FACULTY OF SCIENCE AND TECHNOLOGY

MASTER THESIS

Study programme / specialisation:
Petroleum Engineering / Reservoir Engineering

The spring semester, 2022

Open

Author: **Siavash Heydari**

.....S. Heydari.....
(Signature author)

Supervisor(s): **Aojie Hong, Reidar B. Bratvold**

Thesis title: **Accounting for Model Error in Probabilistic History Matching to Improve Uncertainty Quantification**

Credits (ECTS):30

Keywords:
Model Error
Data Assimilation
Probabilistic History Matching
Ensemble Kalman Filter
Reservoir Modeling
Upscaling Error

Pages: 50

+ appendix: 12

Stavanger, 15th June 2022

Acknowledgments

This thesis concludes my master's study in the Petroleum Engineering – Reservoir Engineering program at the University of Stavanger. I would like to express my gratitude to those who have helped and supported me during this thesis project.

First and foremost, I would like to thank my supervisors, Professor Aojie Hong and Professor Reidar B. Bratvold, for their impeccable advice, vast knowledge, kindness, and encouragement during this project and my master's study. It has been a pleasure and a wonderful experience to work with the experts in computational engineering and decision analysis.

Sincere thanks to Peyman Kor for his valuable support and advice, which was very helpful during my master's degree.

I owe my deepest gratitude to my parents and family for providing me with the best education and endless support throughout my studies and life.

And last but not least, I'd like to thank myself and congratulate myself for my hard work to improve myself during these two years.

Abstract

Model Error in probabilistic history matching is an important topic to study, but calculating the model error is a challenge since the truth is uncertain. In this thesis, sources of model error will be discussed briefly; a novel approach has been proposed, first to define model error by using a high-quality model instead of the truth, then statistical parameters of model error will be calculated, and these parameters will be used to account for model error in EnKF.

Two cases have been tested with this approach. In the first case, decline curve analyses were used to model the production rate. Model error has been calculated and accounted for in updating the model with EnKF. The second case studied model error in a 2D reservoir for the upscaling process. The reservoir has been upscaled, and statistical parameters of model error were obtained to be used in updating the model with EnKF.

Results from these two examples showed the importance of the model error in data assimilation. In both cases, it has been proven that neglecting model error caused biases and overconfidence in the forecasted updated model. Additionally, the proposed approach could mitigate the biases and the overconfidence in the forecasted updated model.

The proposed method is in an early stage, and further study should be done to verify and improve it. Other sources of error should be examined. A proper machine learning algorithm could improve the quality of this method to account for model error.

Table of Contents

List of Figures

List of Tables

Nomenclature

Two Dimensional …….	2D
Kalman Filter……	KF
Stretched Exponential Method……..	SEM
Cumulative production……..	QT
Decline Curve Analyses……	DCA
Standard Deviation……	STD
Data Assimilation……..	DA
Monte Carlo Simulation……	MCS
History Matching……...	HM
Assisted History Matching……	AHM
Bottom-hole Pressure……	BHP
Equation Of State……..	EOS
Initial Production……...	Qi
Root Mean Square Error……	RMSE
Mean Absolute Percentage Error……...	MAPE
80 Presents Confidence Interval……	P80
Probability Density Function…….	PDF
Water Cut……...	WCT
Production Well BHP……	PBHP
Injection Well BHP……...	IBHP
High Quality……..	HQ
Low Quality……...	LQ
Probability Density Function…….	PDF
Uniform Probability Distribution……	$\mathcal{U}$
Normal Probability Distribution……..	$\mathcal{N}$

Chapter 1 Introduction

Bayes theorem is a method that allows us to update the prior knowledge (expressed as probability distribution) about an event with given extra information related to the event(Birnbaum, 1962). The direct application of the Bayes theorem is not computationally approachable due to the high dimensionality of prior and the observation in the oil and gas industry.

Data Assimilation (DA), based on the Bayes theorem, is a recursive procedure that updates the previous model with new information related to the event. In the 1960s, R.E. Kalman developed a statistical approach -called the Kalman Filter (KF), which introduces an equation for the time evolution of the error covariance matrix. In the KF, the foreword model is linear, and uncertainties are modeled using Gaussian distributions (Kalman, 1960).

Monte Carlo Simulation (MCS) is a broad group of computational methods which uses random sampling to calculate the numerical results, and it handles calculations with a probability distribution. The MSC uses a random sample from probability distributions of the uncertain parameters to obtain a numerical solution. These random samples represent the probability distribution, creating an ensemble of realizations (figure 1-1).

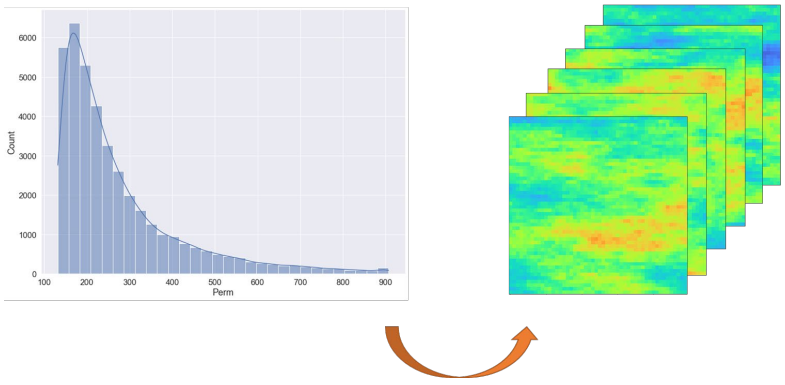


Figure 1-1 A distribution of Permeability and its representative realizations

Later, in 1994 Evensen modified KF to handle non-linearity by combining Monte Carlo simulation and KF. The modified KF was named the Ensemble Kalman Filter (EnKF). Ensemble-Based Data Assimilation is a family of statistical approaches to update and correlate the probability distribution of prior knowledge with given noisy observations from related phenomena. The ensemble-based method refers to random sampling from prior probability distribution. In DA, typical problems of interest are non-linear, non-Gaussian, and/or high-dimensional(Harlim, 2017).

EnKF is an implementation of the Bayesian theorem with the association of MCS in a recursive manner, where the probability of prior knowledge will be updated with additional new observations. KF and EnKF consider probability distribution and error in input variables and observation. Prior knowledge is all available information and understanding about a specific topic at the beginning of a process (Schmidt et al., 2015).

Observation error/noise could be defined as the difference between the actual and observed state of nature. For example, the actual state of nature, the oil rate in a well is 100 bbl per day, and the observed state nature by the flow meter device is 90 bbl per day.

In studying any natural phenomenon, uncertainty should be considered for high-quality decision-making (Bratvold et al., 2010). Uncertainty is due to a lack of human knowledge about that occurrence and uncertainty does not mean ambiguity in natural occurrence. So, instead of explaining the circumstance with a single deterministic model, the probability of occurrence should be used for that natural occurrence.

In the oil and gas (O&G) industry, geological uncertainties are commonly modeled by assigning probability distributions to the parameters of a geological model; for example, a multivariate log-normal distribution is assigned to the permeability field of a geological model.

Numerous uncertainties exist in studying subsurface oil and reservoir, including reservoir engineering characteristics, fluid properties, wellbore design, fluid flow, and geological interpretation.

PVT: PVT models generally try to simplify the fluid behavior to some equations and plots based on small-scale lab experiments. Black-Oil models and Equation-of-state models have their assumptions and simplification. The Black-Oil model can represent gravity segregation, near-wellbore effect, and negative compressibility(Coats, 2000). Equation of State (EOS) models are obtained under specific pressure-volume-temperature conditions.

Wellbore**:** Wellbore geometry errors are related to the errors in the measurements, geological structures in the drilling path, and reservoir. Upscaling the field during simulation also causes changes in grids connected to the well and its trajectory (Rammay, 2020).

Numerical error: Analytical solutions equations used in the oil and gas industry are generally too complex to solve and computationally infeasible; thus, instead, numerical solutions have been used in all reservoir simulators. A numerical solution will add error to the model (Haefner, 2005).

Upscaling: The main idea of upscaling is to replace several heterogeneous fine grid blocks with one equivalent coarse homogeneous grid block. So, the essence of upscaling is arithmetic averaging (Qi and Hesketh, 2005). Because of the averaging in upscaling, an upscaled model is less heterogeneous than its corresponding fine-grid model and the nature of a reservoir.

All these uncertainties and other sources of errors plus engineering assumptions add up and divert the model, change the probability distribution of the model, and reduce the quality of the model. Using a model which does not correctly represent uncertainties might result in biases and wrong variance. Any decision based on this model would not be optimal. Applying the Bayes theorem on a low-quality model with extreme error would obtain a wrong or maybe unphysical solution. So studying model error and controlling model error is essential; therefore, the effects of model error in updating the prior knowledge in DA is important.

The classic approaches to mitigate model error in data assimilation defines the model error as the difference between the truth and the model(Harlim, 2017). However, since the truth is unknown, this approach cannot mitigate the model error. Several studies have been done to understand and quantify the model error and its impact on DA.

The effect of the model error has been considered by introducing an optimal reduced stochastic filter in KF (Harlim, 2017). It has been suggested that the model error is not a white Gaussian distribution, and the biasness of the model error should also be considered. In other studies with Extended Kalman Filter and Short-Time Extended Kalman Filter (Carrassi and Vannitsem, 2016), algorithms were reformulated to consider the effects of model error. The results showed that model error should be considered to update the model more accurately (posterior of Bayes' theorem). Carrassi et al. (2016) tried to find an analytical solution by adding model error to the initial condition, and their numerical case study showed that by neglecting model error, the accuracy of the updated model would be reduced. Harlim (2017) showed that by adding the error in

the initial model in the meteorological system, we could predict the future outcome more accurately.

The main contribution of this thesis is to discuss the importance of accounting for model error in ensemble-based DA and analyze and discuss the impact of model errors on data assimilation and probability forecasts by contrasting the results with and without including model errors in data assimilation.

This thesis consists of 4 chapters. Following this introduction, Chapter 2 review the workflow of EnKF and discusses model error and model error sources. In chapter 3, a novel approach will be proposed to model and assess the model errors and account for them the EnKF updating. Chapter 4 discusses the results of applying the proposed approach based on two examples and concludes this thesis together with recommendations for future works.

Chapter 2 Probabilistic history matching

2.1 History matching

An essential task for a reservoir engineer is analyzing available data, predicting future production, and designing and optimizing production strategies in a specific field. The available subsurface reservoir model, which is built based on prior knowledge, should be adjusted with production data to be able to reproduce previous behavior.

Since the start of the oil industry, several ideas have been developed to accomplish this task, from Decline Curve Analyses (DCA) to reservoir simulation. (Craft and Hawkins, 1991).

In DCA, dynamic and static properties of the reservoir such as porosity, permeability, capillary pressure, and fluid saturations are not considered directly. It assumes that whatever controlled the production rate in the past will continue to maintain its trend in the future similarly (figure 2-1).

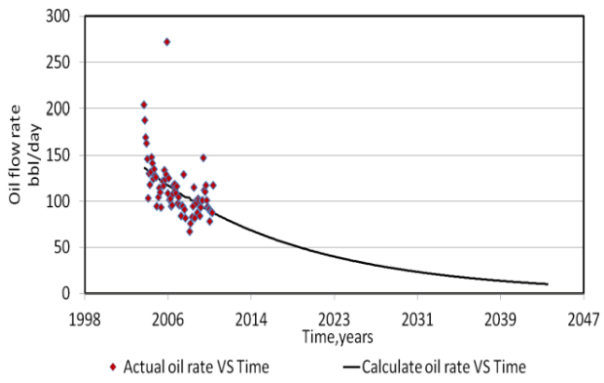


FIGURE 2-1 DECLINE CURVE ANALYSES (RAHUMA ET AL.)

In manual History Matching (HM) with reservoir simulation, one tries to precisely change reservoir properties, including porosity, permeability, fault locations, et cetera, to reduce the discrepancy between simulation and the observation from the subsurface reservoir, as shown in figure 2-2. This method is time-consuming, not accurate, and demands lots of experience, and results might be biased.

In recent decades, a new HM approach called Assisted History Matching (AHM) approaches which use statistical solutions to assist reservoir engineers in considering more uncertainties and variables to perform history matching. AHM has been compared with manual HM, and results showed that AHM is trustable and very successful in updating and adjusting the geological model(Gruenwalder et al., 2007). These methods are based on Kalman Filter (KF).

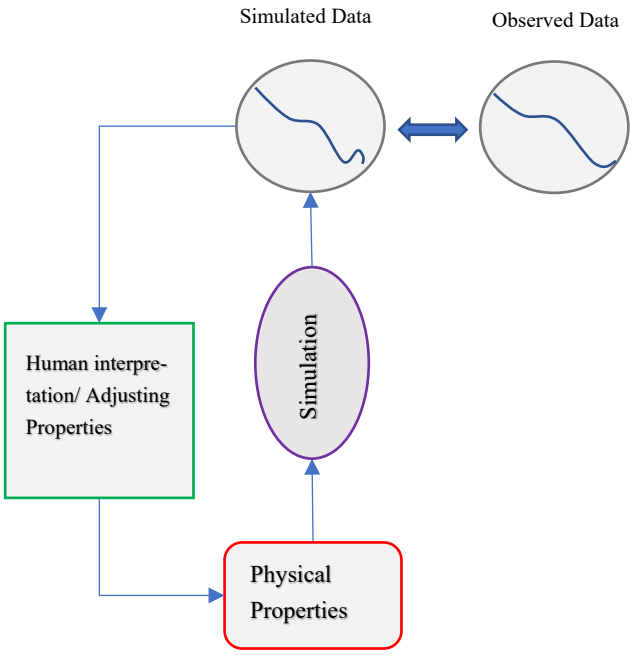


FIGURE 2-2 MANUAL HISTORY MATCHING FLOWCHART

2.2 Kalman Filter:

The Kalman filter is a recursive Bayesian algorithm that updates the model parameters with given observations.

$$\boldsymbol{x_k = Ax_{k-1} + w_{k-1}} \quad \text{2-1}$$

$$\boldsymbol{z_k = Hx_{k-1} + v_{k-1}} \quad \text{2-2}$$

$$\boldsymbol{\omega \sim \mathcal{N}(0, Q)} \quad \text{2-3-A}$$

$$\boldsymbol{v \sim \mathcal{N}(0, R)} \quad \text{2-4-B}$$

Where x is the state vector at the current time, A is a linear transition matrix (forward model), ω is an error in the model with the covariance of $\boldsymbol{Q}$, z is the observation vector, H is the observation operator, v is the observation error vector with the covariance of $\boldsymbol{R}$, and k is referred to time step. KF updates the prior state vector to posterior with the given observation by calculating the error covariance matrix:

$$\boldsymbol{e_k} \equiv \boldsymbol{x_k} - \boldsymbol{\hat{x}_k} \quad 2\text{-}5$$

$$\boldsymbol{P_k} = \boldsymbol{E[e_k\, e_k^T]} = \mathbf{A}\boldsymbol{P_{k-1}A^T} + \boldsymbol{Q} \quad 2\text{-}6$$

Where $\boldsymbol{e_k}$ is estimated error, $\boldsymbol{\hat{x}_k}$ is posterior of the state vector, P is the estimated error covariance matrix, E is the expected value, and T is the transpose operator. Updating step is defined as equations 2-6 and 2-7 when x^f and x^a are forecasted and updated vectors and term k_k is the Kalman gain.

$$\boldsymbol{x_k^a} = \boldsymbol{x_k^f} + \boldsymbol{k_k}(\boldsymbol{z_k} - \boldsymbol{H\hat{x}_k}) \quad 2\text{-}7$$

$$\boldsymbol{K_k} = \boldsymbol{P_k H_k}(\boldsymbol{HP_k H^T} + \boldsymbol{R}) \quad 2\text{-}8$$

While applying the KF algorithm, some assumptions must be considered. Forward model A should be linear, model error and measurement error are time-independent and independent from each other, and they follow Gaussian distribution. However, linearity of the forward model and Gaussianity limit the application of KF, so different variations based on KF have been proposed, such as Extended KF, Unscented KF, et cetera, were developed. These methods try to estimate the derivative of the non-linear forward model or use unscented transformation (Julier and Uhlmann, 2004) to handle highly non-linear forward models. Details of these methods are beyond this thesis.

Evensen, in 1994, introduced a new approach to using Monte Carlo Simulation to overcome some of these assumptions.

2.3 Monte Carlo Simulation

Monte Carlo Simulation is a powerful tool for analyzing uncertainties in the oil and gas industry and geoscience(Bratvold et al., 2010). MCS works perfectly in high dimensions. First, the model needs to be appropriately defined. Second, the expert's beliefs about the uncertainty variables are represented by the probability distributions and defined as inputs in MCS. The variables are assumed to be independent in MCS. Therefore, dependencies need to be included in the model if two or more variables are dependent on each other. Third, MCS uses a random number to sample each input probability distribution and computes the output variables. The result of each trial will be stored.

Moreover, this process is repeated many times, and the stored results will be used to generate histograms and statistics of the output variables. The distributions of the output variables represent the probabilities of their occurrence; the bigger number of samples taken, the more representative the output distributions(Elvaretta, 2021). Bratvold and Begg (2010) summarized the procedure of the MCS in Figure 2.3. One limitation of MCS could be its computational burden(Alerstam et al., 2008)

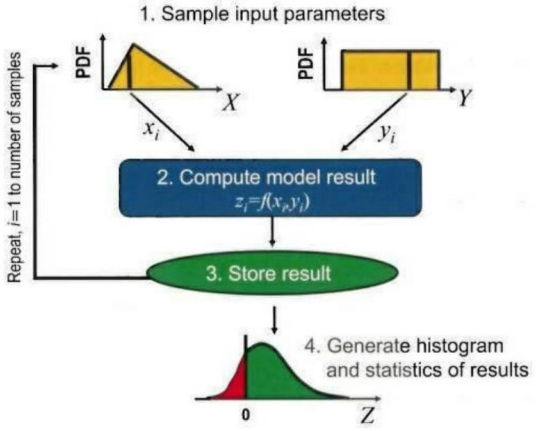


FIGURE 2-3 SCHEMATIC OF MONTE CARLO SIMULATION PROCEDURE (BRATVOLD ET AL., 2010)

2.4 The Ensemble Kalman Filter

The Ensemble Kalman Filter, a combination of KF and MCS, can be used to rectify the assumptions of non-linearity and Gaussianity in KF. In EnKF, having N number of realizations as representative of the distribution will be used recursively with given observation data. The posterior realizations are obtained in each iteration by updating the prior realizations and the observation, and the posterior will be prior for the next time step.

Aanonsen et al. (2009) explained the implementation of EnKF in the reservoir engineering context.

$$\boldsymbol{Y_k^f} = \begin{bmatrix} \boldsymbol{m_k^{n,f}} \\ \boldsymbol{x_k^{n,f}} \\ \boldsymbol{d_k^{n,f}} \end{bmatrix} \qquad 2\text{-}9$$

Where m is static variables (porosity, permeabilities), x is dynamic variables (pressure, saturation), d is observation/data from the model (BHP, water cut, and rates), and f, n superscripts are for forecasted state and realization number and k subscript for timestep.

$$\boldsymbol{x_k = \mathcal{M}(x_{k-1}, m)} \quad 2\text{-}10$$

$$\boldsymbol{z_k = \mathcal{H}(x_{k-1}, m)} \quad 2\text{-}11$$

So, with this definition, the error covariance matrix and Kalman gain will be:

$$\boldsymbol{P_k^f = \frac{1}{N-1}[Y_k^f - E(Y_k^f)][Y_k^f - E(Y_k^f)]^T} \quad 2\text{-}12$$

$$\boldsymbol{K_k = P_k^f H_k^T (H_k P_k^f H_k^T + C_{d_k})^{-1}} \quad 2\text{-}13$$

and

$$\boldsymbol{C_{d_k} = \frac{1}{N-1} v_k v_k^T} \quad 2\text{-}14$$

$$\boldsymbol{\mathcal{H}_k = [g_k^{n,f}] (Y_k^f)^{-1}} \quad 2\text{-}15$$

Where v is measurement noise. Finally, updated variables are Y_k^a when d_k is observed data at timestep k:

$$\boldsymbol{Y_k^a = Y_k^f + K_k(d_k - H_k Y_k^f)} \quad 2\text{-}16$$

2.4.1 Model Error Covariance:

In KF, the term Q is defined as the process noise covariance, which indicates the model error. The model error is generally not known, and a numerical model will therefore evolve (Evensen, 2009). Model error is introduced because of the numerical solution of nonlinear and non-Gaussianity (Jiyuan et al., 2013).

For a linear system(Evensen, 2009):

$$\boldsymbol{x_k^t = A x_{k-1}^t + w_k} \quad 2\text{-}17$$

$$\boldsymbol{x_0^t = X_0 + e} \quad 2\text{-}18$$

When x^t is the true state and x_0^t is the true state at the initial time with error e. Moreover,

the forecasted model is shown as follows:

$$\boldsymbol{x_k^f = Ax_{k-1}^a} \quad 2\text{-}19$$

$$\boldsymbol{x_0^a = X_0} \quad 2\text{-}20$$

Furthermore, by subtracting 2-18 and 2-19:

$$\boldsymbol{x_k^t - x_k^f = A(x_{k-1}^t - x_{k-1}^a) + w_k} \quad 2\text{-}21$$

Error covariance matrix will be calculated as:

$$\begin{aligned} \boldsymbol{P_{K_{xx}}^f} &= \overline{\boldsymbol{(x_k^t - x_k^f)^2}} \\ &= \overline{\boldsymbol{(A(x_k^t - x_{k-1}^a) + w_k)^2}} \\ &= \boldsymbol{AP_k^a A^T + Q_{k-1}} \end{aligned} \quad 2\text{-}22$$

For a nonlinear system, a similar approach will show where $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{Q}$ are matrix forms of the forward model and the covariance of the process noise(Evensen, 2009):

$$\mathbf{P_{K_{xx}}^f = \mathcal{M} P_k^a \mathcal{M}^T + Q_{k-1}} = \overline{\left(\boldsymbol{\psi_k^t - \psi_k^f}\right)^2} \quad 2\text{-}23$$

In EnKF error covariance matrix is shown below if $\boldsymbol{\psi}$ is the state variable(Evensen, 2009). One should be aware that the true state is unknown, so that it will be replaced with the mean of the ensemble $\overline{\boldsymbol{\psi_k}}$. If the ensemble size is infinite, the error covariance matrix will converge to the below equation. In this condition there is an infinite number of realizations converge to P_K^f , so the same error covariance can be approximated with sufficient ensemble size:

$$\boldsymbol{P_{K_{xx}}^f \cong \overline{(\overline{\psi_k} - \psi_k^f)(\overline{\psi_k} - \psi_k^f)^T}} \quad 2\text{-}24$$

2.5 Model error and sources of model error:

In both linear and nonlinear models, modeling error is the main item to achieving high-accuracy state estimation(Lu et al., 2021). As pointed out earlier, the error is a discrepancy between the true state and what can understand from the true state. So, in a condition where the true state is uncertain, for example, a subsurface reservoir, the model error cannot be defined and calculated with this approach.

In different fields of engineering and science, sources of error are various, but in subsurface studies, they are generally due to: (Jiyuan et al., 2013) and (Rammay, 2020)

a. Numerical errors
b. Upscaling errors
c. Physical errors
d. Human errors

2.5.1 Numerical Errors

In reservoir simulation, the forward models (functions) such as Darcy law, Mass conservation law, Buckley-Leverett, et cetera; are highly nonlinear, and solving them in analytical form would be computationally a big challenge. Thus, numerical approaches are being applied to approximate them. The numerical solutions are specified to limited time and space resolution; higher resolution and more nodes cause a smaller error but increase the computational costs. This error arises because an exact analytical solution to the equation is not obtained by the numerically approximated solution(Jiyuan et al., 2013). In figure 2-4, conservation law with a simple non-linear function with a specific initial condition is solved. In the left plot, 50 nodes, and for the right plot, 500 nodes have been used to calculate the solution. Comparing the numerical and exact analytic solutions clarifies the numerical error.

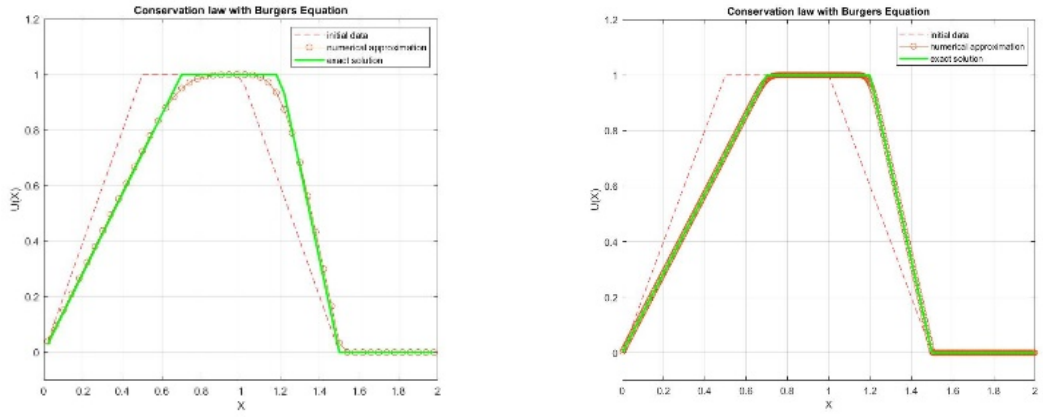


FIGURE 2-4 SOLUTION OF CONSERVATION LAW WITH A NONLINEAR FUNCTION WITH 50 (LEFT) AND 500 (RIGHT)NODES.

2.5.2 Upscaling error:

In reality, a reservoir has continuous properties and different properties in every centimeter or millimeter. For example, pore size and, consequently, porosity and permeability change in each

centimeter of formation; in one specific point, one can be in a void space with high porosity, and a few millimeters aside is noneffective porosity in capsulated in the matrix of rock which not permeable. Also, a geological interpretation measurement like seismic data has its resolution with some uncertainties so that geologists can make fine grids with a specific resolution.

A reservoir can have millions of grids in high-quality fine resolution built by geologists, but simulating a reservoir with millions of grid cells is computationally costly, so reservoir engineers prefer to scarify some details to speed up the simulation by coarsening the grids. Upscaling or grid coarsening is the process of attaching fine cells to create a bigger cell with the average value of all fine cells; heterogeneity will decrease, dimension of one cell will increase, which both cause an error in the model.

In figure 2-5, with the upscaling ratio of 1 to 3, permeability distribution was changed (histogram plot), and values were averaged.

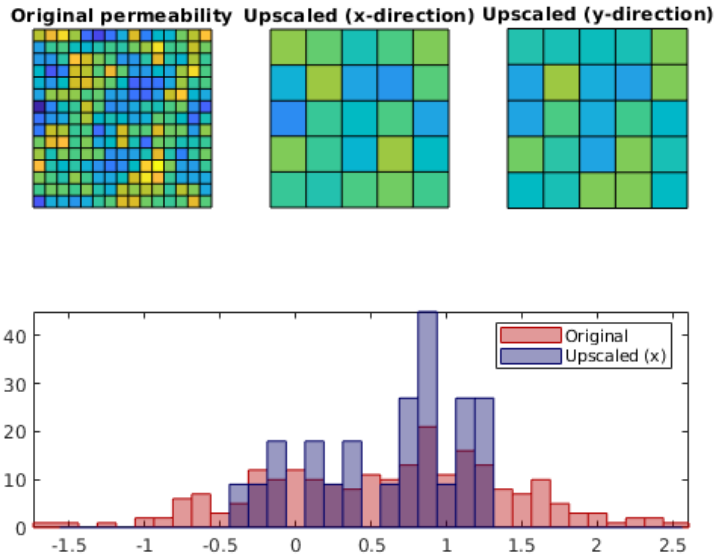


FIGURE 2-5 "UPSCALING: (UPSCALING OF RESERVOIR PROBLEMS, THE MATLAB RESERVOIR SIMULATION TOOLBOX 2019B DOCUMENTATION," 2019)

The upscaling process introduces an error due to shifting connecting cells to the wells and the trajectory of the well. In the blow plot, a well is lactated in a grid with extremely high permeability and surrounded by high permeable grids, but after coarsening, the property of the connected

cell is changed to average also, and the well location and position are now different as shown in figure 2-6.

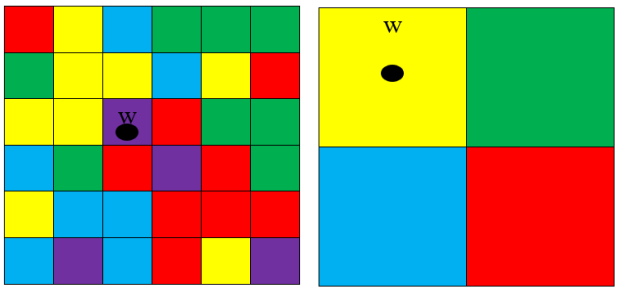


FIGURE 2-6 UPSCALING IN 2D, THE EFFECT OF UPSCALING ON WELLBORE LOCATION AND GEOMETRY.

2.5.3 Physical errors:

Physical errors refer to choosing PVT tables, relative permeability models, and some assumptions while describing the physical properties of the subject field. Black-oil model is too simple to consider all parameters; EOS methods are experimental and valid under specific conditions. Relative permeability is calculated in the lab, and it is an approximation. The rock compressibility factor is kept constant, but it is not a constant value in reality.

2.5.4 Human Errors:

Geological interpretation, assumptions, and initial engineering guesses are always part of reservoir simulation and model interpretation. It is constantly being said that experienced geologists or engineers have better interpretation or understanding. Still, it is just human adventitious biases since there is no clear and confident answer to whether the interpretation and engineering guesses are correct.

2.6 Decline Curve analyses

Before the invention of the modern digital computer, which is capable of handling complex calculations, in 1945, J.J Arps, in a series of studies, introduced a method for analyzing declining production rates and forecasting the future performance of oil and gas wells. This method is based on empirical observation of production decline fitted on exponential, hyperbolic or harmonic functions. Figure 2-7 shows the decline curves for these equations.

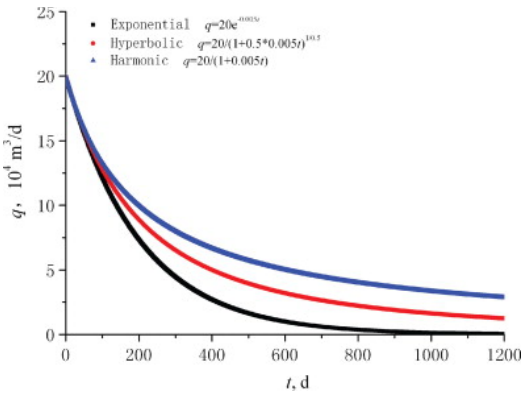


FIGURE 2-7 ARPS DCA CURVES FOR EXPONENTIAL, HYPERBOLIC AND HARMONIC FORM.

Basic characteristics	Decline type	Rate–time	Rate–cumulative production	D_i*t
$b = 0$	Exponential	$q_t = q_i e^{-D_i t}$,	$Q = (q_i - q)/D_i$	$D_i t = \ln(q_i/q)$
$0 < b < 1$	Hyperbolic	$q_t = q_i(1 + bD_i t)^{-1/b}$,	$Q = \frac{q_i}{(1-b)D_i}[1 - (\frac{q}{q_i})^{1-b}]$	$D_i t = \frac{(q_i/q)^b - 1}{b}$
$b = 1$	Harmonic	$q_t = q_i/(1 + bD_i t)$,	$Q = q_i \ln(q_i/q)/D_i$	$D_i t = (q_i/q) - 1$

TABLE 2-1 ARPS DCA TYPES AND PARAMETERS (HEDONG, 2015)

In table 2-1, these equations have been demonstrated, where q_t is well the rate at time t, q_0 is the initial rate at t = 0, Di is the initial decline rate, and b is the decline exponent. With the assumption of boundary-dominated flow, the upper bound of b is unity. For the transient flow regime of unconventional production, b is often greater than unity (Valkó and Lee 2010), and as a result, the cumulative production is unbounded. To avoid unbounded cumulative production, we restrict the range of parameter b to stay in 0 <b < 1 (the cumulative production will also be unbounded when b = 1) in our application.

Production data can be categorized into two-time intervals. In the initial transition, when the well start producing, the pressure of the well at a constant flow rate will not change. During this period which can last from several minutes to several years, depending on the mobility ratio of the reservoir, the flow has not reached the boundary or draining flow of the neighboring well. However, the pressure starts to drop when the flow regime has reached any boundary, and the

well will produce under the boundary-dominant regime. This period lasts for many years, and DCA can be used to model the flow rate.

The Arps model answered three major questions at the time, a method to predict for future based on past production, an appropriate mathematical approach to describe the data trends, and the ability to identify its parameters(Valkó and Lee, 2010).

2.6.1 Stretched Exponential Method:

After decades of applying Arps DCA, with improving understanding of reservoir behavior and effective parameters in reservoir engineering, new questions arose that which Arps model could not answer. Pseudo-steady-state flow, material balance, or unconventional reservoir were beyond the capability of Arps DCA, so several new imperial methods were introduced by the time to answer some of those issues.

P.P Valkó, by studying the tight gas reservoirs, introduced Stretched Exponential Method (SEM) to predict the unconventional reservoir production rate. SEM can consider heterogeneity because SEM is an integral form of exponential decays with a "fat-tailed" probability distribution of the time constants(Valkó and Lee, 2010)

$$\boldsymbol{q_t = q_i e^{-(\frac{t}{\tau})^n}} \qquad 2\text{-}25$$

Where τ , n, and t are the characteristic time parameter, the exponent parameter, and production time, respectively(Valkó, 2009).

2.7 Reservoir simulation:

With having modern computers around that can process and store big data, solve non-linear equations, and perform logic operations, now reservoir engineers are able to study the behavior of a reserve under any physically correct initial condition, monitor fluid flow in porous media, and history matches the production and predict the future of the reservoir.

Generally, reservoir simulators use discretization methods to solve conservation of mass, isothermal fluid phase behavior, a chemical reaction between fluids and porous media, and the Darcy equation for a 3D model. Building a 3D model starts with defining the geometry of the reservoir and shape of grids, assigning petrophysical properties to each grid and then production/injection wells will be connected to grids. In the next step, initial conditions like rates,

pressures, and composition of injected fluid will be introduced. Finally, the model will simulate under proper physical laws in the desired time steps.

The “Matlab Reservoir Simulation Toolbox”(Lie, 2019) and the “Open Porous Media”(Rasmussen et al., 2019) are two famous open-source simulators, and “Schlumberger ECLIPSE”(Schlumberger, 2014) is the most trusted commercial simulator.

Chapter 3 Proposed Approach for Accounting For Model Errors in EnKF Updating

Error and uncertainty are fundamental parts of statistics, and the terms “Error” and “Uncertainty” are being used interchangeably, but these terms should be distinguished in this study. In the American Institute of Aeronautics and Astronautics (AIAA) Guidelines, it has been stated that: Uncertainty is defined as: *"A potential deficiency in any phase or activity of the modeling process that is due to the lack of knowledge."* Error is defined as: *“A recognizable deficiency in any phase or activity of modeling and simulation that is not due to lack of knowledge.”* (“AIAA Guidelines 077,” 1998)

As briefly discussed in chapter 2, there are several sources of error in a subsurface reservoir model, including simplified PVT, grid upscaling, et cetera. In this chapter, a method will be proposed to account for model error in probabilistic HM, which allows us to use the statistical information of model error to account for it while updating the prior knowledge. This method will be examined in 2 case studies from the reservoir engineering context.

DCA methods will be used to model an unconventional tight gas field in the first case. One model will be built as a high-quality model, and the second is a low-quality model which contains errors. Their distribution will be compared, and then by EnKF, these models will be updated in the next step. Then, the suggested approach will be used to calculate the discrepancy between these two models, and a function of Model Error will be proposed to account for the model error. Furthermore, in the last step, statistical information from the Model Error algorithm will be used in the EnKF to update the prior model to account for model error and correlate the prior model with the given observation.

Next, a synthetic 2D reservoir will be upscaled to different grid sizes, and then the high fidelity model will be used as the HQ model. The model error will be calculated. Changes in these models will be discussed, and then EnKF will be used to these models with and without accounting for model error. Finally, the distribution of models and forecasted models will be compared to understand the effect of accounting for model error in probabilistic history matching.

3.1 Methodology:

The Bayesian theorem can consider all uncertainties in the problem, and it could be used to consider model error as probability distribution to help us to update the model containing the error. Bayes theorem is shown in equation (3-1) where m_{prior} and m_{post} are probabilities of prior and posterior models and d_{obs} is observed data:

$$\boldsymbol{P(m_{post}|d_{obs}) \propto P(d_{obs}|m_{prior}).P(m_{prior})} \quad 3\text{-}1$$

Now, if we study high-quality and low-quality models, it could be presented with equation 3-2 and by substituting it in Bayes theorem:

$$\boldsymbol{\epsilon = m_{HQ} - m_{LQ} \implies m_{HQ} = m_{LQ} + \epsilon} \quad 3\text{-}2$$

$$\boldsymbol{P(m_{HQ}|d_{obs}) \propto P(d_{obs}|m_{HQ}).P(m_{HQ})} \quad 3\text{-}3$$

$$\boldsymbol{P(m_{HQ}|d_{obs}) \propto P(d_{obs}|m_{LQ} + \epsilon).P(m_{LQ} + \epsilon)} \quad 3\text{-}4$$

Having a function or algorithm to consider error ϵ in the Bayesian theorem to calculate the posterior would correct the posterior's uncertainty distribution and avoid neglecting the low-quality model's errors.

3.1.1 Model Error function:

In this approach, in contrast with the classical definition of model error, the model error has been defined as the discrepancy between the forecasted high-quality model and the forecasted low-quality model. Note that since a low-quality model (the model with error) can be overestimated or underestimated, the model error value can be positive or negative. Now the main question is how to extract this information from the error in the term of a probability distribution of model error.

3.1.2 Sliding window algorithm:

In an experiment, a sliding window algorithm has been modified and implemented to calculate the statistical information. The sliding window moves from the beginning of input data (discrepancy between forecasted high and low quality models) with the proper given window size. Then it calculates the Mean and the Standard Deviation of the values inside the window and stores them as output values corresponding to the input data. This process will be repeated for all input values one by one. Finally, with linear interpolation, two continuous functions of the mean and the standard deviation regarding the model's forecast with error will be available, assuming

that model error has Gaussian distribution. In figures 3-1 and equations 3-5, 3-6, and 3-7, the mechanism of this method has been demonstrated.

Note that window size should be small enough to capture slight fluctuation and big enough to catch a specific sample size. A machine-learning algorithm could be used to optimize the window size. Figure 3-2 shows all steps of calculating the mean and STD of the model error.

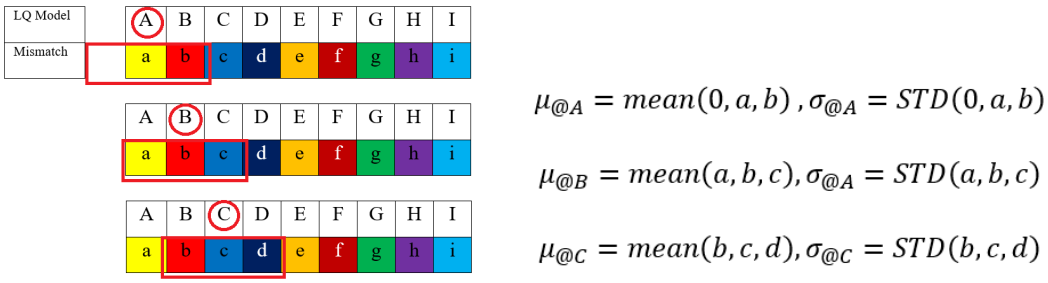


FIGURE 3-1 SLIDING WINDOW SAMPLING AND PARAMETER CALCULATION

$$\boldsymbol{\mu_{@A} = mean(0,a,b)\,, \sigma_{@A} = STD(0,a,b)} \quad 3\text{-}5$$
$$\boldsymbol{\mu_{@B} = mean(a,b,c), \sigma_{@A} = STD(a,b,c)} \quad 3\text{-}6$$
$$\boldsymbol{\mu_{@C} = mean(b,c,d), \sigma_{@C} = STD(b,c,d)} \quad 3\text{-}7$$

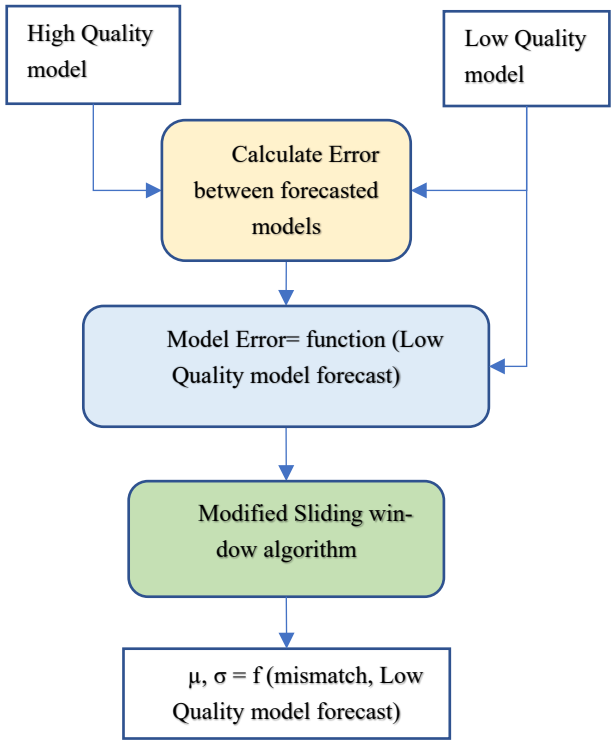


FIGURE 3-2 MODEL ERROR CALCULATION FLOWCHART.

3.1.3 Updating the prior with the consideration of model error:

EnKF algorithm has been discussed earlier in chapter 2. We modify this algorithm to consider the bias and overconfidence in the low-quality prior model while updating it with observation data.

Now, the mean and standard deviation of the model error for the forecasted low-quality model are available. We use these statistical parameters to generate a random sample and add it to the initial forecasted prior; Thus, an interpolator function was used to estimate the values of the forecasted prior model, and for each forecasted value, a Gaussian random sample with the mean and standard deviation of the model error will be generated. It should be noted that accounting for the model error should be calculated just for the first iteration of EnKF. Thus, the corrected prior model containing error information will be updated; otherwise, we add extra error in each iteration. Equation 3-8 shows how equation 2-8 in EnKF has been modified. In equation 3-8, $\boldsymbol{\mu\epsilon_1^{n,f}}$ and $\boldsymbol{\sigma\epsilon_1^{n,f}}$ the mean and STD of model error.

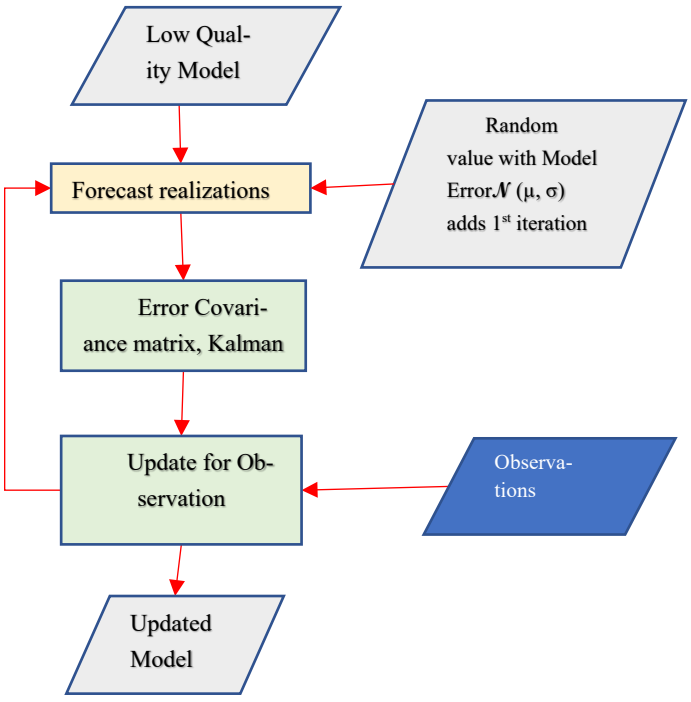


FIGURE 3-3 FLOWCHART OF ENKF ACCOUNTING FOR MODEL ERROR

$$Y_1^f = \begin{bmatrix} m_1^{n,f} \\ x_k^{n,f} + \mathcal{N}(\mu\epsilon_1^{n,f}, \sigma\epsilon_1^{n,f}) \\ d_1^{n,f} \end{bmatrix} \quad 3\text{-}8$$

3.1.4 Forecasting metrics

An accuracy measure is needed to compare the results of the proposed method as forecasts of an ensemble with or without accounting for the model error. Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) will be calculated to compare how updates are close to observations. It has been noticed that these data-fitting measurements alone cannot be an excellent reference to check the quality of the fit (Rammay, 2020), so here RMSE, MAPE, and visual quality check of the forecasted model should be considered.

$$\mathbf{RMSE} = \sqrt{\frac{\sum_{i=}^{N}(x_i - \hat{x})^2}{\mathbf{N}}} \quad 3\text{-}9$$

Where in RMSE, N, x_i and $\hat{x}$ are the number of data points, observed value, and estimated value orderly.

$$MAPE = \frac{1}{n}\sum_{t=}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad 3\text{-}10$$

Which n, A_t, and F_t are the value of times the summation iteration happens, actual value, and forecast value. MAPE works correctly as long as there is no extreme point in the data.

3.2 Case study 1: Decline Curve Analyses

Production data (gas rate) has been generated for the period of 15 years in a synthetic tight gas field, as shown in figure 3-4. The assumption of having a tight gas reservoir will lead to use SEM as a high-quality method to describe the production rate and build a model to predict future production.

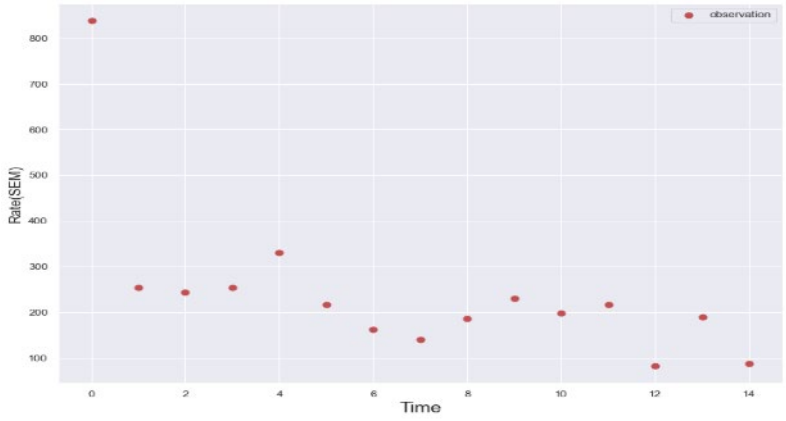


FIGURE 3-4 OBSERVATION OF TIGHT GAS RESERVOIR

3.2.1 Prior realizations

To generate the prior high-quality realizations, SEM parameters τ, n, and q_i will be generated randomly as below, where $\mathcal{U}$ is a uniform probability distribution. One hundred random initial parameters were generated by implementing MCS to calculate the prior model with the SEM equation(equation 2-24). Figure 3-5 shows these realizations for the entire production time.

$$\boldsymbol{\tau \sim \mathcal{U}(min = 0.01, max = 1)} \quad 3\text{-}11$$

$$\boldsymbol{n \sim \mathcal{U}(min = 0.1, max = 1)} \quad 3\text{-}12$$

$$\boldsymbol{q_i \sim \mathcal{U}(min = 100, max = 1500)} \quad 3\text{-}13$$

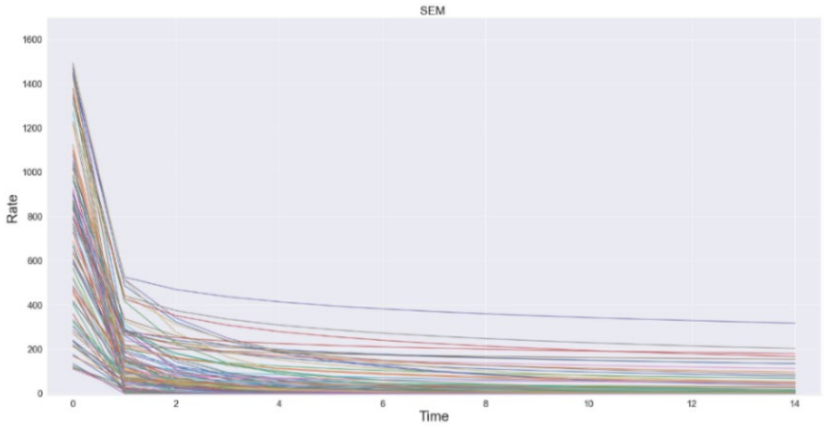


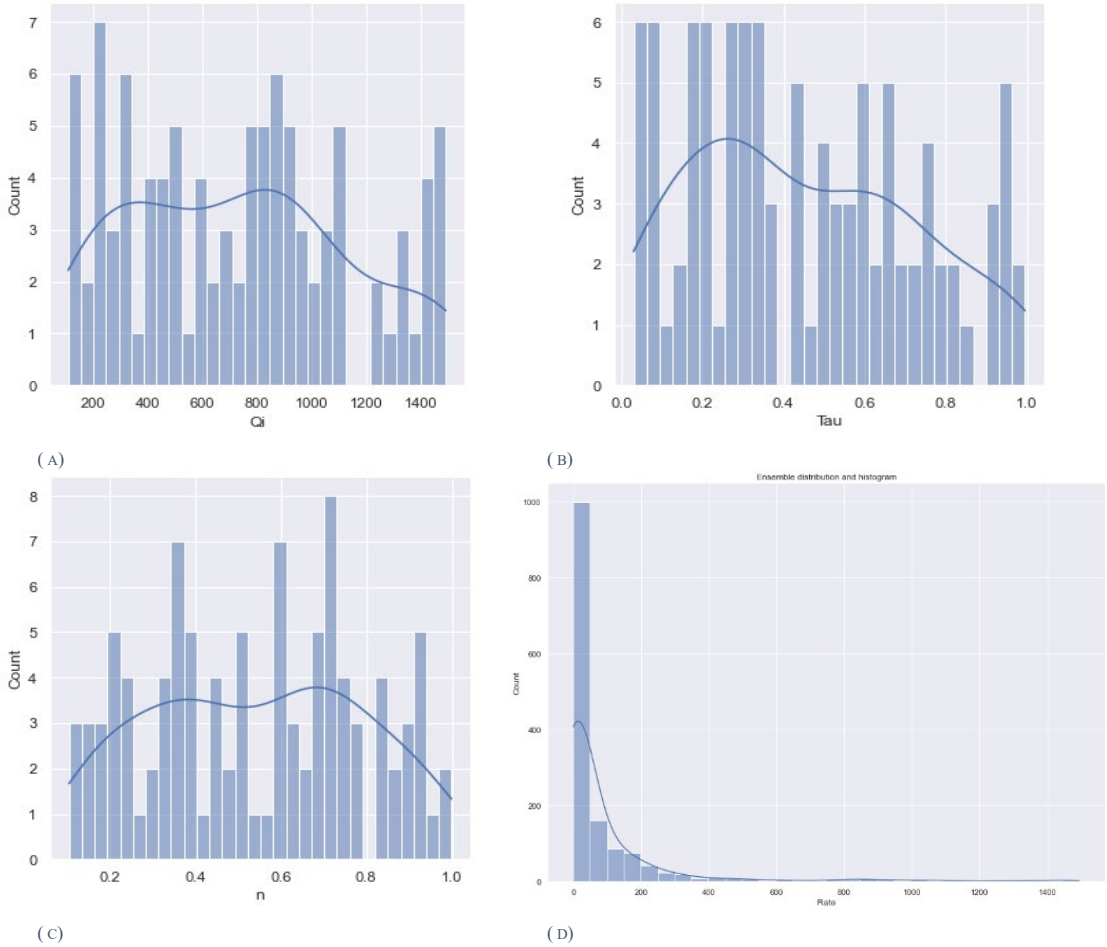
FIGURE 3-5 PRIOR REALIZATIONS WITH SEM

To have a model which can be described as low-quality, Arps DCA has been used. In the main assumptions of the Arps DCA, the decline exponent parameter b is:

$$\boldsymbol{0 \leq b \leq 1} \quad 3\text{-}14$$

However, it has been acknowledged that in the case of using the Arps model for a tight gas reservoir, parameter b can be more than unity(Maley, 1985). So, in this case, parameter b in the hyperbolic formation of Arps was allowed to get bigger values. This change in the Arps model provides maximum fit to SEM realizations.

In the next step, non-linear least squares analyses were applied to fit Arps equations on SEM curves, and at the end of this step, Arps's parameters that can fit SEM realizations curve were obtained. This step has been done using the 'SciPy/optimize/curve_fit' function (Jones et al., 2001). Figure 3-7 clarifies that some fitted curves with Arps equations are a good match while others could not be fitted correctly to the SEM curve, so the error has been introduced in the Arps model.



(A) (B) (C) (D)

FIGURE 3-6 HISTOGRAMS OF SEM PARAMETERS AND QT, THE PROBABILITY DISTRIBUTION OF (A) QI, (B)TAU, (C) N, AND (D) QT.

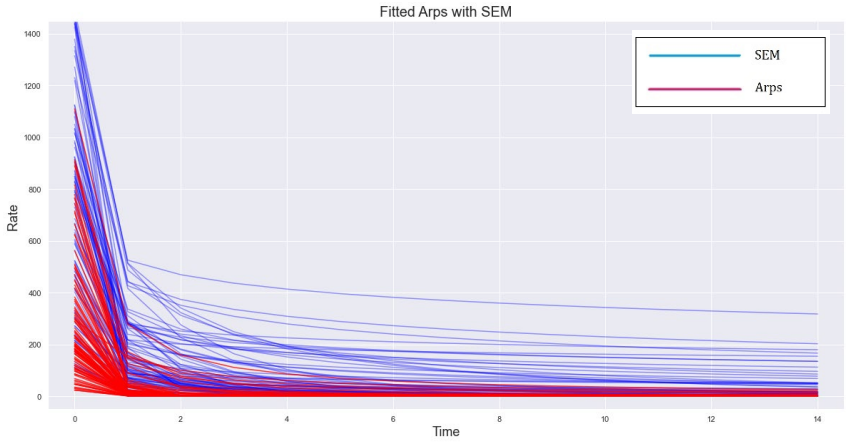


FIGURE 3-5 FITTED ARPS MODEL WITH SEM

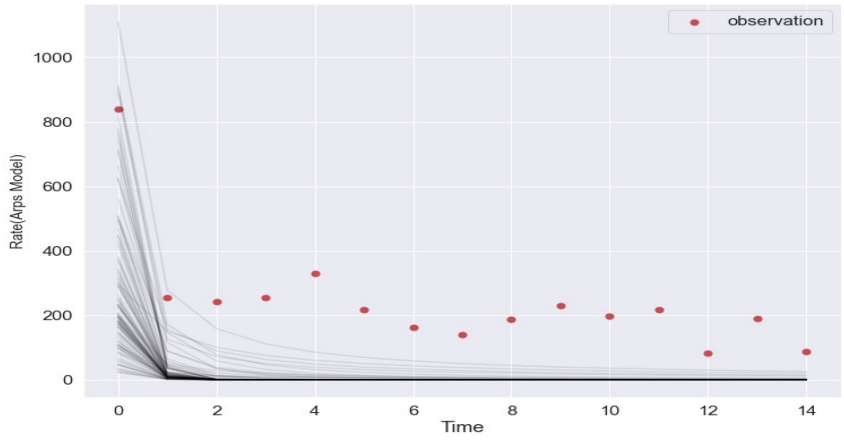


FIGURE 3-6 REALIZATIONS WITH ARPS DCA AND OBSERVATIONS

These distributions demonstrated that the prior Arps parameters b and D_i have a small variance, and based on figure 3-8, the Arps model realizations were not close to the observations, so the Arps model is a low-quality model representing our synthetic tight gas field.

The distribution of Arps's parameter is shown in figure 3-9. It should be considered that parameters of Arps and SEM are defined in different domains and cannot be compared.

Figure 3-10 shows the discrepancy between high and low quality models as black dots. The black box is a schematic of the moving window, and it depicts how the window slides over points and calculates statistical information.

Figure 3-11 shows the mean and the STD of the production rate model error concerning the Arps production rate. In this case study, a cross-validation algorithm has been used to optimize the window size.

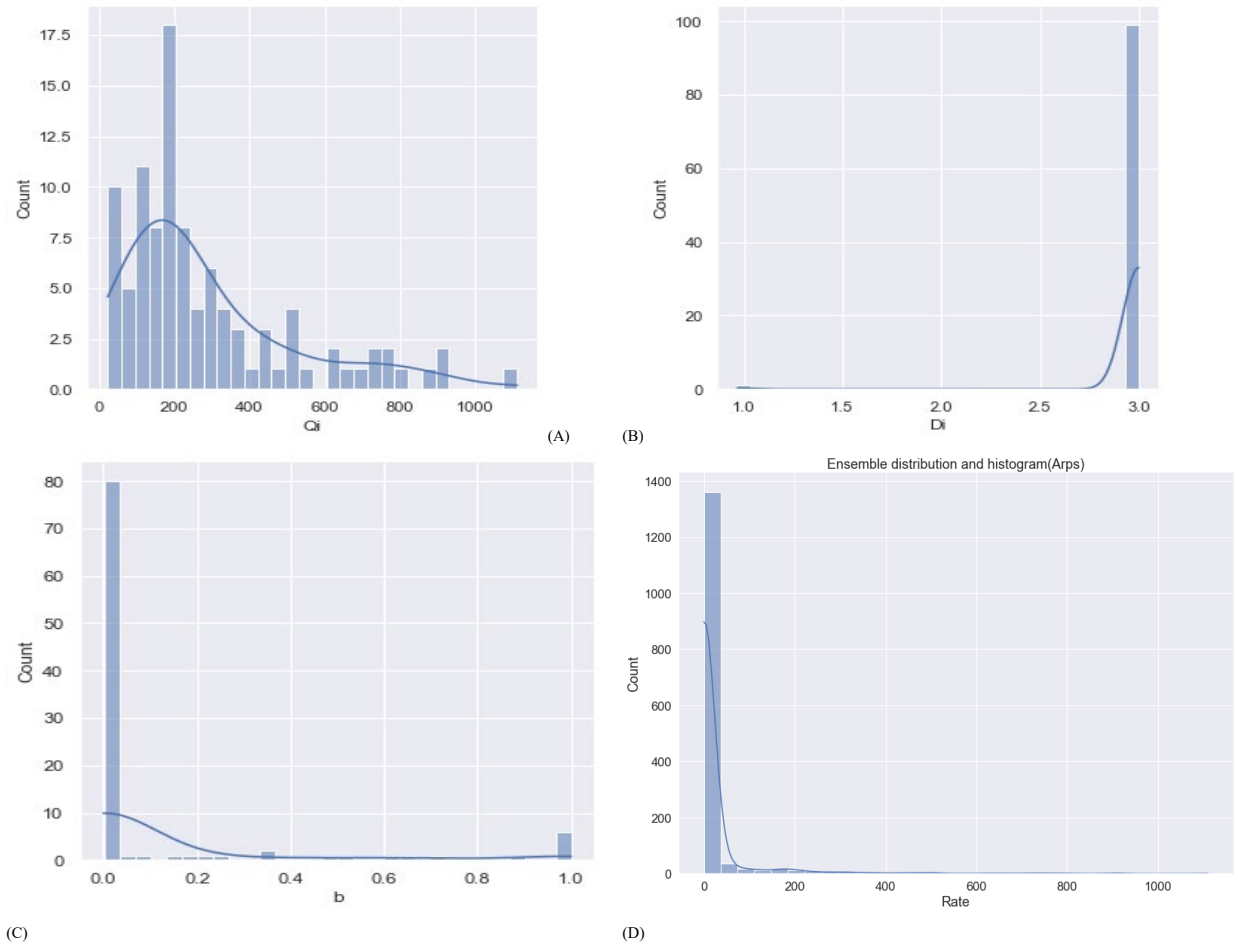


FIGURE 3-7 HISTOGRAMS OF ARPS PARAMETERS AND Q_T, THE PROBABILITY DISTRIBUTION OF (A) Q_I, (B)D_I, (C) B, AND (D) Q_T.

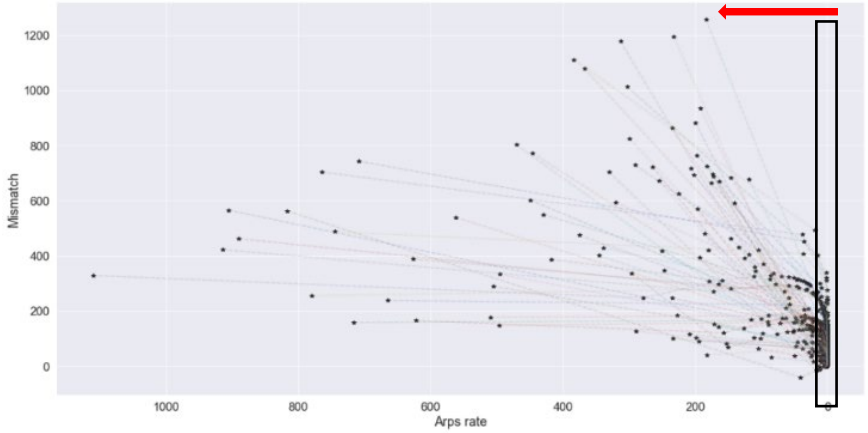


FIGURE 3-8 SLIDING WINDOW SCHEMATIC

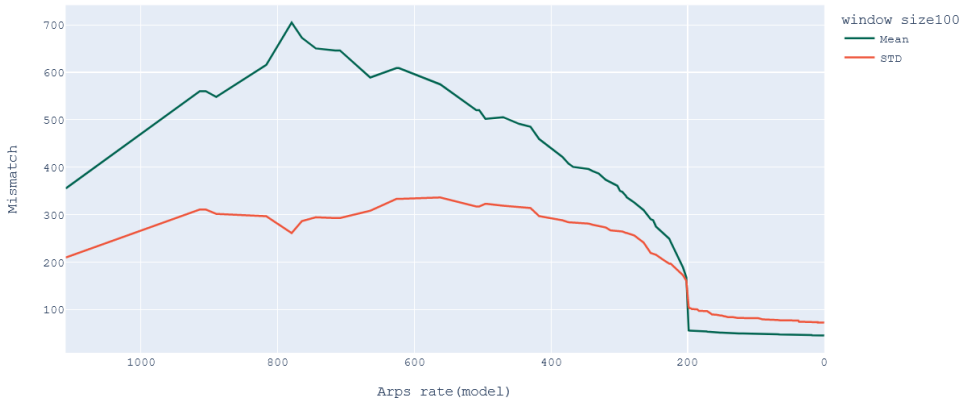


FIGURE 3-9 GRAPHS OF THE MEAN AND STANDARD DEVIATION OF MODEL ERROR

3.2.2 Updating the prior:

Based on figure 3-12, a better fit is observed after updating SEM parameters with observations. Qi at the initial time has a perfect fit, and later, the P80 confidence interval covers the observation points, and the mean is passing in the middle of observations. This plot shows that SEM is a high-quality model and close to the truth, and it was able to model the truth after its parameters were updated with given observations.

From probability distribution and histogram plots in table 3-1, for parameter Q_i, the initial rates are shifted to approximately 809 to 810, which is close to the observation at the initial production time. Also, a narrow distribution for Q_i is obtained, indicating uncertainty reduction. For parameters b and τ, distribution also changed to a wide distribution with bigger variance, and finally, for Q_t, distribution became wider, and a bimodal distribution with the smaller mode at 800 is noticeable.

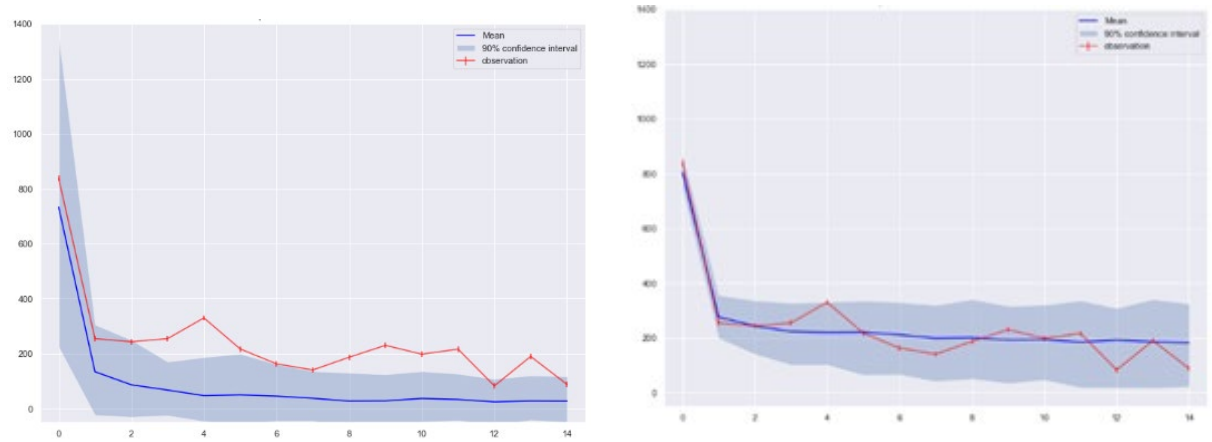


FIGURE 3-10 SEM, PRIOR (LEFT) AND POSTERIOR(RIGHT)

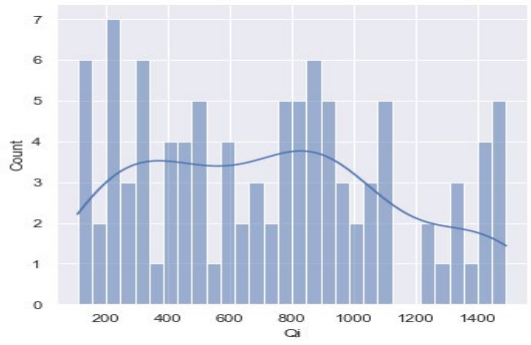
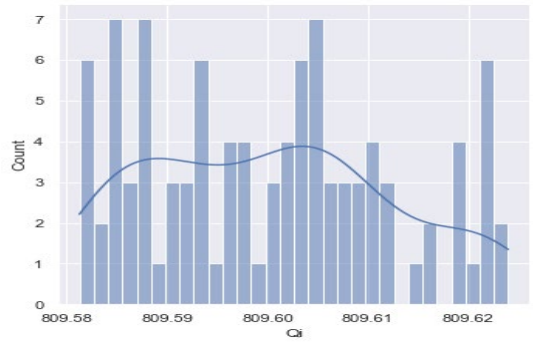
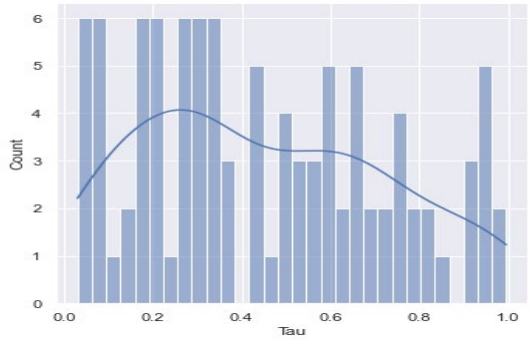
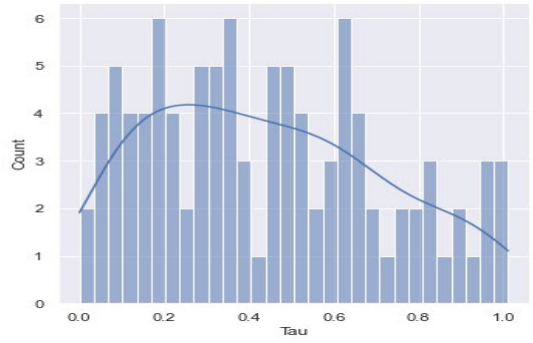
	Prior of SEM Parameters	Posterior of SEM Parameters
Qi		
τ		

TABLE 3-1 HISTOGRAMS OF SEM PARAMETERS AND Q_T FOR PRIOR AND POSTERIOR

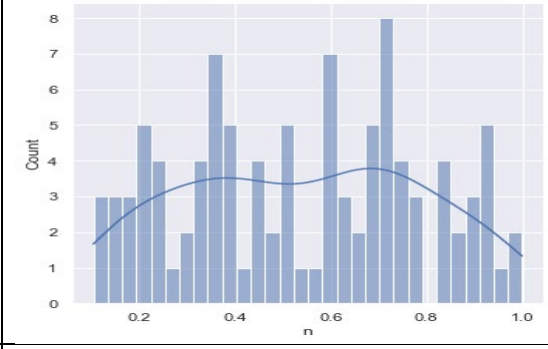
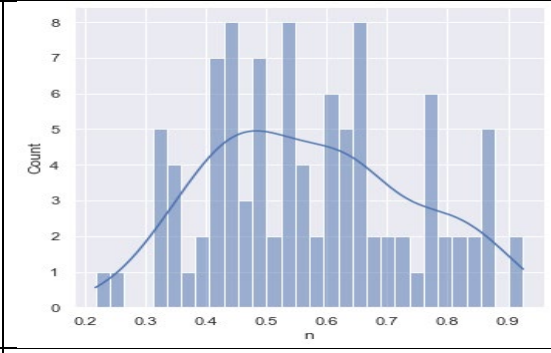
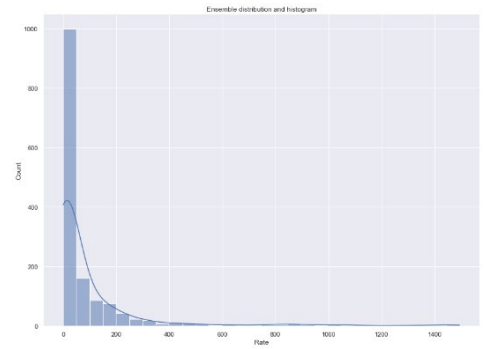
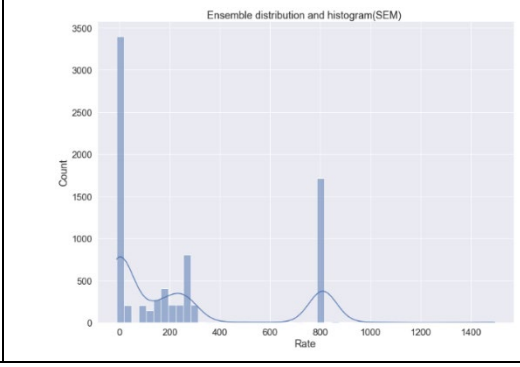
	Prior of SEM Parameters	Posterior of SEM Parameters
n		
Qt		

TABLE 3-1 HISTOGRAMS OF SEM PARAMETERS AND Q_T FOR PRIOR AND POSTERIOR

Then, with EnKF, the Arps model parameters have been updated to correspond to observation points.

From figure 3-13, the Arps model with updated parameter shows a good fit at the beginning of the observation, and between year 2 and year 4, the shape of the curve matches with the observation curve, but the updated model is still underestimating production rate. Since it had been assumed that the reservoir is a tight gas formation (which is not thoroughly describable with Arps DCA), the updated model Arps cannot be a good fit in this case.

Figure 3-14 shows updated Aprs parameters while accounting for model error with EnKF. Distribution is generally wider, but the mean of updated realizations is close to the observation points, and more importantly, the 80 percent confidence interval covers the observation.

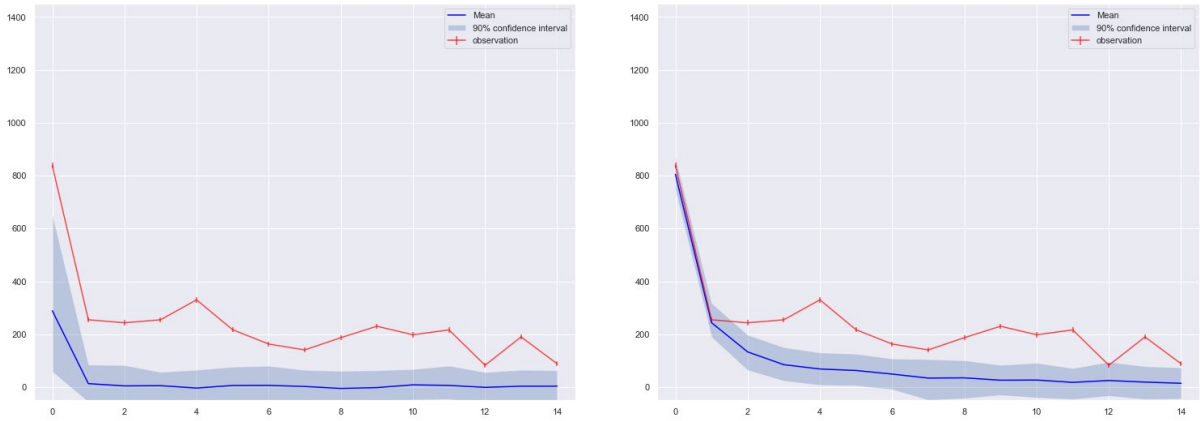


FIGURE 3-11 PRIOR(LEFT) AND POSTERIOR(RIGHT) OF ARPS DCA WITH ENKF. A BETTER FIT IS OBSERVABLE AT THE YEAR 0 TO 2, BUT IT IS STILL UNDERESTIMATED, WHILE A SMALLER VARIANCE IS NOTICEABLE.

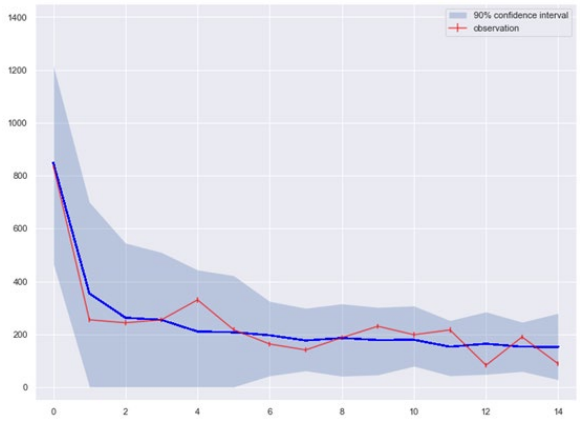


FIGURE 3-14 POSTERIOR OF ARPS DCA WITH FOR MODEL ERROR.

As explained earlier, RMSE and MAPE could not be correct measures to compare the result of updating, but as table 3-2 shows, RMSE and MAPE depict good results when the model error is considered in updating. For the case with accounting for model error, RMSE and MAPE decreased in comparison to the case without model error.

Since SEM and Arps DCA parameters are based on different definitions and equations, it is impossible to compare them, so parameters of initial realizations and parameters updated using EnKF with and without accounting for model error are contrasted.

DCA type		RMSE	MAPE
SEM	Prior	186	3.05
	Posterior	162.1	1.64
Arps	Prior	222.9	2.68
	Without Model Error	133.2	4.22
	With Model Error	44.7	0.24

TABLE 3-2 FITTING MEASURE FOR DCA

From table 3-3, Q_i and D_i have a wide truncated normal distribution for initial realization. The forecasted update without and with model error has a normal distribution. Updated realizations accounting for model error have a higher mean and slightly wider distribution for Q_i and D_i. However, for parameter b, narrower distribution is noticeable when the model error is considered. Finally, for Q_t, the shape of the PDF changed from a long-tailed distribution to log-normal or truncated normal distribution. Based on these probability distributions and PDF plots, wider distribution and shift in the mean of each parameter were observed. These changes in the distribution of the parameters showed that accounting for model error can avoid overconfidence. Also, in figure 3-14, biases of the Arps model has been mitigated.

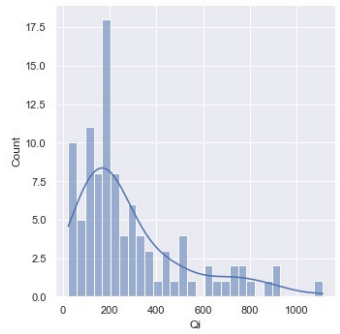
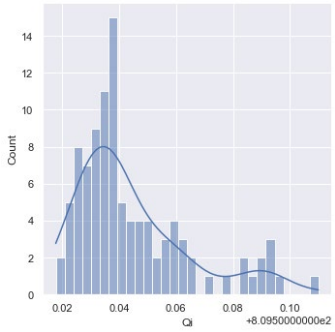
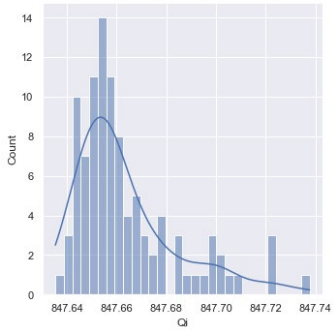
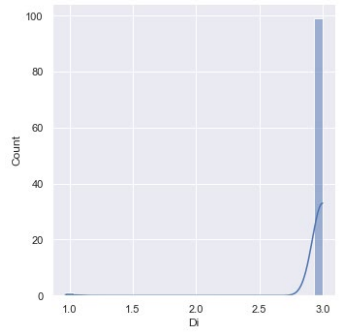
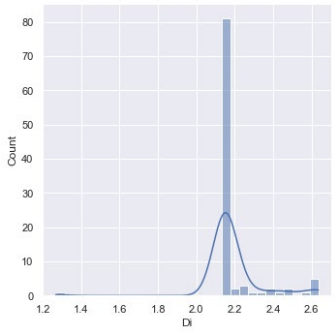
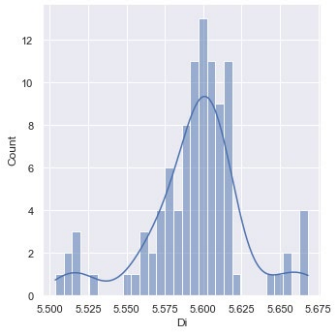
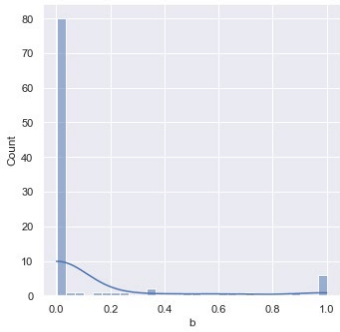
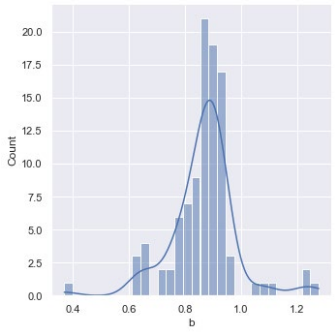
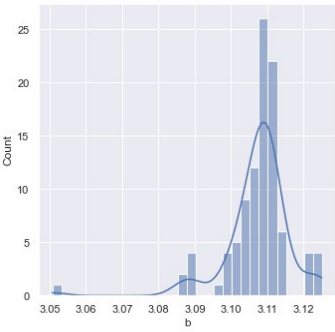
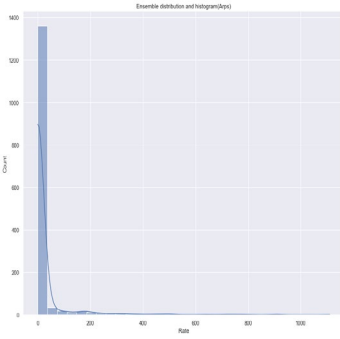
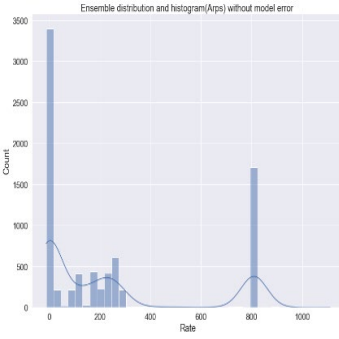
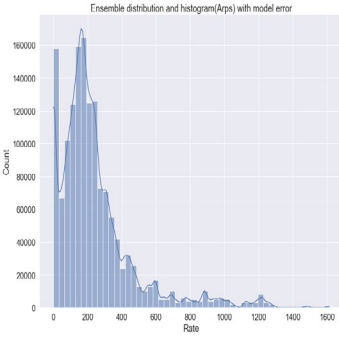
	Initial of Arps Parameters	Updated Arps Parameters without model errors[1]	Updated Arps Parameters with model error
Qi			
Di			
b			
Qt			

TABLE 3-3 HISTOGRAMS OF ARPS PARAMETERS AND QT FOR PRIOR AND POSTERIOR

[1] Note that the scale of this plot is different from other plots due to limitation of Python package which have been used.

3.3 Case Study 2: 2D Reservoir

A synthetic 2D reservoir simulation model with 200 × 200 grid cells was built with a geostatistical model for the permeability field (the only uncertain parameters in this example) with log-normal distribution with the mean 190 mD, standard deviation 300, and spherical variogram (major axis = 3000 minor axis = 1000 and the nugget effect 0.01) with the minimum 100 mD and maximum 900 mD as the true field as shown in figure 3-16. Porosity, initial pressure, and initial water saturation were 20%, 290 bar, and 10 percent. Incompressible two-phase flow of oil and water is considered in porous media. The dimension of the reservoir is 1000 m × 1000 m × 10 m in the x, y, and z directions, respectively (figure 3-16).

A production well is placed in the southeast, and an injector well is in the northwest of the field. Production and water injection have been started in 1970 and continued until 2020. The measurements data include water cut at the production well (WCT) and bottom-hole pressure (PBHP & IBHP) for both wells have been collected at the end of 1970, 1980, 1990, 1995, 2000, 2010, 2015, 2018, and 2020. The liquid rates remained constant during production (figure 3-17).

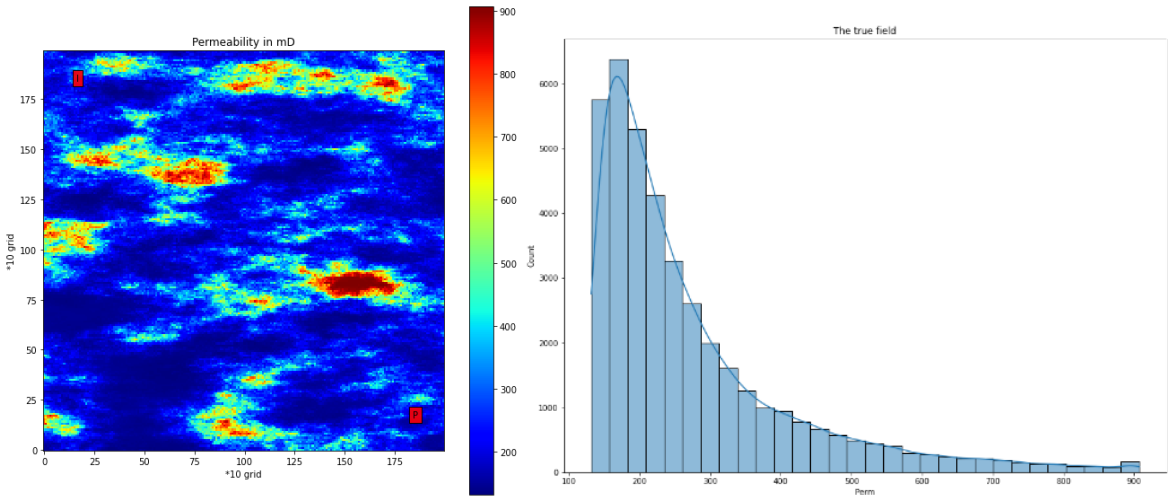


FIGURE 3-12 TRUE FIELD AND HISTOGRAM OF THE FIELD

101 realizations were randomly drawn from the geostatistical model for the prior ensemble, and one realization has been selected and separated as truth. Each realization has 200 × 200 grid cells.

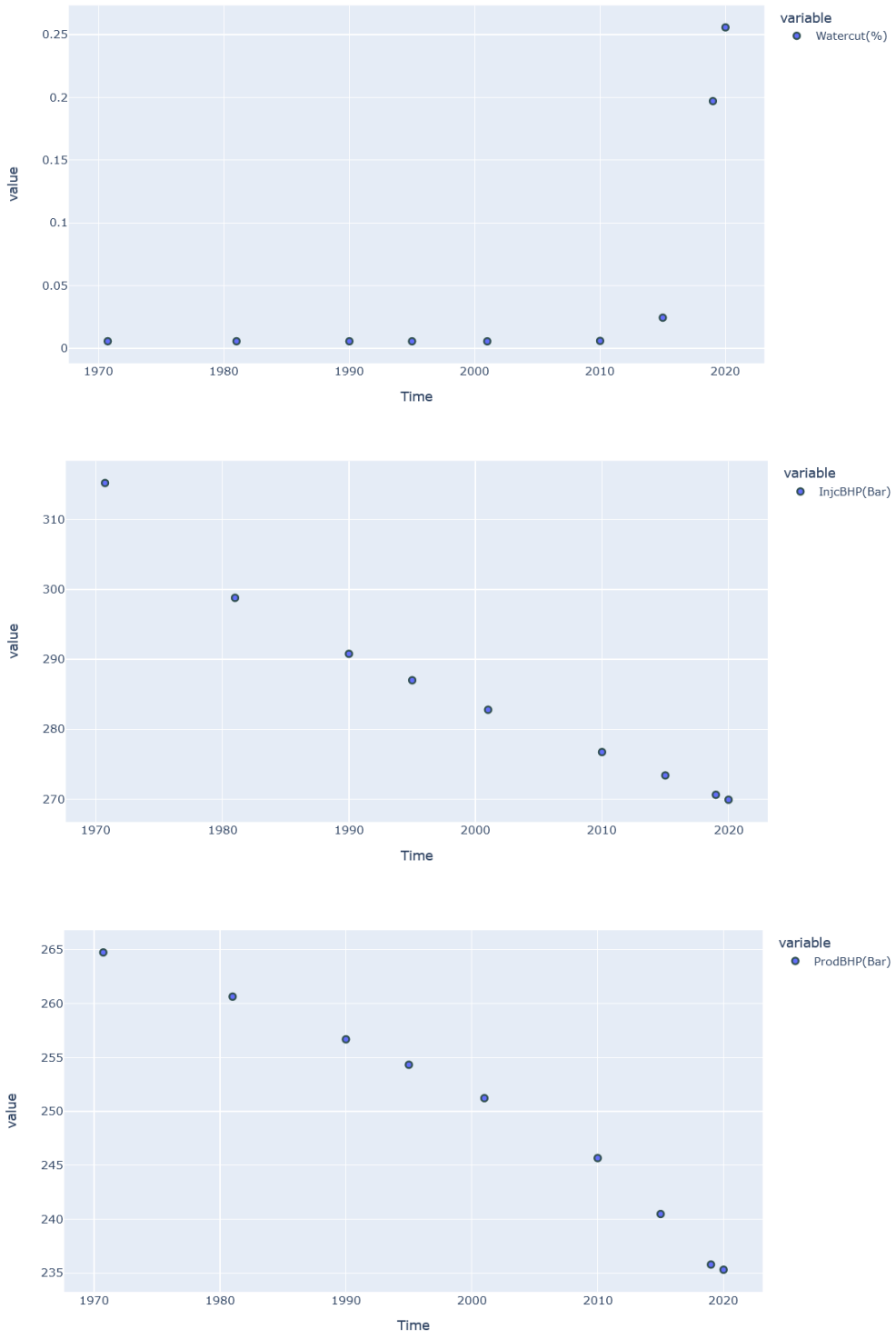


FIGURE 3-13 OBSERVATION DATA FROM THE TRUE MODEL (WCT, IBHP, IBHP)

In this case study, the model error is due to grid upscaling. Upscaling or grid coarsening is a routine part of simulating a reservoir to reduce computation costs. So, the prior realizations will be upscaled from high fidelity(200 × 200 grid cells) to (100 × 100) and lower fidelity, (50 × 50), (25 × 25), and (5 × 5) grid cells and at the end, five ensembles with different grid numbers created. Ensembles with 200 × 200 and 100 × 100 cells were considered high-quality prior models. By upscaling to different levels of fidelity, the mean of the ensemble was almost the same value, and the standard deviation of the field changed, and this process caused the error in the upscaled model. In table 3-4, the realization No.1 and the mean and STD of all realizations for five different ensembles are demonstrated after the upscaling process. In this study, log-permeability has been used, which is noticeable from the histogram/PDF plots of the high-fidelity field.

From table 3-4, clearly, the upscaling process makes the permeability more homogenous. Coarsening does not change the mean of ensembles at different levels. The mean of each ensemble is around 270, while STDs are 183, 90.6, 46.5, 29, and 20 for 100×100, 50×50, 25×25, and 5×5 cells. With different grid upscaling levels, the heterogeneity of realizations has been reduced. So, it can be said that grid upscaling is not biasing the prior with respect to the mean.

These ensembles have been simulated with similar initial conditions except well location grid. As explained earlier, the geometry and location of a well would change in upscaling. To minimize the impact of this error in this study, the location of both production and injection wells was designed to have fix position corresponding to the boundary of the reservoir.

Since BHP plots for production and injection wells are almost straight lines and there are no significant changes, they will not be demonstrated in this chapter, but note that BHP data have been used in updating step.

In table 3-5, as expected, simulated initial ensembles have a wider spread with fine grid size and more heterogeneity in the model. As grid size increases with different levels of upscaling, grids' permeability converges to the ensemble's mean. So, by simulating these ensembles, we will see narrow distributions.

Even though permeability data were not biased with grid upscaling, the production data became biased by simulating those upscaled ensembles. In one realization, coarsening makes the field property more homogenous. Homogeneity makes the fluid flow smoother and faster, so

water breaks through the reservoir faster, and we see higher water cuts. This idea explains the biasness of grid coarsening.

EnKF was implemented to update the prior models with field observations. Note that, in each iteration, dynamic properties (pressure and saturation) were set to initial conditions, and the model has been simulated from the beginning of production time. Also, permeability is truncated to 100 to 900 mD.

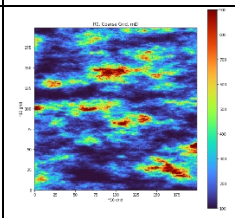
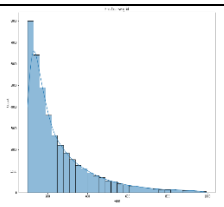
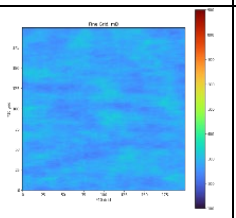
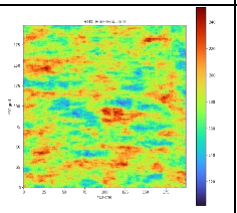
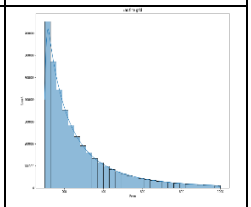
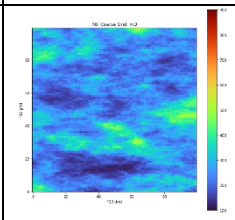
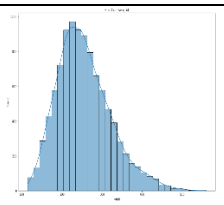
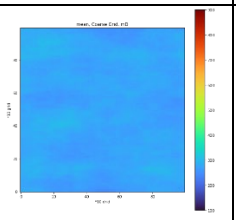
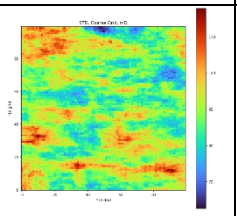
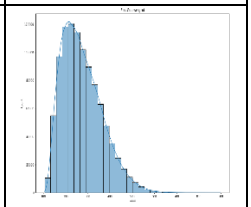
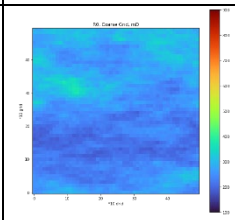
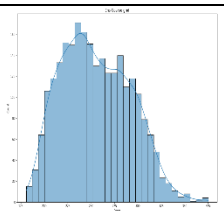
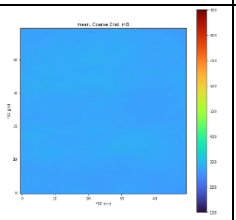
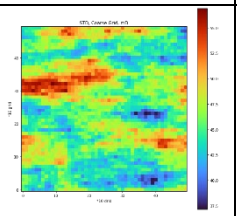
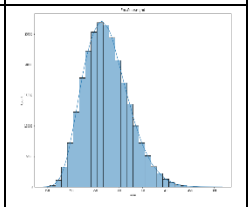
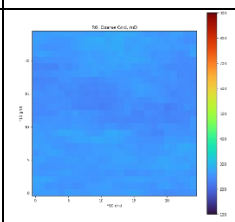
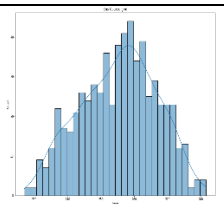
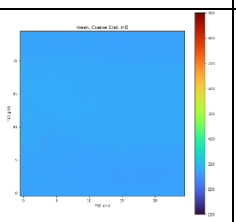
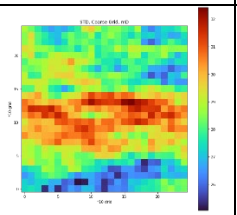
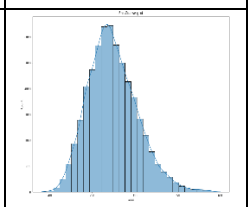
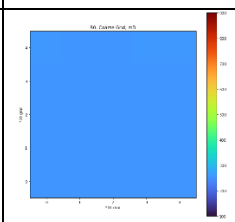
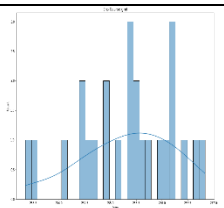
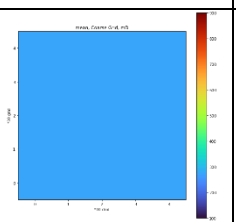
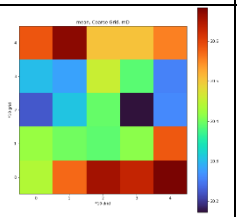
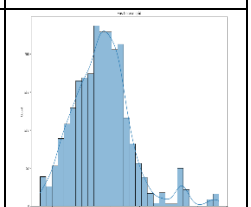
	Permeability of realization No.0	Histogram of realization No.0	Mean of the ensemble	Standard Deviation of the ensemble	Histogram of the ensemble
200 × 200					
100 × 100					
50 ×50					
25 × 25					
5 × 5					

TABLE 3-1 DIFFERENT UPSCALING LEVELS AND MEAN FOR EACH LEVEL OF UPSCALING

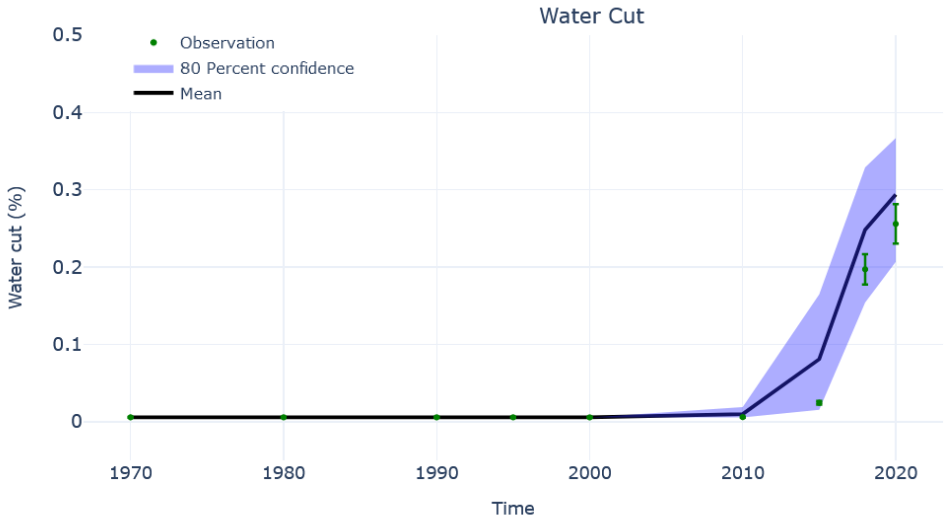
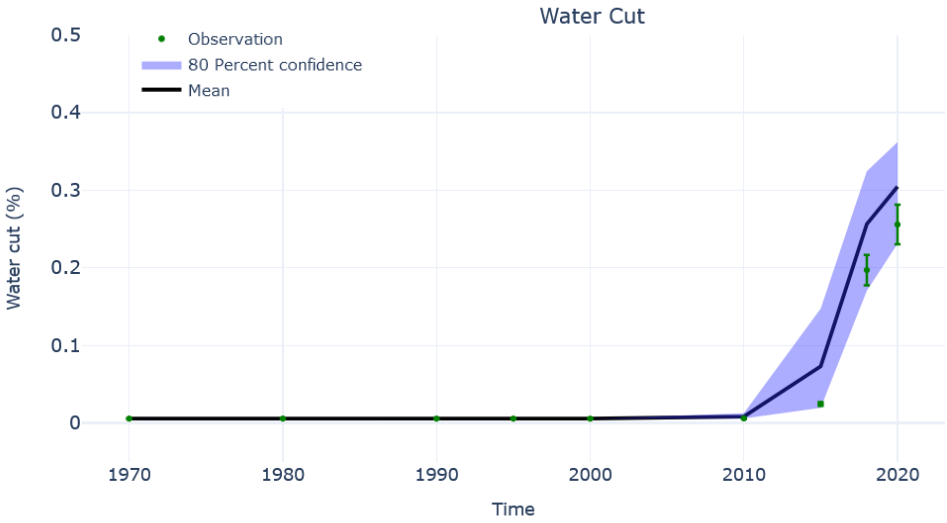
Dim	Water Cut (%)	Measures
200 × 200		RMSE WCT:0.03 P BHP:16.83 I BHP:1.35 MAPE WCT:0.02 P BHP:13.53 I BHP:1.08
100 × 100		RMSE WCT:0.03 P BHP:9.91 I BHP:1.76 MAPE WCT:0.02 P BHP:8.15 I BHP:1.51

TABLE 3-2 FORECASTED HIGH-QUALITY PRIOR MODELS

Model Error of upscaling and statistical information:

Now proposed algorithm has been used to calculate the model error and corresponding statistical data of model error. Various combinations of high-quality and low-quality models arrangements have been tested, as shown in table 3-6. For example, the number '200' indicates the high-

quality model with 200×200, and the number '5' indicates the low-quality model with 5×5 grid cells.

HQ Model	LQ Model	HQ Model	LQ Model	HQ Model	LQ Model	HQ Model	LQ Model
200	50						
200	25	100	25				
200	5	100	5	50	5	25	5

TABLE 3-3 COMBINATIONS OF HIGH-QUALITY AND LOW-QUALITY MODELS ARRANGEMENTS

Figure 3-18 Shows the Model Error function and its mean and STD for Model Error between two models with 200×200 and 5×5 grid cells. In the left plot, each point shows the difference between high and low quality models. Straight lines show linear interpolation between the statistical parameters in the right plot.

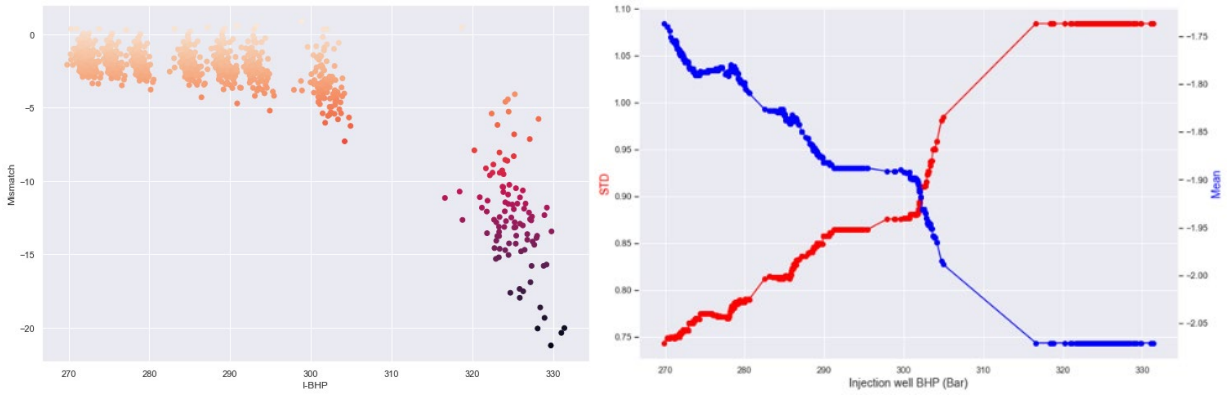


FIGURE 3-14 MODEL ERROR(LEFT) & STATISTICAL PARAMETERS OF MODEL ERROR(RIGHT)

For realizations with 5×5 grid cells, the prior ensemble has been simulated and then updated for observation data as shown in table 3-4 and 3-5. As expected, the forecast distribution has shrunk after updating this ensemble, and we probably faced ensemble collapse. However, since, in this case, the initial realization was not good prior, the forecasted WCT is still far from the observation. The distribution of updated permeability is slightly changed to a more normal distribution. The ensemble collapse can be recognized when the updated ensemble has a close value in the forecasted model.

We see a broader P80 confidence band and a slight shift toward observation by considering model error with the ensemble of 200×200 grid cells as the HQ model and 5×5 grid cells as the LQ model. The wider confidence band is due to model error and shows that if we use a poor model to describe the reservoir behavior, how much uncertainty could be neglected in the process. Also, with this approach, we have mitigated the biasness of the LQ-model to some degree, the confidence band covers some observation, and the mean of the forecast is closer to observation data compared to the case without considering the model error.

Tables 3-8 and 3-9 show forecasted models and the distributions of updated models without and with model error (different levels of model error). As expected, EnKF shrank the confidence band while forecasted realizations are still biased compared to the observation. Considering model error expanded the confidence interval band, and in the case with 200×200 grid cells, some realizations are covering the observation points. RMSE and MAPE are bigger since, with model error, a broader confidence interval is obtained. Updated permeability field plots are still highly homogeneous, and the histogram of permeability became narrower with the decrease in quality of model error. The term model error quality refers to which upscaling levels have been chosen as HQ-model. If, instead of 200×200 grid cells, we use 50×50, the quality of the model would decrease.

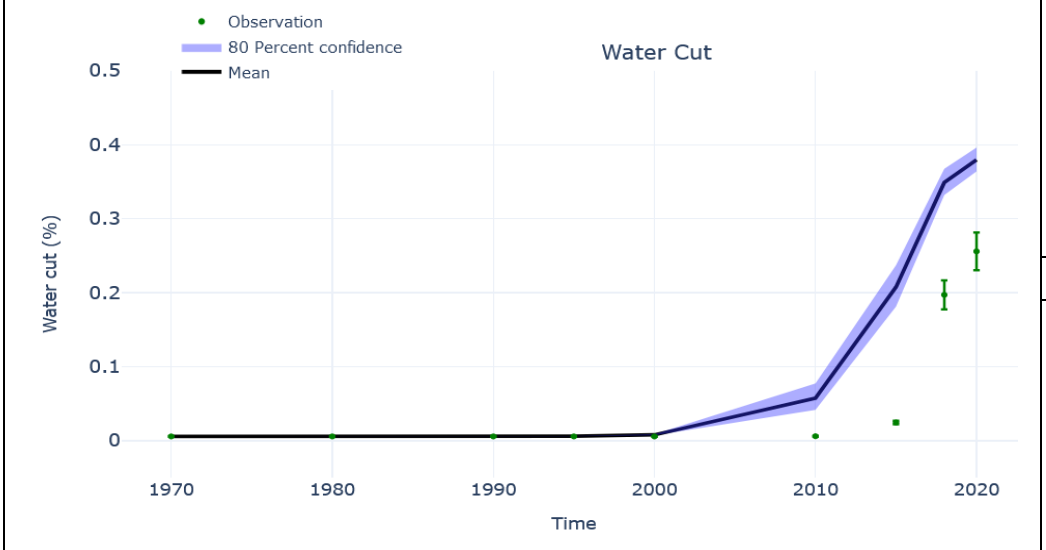
Forecasted initial ensemble with 5×5 cells	Measures
	RMSE WCT:0.02 P BHP:7.6 I BHP:1.49 MAPE WCT:0.02 P BHP:6.76 I BHP:1.27

TABLE 3-4 WCT FOR PRIOR ENSEMBLE WITH 5×5 CELLS

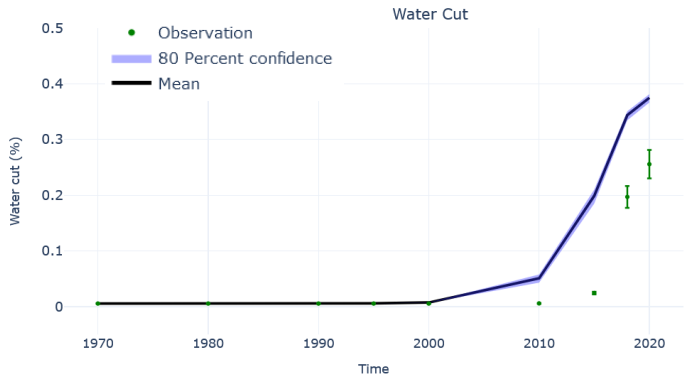
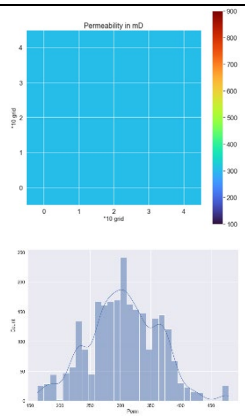
Forecasted updated with EnKF ensemble 5×5 cells	Plot & Histogram	Measures
		RMSE WCT:0.04 P BHP:8.24 I BHP:9.46 MAPE WCT:0.03 P BHP:5.06 I BHP:8.51 Run time: 6694 s

TABLE 3-5 FORECASTED POSTERIOR WITH SIZE 5×5 CELLS

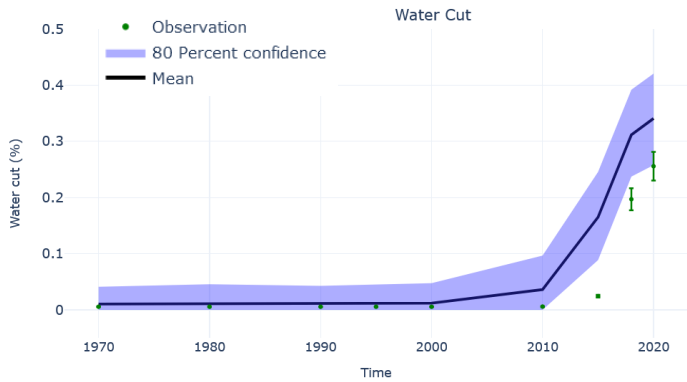
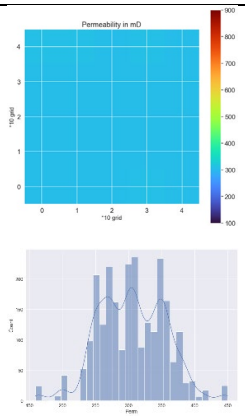
Update accounting for Model Error ensemble 5×5 (200×200)	Plot & Histogram	Measures
		RMSE WCT:0.062 P BHP:13.7 I BHP:8.55 MAPE WCT:0.05 P BHP:10.86 I BHP:7.4 Run time: 6730 s

TABLE 3-6 FORECASTED POSTERIOR WITH SIZE 5×5 CELLS ACCOUNTING FOR MODEL ERROR FROM HQ MODEL (200×200). THE MEAN OF REALIZATIONS AND HISTOGRAM.

In similar steps, the model error between the ensembles with 100×100, 50×50, and 25×25 grid cells and the same low-quality model (5×5 grid cells) were calculated and considered in updating the model. As the quality of model error decreases, the forecasted WCT becomes more biased and shows a smaller standard deviation. As shown in tables 3-10 to 3-3-12, the forecasted WCT shows more biases and a narrower confidence band with the reduction in the quality of model error. From the histogram\PDF plot of permeability, we can see the permeability variance decreased with reduced model error quality. The plot of permeability shows that permeability is still almost homogenous after updating, as shown in the second column of table 3-6.

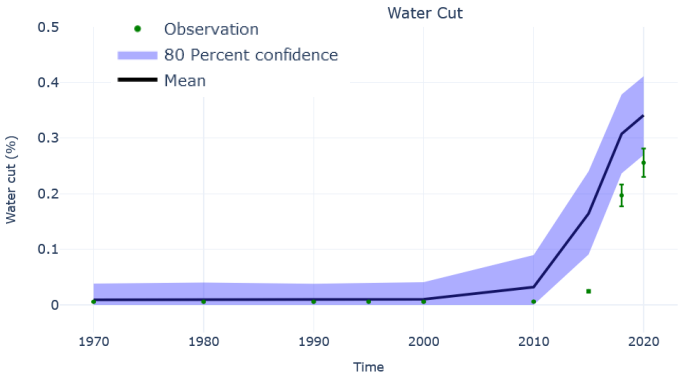
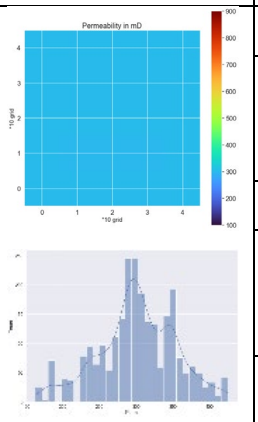
Update accounting for Model Error ensemble 5×5 (100×100)	Plot & Histogram	Measures
		RMSE WCT:0.06 P BHP:10 I BHP:2.36 MAPE WCT:0.48 P BHP:8.24 I BHP:1.95 Run time: 4961 s

TABLE 3-7 FORECASTED POSTERIOR WITH SIZE 5×5 CELLS ACCOUNTING FOR MODEL ERROR FROM HQ MODEL (100×100). THE MEAN OF REALIZATIONS AND HISTOGRAM.

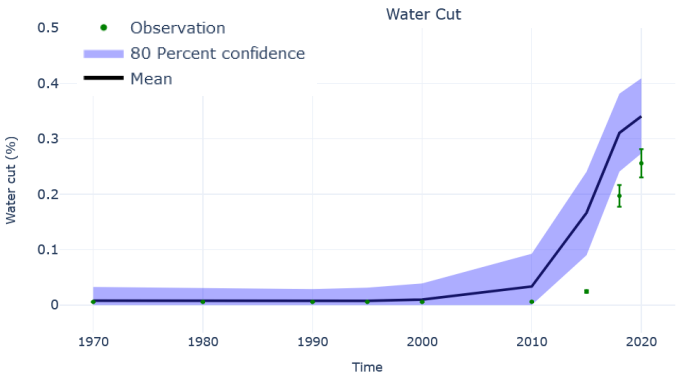
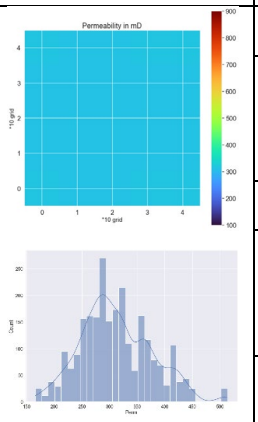
Update accounting for Model Error ensemble 5×5 (50×50)	Plot & Histogram	Measures
		RMSE WCT:0.056 P BHP:8.2 I BHP:2.75 MAPE WCT:0.048 P BHP:6.6 I BHP:2.26 Run time: 4946 s

TABLE 3-8 FORECASTED POSTERIOR WITH SIZE 5×5 CELLS ACCOUNTING FOR MODEL ERROR FROM LQ MODEL (50×50). THE MEAN OF REALIZATIONS AND HISTOGRAM.

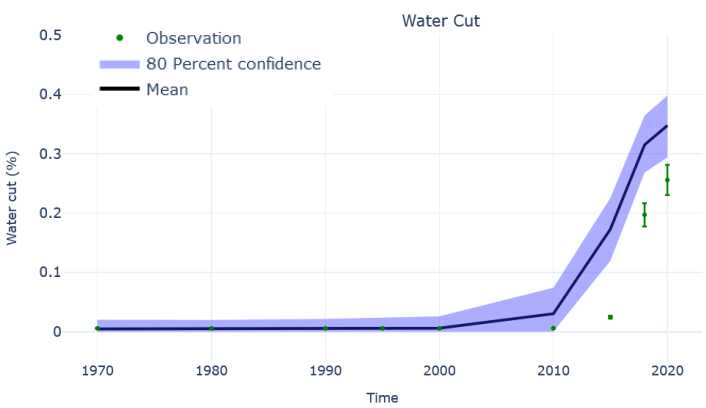
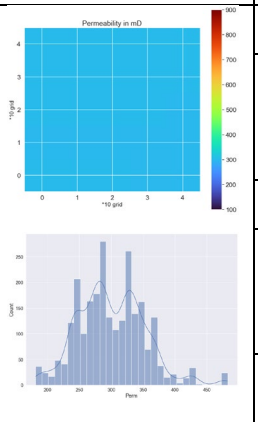
Update accounting for Model Error ensemble 5×5 (25×25)	Plot & Histogram	Measures
		RMSE WCT:0.053 P-BHP:6.7 I-BHP:2.6 MAPE WCT:0.05 P-BHP:5.43 I-BHP:2.15 Run time: 4937 s

TABLE 3-9 FORECASTED POSTERIOR WITH SIZE 5×5 CELLS ACCOUNTING FOR MODEL ERROR FROM LQ MODEL (25×25). THE MEAN OF REALIZATIONS AND HISTOGRAM.

We repeated the same procedure for the ensemble with 25× 25 grids; the results are shown in tables 3-13 to 3-17. The forecasted model (WCT) covers the observation in the initial simulation since the prior ensemble has less error than the case with 5×5 grid cells but is slightly biased. After updating, we see that the mean of realizations is closer to observation and has a smaller variance.

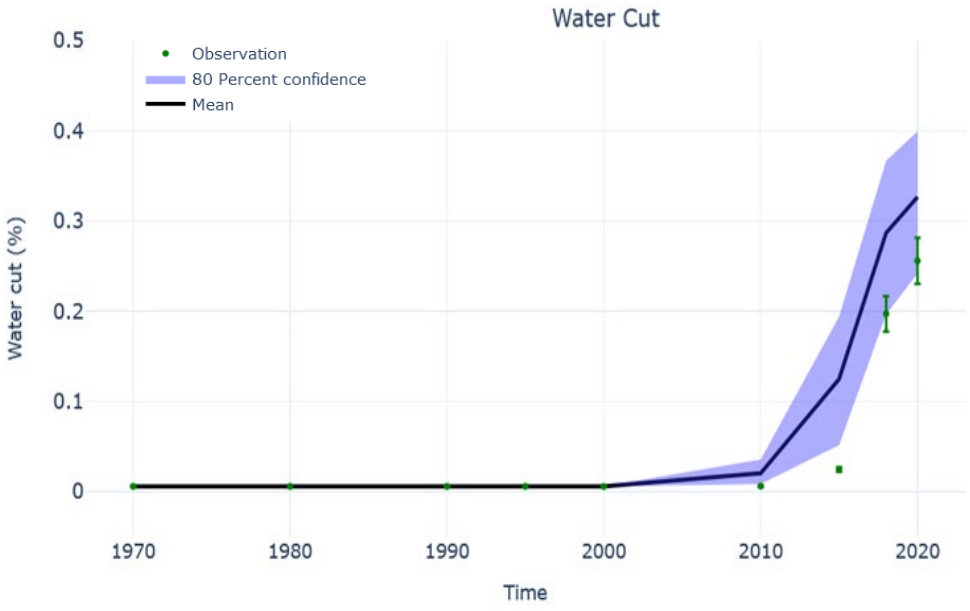
Forecasted initial ensemble of 25×25 cells	Measures
	RMSE WCT:0.04 P BHP:6.12 I BHP:7.3 MAPE WCT:0.03 P BHP:4.8 I BHP:6

TABLE 3-10 WCT FOR PRIOR ENSEMBLE WITH 25×25 CELLS.

The updated permeability field plots in tables 3-14 to 3-17 show two high permeability areas in the north and south. In the EnKF, these high permeable areas are large, while in the case of using 200× 200 grid cells as the HQ model, the high permeability zone at the south is faded while there is a higher permeability at the northeast zone. Other updated permeability plots show similar patterns. The histograms and PDFs are not Gaussian after considering the model error, and the positive skew is noticeable. In the histograms, the first bin on the left and the last bin on the right side of the histogram are due to truncating permeability for less than 100 mD and more than 900 mD.

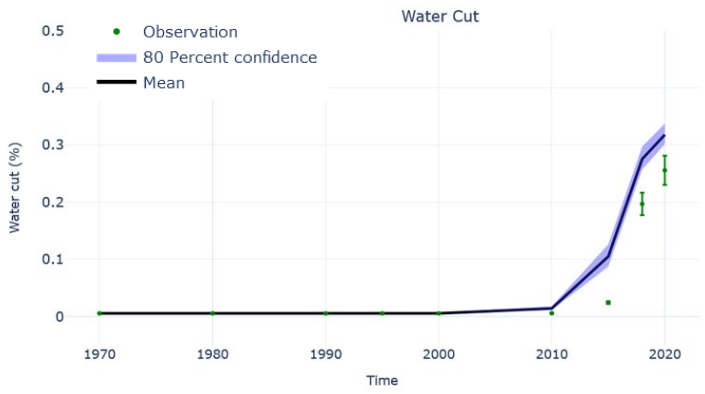
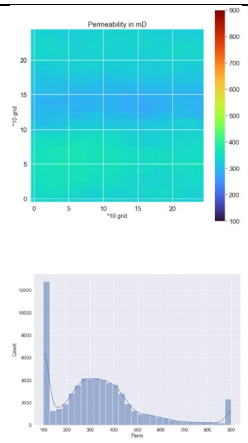
Forecasted update with EnKF ensemble 25×25 cells	Plot & Histogram	Measures
		RMSE WCT:0.03 P BHP:4.8 I BHP:1.2 MAPE WCT:0.03 P BHP:4.2 I BHP:1 Run time: 5010 s

TABLE 3-11 FORECASTED POSTERIOR WITH SIZE 25×25 CELLS. THE MEAN OF REALIZATIONS AND HISTOGRAM.

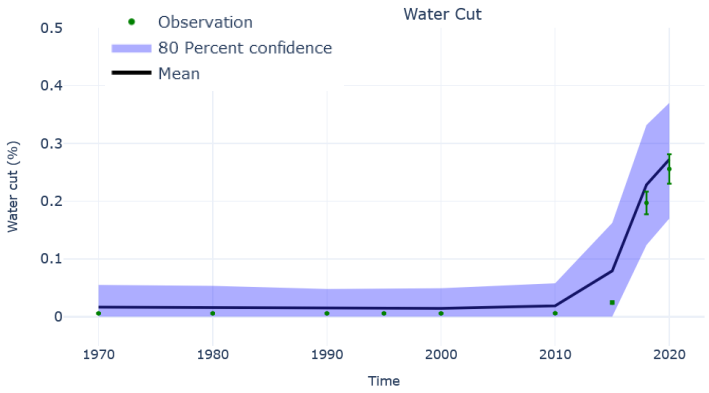
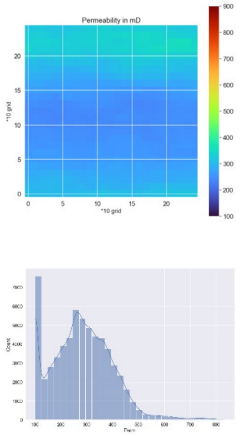
Update accounting for Model Error ensemble 25×25 (200×200)	Plot & Histogram	Measures
		RMSE WCT:0.0441 P BHP:15.7 I BHP:5.8 MAPE WCT:0.032 P BHP:12.63 I BHP:4.54 Run time: 4978 s

TABLE 3-12 FORECASTED POSTERIOR WITH SIZE 25×25 CELLS ACCOUNTING FOR MODEL ERROR FROM HQ MODEL (200×200). THE MEAN OF REALIZATIONS AND HISTOGRAM.

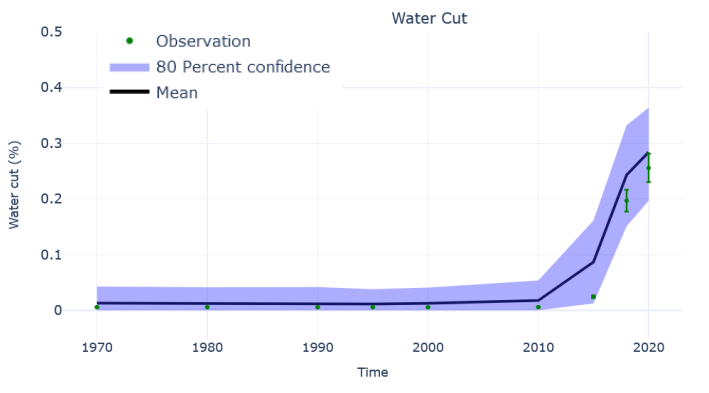
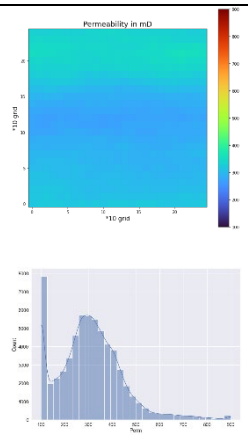
Update accounting for Model Error ensemble 25×25 (100×100)	Plot & Histogram	Measures
		RMSE WCT:0.0405 P BHP:12.07 I BHP:6 MAPE WCT:0.030 P BHP:8.6 I BHP:4.6 Run time: 4988 s

TABLE 3-13 FORECASTED POSTERIOR WITH SIZE 25×25 CELLS ACCOUNTING FOR MODEL ERROR FROM HQ MODEL (100×100). THE MEAN OF REALIZATIONS AND HISTOGRAM.

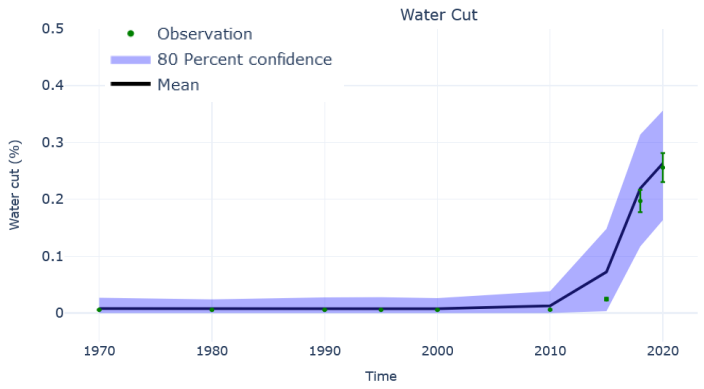
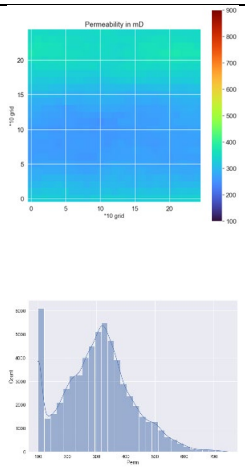
Update accounting for Model Error ensemble 25×25 (50×50)	Plot & Histogram	Measures
		RMSE WCT:0.034 P BHP:9.1 I BHP:5.4 MAPE WCT:0.026 P BHP:6.6 I BHP:4.3 Run time: 4990 s

TABLE 3-14 FORECASTED POSTERIOR WITH SIZE 25×25 CELLS ACCOUNTING FOR MODEL ERROR FROM LQ MODEL (50×50). THE MEAN OF REALIZATIONS AND HISTOGRAM.

The last experiment repeated the same procedure for the ensemble of 50×50 grid cells as the LQ model. As shown in table 3-18, the forecasted prior model has a broad confidence band due to higher heterogeneity in the model. Since this ensemble has higher quality than the two previous cases, forecasted realizations cover the observed WCT in most of the observation points.

Since high-quality and low-quality models are similar, the model error has a smaller mean and standard deviation, so the uncertainty band is smaller than in previous cases.

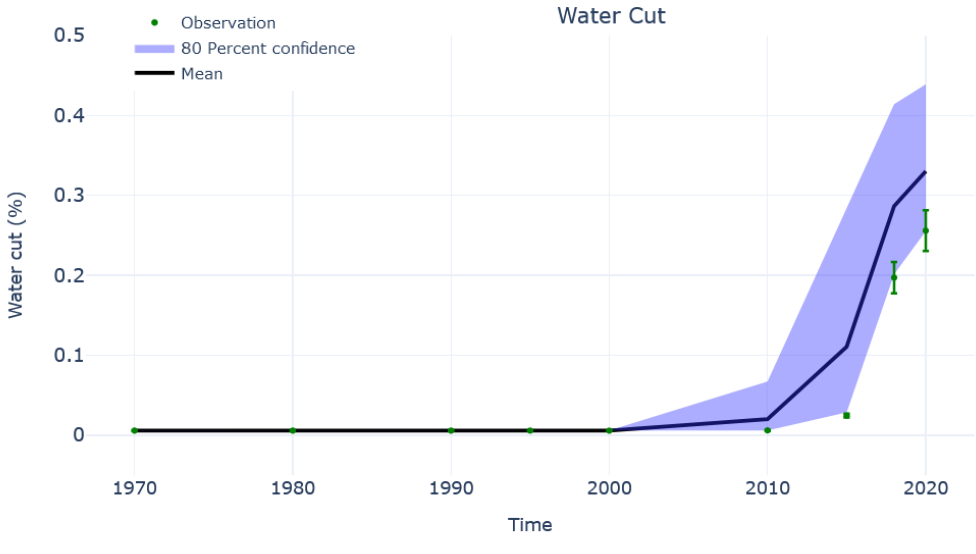
Forecasted initial ensemble 50×50 cells	Measures
	RMSE WCT:0.02 P BHP:7.6 I BHP:1.49 MAPE WCT:0.02 P BHP:6.76 I BHP:1.27

TABLE 3-15 WCT FOR PRIOR ENSEMBLE WITH 50×50 CELLS.

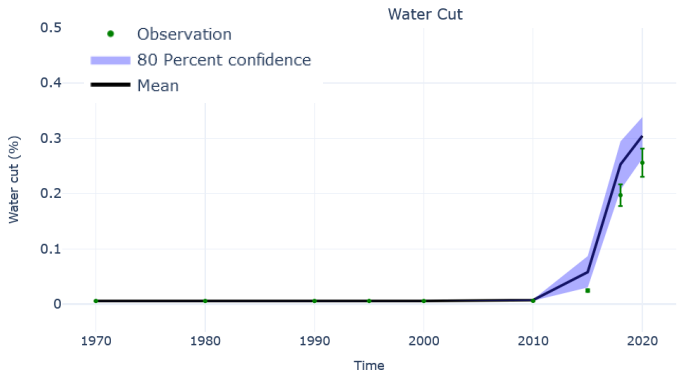
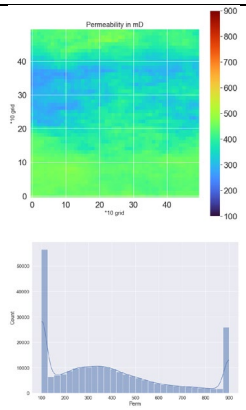
Update with EnKF ensemble 50×50	Plot & Histogram	Measures
		RMSE WCT:0.02 P BHP:7.6 I BHP:1.49 MAPE WCT:0.02 P BHP:6.76 I BHP:1.27 Run time: 6694 s

TABLE 3-16 FORECASTED POSTERIOR WITH SIZE 50×50 CELLS

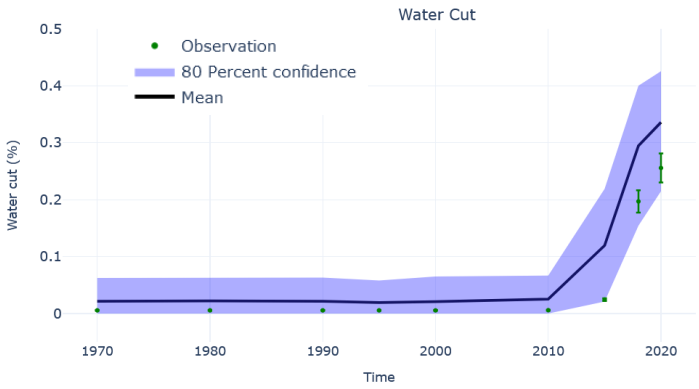
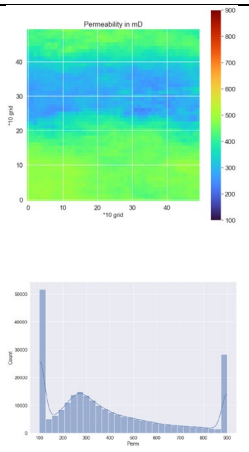
Update accounting for Model Error ensemble 50×50 (200×200)	Plot & Histogram	Measures
		RMSE WCT:0.062 P BHP:13.7 I BHP:8.55 MAPE WCT:0.05 P BHP:10.86 I BHP:7.4 Run time: 6730 s

TABLE 3-17 FORECASTED POSTERIOR WITH SIZE 50×50 CELLS ACCOUNTING FOR MODEL ERROR FROM HQ MODEL (200×200).

In the case with 50×50 grid cells, again, two high permeable zones at the north and south of the field; however, in this case, higher and broader high permeable are is at the south. Histogram and PDF of permeability with accounting for model error is close to a Gaussian distribution with a positive skew.

The True field has been upscaled compared with the updated permeability fields in table 3-21. With upscaling the true synthetic model to 100×100 and 50×50, one high permeability area in the north and another high permeability area in the southwest of the field appeared. However, in 25×25 grids, the high permeability area is almost in the middle south of the field, and for 5×5 grids, we have a uniform permeability in the reservoir. In updated ensembles of 50×50, a wide high permeability zone in the south and a smaller high permeability zone is noticed.

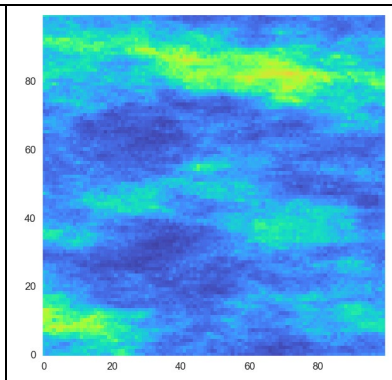
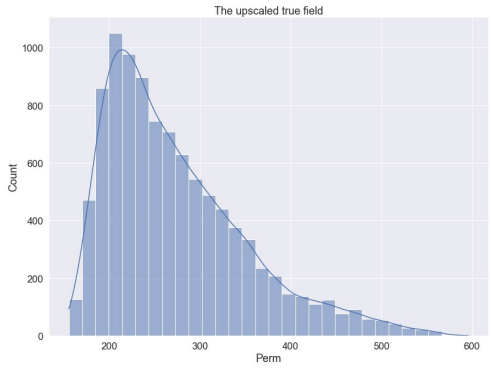
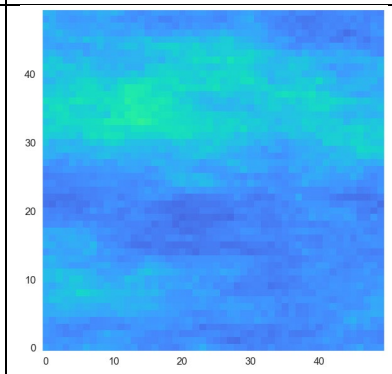
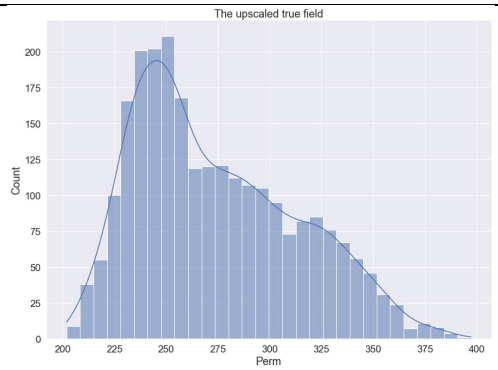
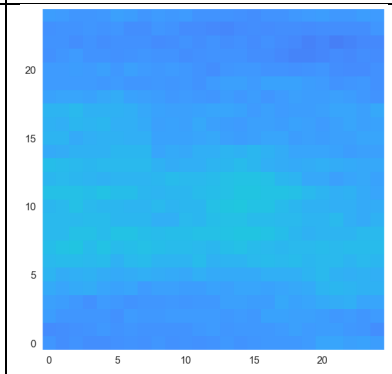
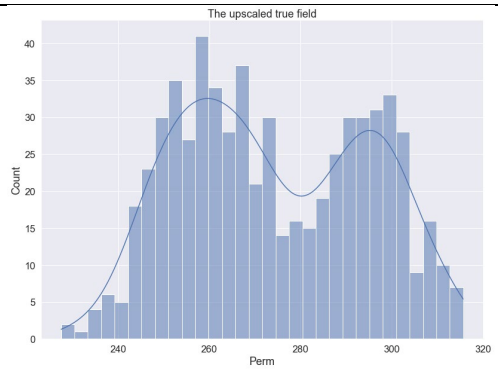
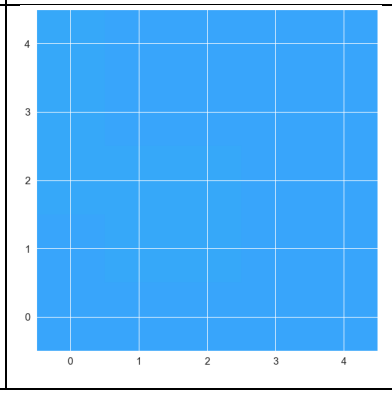
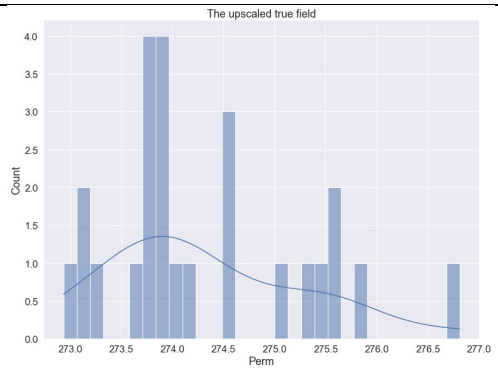
100×100		
50×50		
25×25		
5×5		

TABLE 3-18 UPSCALED TRUE MODEL

Running time:

All steps of this study have been coded in python language version 3.9 with Jupiter notebook IDE. Eclipse from Schlumberger version 2014 was used as a forward model coupled with python to import data and run and export the static and dynamic outputs. A computer with CPU Intel Xeon E5-2609 at 2.4GHz with 64 GB ram has been used to run the code.

In figure 3-19, the average running time for each ensemble is shown, and obviously, with fewer grid cells running time decreases. As discussed earlier, upscaling is scarifying the quality of the model to reduce the computational costs, which is a massive burden in EnKF.

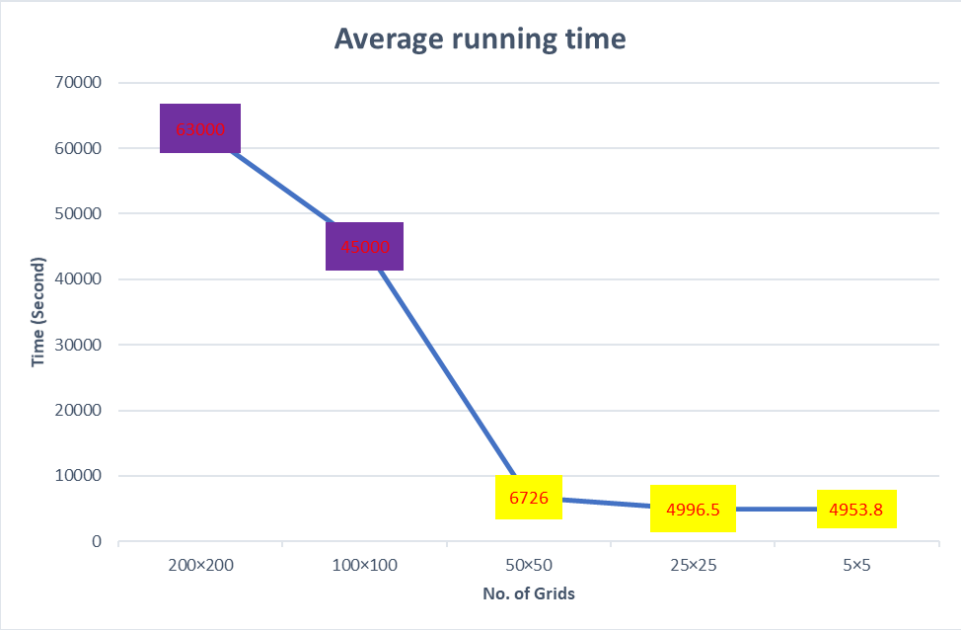


FIGURE 3-15 AVERAGE RUNNING TIME ACCOUNTING FOR MODEL ERROR VS. GRID NUMBER. (POINTS IN PURPLE COLOR ARE APPROXIMATED)

Chapter 4 Conclusion and recommendations:

Ensemble-based data assimilation has been applied in different industries, and recently more companies and institutes are implementing it in their modeling, optimization, and decision analyses process(Hanea et al., 2015).

Ensemble-based data assimilation provides a framework to update and calibrate the uncertainties with additional observations in a probabilistic and Bayesian manner

Classically, model error in DA has been defined as the discrepancy between the truth and the model; however, we can never calculate model error directly because the truth is never known.

The model error has various sources that affect the updated model's outcome. This study proposes a workflow to represent model error for grid coarsening in reservoir simulation. The model error has been assumed as the mismatch between the forecast of the fine grid and coarse grid models. The discrepancy's mean and standard deviation was calculated as a function of the forecast of the coarse grid model, and parameters represent the model error in term of a probability distribution.

In a simple case study with DCA, assimilating the low-quality model showed biases and overconfidence results compared to the assimilated high-quality model. Any decision based on the low-quality model, which is biased and overconfident, could result in choosing an incompatible probability distribution to model the truth. With accounting for model error in assimilating the low-quality prior, the biasness of the forecasted model decreased, and the updated model was not overconfident.

For the second case study, ensembles of synthetic 2D permeability fields have been generated and forecasted with a reservoir simulator as a highly non-linear forward model. The process of upscaling as the source of model error was studied. It has been explained that grid upscaling decreased heterogeneity or the standard deviation of the ensemble while the mean was almost constant. Forecasting models with different grid upscaling levels showed overconfident and biased results for low-quality models when the model error is not considered in EnKF updating.

This study showed that model error is an inherent part of modeling that affects DA outcome, and the model error should be considered in DA for debiasing probabilistic forecasts. Using a prior model with ignoring model error will result in biasness and overconfidence. The 2D reservoir model example in this study has clarified that upscaling caused less heterogeneity in the permeability and consequently caused more uniform fluid flow resulting in faster water breakthrough. Whether upscaling will cause to under or overestimation in oil production should be studied case by case. Also, upscaling or other simplifications that reduce heterogeneity make a prior ensemble's forecasts overconfident, and updating overconfident results in a narrower distribution of posterior or even ensemble collapse.

This study can be extended to other sources of error such as wellbore geometry, PVT modeling, or combination to analyze the effect of the model error in DA. Also, the approach should be modified and developed for other ensemble-based algorithms such as Ensemble Smoother with Multiple Data Assimilation.

A recommendation for future works is as the following. To have better results, it might be better to use artificial intelligence algorithms or develop a new algorithm to assess the statistical parameters of model error as a function of other feathers (e.g., geological properties) than the features (BHP, WCT, or production rate) used in this thesis work.

An important observation from this study was the correction of biasness in assimilating low-quality models, so in the case of using artificial intelligence to assess statistics of model error and predict model error for unseen observation. Then one might be able to build a general model error function with real data from several subsurface reservoirs and utilize it to account for model error. Additionally, the generalized model error function could allow us to assimilate low-quality models and reduce the computation cost of simulations in decision analyses.

References

Aanonsen, S. I., G. Nœvdal, D. S. Oliver, A. C. Reynolds, and B. Vallès, 2009, The ensemble Kalman filter in reservoir engineering-A Review: SPE Journal, v. 14, no. 3, p. 393–412, doi:10.2118/117274-PA.

Guide for the Verification and Validation of Computational Fluid Dynamics Simulations, 1998: accessed May 10, 2022, https://arc.aiaa.org/doi/book/10.2514/4.472855.

Alerstam, E., T. Svensson, and S. Andersson-Engels, 2008, Parallel computing with graphics processing units for high-speed Monte Carlo simulation of photon migration: Journal of Biomedical Optics, v. 13, no. 6, p. 060504, doi:10.1117/1.3041496.

Birnbaum, A., 1962, On the Foundations of Statistical Inference: Journal of the American Statistical Association, v. 57, no. 298, p. 269–306, doi:10.1080/01621459.1962.10480660.

Bratvold, R. B., S. Begg, and Society of Petroleum Engineers (U.S.), 2010, Making good decisions: Richardson, TX, Society of Petroleum Engineers.

Carrassi, A., and S. Vannitsem, 2016, Deterministic Treatment of Model Error in Geophysical Data Assimilation, *in* Mathematical Paradigms of Climate Science: Springer International Publishing, p. 175–213, doi:10.1007/978-3-319-39092-5_9.

Coats, K. H., 2000, A Note on IMPES and Some IMPES-Based Simulation Models: SPE Journal, v. 5, no. 03, p. 245–251, doi:10.2118/65092-PA.

Craft, B. C., and M. F. Hawkins, 1991, History Matching, second ed. (revised), *in* Applied Petroleum Reservoir Engineering: USA, Prentice Hall PTR.

Elvaretta, S., 2021, Value of Information Analysis in CO2 Sequestration Projects, Master thesis: University of Stavanger, Stavanger, Norway, 62 p.

Evensen, G., 2009, Data Assimilation:The Ensemble Kalman Filter, Second ed. (revised): Dordrecht Heidelberg London New York, Springer, 1–291 p.

Gruenwalder, M., S. Poellitzer, and T. Clemens, 2007, Assisted and manual history matching of a reservoir with 120 wells, 58 years production history and multiple well re-completions, *in* 69th European Association of Geoscientists and Engineers Conference and Exhibition 2007: Securing The Future. Incorporating SPE EUROPEC 2007: p. 965–976, doi:10.2523/106039-ms.

Haefner, J. W., 2005, Modeling biological systems: principles and applications, 2nd ed ed. (revised): New York, Springer, 475 p.

Hanea, R., G. Evensen, L. Hustoft, T. Ek, A. Chitu, and F. Wilschut, 2015, Reservoir management under geological uncertainty using fast model update: Society of Petroleum Engineers - SPE Reservoir Simulation Symposium 2015, v. 3, p. 1912.

Harlim, J., 2017, Model Error in Data Assimilation, *in* Nonlinear and Stochastic Climate Dynamics: Cambridge University Press, p. 276–317, doi:10.1017/9781316339251.011.

Jiyuan, T., Y. Guan-Heng, and L. Chaoqun, 2013, CFD Solution Analysis—Essentials, Second Edition ed. (revised), *in* Computational Fluid Dynamics,A Practical Approach: Butterworth-Heinemann, p. 177–217.

Jones, E., T. Oilphabt, P. Peterson, and others, 2001, SciPy: Open Source Scientific Tools for Python.

Julier, S. J., and J. K. Uhlmann, 2004, Unscented filtering and nonlinear estimation: Proceedings of the IEEE, v. 92, no. 3, p. 401–422, doi:10.1109/JPROC.2003.823141.

Kalman, R. E., 1960, A New Approach to Linear Filtering and Prediction Problems: Transactions of the ASME--Journal of Basic Engineering, v. 82, no. Series D, p. 35--45.

Lie, K.-A., 2019, An Introduction to Reservoir Simulation Using MATLAB/GNU Octave: User Guide for the MATLAB Reservoir Simulation Toolbox (MRST): Cambridge, Cambridge University Press, doi:10.1017/9781108591416.

Lu, C., C. Xiaoqian, and X. Bing, 2021, Cubature predictive filter, *in* Predictive Filtering for Microsatellite Control System: Academic Press, p. 141–164.

Maley, S., 1985, The Use of Conventional Decline Curve Analysis in Tight Gas Well Applications, *in* SPE/DOE Low Permeability Gas Reservoirs Symposium: OnePetro, doi:10.2118/13898-MS.

Qi, D., and T. Hesketh, 2005, An Analysis of Upscaling Techniques for Reservoir Simulation: Petroleum Science and Technology, v. 23, no. 7–8, p. 827–842, doi:10.1081/LFT-200033132.

Rammay, M. H., 2020, Calibration and prediction improvement of imperfect subsurface flow models, PhD thesis: Heriot Watt University, Edinburgh, 179 p.

Rasmussen, A. F. et al., 2019, The Open Porous Media Flow Reservoir Simulator, arXiv:1910.06059: arXiv, doi:10.48550/arXiv.1910.06059.

Schlumberger, 2014, Eclipse reservoir simulation software v2014. Technical Description Manual (2014): Schlumberger.

Schmidt, H., M. Rothgangel, and D. Grube, 2015, Prior knowledge in recalling arguments in bioethical dilemmas: Frontiers in Psychology, v. 6.

Valkó, P. P., 2009, Assigning value to stimulation in the Barnett Shale: a simultaneous analysis of 7000 plus production hystories and well completion records, *in* All Days, SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas: SPE, p. SPE-119369-MS, doi:10.2118/119369-MS.

Valkó, P. P., and W. J. Lee, 2010, A Better Way to Forecast Production from Unconventional Gas Wells, *in* All Days, SPE Annual Technical Conference and Exhibition, Florence, Italy: SPE, p. SPE-134231-MS, doi:10.2118/134231-MS.

Appendix

Appendix A: Pseudocodes

Model Error Function

$Y|_{2D}$ = HQ-model – LQ-model : for each measurement vector of forecasted realizations

$X|_{2D}$ = LQ-model : measurement vector of forecasted realizations

Fx ← $X|_{2D}$ to $X|_{1D}$, Fy ← $Y|_{2D}$ to $Y|_{1D}$ Reducing dimension to 1D

for i^{th} in elements Fx:

Create a list to save item i^{th} in window

for j^{th} in length Fx: Iterate over each item to increase accuracy

If Fx[j] inside window:

save Fy[j] to Wi

end if

end For

μ[i] ← mean (Wi)

σ[i] ← Standard deviation (Wi)

save (i, μ [i] & σ [i])

Mean = $Function_1$ (i, μ[i])

STD = $Function_2$ (i, σ [i])

EnKF:

$$x_k^{if} \leftarrow \mathcal{M}(x_{k-1}^{ia}, m_{k-1}^{ia})$$

$$z_k^{if} \leftarrow \mathcal{G}(x_{k-1}^{ia}, m_{k-1}^{ia})$$

$$y_k^f \leftarrow \begin{bmatrix} m_{k-1}^{ia} & \dots & m_{k-1}^{Nea} \\ x_k^{if} & \dots & x_k^{Nef} \\ z_k^{if} & \dots & z_k^{Nef} \end{bmatrix}$$

$$P_k^f \leftarrow \frac{1}{N-1}[y_k^f - E(y_k^f)][y_k^f - E(y_k^f)]^T$$

$$d_k \leftarrow [d_j^{obs} + v_k^i, \dots, d_J^{obs} + v_k^{Ne}]$$

$$R_k \leftarrow \frac{1}{N-1} v_k {v_k}^T$$

$$\mathcal{H}_k \leftarrow \left[z_k^{if} \quad \dots \quad z_k^{Nef}\right](y_k^f)^{-1}$$

$$\mathcal{K}_k \leftarrow P_k^f \mathcal{H}_k^T (\mathcal{H}_k P_k^T \mathcal{H}_k^T + R_k)$$

$$x_{k+1}^{ia} \leftarrow x_k^{if} + \mathcal{K}_k (z_k^i - (\mathcal{H}_k y_k^f + \mathcal{N}(0, R)))$$

Modified EnKF to consider model error:

if k = 0: first iteration in EnKF

$z_0^{i^{initial}} \leftarrow \mathcal{G}(x_0^{i^a}, m_0^{i^a})$ $\mathcal{G}$ is observation operator

$\varepsilon_{ME} \sim \mathcal{N}(\mu_{ME}^{z_0^{i^{initial}}}, (\sigma_{ME}^{z_0^{i^{initial}}})^2)$

$z_0^{i^f} \leftarrow z_0^{i^{initial}} + \varepsilon_{mismatch}$

if K > 0:

$x_k^{i^f} \leftarrow \mathcal{M}(x_{k-1}^{i^a}, m_{k-1}^{i^a})$

$z_k^{i^f} \leftarrow \mathcal{G}(x_{k-1}^{i^a}, m_{k-1}^{i^a})$

$$y_k^f \leftarrow \begin{bmatrix} m_{k-1}^{i^a} & \dots & m_{k-1}^{Ne^a} \\ x_k^{i^f} & \dots & x_k^{Ne^f} \\ z_k^{i^f} & \dots & z_k^{Ne^f} \end{bmatrix}$$

$$P_k^f \leftarrow \frac{1}{N-1}[y_k^f - E(y_k^f)][y_k^f - E(y_k^f)]^T$$

$$d_k \leftarrow [d_j^{obs} + v_k^i, \dots, d_J^{obs} + v_k^{Ne}]$$

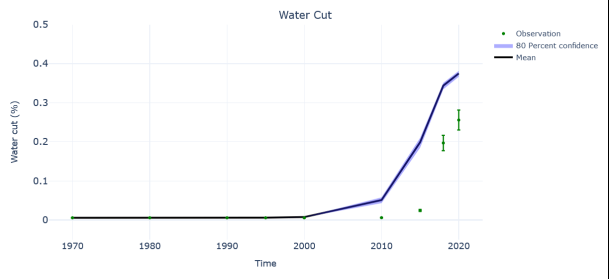
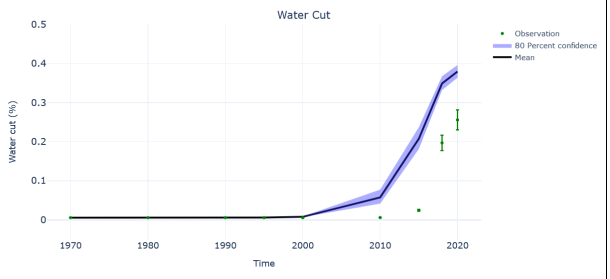
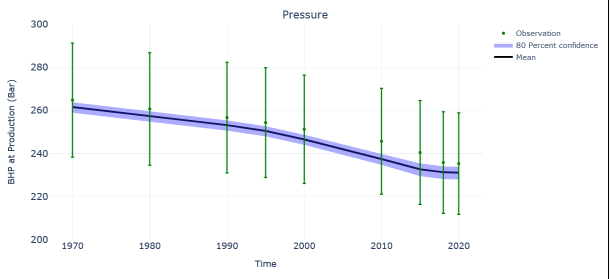
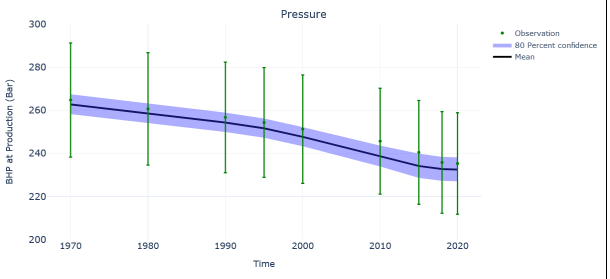
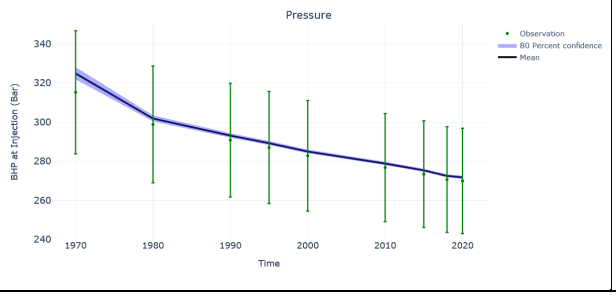
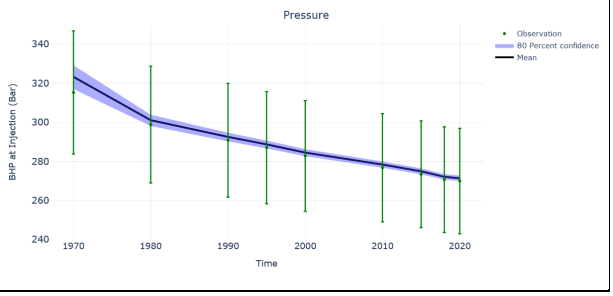
$$R_k \leftarrow \frac{1}{N-1} v_k {v_k}^T$$

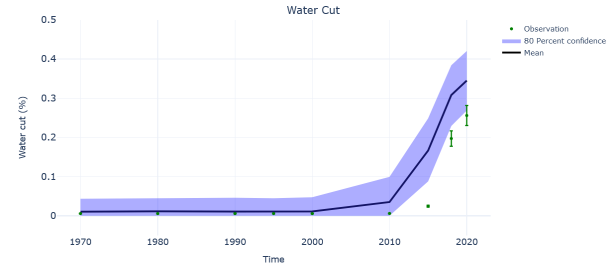
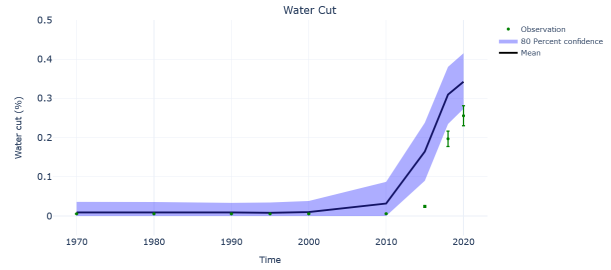
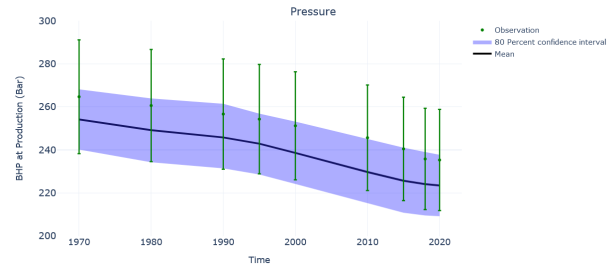
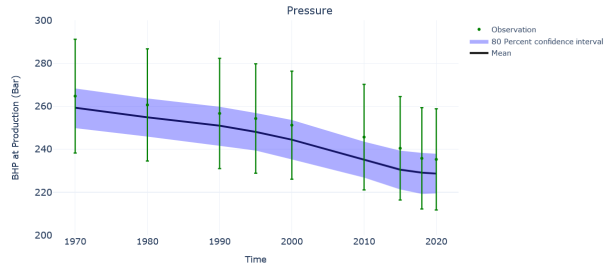
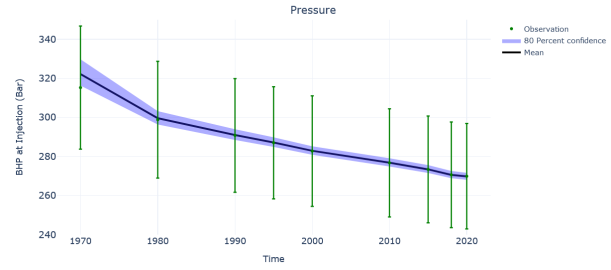
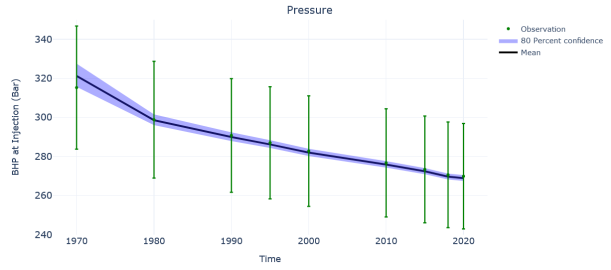
$$\mathcal{H}_k \leftarrow \left[z_k^{i^f} \quad \dots \quad z_k^{Ne^f}\right](y_k^f)^{-1}$$

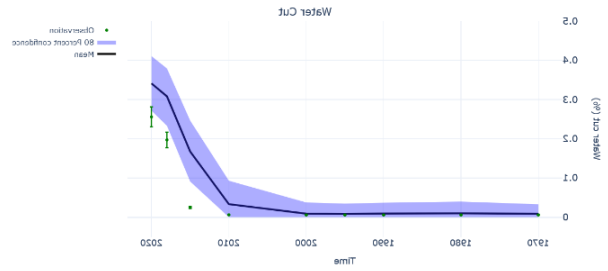
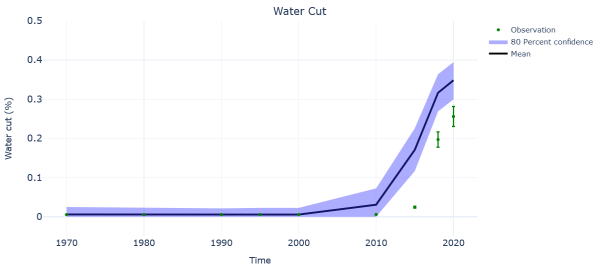
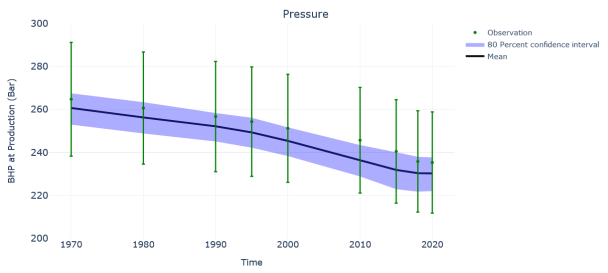
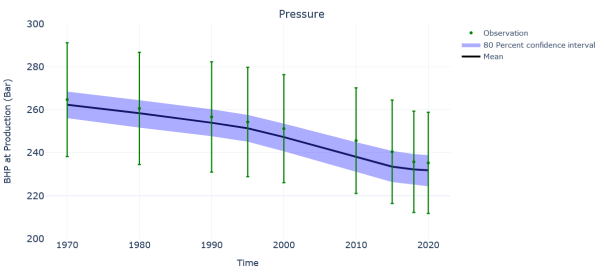
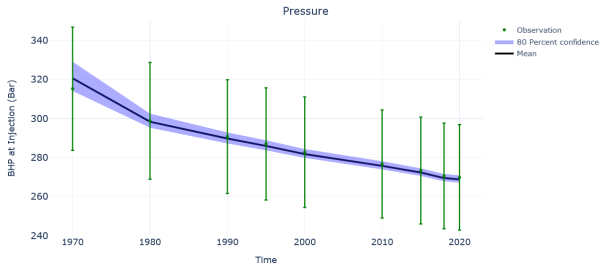
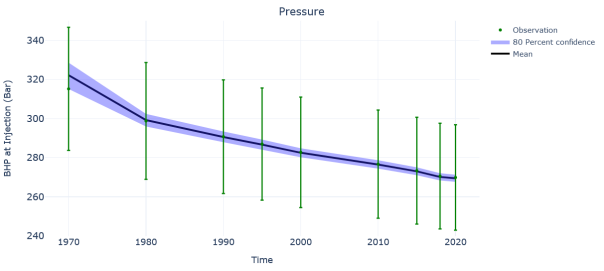
$$\mathcal{K}_k \leftarrow P_k^f \mathcal{H}_k^T (\mathcal{H}_k P_k^T \mathcal{H}_k^T + R_k)$$

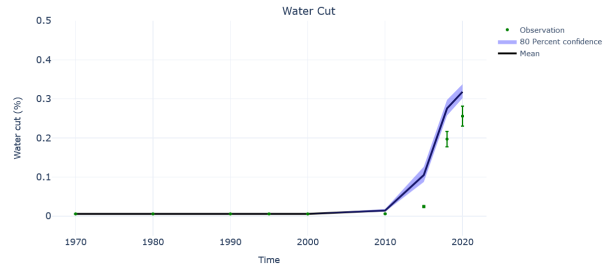
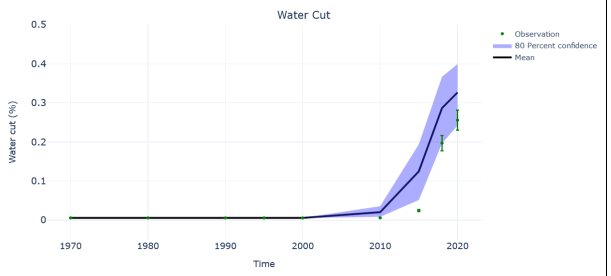
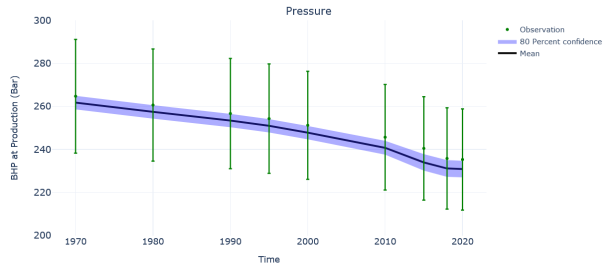
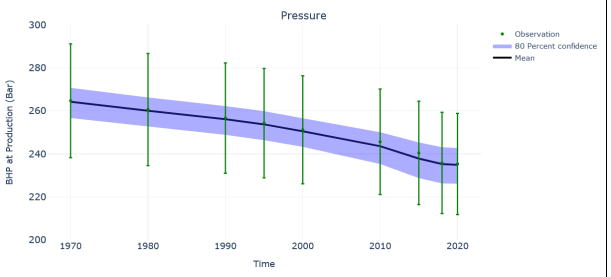
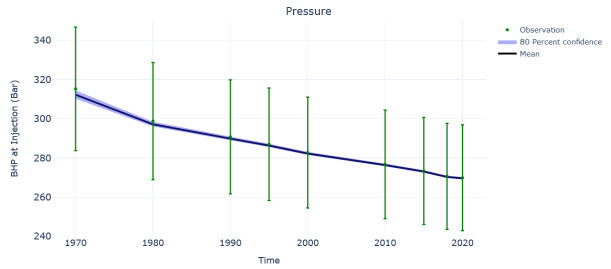
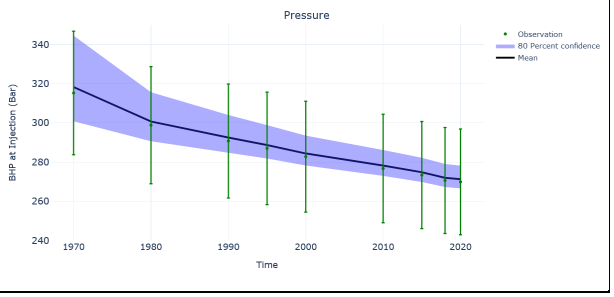
$$x_{k+1}^{i^a} \leftarrow x_k^{i^f} + \mathcal{K}_k (z_k^i - (\mathcal{H}_k y_k^f + \mathcal{N}(0, R)))$$

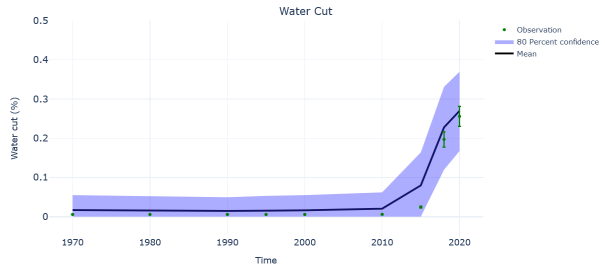
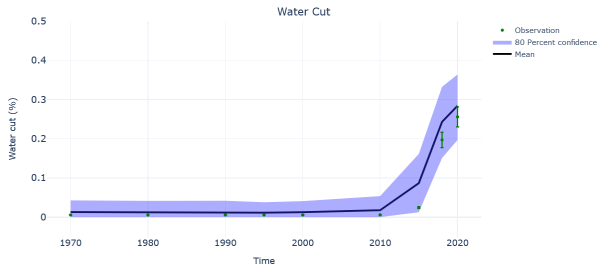
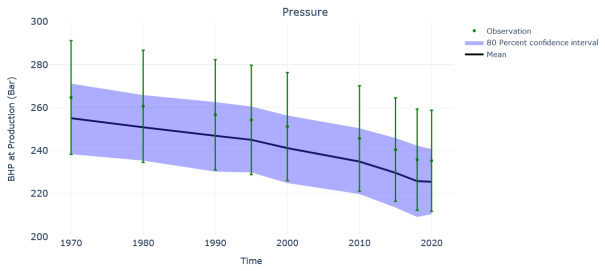
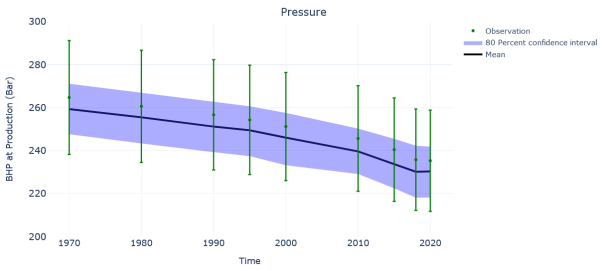
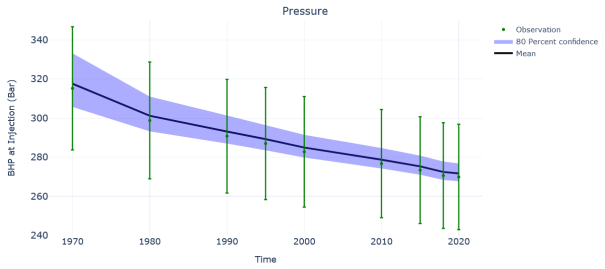
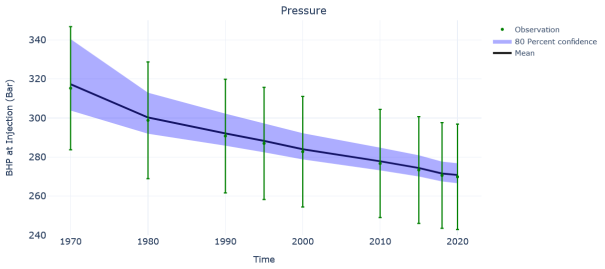
Appendix B: Plots

Forecasted prior ensemble with 5×5 cells	Forecasted update with EnKF ensemble 5×5 cells
	
	
	

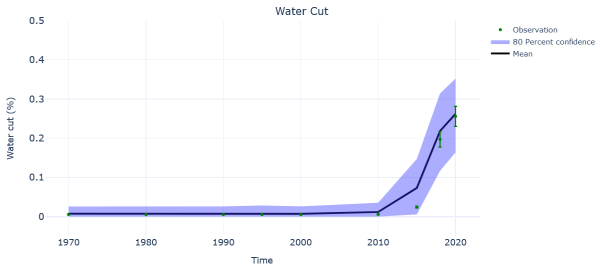
Forecasted update accounting model ensemble with 5×5 cells (model error 200×200)	Forecasted update accounting model ensemble with 5×5 cells (model error 100×100)
	
	
	

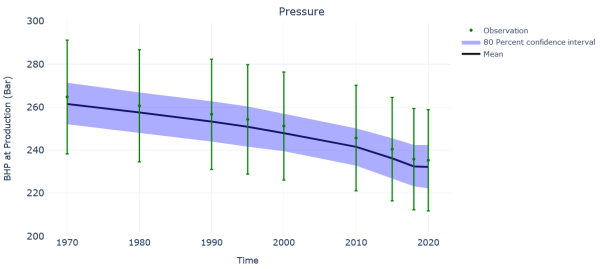
Forecasted update accounting model ensemble with 5×5 cells (model error 50×50)	Forecasted update accounting model ensemble with 5×5 cells (model error 25×25)
	
	
	

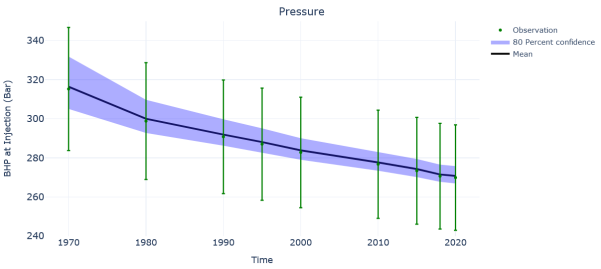
Forecasted prior ensemble with 25×25 cells	Forecasted update with EnKF ensemble 25×25 cells
	
	
	

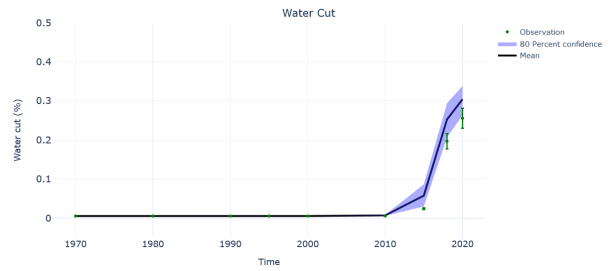
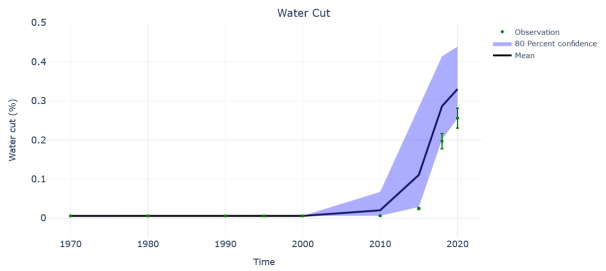
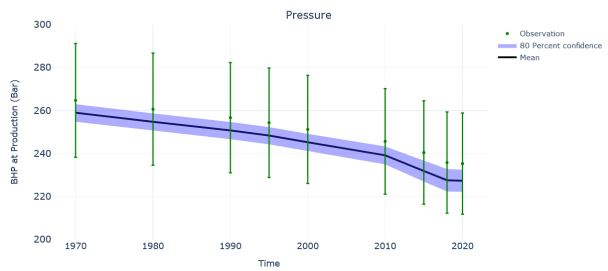
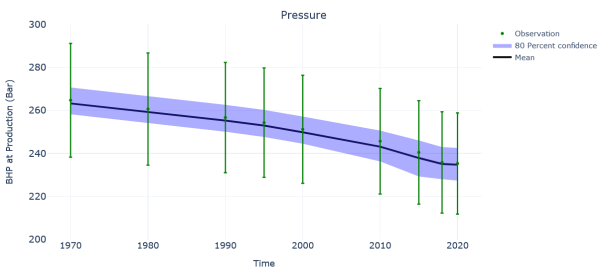
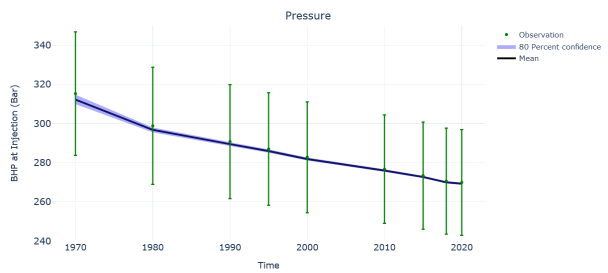
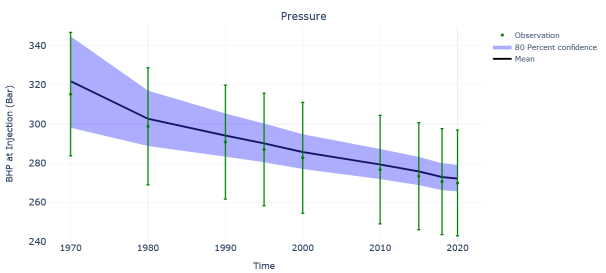
Forecasted update accounting model ensemble with 25×25 cells (model error 200×200)	Forecasted update accounting model ensemble with 25×25 cells (model error 100×100)
	
	
	

Forecasted update accounting model ensemble with 25×25 cells (model error 50×50)

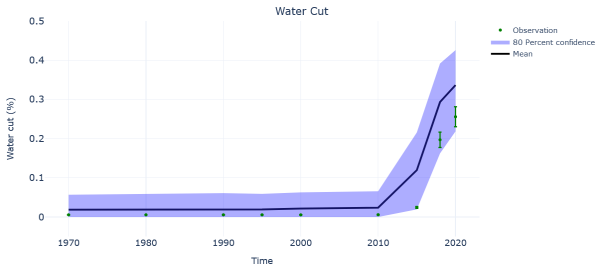


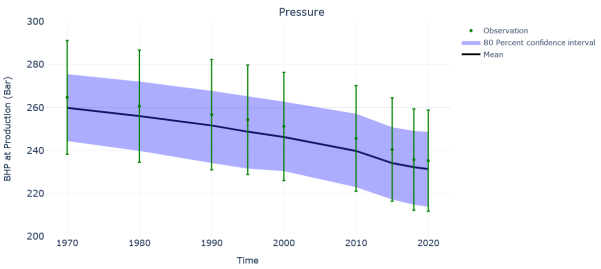


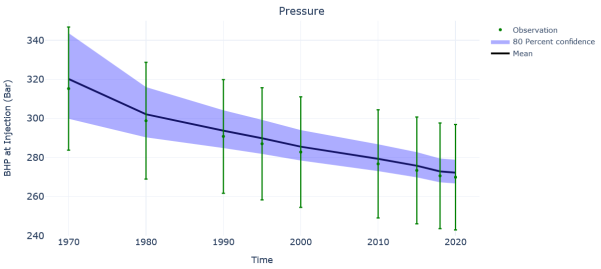


Forecasted prior ensemble 50×50 cells	Forecasted update with EnKF ensemble 50×50 cells
	
	
	

Update accounting for Model Error ensemble 50×50 (model error 200×200)







Forecasted prior ensemble with 100×100 cells HQ models	Forecasted prior ensemble with 200×200 cells HQ models