *paper SPE*, 75223.

[37] C. Branco, A. Capeleiro Pinto, P.M. Tinoco, P. Vieira, A. Sayd, R.L. Santos, and F. Prais. The role of the value of information and long horizontal wells in the appraisal and development studies of a brazilian offshore heavy oil reservoir. 2005.

[38] M. Zafari and A. Reynolds. Assessing the uncertainty in reservoir description and performance predictions with the ensemble kalman filter. 2005.

[39] D. Steagall, J. Gomes, R. De Oliveira, N. Ribeiro, R. Queiroz, M. Carvalho, and C.C. Souza. How to estimate the value of the information (voi) of a 4d seismic survey in one offshore giant field. 2005.

[40] R. Kumar and S. Hara. Appraisal voi when multiple prospects are dependent-with the aid of recursion for generating numerous economic scenarios. 2005.

[41] G. Gao, M. Zafari, and A.C. Reynolds. Quantifying uncertainty for the punq-s3 problem in a bayesian setting with rml and enkf. 2005.

[42] N. Liu and D.S. Oliver. Critical evaluation of the ensemble kalman filter on history matching of geologic facies. 2005.

[43] M. Cuypers, O. Dubrule, P. Lamy, and R. Bissell. Optimal choice of inversion parameters for history matching with the pilot point method. In *Proceedings of the ECMOR VI conference, Peebles*, pages 8–11, 1998.

[44] W.E. Walker, P. Harremoes, J. Rotmans, JP Van der Sluijs, MBA Van Asselt, P. Janssen, and M.P.K. Von Krauss. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1):5–17, 2003.

[45] M. Christie, S. Subbey, M. Sambridge, and M. Thiele. Quantifying prediction uncertainty in reservoir modelling using streamline simulation. 2002.

[46] R. Bratvold, J.E. Bickel, and H.P. Lohne. Value of information in the oil and gas industry: past, present, and future. *SPE Reservoir Evaluation & Engineering*, 12(4):630–638, 2009.

[47] R. Baker. Streamline technology: reservoir history matching and forecasting= its success, limitations, and future. *Journal of Canadian Petroleum Technology*, 40(4), 2001.

[48] D.S. Oliver, L.B. Cunha, and A.C. Reynolds. Markov chain monte carlo methods for conditioning a permeability field to pressure data. *Mathematical Geology*, 29(1):61–91, 1997.

[49] D.S. Oliver, N. He, and A.C. Reynolds. Conditioning permeability fields to pressure data. pages 1–11, 1996.

[50] Y. Gu and D.S. Oliver. The ensemble kalman filter for continuous updating of reservoir simulation models. *Journal of Energy Resources Technology*, 128:79, 2006.

[51] A.A. Grimstad, T. Mannseth, G. Nævdal, and H. Urkedal. Adaptive multiscale permeability estimation. *Computational Geosciences*, 7(1):1–25, 2003.

[52] F. Knight. *Risk, Uncertainty and Profit*. 1921.

[53] J.P. Jensen. Ensemble kalman filtering for state and parameter estimation on a reservoir model, 2007.

[54] Y. Chen. Ensemble-based closed-loop production optimization, 2009.

Appendix A

Problem 1: Additional information

Attached is porosity and permeability representations for layers 2-5 in Problem 1.

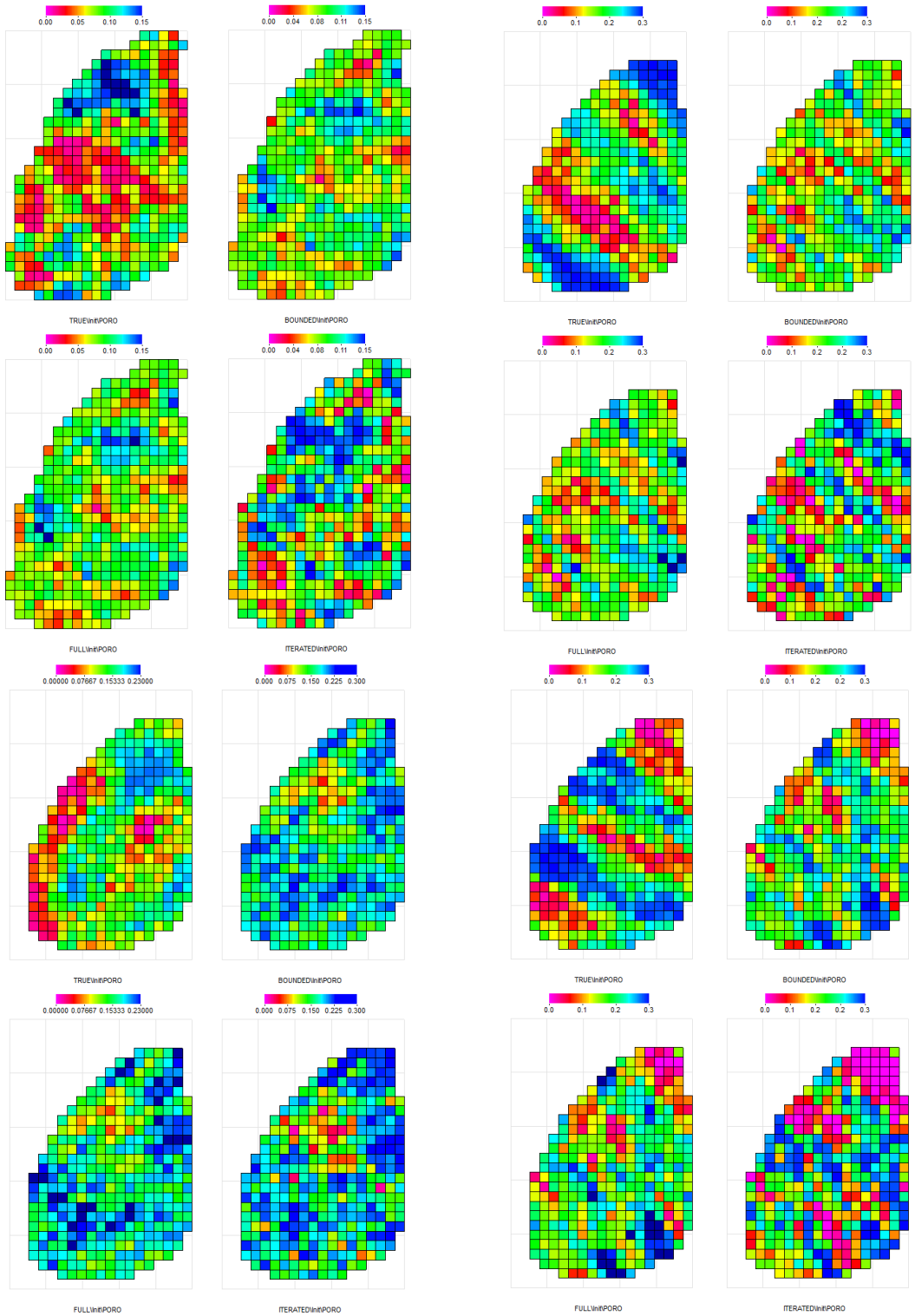

Figure A.1: Porosity for layers 2-5

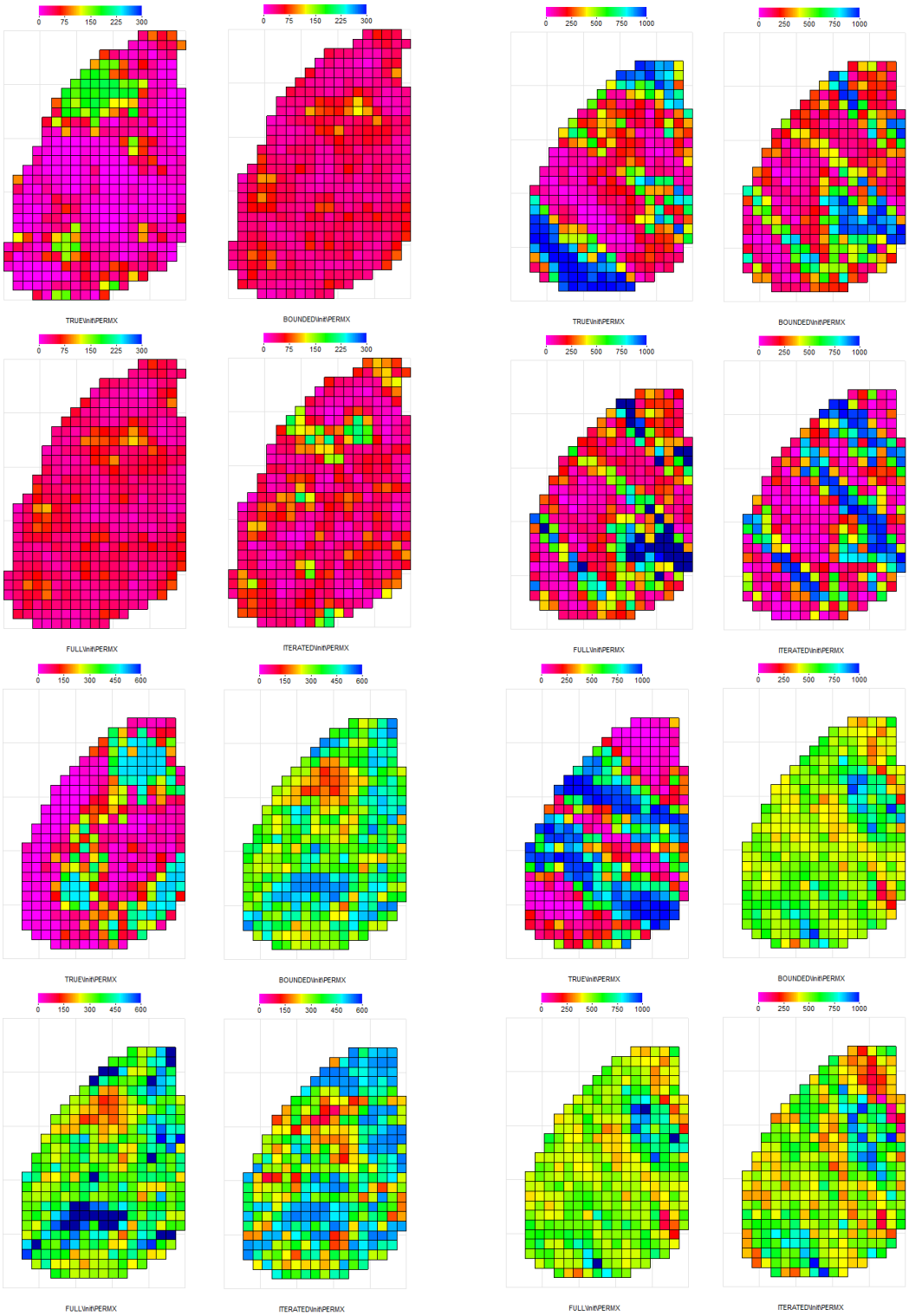

Figure A.2: Permeability in x-direction for layers 2-5

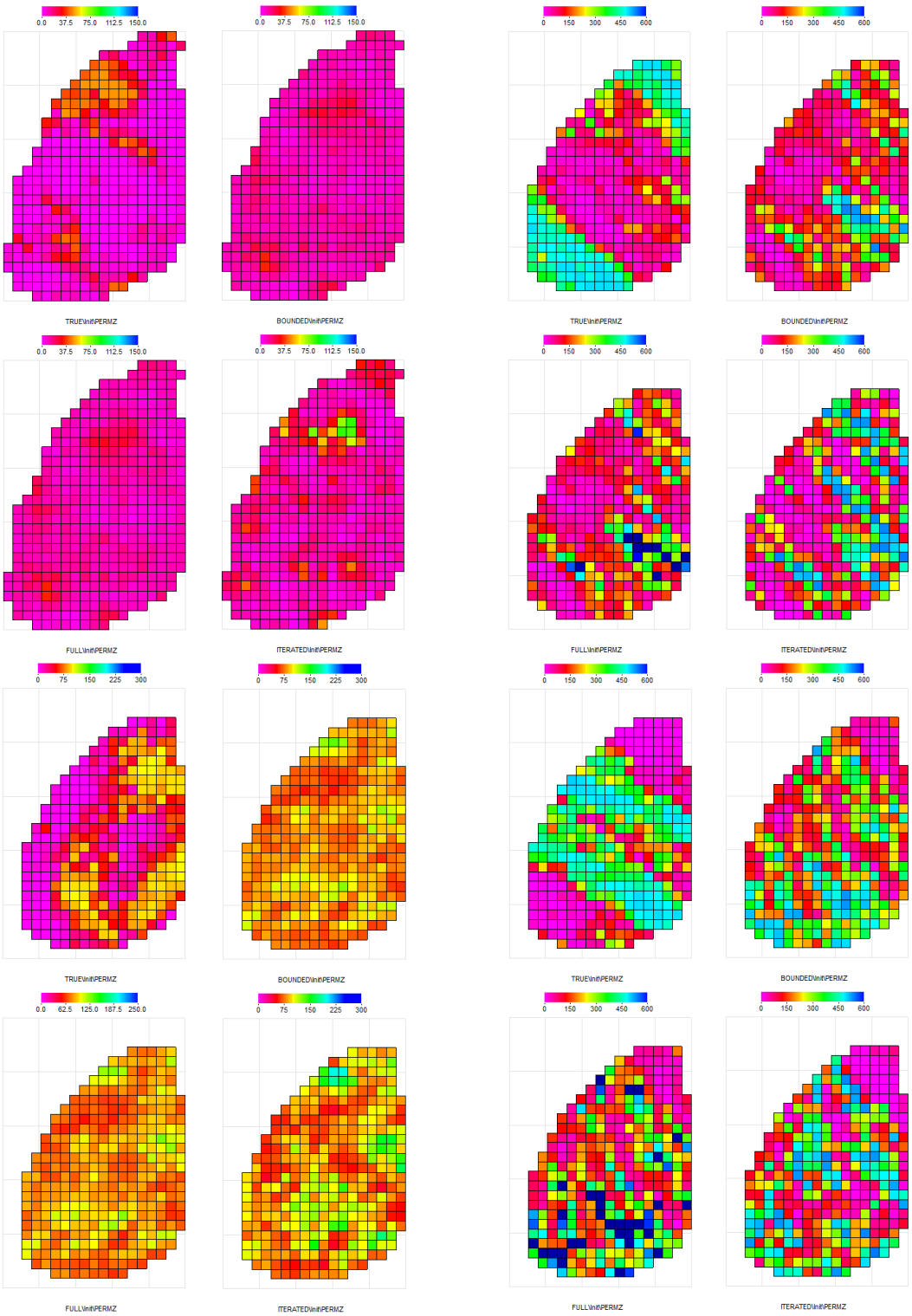

Figure A.3: Permeability in z-direction for layers 2-5

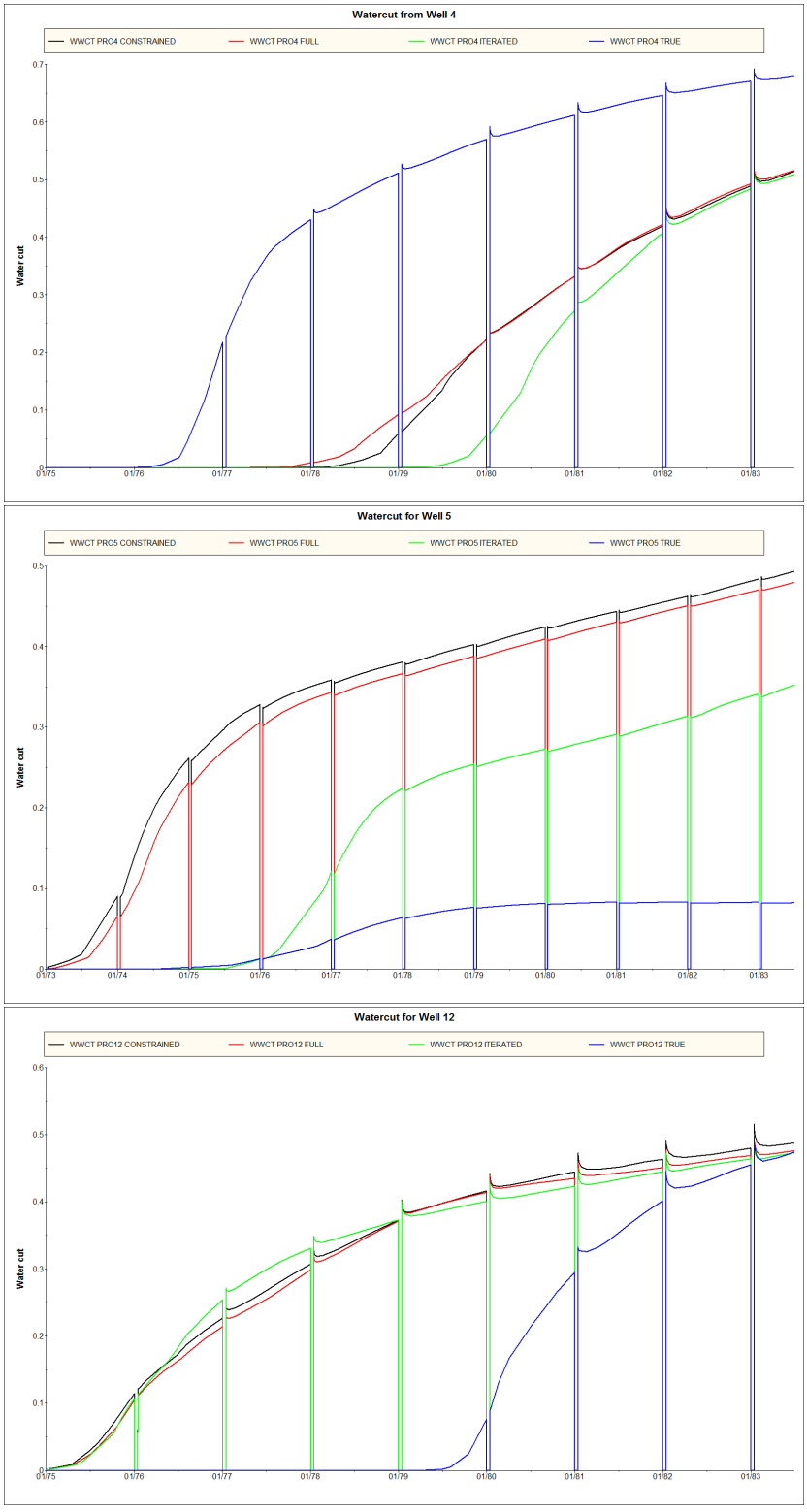

Figure A.4: Watercut for selected wells from Problem 1. These wells showed a clear separation from the true case.

Appendix B

Problem 1: Statistical parameters

Attached is an overview of the deviation compared to the true parameters for the cases in Problem 1.

Case	Porosity	Perm-X	Perm-Z
CONSTRAINED	0.07606	1.49109	1.41740
ITERATED	0.08489	1.53069	1.50941
FULL	0.07793	1.50695	1.42672

Table B.1: Sum of absolute deviation compared to true solution

Constrained case			
Layer	Porosity	Perm X	Perm Z
1	0.115	1.988	1.958
2	0.039	1.280	1.232
3	0.083	1.192	1.172
4	0.063	1.715	1.343
5	0.080	1.217	1.302
Iterated case			
Layer	Porosity	Perm X	Perm Z
1	0.122	1.987	1.952
2	0.042	1.336	1.314
3	0.091	1.330	1.377
4	0.069	1.733	1.405
5	0.102	1.207	1.443
Full case			
Layer	Porosity	Perm X	Perm Z
1	0.121	2.001	1.951
2	0.038	1.278	1.234
3	0.083	1.247	1.187
4	0.065	1.735	1.341
5	0.082	1.214	1.346

Table B.2: Absolute deviation compared to true solution

Appendix C

Problem 2: Additional information

Attached in this Appendix is porosity and permeability plots for all layers in Problem 2. The first figure show all plots for layer 1, while the remaining is organized by parameter, showing the porosity for layers 2 to 5 first.

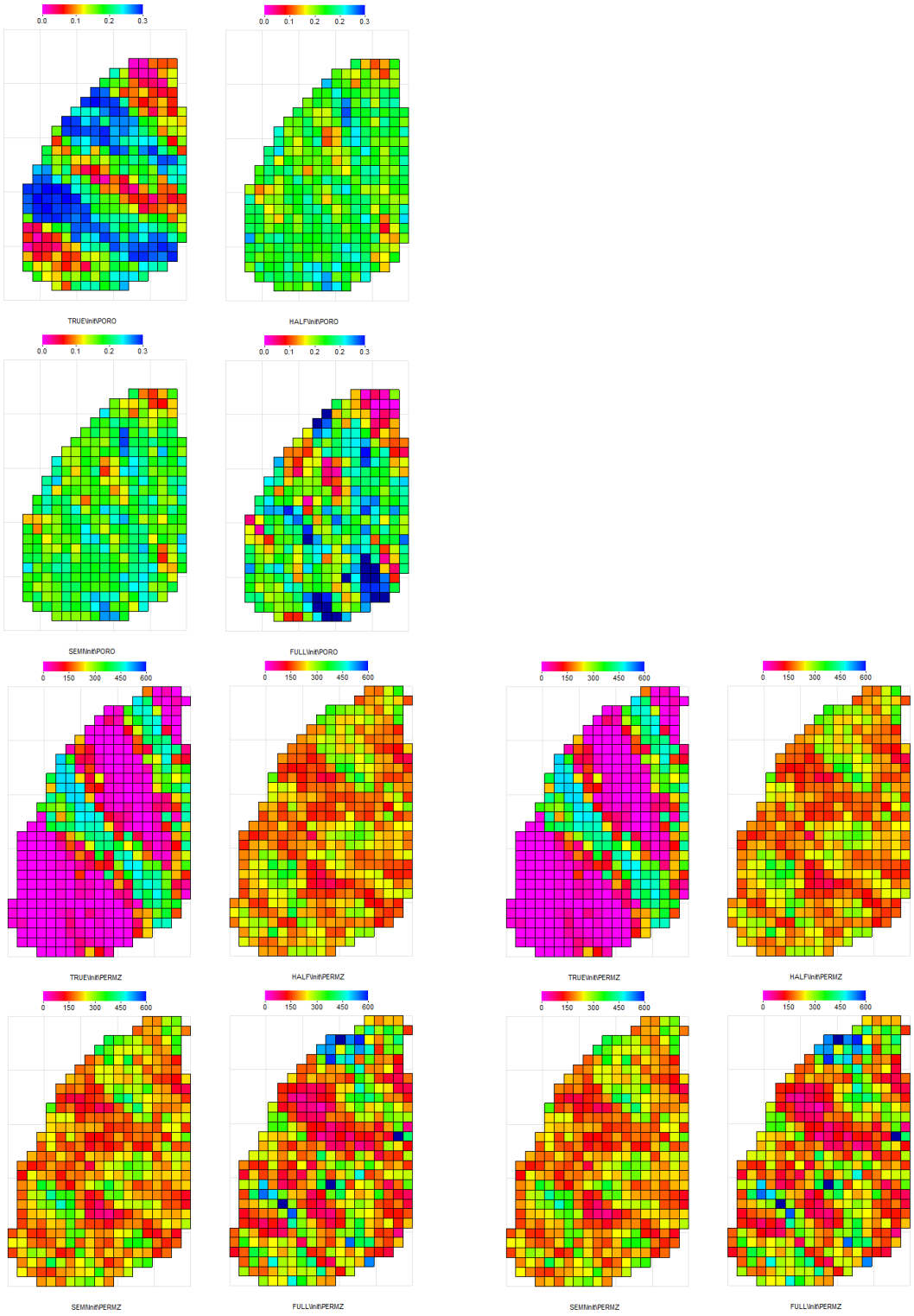

Figure C.1: Porosity and permeabilities for layer 1

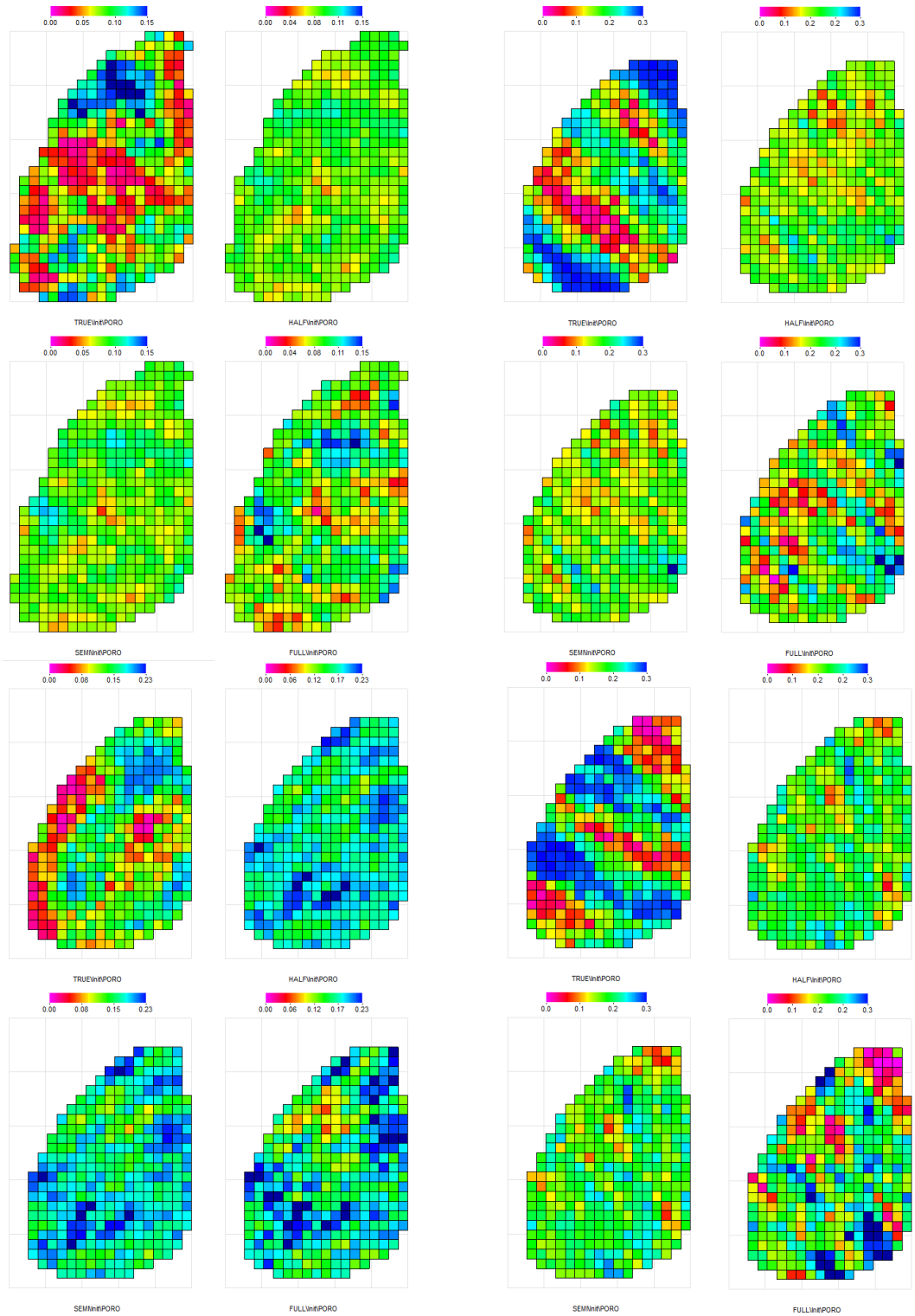

Figure C.2: Porosity for layers 2-5

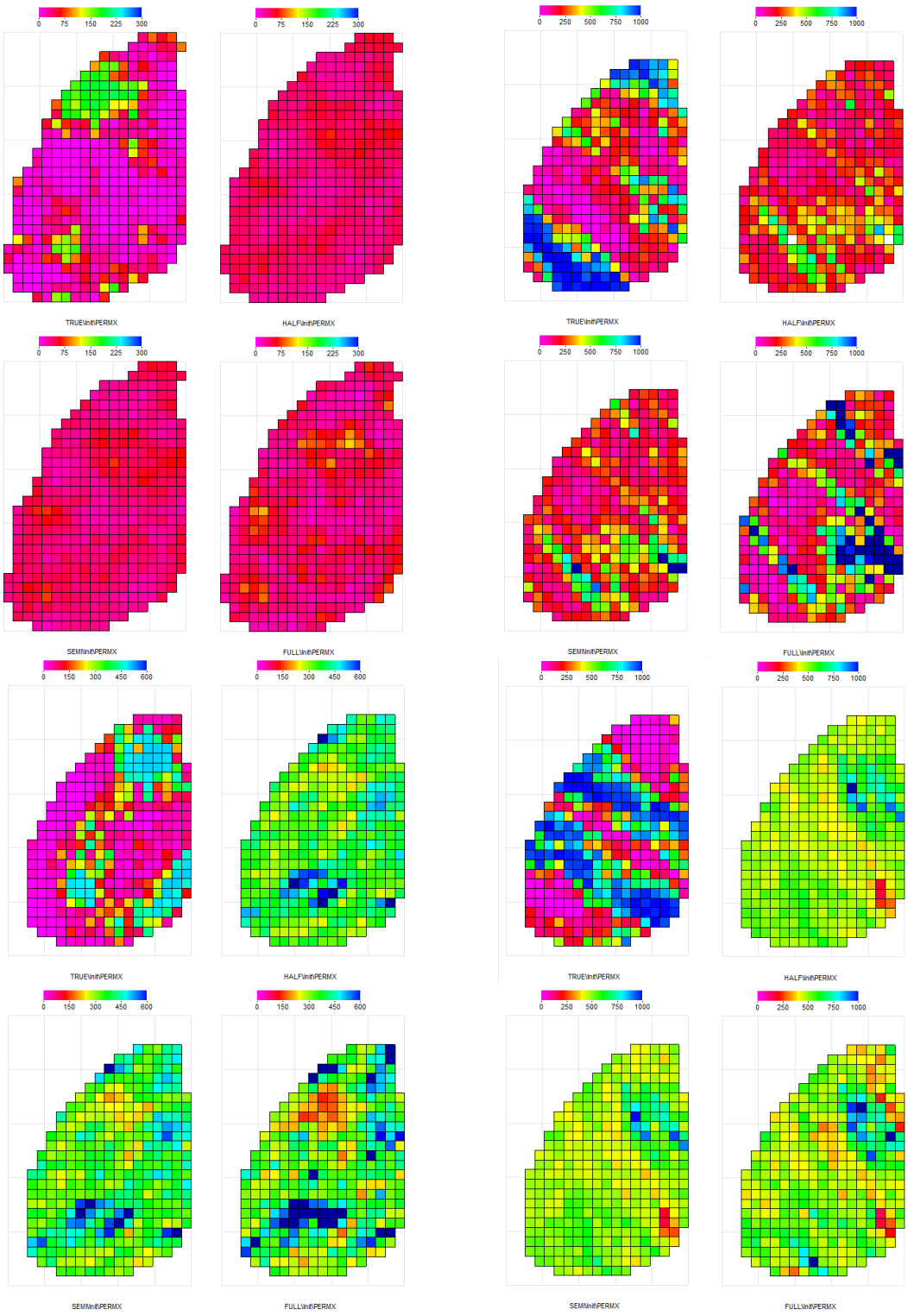

Figure C.3: Permeability in x-direction for layers 2-5

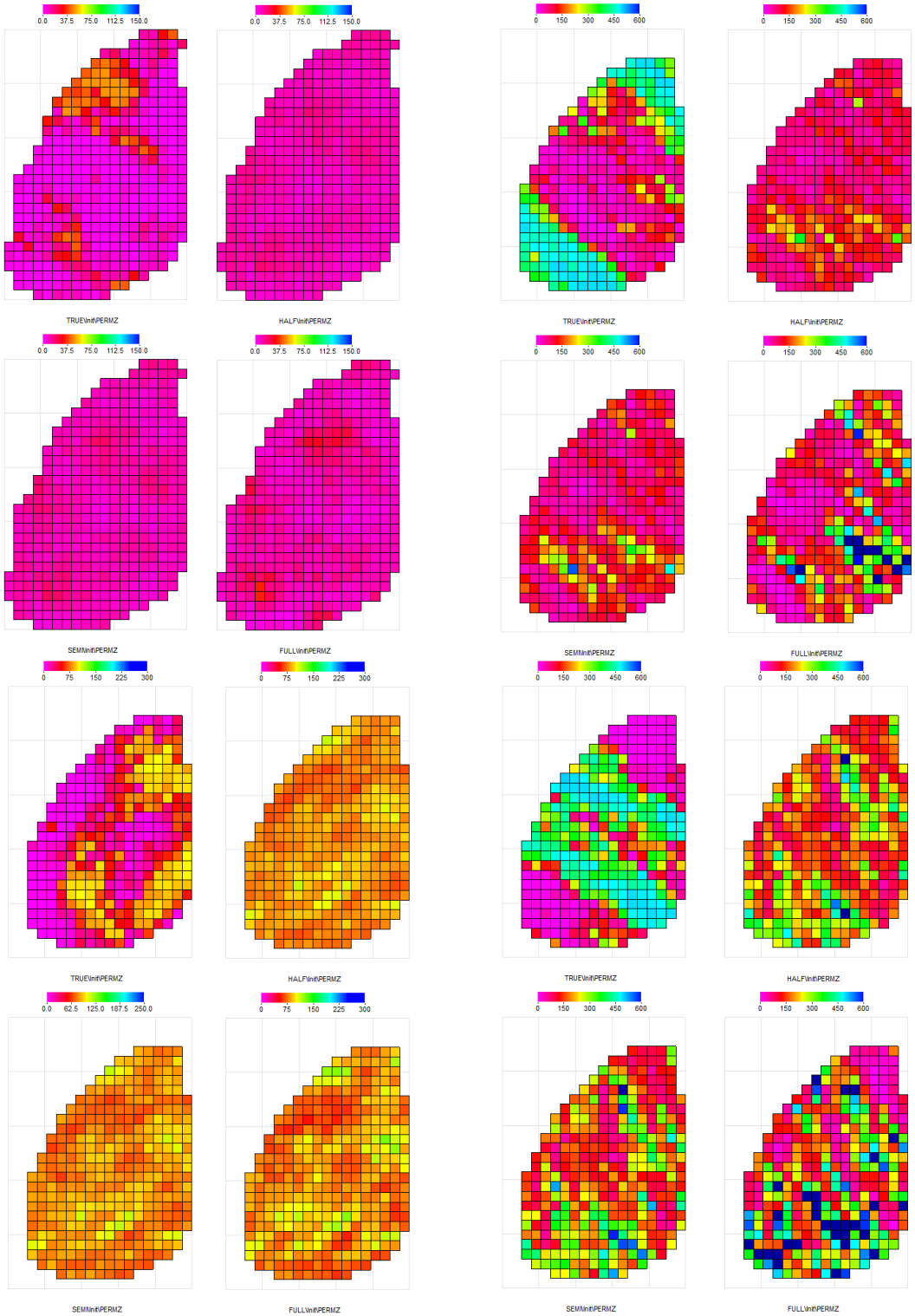

Figure C.4: Permeability in z-direction for layers 2-5

Appendix D

Problem 2: Statistical parameters

Attached is some of the calculations done on the deviation compared to the true parameters and the deviation internally between the cases for Problem 2.

Absolute deviation			
Case	Porosity	Perm-X	Perm-Z
HALF	0.07103	1.46377	1.42052
SEMI	0.07186	1.46718	1.41134
FULL	0.07793	1.50695	1.42672
Squared deviation			
Case	Porosity	Perm-X	Perm-Z
HALF	0.00817	3.79340	3.44147
SEMI	0.00843	3.80846	3.41500
FULL	0.01037	3.92394	3.53593

Table D.1: Sum of absolute deviation and squared deviation compared to the true case.

Absolute deviation			
Case	Porosity	Perm-X	Perm-Z
Half vs Semi	0.0108	0.0847	0.0997
Semi vs Full	0.0301	0.2463	0.3060
Half vs Full	0.0348	0.2844	0.3548
Squared deviation			
Case	Porosity	Perm-X	Perm-Z
Half vs Semi	0.00024	0.0152	0.0198
Semi vs Full	0.00163	0.1259	0.1938
Half vs Full	0.00220	0.1700	0.2602

Table D.2: Sum of absolute deviation and squared deviation relatively between the cases.

Half case			
Layer	Porosity	Perm X	Perm Z
1	0.106	2.011	1.971
2	0.034	1.249	1.228
3	0.075	0.993	1.082
4	0.063	1.755	1.346
5	0.077	1.235	1.394
Semi case			
Layer	Porosity	Perm X	Perm Z
1	0.108	2.004	1.965
2	0.035	1.253	1.224
3	0.076	1.031	1.092
4	0.064	1.749	1.346
5	0.076	1.227	1.346
Full case			
Layer	Porosity	Perm X	Perm Z
1	0.121	2.001	1.951
2	0.038	1.278	1.234
3	0.083	1.247	1.187
4	0.065	1.735	1.341
5	0.082	1.214	1.346

Table D.3: Absolute deviation organized by layers compared to the true case.

Half case			
Layer	Porosity	Perm X	Perm Z
1	0.0198	6.619	6.433
2	0.0022	2.413	2.075
3	0.0101	2.190	2.197
4	0.0059	4.914	3.570
5	0.0101	2.950	2.942
Semi case			
Layer	Porosity	Perm X	Perm Z
1	0.0222	6.686	6.400
2	0.0021	2.399	2.074
3	0.0099	2.342	2.254
4	0.0063	4.973	3.559
5	0.0104	2.940	3.075
Full case			
Layer	Porosity	Perm X	Perm Z
1	0.0216	6.589	6.395
2	0.0029	2.727	2.454
3	0.0124	2.747	2.945
4	0.0068	4.984	3.685
5	0.0159	2.829	3.521

Table D.4: Squared deviation organized by layers compared to the true case.