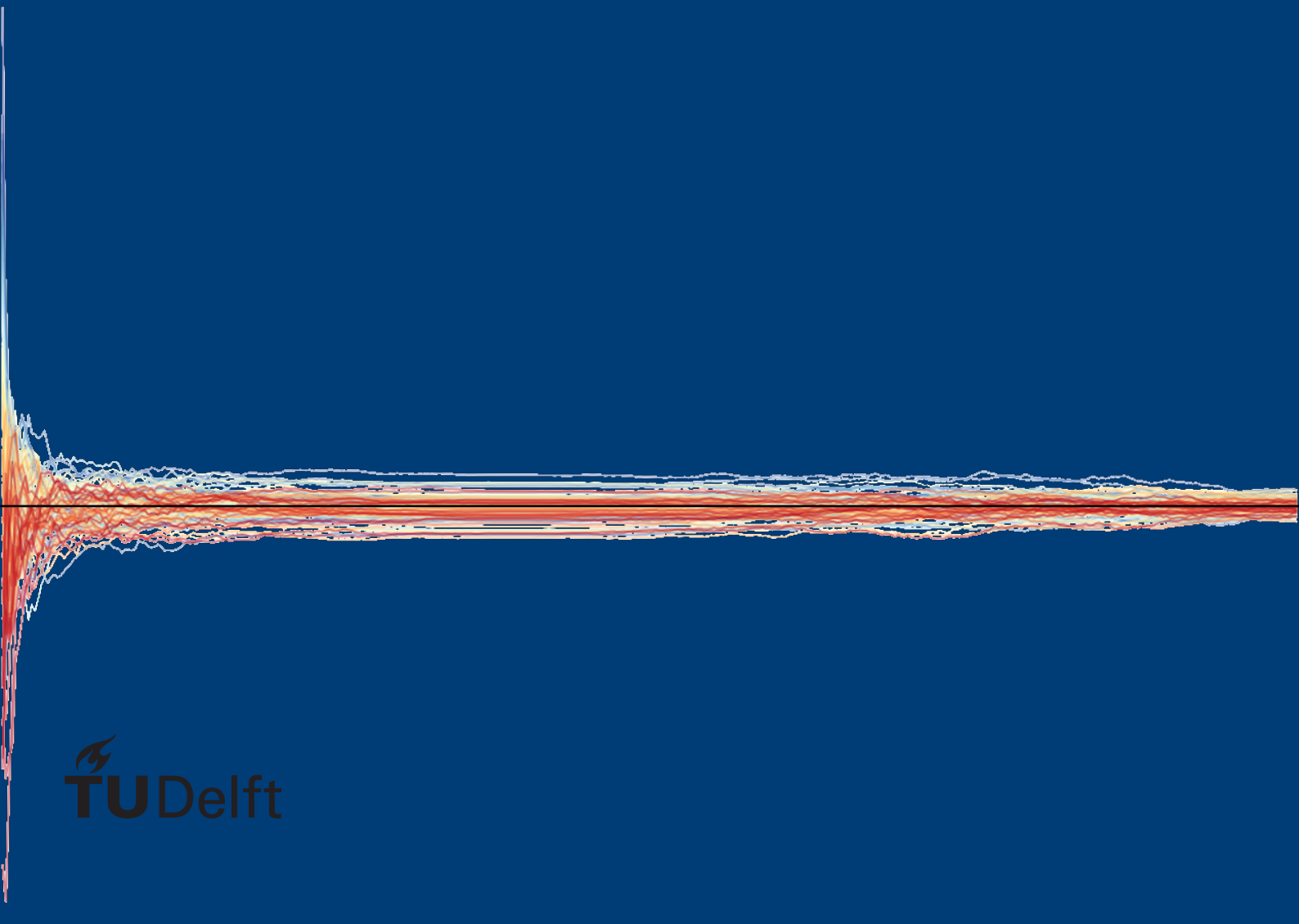
Parameter Estimation and Uncertainty Quantification for Core Flooding

K.C. O'Hara



TU Delft

Parameter Estimation and Uncertainty Quantification for Core Flooding

by

K.C. O'Hara

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday December 21, 2021 at 15:00.

Student number:	4160886	
Project duration:	February, 2020 – December, 2021	
Thesis committee:	Prof. dr. ir. A. W. Heemink,	TU Delft, supervisor
	Dr. ir. J. Romate,	Shell
	Prof. dr. ir. C. Vuik,	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Abstract

In light of our depleting fossil fuel reserves and the relatively 'cheap' extraction of oil and in spite of the highly nonlinear nature of reservoirs, waterflooding has become big business. In recent times, the use of numerical reservoir simulation has not only become possible but has increasingly been used in the petroleum industry in the forecasting of output and money to be made. However, this numerical modelling and automated history matching is not without its problems. The inner workings of sophisticated commercial reservoir simulators are often taken for granted, i.e., "black boxes". These simulators are constructed around a numerical method, with its advantages and disadvantages. Herein, the input settings play a role in the stability and precision of results. For example, the chosen iteration method, grid spacing and time step size or even the choice of iteration parameters, all based on insufficient data, leads simulators to be unreliable and inefficient. Moreover, even under the assumption that such "black boxes" are able to produce a true prediction, this is entirely conditional on correctly establishing the current state and conditions. Thus, practitioners have many global settings given, many of which inaccurate. These many uncertainties can lead to costly mistakes. The aim of this study therefore: is to develop a tool that can quantify the uncertainty of a core flood model and a parameter estimation routine. Specifically, it investigates whether, given limited incoming data, an uncertain parameter can be estimated and then used to simulate and quantify the uncertainty of the water saturation and oil pressure in the core sample. In this context, we question if it be used to provide a history match of the core sample. To see how the uncertain input parameters are reflected in a model output, the tool, based on, the IMPES scheme simulates the two-phase flow, and the Ensemble Kalman filter (EnKF) to estimate the parameters. Building on the base twin experiment, a variety of twin experiments were performed to understand the parameter estimation, by presenting and visualizing the uncertainties in the data and states. We investigate the use of the EnKF for history matching and ways to improve are also explored. Based on the results of this study, it is concluded that the Ensemble Kalman Filter is capable of effective parameter estimation. With modification, it can also be used for history matching and uncertainty quantification. It clearly suggests that the utility of numerical modelling and automated history matching will continue to make contributions to the success of the exploration and extraction of fossil hydrocarbons.

K.C. O'Hara
Delft, December 14, 2021

Contents

Introduction

Throughout history, humankind has relied on hydrocarbon fuels, initially as a source of light and heating. Then the advent of the industrial revolution, exchanged human labor for the combustion engine and more. Even in today's world, hydrocarbons are cardinal to the global energy supply and underpin the global economy. With the depletion of fossil hydrocarbon reserves and the resulting climate change, new solutions will need to be sought out. Currently however, world-wide we obtain approximately 80% of our energy from hydrocarbons[2]. A large variety of fossil energy hydrocarbons are available with crude oil being the largest source of energy consumption. In the year 2017 crude oil accounted for approximately 32% of total energy produced worldwide[2].

An Introduction to Oil Extraction

Oil extraction is no easy task. Initially, a team of petroleum geologists survey the structural and sedimentary specifics of the earth's rock as well as seismic data in a designated area. Their goal is to locate possible petroleum reservoirs, as well as pinpoint the best drill site. Crude oil is typically found in deserts, arctic areas, river deltas and continental margins offshore. More specifically, it is found in subsurface cavities or structural traps in reservoirs between layers of impermeable rock. There are various types of impermeable layers between which we find these fossil hydrocarbon reservoirs: structural geologic structures in deformed layers of rock like faults and folds whose geometries permit retention of hydrocarbons as well as strati-graphic traps formed by changes in rock type or pinch-outs, unconformities, or sedimentary features such as reefs. Different layers of rock, can have different amounts of porosity. Oil flows easily through a layer of sandstone, for example, but would be trapped beneath a layer of shale. It is thus both this permeability and porosity that are important in oil extraction [5]. To extract this oil, a hole is drilled into the earth with an oil rig to create a production well. To maintain structural strength a steel pipe is then inserted. To allow the oil to pass through the base, holes are made at the bottom of the well. At the top collection valves are inserted to maintain pressure during the pumping of the oil. If the well is correctly placed then oil will now start flowing upwards moving from an area of high pressure to low pressure which can then be collected at the surface. Pumps are installed at this point to facilitate extraction. This type of extraction is called primary extraction and usually leaves 70%-85% of hydrocarbons in the reservoir[5, 10]. When this technique no longer yields results, there are two more main oil extracting techniques: secondary and tertiary recovery, which may then be employed to ensure optimum extraction of any fossil hydrocarbons. Tertiary recovery is expensive as it requires significantly more infrastructure and capital, and is not relevant to this study, consult [10, 15] for more details. Alternatively, secondary recovery extraction can be applied to retrieve the remaining oil which is still tightly trapped in the underground reservoir because once pressure drops, there is less oil flowing out. The most often used method in secondary recovery extraction is water flooding which involves drilling injection wells into a reservoir to maintain pressure and push the oil out. That is, water is pumped into the reservoir to increase the pressure difference. This will consequently, flood the reservoir and help to push the remaining oil to the production wells so it can be extracted. The flow of water and oil is called 'flow in porous medium'. 35% to 45% of the oil in the reservoir can thus be extracted [10, 13, 15].

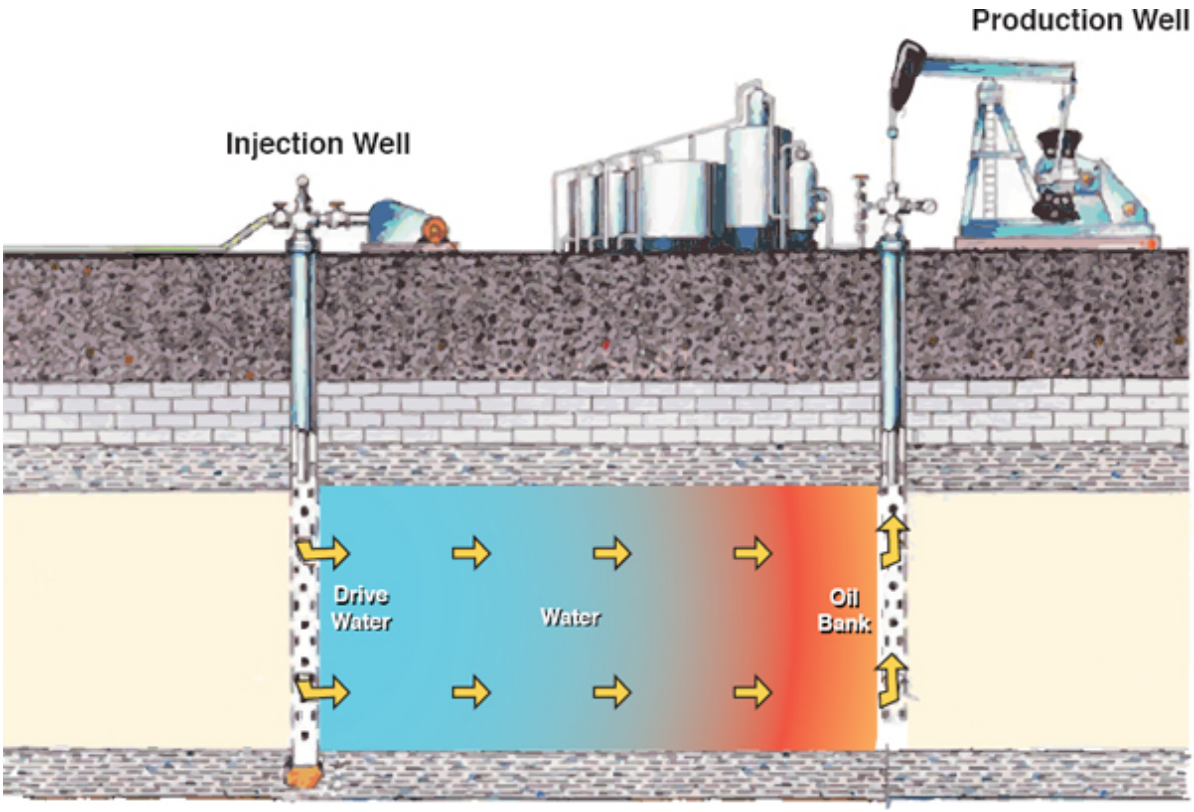


Figure 1: Illustrating waterflooding technique of secondary recovery [19][24]

The entire process of oil extraction is not as easy as it seems. For one the pressure difference may be hard to control when oil still flows naturally. Or extraction may be very difficult when the oil reservoir is in hard to access places such as deep under the earth surface, or below the water in the ocean. It comes as no surprise that multiphase-flow simulation results are used extensively as reservoir performance predictions upon which to base economics for reservoir management decisions. However, accurate modelling requires extensive knowledge of the domain, i.e the porous medium. Thereafter field measurements can be subject to certain limitations. For example, permeability derived from well test data may be reduced by localized formation damage (skin effects) and increased by fractures. As a result, laboratory data are used to support field measurements. This data is gathered from Core Analysis experiments, which can be distinguished into 2 types: Routine (RCAL) and Special (SCAL). Special Core Analysis (SCAL) attempts to extend the data provided by routine measurements to situations more representative of reservoir conditions. SCAL data is then used to support log and well test data (field measurements). However, SCAL measurements are more expensive, and are commonly only done on a small selected group of samples, or if a difficult strategic reservoir management decision has to be made (e.g. to waterflood, or not to waterflood). None the less, even after extensive lab experiments, reservoir properties will remain uncertain. In data assimilation, one combines the information present in the observations and an existing numerical model, as well as the uncertainties in the observations and the models, to produce more realistic results. In petroleum engineering this process is known as “history matching”. Regardless, given that numerical models are only crude approximation of reality, together with uncertain properties and data, subsequent predictions of the reservoir production are anything but certain. But what does this actually mean to be certain or uncertain?

What is Uncertainty?
During any type of uncertainty quantification, uncertainty is subdivided into different types. In the Bayesian framework, the notion of uncertainty is divided into 2 types:

- Aleatoric: uncertainty about an inherently variable phenomenon.
- Epistemic: uncertainty arising from lack of knowledge.

Here the epistemic uncertainty can then be further subdivided into 2 sub-types:

- Model Form: uncertainty that the model is 'structurally correct'.
- Parametric: uncertainty about the correct values to use for particular parameters in the model.

For example, a pressure sensor may have a prescribed accuracy which although deterministic has a very slightly variability in measurements, which we characterize as noise. One could argue, given enough information about the sensor and its environment then this variability could be characterized by a distribution and the uncertainty would be epistemic and not aleatoric.

Problem Statement

Nowadays, reservoir engineers regularly rely on mathematical modelling. However, these models are simple representations, and as such contain errors and uncertainties resulting from assumptions, approximations and inaccurate parameterizations of physical phenomena. These errors and uncertainties propagate into subsequent forecasts. Moreover, the very data used to calibrate model forecasts are imperfectly measured and only at a few locations on a domain. It is thus the reduction of uncertainties that is key to the success of reservoir modelling and consequently to oil exploration, e.g. those pertaining to uncertainty quantification (UQ) in core flood simulations. In this thesis, the aim is to develop a tool that can quantify the uncertainty of a core flood model and a parameter estimation routine. We assume our two-phase flow model $\mathcal{G}$ to be perfect. That is, there is no aleatoric uncertainty present in our nonlinear physical model, $\mathcal{G}$. So, when using the true value of a parameter $\lambda = \lambda^{true}$, it can be used to generate synthetic data d that reflects "reality":

$$d_t = \mathcal{G}(d_{t-1}; \lambda^{true}) \tag{1}$$

However, this synthetic data only consists of the oil pressure in the first and last "control volume", not the entire domain, Ω. As such, Observation Operator $\mathcal{H}$ is used to relate the data d to the forecasts ψ:

$$d_t = \mathcal{H}(\psi_t) + \epsilon \tag{2}$$

Here we assume the measurement noise $\epsilon \sim \mathcal{N}\left(0, \sigma^2_{p^o}\right)$. It could be decried that the noise in (2), which is purely aleatoric, should also be present in the synthetic model. However, we assume that this noise is purely the result of the measurement process and not the model. That is, forecasts, ψ, are generated using the numerical (IMPES) simulator, and are always assumed to be deterministic with no stochastic uncertainty. Then using the initial condition, which is assumed to be exact, ψ^0, in the deterministic model:

$$\psi_t = \mathcal{G}(\psi_{t-1}; \lambda) \tag{3}$$

Here, λ is the unknown parameter, an important component of relative permeability and capillary pressure models, both of which heavily impact two-phase flow. This parameter is uncertain, and so introduces parametric uncertainty into the forecasts, ψ_t. As such, we wish to understand the effect of the uncertainty of the unknown parameter, λ, on the model, $\mathcal{G}$.

Research Objectives

The aim is to develop a tool that can quantify the uncertainty of a core flood model and a parameter estimation routine. We are interested in modelling the two-phase flow in porous media given a relative permeabilities and capillary pressure model, which determines the flow of oil and water through a Core sample. It was found that the following intermediate objectives have to be gained in order to achieve the major goal:

1. analyze the features and physical phenomena which influence two phase flow. From which we construct a mathematical model that simulates two-phase flow and investigate the effect: of the inherent nonlinearity in two-phase flow forecasts, as well as, the intractability and ill-posedness of estimating parameters of two-phase flow models. We explore how the uncertainty of the unknown parameter λ introduces dispersion into the forecast of the two-phase flow.

2. introduce history matching, study the data assimilation framework and implement the Ensemble Kalman filter (EnKF). Discuss some of the foreseeable flaws of the EnKF: probabilistic, numerical, and physical; specifically the issues with nonlinearity as well as conservation of mass.

3. construct an experimental framework from a base twin experiment and determine the behavior of the EnKF wherein the effects of the initial sample, ensemble size, simulator accuracy (via grid refinement) and measurement noise are explored. We closely investigate the issues with nonlinearity as well as conservation of mass.

4. investigate the applicability of the EnKF as a history matching algorithm, whether it is capable of providing a physically accurate estimate, and if we can amend issues that we encounter. We perform twin experiments to compare findings.

5. investigate whether data affect the history matching. Specifically, the quantity, timing, location and type of data.

Thesis Outline

A summary of the chapters used to perform this analysis is listed below:

Chapter 1 provides the background necessary to analyze the main effects which influence core flooding performance, and we construct a mathematical and numerical model simulating the two-phase fluid flow. Preliminary investigations are conducted to better understand some of the geophysics involved, whilst also introducing the problem of the uncertain parameter, λ. In addition, we discuss the non-linearity of the model, the ill-posedness of the parameter estimation problem, and the initial parameteric uncertainty.

Chapter 2 introduces the history matching problem. We start with the general data assimilation model, in which we define the observable state and data, as well as any uncertainties or parameterizations. Using Bayes, we then proceed to derive the associated cost function in order to define the base history matching algorithm: the ensemble Kalman filter. We then analytically investigate the effect of the static state parameterization. We conclude by defining the performance metrics that will be used throughout this thesis.

Chapters 3 from a base twin experiment multiple experiments are conducted using the EnKF in order to determine if it is suitable as a parameter estimation and history matching routine: we aim to understand if, given the available measurement data, the EnKF is capable of providing a parameter estimate. In the second section, we investigate the effect of the initial sample given the initial mean guess and spread. In doing so we aim to discern their effects on the final parameter estimate. In the third section, we will check the robustness of the EnKF by mis-specifying the value of a fixed laboratory environment-controlled parameter. The last section focuses on how some filter settings such as ensemble size, the measurement noise and the grid size may affect the final estimate.

Chapter 4 several twin experiments are conducted, we evaluate their outcome, and aim to introduce and explore solutions for some of the problems that may be encountered. In addition, we evaluate the effects of the data on the final estimate.

Chapter 5 we present an evaluation and conclusions are drawn. We query the use of the EnKF, EnKC: successes, failures, problems, and solutions examined as well as the overall approach. Recommendations are made for further research and conclusions drawn regarding the probable workability and efficiency of these filtering techniques in parameter estimation and history matching in core flooding.

1

Modelling Core Flooding

When determining the suitability of a reservoir for waterflooding there are a number of things that need to be considered. Therefore, the first step is to determine the parameters, and their uncertainties, which determine the flow. Given enough knowledge of the reservoir properties, the underlying physical processes taking place in the reservoir can be modelled. This model can then be solved to estimate or even predict the state of the reservoir. Due to the vast complexity of reservoirs, such models must be solved numerically and as such there are many commerical solvers available. However, even with the most sophisticated solver, if anything that characterizes the reservoir is misspecified or uncertain, the resulting solution will not represent the reservoir of interest. Due to the inherent nonlinearity of the reservoir physics, it may not even be feasible to directly invert simulations to identify where inaccuracies originate.

In this chapter, we begin in the first section by reviewing the known domain properties of the core sample as well as the known properties of the two phases present, i.e. oil and water. We then construct a physical flow model applying the appropriate physical constraints and use physical parameterizations when necessary. We then present a numerical formulation (IMPES) and identify the necessary physical initial and boundary conditions. Finally, we define the physical parameterizations wherein one parameter λ is unknown and needs to be estimated. From this, in the second section, we construct our own numerical solver, detailing the necessary assumptions to be physically accurate. We then run some preliminary simulations to investigate the inherent nonlinearity of the reservoir physics and subsequently the effect of the uncertainty of λ. We conclude the chapter with a short summary and discussion of the numerical solver.

1.1. Two-Phase Flow Model

We begin by stating in the first section physical properties of the domain/rock and then the two phases (oil and water), the second section covers the governing physical model as well as the derived implicit formulation and physical parameterization that we focus on in this work. Take note all subsequent computations are handled in SI units. If one knows the relevant physical data such as rock and fluid properties, it is possible to predict the resulting water and oil production through a simulation model.

1.1.1. Domain Properties

The reservoir is a porous medium whose pores are filled with water and hydrocarbons. How far these pores can be filled with oil or water is determined by the medium's: absolute permeability, porosity, the reservoir pressure, and the inflow speed. However, it is prohibitively expensive and risky to perform experiments on the entirety of the reservoir as such small cylindrical samples are taken. These core samples are taken from an existing well and then used to determine the reservoir medium's properties. In this thesis, we assume that our domain is that of an in-compressible homogeneous core sample.

In this subsection we define the domain properties of the core sample. As "there are some quantities for which the unit standards have not been clarified to the satisfaction of all parties"[34], we converted all properties into SI units for consistency and provide a short description of each one.

Material Parameters	Notation	Value	Measured Unit	Domain	SI Value	SI Unit
Absolute Permeability	K	0.5	$Darcy$ [1]	$\mathbb{R}_{\geq 0}$	4.9346165E-13 ≈ 5E-13	m^2
Porosity	ϕ	0.3	NA[2](%)	$[0,1]$	0.3	NA(%)
Core length	L	6	cm	$\mathbb{R}_{\geq 0}$	6E-2	m
Core cross-section	A	11.4	cm^2	$\mathbb{R}_{\geq 0}$	11.4E-4	m^2
Injection Rate	$R = A \cdot c$	0.57	cm^3/s	$\mathbb{R}_{\geq 0}$	5.7E-7	m^3/s
Inflow Speed	c	0.05	cm/s	$\mathbb{R}_{\geq 0}$	5E-4	m/s
Reservoir Pressure	p^r	1	bar	$\mathbb{R}_{\geq 0}$	1E5	$kg/(ms^2)$

Table 1.1: Core Sample Properties in SI units

Absolute Permeability : The measurement of the permeability, or ability to flow or transmit fluids through a rock, conducted when a single fluid, or phase, is present in the rock.

Porosity: The percentage of pore volume or void space, or that volume within rock that can contain fluids.

Core length : cross-section A complete section of a conventionally drilled core. The section may be up to about 0.6m in length, with typical core diameters lying between 4.4-13.3 cm. The term full-diameter core is also used, but generally refers to shorter sections of about 15cm. The advantage of whole core analysis is that it measures properties on a larger scale, closer to that of the reservoir.

Injection Rate: Volume of fluid injected in a well during hydraulic pumping.

Reservoir: Pressure The pressure of fluids within the pores of a reservoir, usually hydrostatic pressure, or the pressure exerted by a column of water from the formation's depth to sea level.

1.1.2. Phase Properties

Typically there are many different types of hydrocarbons found in a reservoir. However, a full multi-component model is usually prohibitively expensive computationally. Therefore simplifications are usually made to the reservoir fluid flow model in order to obtain a computationally acceptable numerical model. A commonly made assumption is to model phases by very limited number of chemical components. In this thesis we only make use of 2 phases, oil and water. The primary variables or dynamic state variables are the 2 phase pressures and saturations. In addition, we assume the phases to be viscose and in-compressible, i.e. constant density. In this subsection we first state a generic ι phase's physical properties, which in order to be consistent, are all converted into SI units and provided with descriptions. Then we summarize the properties in SI values of the two phases of interest, i.e. oil and water.

ι Phase Parameters	Notation	Measured Unit	Domain	SI Unit
Viscosity	μ^ι	$centiPoise$	$\mathbb{R}$	$kg/(m/s) = Pa \cdot s$
Density	ρ^ι	kg/m^3	$\mathbb{R}_{\geq 0}$	$1kg/m^3$
ι Pressure	p^ι	bar	$\mathbb{R}_{\geq 0}$	$kg/(ms^2)$
ι Saturation	p^ι	(%)	$[0,1]$	(%)
Residual ι Saturation	$S_{r\iota}$	(%)	$[0,1]$	(%)

Table 1.2: SI unit conversion table for a ι phase

Viscosity: The resistance to flow, defined as the ratio of shear stress to shear rate.[3]

Density: The mass per unit of volume of a phase.

ι Pressure: The force per unit area of a phase on the medium.

ι Saturation: The fraction of the void volume of a porous medium filled by this phase

Residual ι Saturation : The minimum fraction of the void volume of a porous medium filled by this, at which point a phase ceases to flow.

The following physical properties will be fixed during entire twin experiment:

ι Phase $[\iota \in \{w,o\}]$	Viscosity $(\mu^\iota)[kg \cdot s/m]$	Density $(\rho^\iota)[kg/m^3]$	Residual ι Saturation $(S_{r\iota})[\%]$	Initial ι Saturation $(S_\iota^0)[\%]$
Water (w)	1E-3	1000	0.2	0.29
Oil (o)	1E-2	730	0.2	0.71

Table 1.3: Fixed physical properties of oil and water phases

[3] Note that in older texts may not taken into account that "the terms poise, centipoise, stokes, and centistokes are no longer used . . . "[34].

1.1.3. Modelling flow in Porous Media

A mathematical model is simply a set of equations which describe certain physical processes occurring in the reservoir. In nearly all cases of interest these equations express conservation of some quantity which is flowing or being transported through the reservoir. The term two-phase flow denotes here the simultaneous movement of two or more immiscible phases through a porous medium.

Here we consider waterflooding to be in a one dimensional domain Ω, where at one end water is injected, in order to drain out oil. In this thesis, we will assume that this simplistic model is exact and mirrors the truth. That is, given the correct domain and phase properties, initial conditions, and subsequent parameterizations and conditions, the model generates the true state.[4]

In what follows is the initial derivation of a two-phase flow model of in-compressible fluid flows through the porous medium which is given by the transport equations for the phase masses, i.e. oil and water (w=water, o=oil).

$$\frac{\partial \varphi \rho^w S^w}{\partial t} = -\nabla \cdot \left(\varphi \rho^w v^w\right) \tag{1.1a}$$

$$\frac{\partial \varphi \rho^o S^o}{\partial t} = -\nabla \cdot \left(\varphi \rho^o v^o\right) \tag{1.1b}$$

wherein Darcy's law relates the volumetric velocity to the pressure gradient:

$$q^o = -\frac{k_{ro}}{\mu_o} K \cdot \nabla p^o = -\lambda^o \cdot \nabla p^o \tag{1.2a}$$

$$q^w = -\frac{k_{rw}}{\mu_w} K \cdot \nabla p^w = -\lambda^w \cdot \nabla p^w \tag{1.2b}$$

These equations are written in terms of the fluid phase pressures (p^o and p^w) and the fluid phase volume fractions (the saturations S^o and S^w), which need to be solved from these equations given appropriate data. The actual velocities v^ι are related to the Darcy (or superficial) velocities q^ι by $q^\iota = \varphi v^\iota$. Here we only state:

$$\lambda^\iota := -\frac{k_{r\iota}}{\mu_\iota} K$$

as the transmissibility for a given ι phase. Herein $k_{r\iota}$ is the relative permeability of phase ι see subsection 1.1.5. In addition, oil, water, rock are assumed to be in-compressible, and hence the porosity and phase densities are considered constants. As such the density terms can dropout of the model:

$$\varphi \frac{\partial S^w}{\partial t} = \nabla \cdot \left(\lambda^w \cdot \nabla p^w\right) \tag{1.3a}$$

$$\varphi \frac{\partial S^o}{\partial t} = \nabla \cdot \left(\lambda^o \cdot \nabla p^o\right) \tag{1.3b}$$

Since the two fluids jointly fill the void space, we have :

$$S^w + S^o = 1 \tag{1.4a}$$

Because of surface tension and the curvature of the interfaces between the two fluids within the small pores, the pressure in the nonwetting fluid is higher than the pressure in the wetting fluid. The difference between these two pressures is the capillary pressure, p^c:

$$p^o - p^w = p^c \tag{1.4b}$$

Notice that the water pressure can be larger than the oil pressure, allowing the capillary pressure to produce negative pressures. This observation implies the following properties:

Capillary Pressure	Notation	Measured Unit	Domain	SI Unit
	p^c	bar	$\mathbb{R}$	$kg/(ms^2)$

Table 1.4: SI unit conversion table for Capillary Pressure, p^c

In layman terms, the rock can "suck" and hold onto fluids[5].

[4] Later in Chapter 2 we will use this model to generate synthetic data and simulated forecasts.

[5] Like a dry kitchen sponge absorbing liquid.

1.1.4. Implicit Pressure Explicit Saturation Formulation

Now we wish to use our constraints (1.4a), (1.4b) to reformulate the (1.3a),(1.3b) into a *Pressure-Saturation Formulation*.

Summing together (1.3a) and (1.3b)

$$\varphi\left(\frac{\partial S^w}{\partial t}+\frac{\partial S^o}{\partial t}\right)=\nabla\cdot\left(\lambda^w\cdot\nabla p^w\right)+\nabla\cdot\left(\lambda^o\cdot\nabla p^o\right)$$

By the derivative addition the LHS becomes

$$\varphi\frac{\partial}{\partial t}\left(S^w+S^o\right)=\varphi\frac{\partial 1}{\partial t}=0\implies\nabla\cdot\left(\lambda^w\cdot\nabla p^w\right)+\nabla\cdot\left(\lambda^o\cdot\nabla p^o\right)=0 \tag{1.5}$$

Now we use the capillary pressure $p^c=p^o-p^w$ and replace $p^w=p^o-p^c$ in (1.5):

$$\nabla\cdot\left(\left(\lambda^o+\lambda^w\right)\cdot\nabla p^o\right)=\nabla\cdot\left(\lambda^w\cdot\nabla p^c\right) \tag{1.6}$$

This is known as the *Pressure Equation* which needs to be solved implicitly! Hence, again using $p^w=p^o-p^c$ into (1.3a) we obtain the *Saturation Equation*

$$\phi\frac{\partial S^w}{\partial t}=\nabla\cdot\left(\lambda^w\cdot\nabla p^o\right)-\nabla\cdot\left(\lambda^w\cdot\nabla p^c\right) \tag{1.7}$$

Thus from (1.6) and (1.7) we see that our formulation is only in terms of S^w, p^o. Hence, the only boundary conditions required are:

$$p^o(t,x=L)=p^r \tag{1.8a}$$

$$\left[-\lambda^o\cdot\nabla p^o-\lambda^w\cdot\nabla p^w\right]_{(t,x=0)}=c \tag{1.8b}$$

$$S^w(t,x=0)=1 \tag{1.8c}$$

Typically, we would assume that $p^o(t=0,x)=p^r$. However, given that $p^c=f(S^w)$ we are only required 1 initial condition:

$$S^w(t=0,x)=0.29$$

where using the pressure equation, $(p^o)^0$ is implicitly (re)computed as a function of $(S^w)^0$. Given that we ignore any change of the oil and capillary pressures over the time step, the pressure equation is said to be elliptic in nature. Whilst the saturation equation may be either parabolic or hyperbolic, depending on the importance of the capillary pressure term relative to the convection term. When capillary pressure effects dominate it will behave parabolic-ally. Whereas when capillary pressure effects are small or absent or, more importantly sometimes, when velocities are large, then the convection term dominates, and thus it approaches a first-order nonlinear hyperbolic equation.

1.1.5. Physical Parameterizations

In water flooding, very complex physical and chemical phenomena occur between the reservoir rock and fluids. These phenomena are often dependent upon the direction of change in wetting phase saturation. Previously in (1.2) and (1.4b), we stated the relative permeability and capillary pressure as parameters of the mathematical model, without stating their properties or physical interpretations. Here we will parameterize these physical phenomena as functions of the saturation, but will also make use of (hyper) parameters.

Specifically, in this thesis, we use the Brooks-Corey[8] framework to define the relative permeabilities and the capillary pressure. In addition, we assume that one such parameter, i.e. the Corey parameter λ, is uncertain and needs to be estimated. In this subsection we first present the physical parameterizations of the Capillary Pressure and Relative Permeabilities before then introducing the uncertain Corey Parameter λ and under which uncertainties.

Recall that previously we defined the capillary pressure as the difference between the oil and water pressure. However, the Brooks-Corey formulation requires the knowledge of p_b, a measure of the maximum pore-size forming a continuous network of flow channels within the medium:

$$p^c=p_b^{\lambda}S^{\frac{-1}{\lambda}}\quad\text{for } p^c\geq p_b \tag{1.9}$$

Instead of including an additional unknown p_b, we take as an empirical fact, the capillary pressure is a unique function of saturation:

$$p^o-p^w=p^c(S)$$

Subsequently, we turn to a Leverett J-function [29], a dimensionless function of water saturation, to model the capillary pressure, p^c, we employ the Leverett J-function, a dimensionless function of water saturation:

$$J(S) = \frac{p^c}{\sigma \cos\theta}\sqrt{K\phi^{-1}}$$

Surface Tension (σ): the property of the surface of a liquid that allows it to resist an external force, due to the cohesive nature of the water molecules.

In general, the J-function would also depend on the contact angle θ with the rock surface, but in this thesis we assume a one dimensional domain and so only effects parallel to the domain are relevant. That is, we ignore the contact angle between phases, such that the surface tension is equivalent to the interfacial tension between oil and water phases, which is the force per unit length. As a result $\theta = 0$:

$$p^c = \sigma\sqrt{\phi K^{-1}}J(S)$$

This together with (1.9) implies:

$$p^c = \sigma\sqrt{\phi K^{-1}} \cdot J^0 S^{\frac{-1}{\lambda}}$$

Leverett Threshold (J^0): is the relative pressure necessary to start displacement of the nonwetting phase (e.g. oil) by the wetting phase (e.g. water)

the following parameters are taken as constant:

Parameter:	Notation	Value	Measured Unit	Domain	SI Value	SI Unit
Surface Tension	σ	30	$Dyne/cm$	$\mathbb{R}$	30E-3	kg/s^2
Leverett Threshold	J^0	1	NA	$\mathbb{R}$	1	NA

Table 1.5: Capillary Pressure Constant Parameters in SI Units

Relative permeabilities $k_{r\iota}$ are required when several phases in the porous media are modeled. If several phases are flowing simultaneously through a porous medium, then the flow of each phase is slowed down by the resistance of the pore structure and additionally by the presence of the other phases. Typically relative permeabilities must be determined empirically or experimentally for each particular porous medium of interest. However, the literature is rich on analytical expressions for the relationship between relative permeabilities and the saturation of the wetting phase [12]. These expressions were usually obtained from simplified porous media models. For this reason, the choice of parameterization is one of finding the best fit.

However, irreducible (water) saturations can vary from approximately 5% to almost 100%. If the rock wettability is known from offset pool data, some limits can be set on the saturation[6]. This can then be used to re-scale the saturation. Thus, in this thesis, we make use of the normalized water saturation:

$$S_n = \frac{S^w - S_{rw}}{1 - S_{ro} - S_{rw}}$$

In this thesis we apply the Brooks-Corey model for the relative permeabilities and a Leverett J-function for the capillary pressure:

$$p^c(S^w) = \sigma\sqrt{\phi K^{-1}} \cdot J^0 S_n^{\frac{-1}{\lambda}} \tag{1.10a}$$

$$k_{ro} = (1 - S_n)^2\left(1 - S_n^{\frac{2+\lambda}{\lambda}}\right) \tag{1.10b}$$

$$k_{rw} = S_n^{\frac{2+3\lambda}{\lambda}} \tag{1.10c}$$

Clearly, all three parameterizations depend on the saturation and uncertain Corey parameter λ. Notice that the current parameterizations of the relative permeabilities are a power law of the normalized saturation. Recall that for a power-law x^{-k} has a well-defined mean $x \in [1,\infty)$ only if $k > 2$, and it has a finite variance only if $k > 3$. For example in the case of k_{rw}, for the normalized saturation to have a well defined mean we need $\lambda < -2$ or $\lambda > 0$, and to have finite variance $\lambda > 0$, [4]. This is important to note, as it maybe one of the sources of failure when estimating values of λ in the region of $\lambda \in (0, 2+\epsilon)$, for some $\epsilon > 0$.

In order to get some intuition for the Corey Parameter λ, in the figure below we demonstrate the sensitivity of the capillary pressure and relative permeabilities given different values of λ.

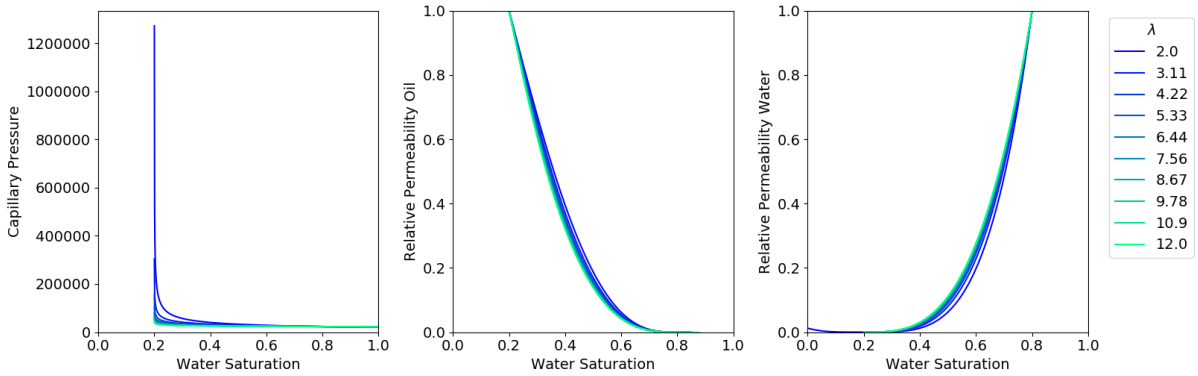


Figure 1.1: Brooks-Corey Curves

The Corey Parameter λ is an index of pore-size distribution and as such is dimensionless. Rather, Brooks and Corey "reasoned that for media having a uniform pore-size, the index would be a large number that theoretically could approach infinity. On the other hand, media with a very wide range of pore sizes should have a small value of λ that theoretically could approach zero. They found that typical porous media, have values of $\lambda \approx 2$"[12]. Now, λ is empirically measured and is simply a shape parameter. This typically is obtained from measured data either by optimizing to analytical interpretation of measured data, or by optimizing using a history matching procedure. In this thesis, we take the Corey Parameter λ to be uncertain and will be estimated via a history matching procedure using synthetic data. That is, we will sample data generated from the same physical model as the history matching procedure but with a true Corey value, $\lambda = 2.5$. In this thesis, unless explicitly stated otherwise, the true Corey parameter value is $\lambda = 2.5$. Experiments show for the case of naturally occurring sand deposits, λ is often about 5 or 6, especially if the material is thoroughly mixed and densely packed[12]. As such in this thesis, the history matching procedure we will take an mean guess, $\mathbb{E}\left[\lambda^0\right] = \lambda_{guess} = 6.5$. However, for soils in an undisturbed state $\lambda < 1$ is not uncommon[12]. For this reason in this thesis we assume that the range of likely values for $\lambda \in (0,10)$, and as such assume that the initial variance of $\text{Var}\left[\lambda^0\right] = \mathbb{E}\left[(\lambda^0 - bound)^2\right] = \sigma^2_{\lambda\lambda} = 27.5$.

Corey Parameter	Notation	Value	Domain	Extra information
True Value	λ	2.5	$[0,\infty)$	Dimensionless
Initial Guess	λ_{guess}	6.5	$(0,10)$	$\sigma^2_{\lambda\lambda} = 27.5$

Table 1.6: Physical Parameterization Properties in SI Units

Later when we apply the history matching procedure, we will need an ensemble of initial guesses for λ. Given that in theory $\lambda \in [0,\infty)$ with the aforementioned anecdotale evidence, (see [12]), we select a distribution whose properties reflects this best. That is, we assume $\lambda \sim Gamma(k,\theta)$. From the method of moments for the gamma distribution we compute; the scale parameter $\theta = \frac{\sigma^2_{\lambda\lambda}}{\lambda_{guess}}$, and the shape parameter $k = \frac{\lambda_{guess}}{\theta}$. In this thesis we will reuse this method of computing the gamma distribution parameters to sample an initial ensemble of λ:

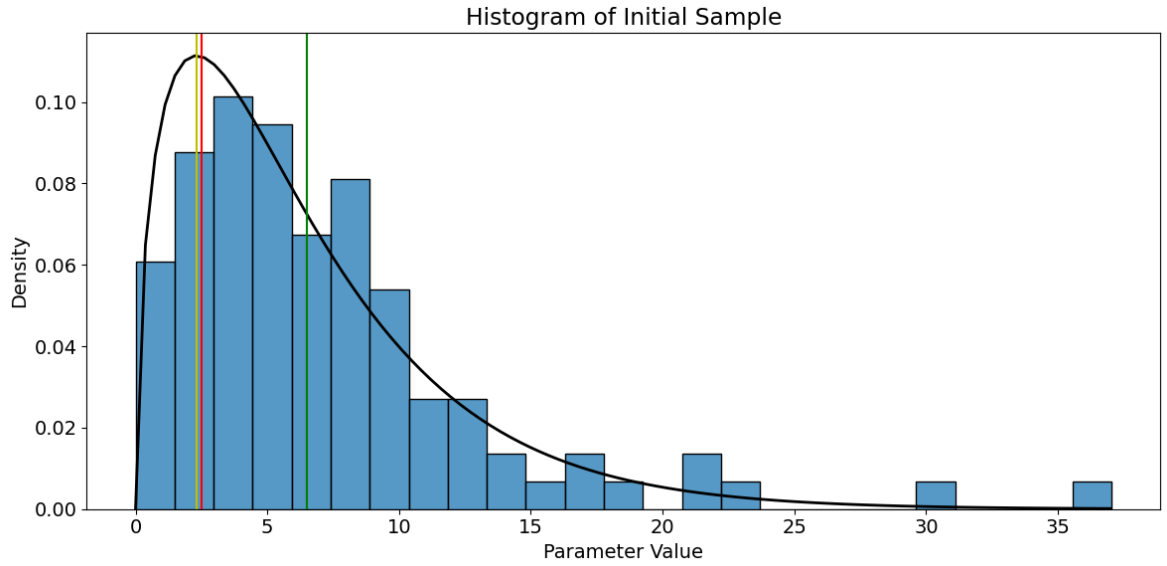


Figure 1.2: 100 samples of unknown parameter λ. (green) initial guess/mean, (yellow) mode, (red) true value

1.2. Numerical Discretization

Typically the geometry of a reservoir is not smooth or convex, and requires complex mesh generation software to partition into control volumes. To guarantee an accurate solution of the simulation the control volumes should be small enough to accurately represent the reservoir heterogeneities, but large enough to assure that the computational time of the model stays within practical limits. In this thesis, we only deal with a cylindrical core sample which can be modelled as a one-dimensional domain. In this section we will construct the finite volume discretization used to solve the two-phase flow problem.

The one-dimensional domain $[0, L]$ is discretized into N_g equal sized control volumes Ω_i, $i = 1, \ldots, N_g$, $|\Omega_i| = \Delta x$, using a cell-centered discretization with cell centers denoted by x_i:

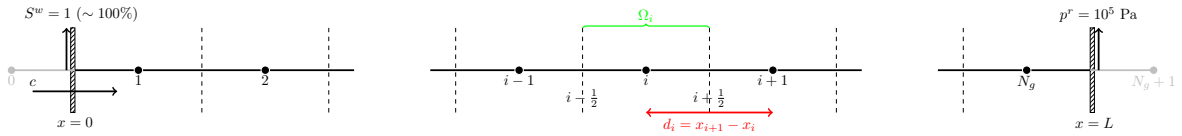


Figure 1.3: Cell Centered Domain Discretization

In this thesis, unless stated otherwise, we will take there to be $N_g = 50$ Control volumes, all of which are equidistant apart, $d_i = d_j$, $\forall 1 \leq i, j \leq N_g$.

1.2.1. Pressure Equation Discretization

Now we integrate the pressure equation (1.6) over the ith control volume (Ω_i)

$$\int_{\Omega_i} \nabla \cdot \left(\left(\lambda^w + \lambda^o \right) \cdot \nabla p^o \right) d\Omega = \int_{\Omega_i} \nabla \cdot \left(\lambda^w \cdot \nabla p^c \right) d\Omega$$

Applying Gauss' Theorem

$$\underbrace{\int_{\partial\Omega_i} \left(\left(\lambda^w + \lambda^o \right) \cdot \nabla p^o \right) \cdot \vec{n} \, d\Gamma}_{I^o} = \underbrace{\int_{\partial\Omega_i} \left(\lambda^w \cdot \nabla p^c \right) \cdot \vec{n} \, d\Gamma}_{I^c} \tag{1.11}$$

Note we upwind $\lambda_{i+\frac{1}{2}} = \lambda_i$ as the flow $i \to i+1$. In addition, ∇p in the flux terms in approximated by a finite difference scheme.

$$I^c := \underbrace{\lambda_i^w \left(\frac{p_{i+1}^c - p_i^c}{d_i} \right)}_{F_{i+\frac{1}{2}}} - \underbrace{\lambda_{i-1}^w \left(\frac{p_i^c - p_{i-1}^c}{d_{i-1}} \right)}_{F_{i-\frac{1}{2}}} = \begin{bmatrix} \frac{\lambda_{i-1}^w}{d_{i-1}} & -\left(\frac{\lambda_{i-1}^w}{d_{i-1}} + \frac{\lambda_i^w}{d_i} \right) & \frac{\lambda_i^w}{d_i} \end{bmatrix} \begin{bmatrix} p_{i-1}^c \\ p_i^c \\ p_{i+1}^c \end{bmatrix}$$

and similarly

$$I^o := \ldots = \begin{bmatrix} \frac{(\lambda_{i-1}^w + \lambda_{i-1}^o)}{d_{i-1}} & -\left(\frac{(\lambda_{i-1}^w + \lambda_{i-1}^o)}{d_{i-1}} + \frac{(\lambda_i^w + \lambda_i^o)}{d_i} \right) & \frac{(\lambda_i^w + \lambda_i^o)}{d_i} \end{bmatrix} \begin{bmatrix} p_{i-1}^o \\ p_i^o \\ p_{i+1}^o \end{bmatrix}$$

Now we will first apply the right Dirichlet Boundary Condition (1.8a). To do this we must linearly extrapolate the value on the boundary. In 1D we are able to expand a Taylor series around $N_g + \frac{1}{2}$ to N_g in order to find a first order accurate boundary condition. That is,

$$p_{N_g}^o = p_{N_g+\frac{1}{2}}^o - \left. \frac{\partial p^o}{\partial x} \right|_{N_g+\frac{1}{2}} \cdot \left(\frac{d_{N_g}}{2} \right) + \mathcal{O}\left(\left(\frac{d_{N_g}}{2} \right)^2 \right)$$

$$\left. \frac{\partial p^o}{\partial x} \right|_{N_g+\frac{1}{2}} = \frac{2}{d_{N_g}} \left(p_{N_g+\frac{1}{2}}^o - p_{N_g}^o \right) - \mathcal{O}\left(\frac{d_{N_g}}{2} \right)$$

Setting $p_{N_g+\frac{1}{2}}^o = p^r$

$$\implies \left. \frac{\partial p^o}{\partial x} \right|_{N_g+\frac{1}{2}} = \frac{2}{d_{N_g}} \left(p^r - p_{N_g}^o \right) - \mathcal{O}\left(\frac{d_{N_g}}{2} \right) \tag{1.12}$$

We apply the same technique on p^c, however we assume that $p_{N_g+\frac{1}{2}}^c = 0$ as otherwise $p^c \to -\infty$ as $S^o \to S^{or}$. That is, in this thesis we are only interested in the case of drainage, and so imposing this conditions on the capillary pressure we avoid imbibition. Then replacing these into the flux terms:

$$\begin{aligned} I^o :=& \frac{2}{d_{N_g}} \left[\left(\lambda_{N_g}^w + \lambda_{N_g}^o \right) \left(p^r - p_{N_g}^o \right) \right] - \left(\lambda_{N-1}^w + \lambda_{N-1}^o \right) \left(\frac{p_{N_g}^o - p_{N_g-1}^o}{d_{N_g-1}} \right) \\ =& \begin{bmatrix} \frac{\left(\lambda_{N_g-1}^w + \lambda_{N_g-1}^o \right)}{d_{N_g-1}} & -\left(\frac{\left(\lambda_{N_g-1}^w + \lambda_{N_g-1}^o \right)}{d_{N_g-1}} + \frac{2\left(\lambda_{N_g}^w + \lambda_{N_g}^o \right)}{d_{N_g}} \right) \end{bmatrix} \begin{bmatrix} p_{N_g-1}^o \\ p_{N_g}^o \end{bmatrix} + \frac{2\left(\lambda_{N_g}^w + \lambda_{N_g}^o \right) p^r}{d_{N_g}} \end{aligned}$$

and

$$\begin{aligned} I^c :=& \frac{2}{d_{N_g}} \left[\lambda_{N_g}^w \left(0 - p_{N_g}^c \right) \right] - \lambda_{N-1}^w \left(\frac{p_{N_g}^c - p_{N_g-1}^c}{d_{N_g-1}} \right) \\ =& \begin{bmatrix} \frac{\lambda_{N_g-1}^w}{d_{N_g-1}} & -\left(\frac{\lambda_{N_g-1}^w}{d_{N_g-1}} + \frac{2\lambda_{N_g}^w}{d_{N_g}} \right) \end{bmatrix} \begin{bmatrix} p_{N_g-1}^c \\ p_{N_g}^c \end{bmatrix} \end{aligned}$$

Second we apply the left Neumann Boundary Condition (1.8b), which states that the total Darcy Velocity is constant, c, at the inlet (i.e. $x = 0$):

$$c = q = q^o + q^w = -\lambda^o \cdot \nabla p^o - \lambda^w \cdot \nabla p^w = -\left(\lambda^w + \lambda^o \right) \cdot \nabla p^o + \lambda^w \cdot \nabla p^c$$

Hence, $(\lambda^w + \lambda^o) \cdot \nabla p^o = \lambda^w \cdot \nabla p^c - c$. Thus, the first control volume:

$$I^o := \left(\lambda_1^w + \lambda_1^o \right) \left(\frac{p_2^o - p_1^o}{d_1} \right) - \left[\lambda_0^w \left(\frac{p_1^c - p_0^c}{d_0} \right) - c \right]$$

and

$$I^c := \lambda_1^w \left(\frac{p_2^c - p_1^c}{d_1} \right) - \lambda_0^w \left(\frac{p_1^c - p_0^c}{d_0} \right)$$

Notice how at the inlet, the capillary pressure terms in I^o and I^c cancel each other out. This implies that the left Dirichlet condition (1.8c) is not relevant to the pressure equation and results in the first row terms:

$$I^o := \begin{bmatrix} -\frac{(\lambda_1^w+\lambda_1^o)}{d_1} & \frac{(\lambda_1^w+\lambda_1^o)}{d_1} \end{bmatrix} \begin{bmatrix} p_1^o \\ p_2^o \end{bmatrix}$$

and

$$I^c := \begin{bmatrix} -\frac{\lambda_1^w}{d_1} & \frac{\lambda_1^w}{d_1} \end{bmatrix} \begin{bmatrix} p_1^c \\ p_2^c \end{bmatrix} - c$$

Hence the resulting Matrix System:

$$T\vec{p^o} = M\vec{p^c} + \vec{b} \qquad \text{where } \vec{b} = \begin{bmatrix} -c, 0, \ldots, 0, -\dfrac{2\left(\lambda_{N_g}^w + \lambda_{N_g}^o\right)p^r}{d_{N_g}} \end{bmatrix}^\mathsf{T} \tag{1.13}$$

where M and B are tridiagonal symmetric matrices with $\sigma(T) \subset \mathbb{R}^{-1}$, $\sigma(M) \subset \mathbb{R}^{-1}$. Since $p_i^c = f(S_i^w)$, then $M\vec{p^c} + \vec{b} = \vec{B}((\vec{S^w})^n, x) \in \mathbb{R}^{N_g}$:

$$T\vec{p^o} = \vec{B} \tag{1.14}$$

This can then be explicitly solved for $\vec{p^o}$, for example by Gaussian elimination. In this thesis, we relying on Tensorflow's built-in linear operators[3], which when run on a GPU relies on Nvidia's CUDA cuSPARSE library[33].... this library is closed source. So we encourage to read [21, 35, 40] for implementation inspiration.

1.2.2. Saturation Equation Discretization

Similarly, we integrate the Saturation Equation (1.7) over the ith control volume (Ω_i):

$$\int_{\Omega_i} \phi \frac{\partial S^w}{\partial t} d\Omega = \int_{\Omega_i} \nabla \cdot \left(\lambda^w \cdot \nabla p^o\right) d\Omega - \int_{\Omega_i} \nabla \cdot \left(\lambda^w \cdot \nabla p^c\right) d\Omega$$

Now we apply Gauss' theorem to the RHS and 1 point integration to the LHS

$$\frac{1}{2}\left(d_{i-1}+d_i\right)\phi \frac{\partial S^w}{\partial t} = \underbrace{\int_{\partial\Omega_i} \left(\lambda^w \cdot \nabla p^o\right) \cdot \vec{n}\, d\Gamma}_{I^p} - \underbrace{\int_{\partial\Omega_i} \left(\lambda^w \cdot \nabla p^c\right) \cdot \vec{n}\, d\Gamma}_{I^c}$$

Firstly, as we are in 1D we only integrate in the x direction. In addition we upwind transmissibilities $\lambda^\alpha_{i+\frac{1}{2}} = \lambda^\alpha_i$, then first for the oil pressure, p^o:

$$I^p := \begin{bmatrix} \frac{\lambda^w_{i-1}}{d_{i-1}} & -\left(\frac{\lambda^w_{i-1}}{d_{i-1}} + \frac{\lambda^w_i}{d_i}\right) & \frac{\lambda^w_i}{d_i} \end{bmatrix} \begin{bmatrix} p^o_{i-1} \\ p^o_i \\ p^o_{i+1} \end{bmatrix}$$

Identically for the capillary pressure, p^c:

$$I^c := \begin{bmatrix} \frac{\lambda^w_{i-1}}{d_{i-1}} & -\left(\frac{\lambda^w_{i-1}}{d_{i-1}} + \frac{\lambda^w_i}{d_i}\right) & \frac{\lambda^w_i}{d_i} \end{bmatrix} \begin{bmatrix} p^c_{i-1} \\ p^c_i \\ p^c_{i+1} \end{bmatrix}$$

Now we apply boundary conditions. Beginning with the RHS Dirichlet Boundary condition, where, like the in the pressure equation, we interpolate the pressure at the wall. For the oil pressure this results in $p^o_{N_g+\frac{1}{2}} = p^r$.

$$\begin{aligned} I^p &:= \left[\lambda^w_{N_g}\left(\frac{2(p^r - p^o_{N_g})}{d_{N_g}}\right)\right] - \left[\lambda^w_{N_g-1}\left(\frac{p^o_{N_g} - p^o_{N_g-1}}{d_{N_g-1}}\right)\right] \\ &= \begin{bmatrix} \frac{\lambda^w_{N_g-1}}{d_{N_g-1}} & -\left(\frac{\lambda^w_{N_g-1}}{d_{N_g-1}} + \frac{2\lambda^w_{N_g}}{d_{N_g}}\right) \end{bmatrix} \begin{bmatrix} p^o_{N_g-1} \\ p^o_{N_g} \end{bmatrix} + 2\frac{\lambda^w_{N_g}}{d_{N_g}} p^r \end{aligned}$$

Here the p^r term is sent to the boundary vector, $\vec{f}$. However, in the case of the capillary pressure we instead assume $p^c_{N_g+\frac{1}{2}} = p^c_{N_g}$ in order to avoid an unphysical"end-effect" and remain in a pure drainage scenario :

$$\begin{aligned} I^c &:= \left[\lambda^w_{N_g}\left(\frac{2(p^c_{N_g} - p^c_{N_g})}{d_{N_g}}\right)\right] + \left[\lambda^w_{N_g-1}\left(\frac{p^c_{N_g} - p^c_{N_g-1}}{d_{N_g-1}}\right)\right] \\ &= \begin{bmatrix} \frac{\lambda^w_{N_g-1}}{d_{N_g-1}} & -\left(\frac{\lambda^w_{N_g-1}}{d_{N_g-1}} + \frac{2\lambda^w_{N_g}}{d_{N_g}}\right) \end{bmatrix} \begin{bmatrix} p^c_{N_g-1} \\ p^c_{N_g} \end{bmatrix} + 2\frac{\lambda^w_{N_g}}{d_{N_g}} p^c_{N_g} \end{aligned}$$

Notice that to keep the discretization matrices of p^o and p^c to be identical, we send this extra $p^c_{N_g}$ term to the boundary vector, $\vec{f}$. Now we apply the LHS Boundary conditions, recall that

$$c = -\lambda^o \cdot \nabla p^o - \lambda^w \cdot \nabla p^w \iff \lambda^o \cdot \nabla p^o + c = -\left(\lambda^w \cdot \nabla p^o - \lambda^w \cdot \nabla p^c\right)$$

Now (1.8c) implies that $\lambda^o_0 = 0$:

$$\implies 0 + c = -\left(\lambda^w \cdot \nabla p^o - \lambda^w \cdot \nabla p^c\right)$$

and so very naturally we can impose (1.8b)

$$\begin{aligned} I^p - I^c &= \left[\lambda^w_1\left(\frac{p^o_2 - p^o_1}{d_1}\right) - \lambda^w_1\left(\frac{p^c_2 - p^c_1}{d_1}\right)\right] + c \\ &= \begin{bmatrix} -\frac{\lambda^w_1}{d_1} & \frac{\lambda^w_1}{d_1} \end{bmatrix} \begin{bmatrix} p^o_1 \\ p^o_2 \end{bmatrix} - \begin{bmatrix} -\frac{\lambda^w_1}{d_1} & \frac{\lambda^w_1}{d_1} \end{bmatrix} \begin{bmatrix} p^c_1 \\ p^c_2 \end{bmatrix} + c \end{aligned}$$

Notice immediately that the resulting discretization matrices for p^o and p^c are identical to M found in (1.13). On a side note, it is easier and more numerically efficient to compute $\nabla p^w = \nabla(p^o - p^c)$ such that $M(p^o - p^c) = I^p - I^c$. We then result in the following system of equations:

$$D\phi \frac{\partial \vec{S^w}}{\partial t} = M\left(\vec{p^o} - \vec{p^c}\right) + \vec{f}$$

where

$$D = \text{diag}\left(\frac{1}{2}(d_{i-1} + d_i)\right) \text{ and } \vec{f} = \begin{bmatrix} c & 0 & \cdots & 0 & 2\frac{\lambda^w_{N_g}}{d_{N_g}}(p^r - p^c_{N_g}) \end{bmatrix}^\mathsf{T}$$

Finally we apply Forward Euler time integration. The resulting system is known as the saturation equation:

$$\vec{Sw}^{n+1} = \vec{Sw}^n + \delta t \underbrace{\phi^{-1} D^{-1} \left[M\left(\vec{po}^n - \vec{pc}^n\right) + \vec{f}^n \right]}_{E^n \in \mathbb{R}^n}$$

where E^n is can be seen as the evolution:

$$\implies \vec{Sw}^{n+1} = \vec{Sw}^n + \delta t E^n$$

1.2.3. IMPES Summary

In summary we formulated the two-phase flow in porous media model (1.1) as a *Pressure-Saturation Formulation*, and then discretized it using the Implicit Pressure and Explicit Saturation scheme (IMPES). This results in three spatial discretization operators T, M, D and two boundary vectors b, f which are used in the following equations:

Pressure Equation:

$$T\vec{po} = M\vec{pc} + \vec{b} \tag{1.15a}$$

Saturation Equation:

$$\vec{Sw}^{n+1} = \vec{Sw}^n + \delta t \phi^{-1} D^{-1} \left[M\left(\vec{po}^n - \vec{pc}^n\right) + \vec{f}^n \right] \tag{1.15b}$$

In industry, commercial solvers are often, if not always, "black boxes". Thus, rather than giving an upper bound to computational complexity of the IMPES scheme, we simply claim it has a complexity of $\mathcal{O}\left(\chi(N_g)\right)$.

1.3. Preliminary Physical Model Investigations

In Section 1.1, we defined the domain, phase and parameterizations properties, yet we still are uncertain of what observable phenomena each property is responsible for. Having just defined in Section 1.2 a numerical model, we can now investigate the effects of each property. In this section we run some preliminary simulations to investigate: the effects of the constant parameters, the inherent nonlinearity of the reservoir physics, and the effect of the uncertainty of λ.

1.3.1. Effects of the Known (Constant) Properties

Below we investigate the effects of the domain properties ϕ, K, the phase properties $S_{ro}, S_{rw}, \mu_o, \mu_w$, and the parameterization properties, σ, J_o. To do this we recall that our base experiment has the following constant properties:

$$\phi = 0.3,\ K = 5E-13,\ S_{ro} = 0.2,\ S_{rw} = 0.2,\ \mu_o = 1E-2,\ \mu_w = 1E-3,\ \sigma = 30E-3,\ J_o = 1.0$$

Then from this base set we deterministically change a property value in order to investigate the subsequent observable phenomena. What we hope to achieve is intuition for the overall nonlinearity of the physical model. The results are plotted in Figure 1.4 below:

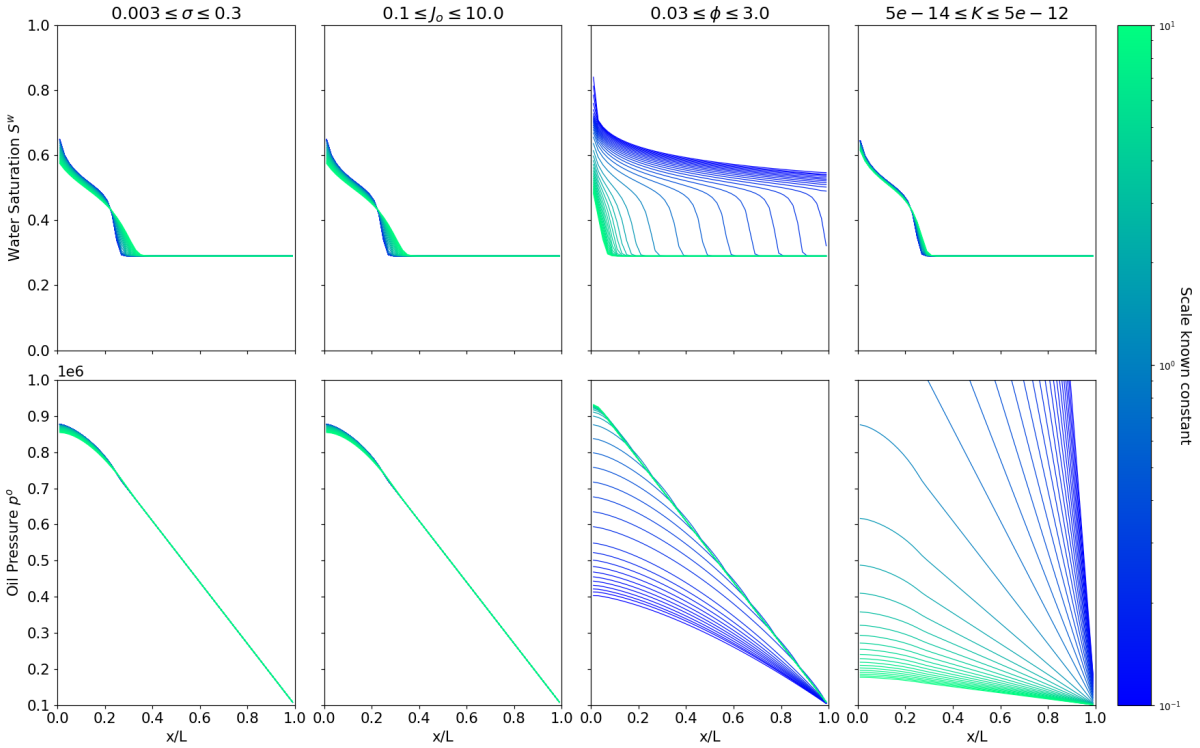


Figure 1.4: The effect parameters on the Water Saturation and Oil Pressure with $N_t = 100$, $N_x = 50$. Base Parameters: $\lambda = 2.5$, $\sigma = 0.03$, $J_o = 1.0$, $\phi = 0.3$, $K = 5e-13$, $S_{ro} = 0.2$, $S_{rw} = 0.2$, $\mu_o = 0.01$, $\mu_w = 0.001$

Notice that we have omitted any changes to the domain dimensions, phase properties, or initial and boundary conditions, i.e. A, L, μ_*, S_{r*}, S^{w0}, p^{o0}, c, or p^r. These are controllable properties that in a lab environment are predetermined by the available sample, whilst we are only interested in the uncontrollable properties specific to the relative permeabilities and capillary pressure. From the Figure 1.4 above we see the following:

σ & J_o : Neither can be too large otherwise it results in oscillations present in both the water saturation and oil pressure. Furthermore, we see that increasing σ diffuses the water saturation wavefront and keeps the oil pressure low. Neither appear to be critical in this context.

ϕ : the value of ϕ inversely controls the speed of the profiles, e.g. ϕ increases the speed decreases.

K : S^w is quite insensitive to K, contrary to p^o which is highly sensitive.

1.3.2. (Non)Linearity of the Physical Model

In the previous subsection, we observed the effects of the relationships between known (constant) properties by how they affected the dynamic state variables. In this subsection we investigate the "inherent nonlinearity" of the physical model caused by the Corey parameter λ. A very simple check for this is by verifying if it is a linear mapping.

Definition 1.3.1 (Linear Mappings)**.** *A linear map $f(x)$ is a function which satisfies both of the following properties:*

- *Homogeneity:* $f(\varsigma x) = \varsigma f(x) \quad \forall \varsigma \in \mathbb{R}$
- *Additivity:* $f(x+y) = f(x) + f(y)$

otherwise the mapping is termed nonlinear.

An example of such a linear mapping is integration. In addition, recall that compositions of linear mappings are also linear mappings. Thus, given that integrations are linear, if the spatial mean water saturation is linear with respect to the unknown parameter then the water saturation is a linear mapping of the unknown parameter itself. Specifically, we recall that the generalized arithmetic mean of a function is:

$$\bar{f} = \frac{1}{Vol(\Omega)} \int_{\Omega} f d\Omega$$

Thus in the case of the discretized S^w we can interpret this as:

$$\mathbb{E}\left[S^w(\lambda)\right] = \frac{1}{N} \sum_{\Omega_i} S_i^w(\lambda)$$

Hence, what is to be shown is that:

- $\mathbb{E}[S^w(\varsigma\lambda)] = \varsigma \mathbb{E}[S^w(\lambda)]$
- $\mathbb{E}[S^w(\lambda+\delta\lambda)] = \mathbb{E}[S^w(\lambda)] + \mathbb{E}[S^w(\delta\lambda)]$

Thus, we run the simulator for given constants ς and $\delta\lambda$ and take the difference of the LHS and RHS which if linear should be 0:

- $\mathbb{E}[S^w(\varsigma\lambda)] - \varsigma \mathbb{E}[S^w(\lambda)] = 0$
- $\mathbb{E}[S^w(\lambda+\delta\lambda)] - (\mathbb{E}[S^w(\lambda)] + \mathbb{E}[S^w(\delta\lambda)]) = 0$

In Figure 1.5, we plot the results of these tests at each time step with parameters $\varsigma = 5, \delta\lambda = +2$ and then repeat with their inverses $\varsigma = \frac{1}{5}, \delta\lambda = -2$. What we see is that as the simulation progresses the difference increases. Moreover, there appear to be 2 clear states with a third intermediary transition state. These reflect the 3 states above where the two clear states are injection flood front formation and post-breakthrough, with the third transition state being the pre-breakthrough to post-breakthrough transition. These can be identified by the change in the growth of the difference. Furthermore, we see that for $\varsigma = 5$ and $\delta\lambda = +2$ (i.e. larger values of λ), there is a greater nonlinear effect than for $\varsigma = \frac{1}{5}$ and $\delta\lambda = -2$ (i.e. smaller values of λ).

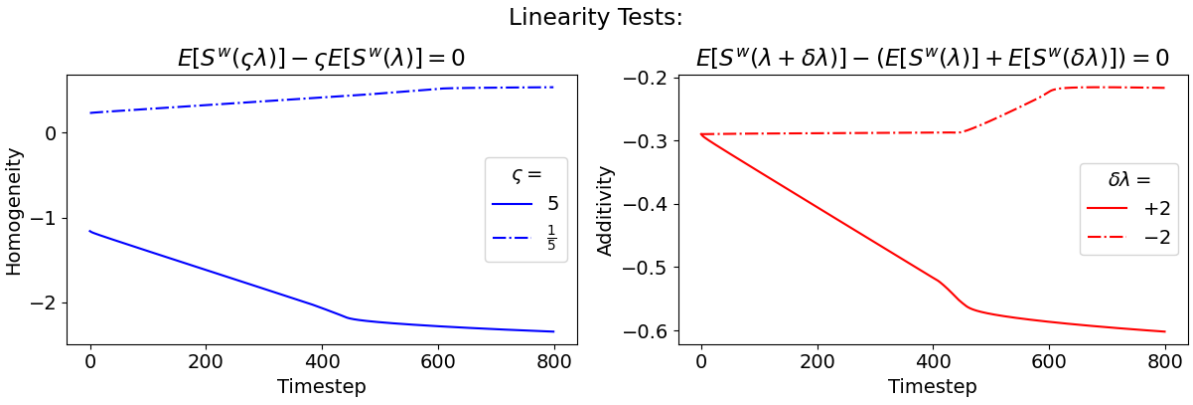


Figure 1.5: Linearity of Physical Model

Notice that $\mathbb{E}[S^w(\delta\lambda = -2)])$ is not physically possible as $\delta\lambda \notin (0,\infty)$ and interestingly renders the relative permeability of the oil $k_{ro} = 0$, for any level of water saturation. This means that relative to the water flow the oil flow is relatively immobile. More concretely, $k_{ro} = 0 \implies \lambda^o = 0$ then returning to the pressure equation:

$$\begin{aligned} &\implies \nabla\cdot\left(\left(0+\lambda^w\right)\cdot\nabla p^o\right) = \nabla\cdot\left(\lambda^w\cdot\nabla p^c\right) \\ &\implies \nabla\cdot\left(\lambda^w\cdot\nabla(p^o - p^c)\right) = 0 \end{aligned}$$

[6]. Then replacing this into the Saturation Equation we see that:

$$\phi\frac{\partial S^w}{\partial t} = 0$$

and so clearly the water saturation never changes regardless of the boundary conditions imposed. Physically this means that changing the injection rate will not improve production. Rather by closely examining the left boundary conditions we see that:

$$\left[0 - \lambda^w\cdot\nabla p^w\right]_{(t,x=0)} = c \tag{1.16}$$

so the water pressure gradient will change with respect to the injection rate.

1.3.3. Intractability of Nonlinear Physical Model

Given that we have shown the physical model is nonlinear why is this even a problem? Recall currently our objective is to identify how the uncertainty of the parameter affects the dynamic state. With this knowledge our main objective, estimating the unknown parameter, is equivalent to identifying the state dynamics from the data. However, even in the linear case this is not trivial if the problem is "ill-conditioned", that is if there does not exist an unique solution. On the other hand,the case when a linear problem is "well-conditioned", then the problem can be solved by direct inversion, e.g. $x = A^{-1}b$. However, as there does not exist a closed form solution for multiphase flow in porous media, we cannot even attempt a direct inversion. Moreover, even if there existed a nonlinear closed form solution the inverse would not necessarily be correct. J.S. Hadamard[7] believed that mathematical models of physical phenomena should have[22]:

- a solution exists,
- the solution is unique,
- the solution's behavior changes continuously with the initial conditions.

Problems that satisfied these conditions came to be known as "well-posed", whilst those that do not are known as "ill-posed". Thus in this thesis, the parameter estimation problem is said to be "ill-posed". That is, as the behavior of the dynamic state does not change continuously given the initial λ estimate, which is a result of the nonlinearity of the dynamic state. Thus, there may not exist or be an unique solution to the dynamic state, and this then implies that the parameter estimation problem of λ may be "ill-conditioned". In that, for each estimate λ there does not necessarily exist an unique solution, and so the estimation problem is not "well-posed". To illustrate this problem we take a simple example, take nonlinear function x^2 with prior $x \sim \mathcal{N}(0,1)$ this then has posterior distribution as chi squared distribution. Now the direct inversion of the posterior using the exact inverse, i.e. the square root, would yield the chi distribution not the $\mathcal{N}(0,1)$.

[6] so $p^o - p^c$ is a harmonic, neat right?

[7] (8 December 1865 - 17 October 1963) a French mathematician known for major contributions in number theory, complex analysis, differential geometry and partial differential equations.

1.3.4. Dispersion of the initial parameter uncertainty through the physical model

In order to gauge the effect of the uncertainty of an initial ensemble of parameters we propagate the dynamic state, i.e. water saturation and oil pressure, from a fixed initial state, then by observing how the uncertainty propagates into the dynamic state we can determine the effect of the initial uncertainty. Below in Figure 1.6 we plot the results at 3 distinct time steps; 100, 400, 801. These represent 3 distinct time points in core flooding; injection flood front formation, pre-breakthrough, post-breakthrough.

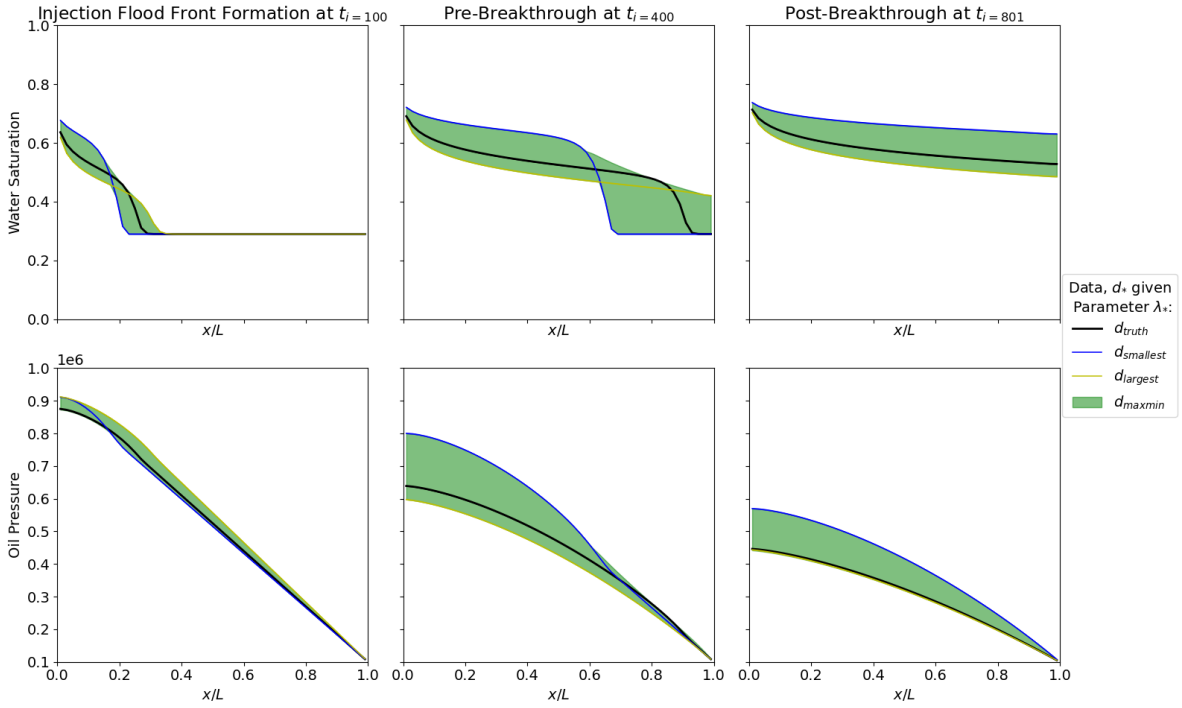


Figure 1.6: Initial parameter uncertainty propagation

In green we observe the *forward propagation of uncertainty*[38] caused by the uncertainty of parameter λ propagating into the dynamic state variables, S^w, p^o. That is, the water saturation, S^w, suffers a smearing effect where small values of λ increase the wave height and slow down the wave front, and conversely large values of λ decrease the wave height and speed up the wave front. [8] An explanation results from Brooks and Corey's Interpretation which defines λ as an index for the distribution of pore sizes in the rock. For narrow distributions, λ is greater than 2; for wide distributions, λ is less than 2. In terms of the effects on the dynamic state variables, when λ decreases less than 2 the pore sizes in the rock can greatly vary, this means that water will need to fill small and large pores, thereby preventing the water phase from distributing itself across the domain easily. Whilst, when λ increases above 2 the water phase is able to distribute itself more evenly across the domain as the sizes of the pores are more homogeneous.

What this also implies is that, for large values of λ, the core *behaves* as if it were "water-wet". When "water-wet", oil predominantly resides in larger pores of the rock matrix and is relatively more mobile than if it resided in the smaller pores[14]. Moreover, being "water-wet", the water naturally imbibes into the smaller pores, and when injected with water is also displaced[14]. This then in turn displaces the oil from the larger pores towards the production well. This is clearly seen for the largest λ (yellow), where the water saturation front is less sharp as the water imbibes ahead of the front as a function of the imbibition capillaries' pressure relationship with respect to distance ($\frac{dP_c}{dx}$)[14]. On the other hand, this also implies that, for small values of λ, the core behaves as if it were "oil-wet". That is, there will be an opposite effect as some oil trails behind the front because the threshold capillary pressure required for entry of water into the pores is much higher. This is clearly seen for the smallest λ (blue) oil pressure which demonstrates a clear kink for the change in pressure.

Note that our interpretation of the behaviour follows analytically from the mathematical model construction. However, in reality this may not be what physically occcurs. Remember, the mathematical model is

[8] Note only 20/100 of the initial parameter samples where smaller that the true value[9], and furthermore the smallest parameter was ≈ 0.4 whilst the largest was ≈ 30.

just a representation of the physics, but it doesn't necessarily capture the physical phenomena accurately as a whole. So by construction, there is an inherent bias. As a result, this interpretation of the behaviour is a perfect example of what "Model Form" uncertainty could originate from.

1.4. Chapter Summary

A background necessary to analyze the main effects which influence core flooding performance was given, and a *Pressure-Saturation Formulation* of the two-phase fluid flow was constructed:

$$\nabla \cdot \left(\lambda^w \cdot \nabla p^c\right) = \nabla \cdot \left(\left(\lambda^o + \lambda^w\right) \cdot \nabla p^o\right) \tag{1.17a}$$

$$\phi \frac{\partial S^w}{\partial t} = \nabla \cdot \left(\lambda^w \cdot \nabla \left(p^o - p^c\right)\right) \tag{1.17b}$$

with the following boundary and initial conditions:

$$p^o(t, x = L) = p^r \tag{1.17c}$$

$$\left[-\lambda^o \cdot \nabla p^o - \lambda^w \cdot \nabla p^w\right]_{(t,x=0)} = c \tag{1.17d}$$

$$S^w(t, x = 0) = 1 \tag{1.17e}$$

$$S^w(t = 0, x) = 0.29 \text{ and } p^o(t = 0, x) = p^r \tag{1.17f}$$

This is then discretized by an Implicit Pressure Explicit Saturation (IMPES) scheme(1.15), which implicitly recomputes $(p^o)^0$ using the pressure equation as a function of $(S^w)^0$. By conducting some preliminary investigations on the known constant properties, we were able to grasp the scale of the inherent nonlinearity of the problem. Using the model $\mathcal{G}$ we then proceeded to show this nonlinearity, experimentally by contradiction, as it failed to uphold the properties of a linear mapping. Subsequently, we then explained why the problem is intractable and thus ill-conditioned. Finally, we investigated how the uncertainty of the unknown parameter λ disperses through the model. We identified why this may be the case through the construction of the model, at the same time highlighting how the model biases our reasoning of how the physics actually work.

2

History Matching

Even today many phenomena that occur during water flooding are not fully understood, but we keep getting closer. In the previous chapter 1, we saw the large number of properties that go into describing fluid flow in a porous medium. Determining the values of each property and/or dynamic state variables is no easy feat as each property must be estimated, either by direct measurements or must be indirectly solved for. That is, often properties must be inferred from their effects on the observable dynamic state for which measurements are available. However, as measurements maybe uncertain this introduces uncertainty in those geological properties (or their parameters) which then directly translates into uncertainty in the forecasts of dynamic state variables. As such, often only uncertain estimates of many properties can be given.

In practice, data is the result of multiple types of costly routine lab experiments, for example one such type is Special Core Analysis (SCAL). These experiments yield varying types of data, such as oil pressure measurements, water saturation snapshots, x-ray micro-tomographic scans, etc. This data is compared against simulations run with different parameter values in order to find the estimate yielding the "best fit". In the past history matches were done manually; in other words, an engineer would select the values, run a simulation, then check the quality of fit. This, process would be manually repeated to improve the fit, and as such were extremely time-consuming.

Automatic or assisted history matching, however, uses an optimization algorithm which rapidly runs several simulations while changing the parameter value to find the best fit. It follows from the No Free Lunch Theorem (NFLT), that an arbitrary search algorithm may achieve superior results on some problems, but then must pay with inferiority on other problems. No solution therefore offers a general "short cut", and implies that tailored solutions must be constructed. However, this has yielded over optimized solutions which are not easily adaptable to various reservoir simulators, and so the current challenge is to find sub-optimal adaptable methods that yield reasonable results. All in all, this makes History matching a challenging and integral aspect of the petroleum extraction industry.

In this chapter, we begin in the first section by defining the available data and observation functional, as well as the general nonlinear model as a function of the dynamic state and uncertain static parameter(s), and any necessary parameterizations there off. Thereafter, in the second section, the basic history matching problem is formulated in the Bayesian framework and the cost function of the basic data assimilation scheme is derived. Then in the third section we are lead to the formulation of the basic ensemble Kalman filter, and for which we define it's computational costs and some foreseeable issues. In the fourth section we analytically show the effect of the parameterizations of the static parameters analysis. The chapter concludes with the metrics that will be used to determine the effectiveness of subsequent ensemble methods as history matching procedures.

2.1. Data Assimilation Model

We begin by defining the model's dynamic state $\psi(x,t)$, which is comprised of the dynamic variables S^w, p^o each of which is a function of space and time:

$$\psi(x,t_i) = \begin{bmatrix} \vec{S^w}^i \\ \vec{p^o}^i \end{bmatrix} \in \mathbb{R}^{n_\psi} \text{ where } n_\psi = 2N_g$$

Thereafter, the model has a set of static parameters α, which can be parameterized. Here we apply a ln-parameterization to the Corey Parameter λ:

$$\alpha = \left[\ln\lambda\right] \in \mathbb{R}^{n_\alpha}, \quad \text{where } n_\alpha = 1$$

Here we opt to ln-parameterize λ in order to force that $\lambda \in [0,\infty)$. In Section 2.4, we will investigate the effect of this choice. Anecdotally, it is known that, spurious correlations in the data assimilation process yield $\lambda < 0$, which then corrupt (and crash) subsequent simulations. In (2.10) we will augment the dynamic state with the parameters states to form an ensemble member state:

$$\Psi(x,t_i) = \Psi_{t_i} = \begin{bmatrix} \psi(x,t_i) \\ \alpha \end{bmatrix} = \begin{bmatrix} \vec{S^w}^i \\ \vec{p^o}^i \\ \ln\lambda \end{bmatrix} \in \mathbb{R}^n, \quad \text{where } n = n_\psi + n_\alpha.$$

Each type of state is initialized with it's own "uncertainty". In the case of the dynamic state ψ, each component is perfectly known from their initial conditions. This means that we "draw" samples from a degenerate deterministic distribution, in which the known initial conditions have probability 1. In the case of the static state α, from Subsection 1.1.5, we reuse the $\lambda \sim Gamma(k,\theta)$ and then apply the ln-parameterization.

Component	Distribution	Parameters
$\psi = \begin{bmatrix} \vec{S^w} \\ \vec{p^o} \end{bmatrix}$	Deterministic	$\psi^0 = \begin{bmatrix} \vec{S^w}^0 \\ \vec{p^o}^0 \end{bmatrix} = \begin{bmatrix} 0.29 \\ p^r = 10^5 \end{bmatrix}$
$\alpha = \left[\ln\lambda\right]$	(log-)Gamma	$k = \frac{(\lambda^0_{guess})^2}{\sigma^2_{\lambda\lambda}}, \theta = \frac{\lambda^0_{guess}}{\sigma^2_{\lambda\lambda}}$

Table 2.1: Initial Uncertainties in the Dynamic and Static States of the Data Assimilation Model

Alternatively, for the dynamic state we could assume the sampling distribution to be (degenerate) Normal with mean equal to the initial conditions and variance 0. Similarly, for the static state we could directly have sampled from a log-gamma distribution (also known as exp-gamma) instead of first sampling from a gamma distribution and then applying the ln-parameterization. Technicalities aside, both formulations are equivalent for sampling the initial state. We define the nonlinear model operator $\mathcal{G}$ to produce forecasts, which in our case is the our simulator is discretized by the IMPES scheme (1.15). Typically, model operators have errors represented by an additive stochastic term $q(x,t) \sim \mathcal{N}\left(0, C_{qq}\right)$ wherein the covariance is known. However, we assume that the IMPES generates perfect forecasts which are devoid of forecast errors, i.e. $q(x,t) = 0 \; \forall x,t$.

$$\psi(x,t_i) = \mathcal{G}(\psi(x,t_{i-1}),\lambda) + 0 \tag{2.1}$$

Equivalently, we could assume that the transition distribution is, as above, a degenerate deterministic distribution, or alternatively, a degenerate multivariate normal with variance equal to 0. In this thesis we will generate synthetic data $d \in \mathbb{R}^m$ from the same nonlinear physical model $\mathcal{G}$:

$$d_{t_i} = \mathcal{G}(d_{t_{i-1}}, \lambda^{true}) \tag{2.2}$$

where $\lambda^{true} = 2.5$ is the true Corey Parameter value which we wish to estimate. However, although data is available at every time step, the observable data only measures the oil pressure in the first and last grid cell, see Figure 1.3. i.e. $m = 2$. Thus we assume that these data observations, d, can be related to a simulated realization,ψ, through the observation operator[1], $\mathcal{H} : \mathbb{R}^n \to \mathbb{R}^m$:

$$\mathcal{H}_i[\psi,\alpha] = \int_\Omega \int_0^T \psi(x)\delta_{\psi_i}\delta(t-t_i)\delta(x-x_i)\,dt\,dx$$

[1] also known as measurement functional

Given that in this thesis, we will only be measuring the oil pressure in the first grid cell and last grid cell at every time step, this reduces the observation operator to :

$$\mathscr{H}_i[\psi,\alpha] = \begin{bmatrix} p^o(x_1,t_i) \\ p^o(x_{N_g},t_i) \end{bmatrix} \in \mathbb{R}^m \quad \text{where } m=2$$

So at any one time, we have observations (measurements) at m locations of the true state $\psi^{true} = \psi(\lambda^{true})$ these are stored in the data vector $d \in \mathbb{R}^m$.

Given the simplicity of the observations, we will allow the abuse of the observation operator notation: $\mathscr{H}_i[\psi,\alpha] = \mathscr{H}_i[\Psi]$, as well as on multiple independent samples:

$$\mathscr{H}\left[\begin{pmatrix}\Psi^1 & \dots & \Psi^N\end{pmatrix}\right] = \begin{bmatrix}\mathscr{H}\left[\Psi^1\right] & \dots & \mathscr{H}\left[\Psi^N\right]\end{bmatrix} \in \mathbb{R}^{m\times N}$$

In addition, it is possible to reframe $\mathscr{H}$ as a matrix $H \in \mathbb{R}^{m\times n}$, in such a way that when applied to the abuse above:

$$H\begin{pmatrix}\Psi^1 & \dots & \Psi^N\end{pmatrix} = \begin{bmatrix}H\Psi^1 & \dots & H\Psi^N\end{bmatrix} = \begin{bmatrix}\mathscr{H}\left[\Psi^1\right] & \dots & \mathscr{H}\left[\Psi^N\right]\end{bmatrix} \tag{2.3}$$

Typically a measurement error is a combination of the error in the data and the error of the observation operator. However, in this thesis, as we have assumed, our synthetic data to be "perfect", i.e. error in the data is 0, the measurement error is equivalent to the observation error. Regardless, below we show that measurement error represent the error in the data and the observation operator. Given that, in general the data may not be perfect, (aleatoric) errors during collection are not uncommon[2], we define ϵ_{data} as the error in the data:

$$d = d^{true} + \epsilon_{data}, \qquad \epsilon_{data} \sim \mathscr{N}\left(0, C_{\epsilon_{data}\epsilon_{data}}\right)$$

In addition, there may also be (epistemic) error resulting from the observation operator $\mathscr{H}$. We denote such errors as $\epsilon_{observation}$:

$$\mathscr{H}[\psi^{true}] = d^{true} + \epsilon_{observation}, \qquad \epsilon_{observation} \sim \mathscr{N}\left(0, C_{\epsilon_{obs}\epsilon_{obs}}\right)$$

thus we can say that the data is related to the true state so long as $\epsilon = \epsilon_{data} - \epsilon_{observation}$:

$$d = \mathscr{H}[\psi^{true}] + \epsilon_{data} - \epsilon_{observation}$$

and so we assume $\epsilon \sim \mathscr{N}(0, C_{\epsilon\epsilon})$:

$$\implies d = \mathscr{H}[\psi^{true}] + \epsilon \tag{2.4}$$

Thus, the random measurement error ϵ represents both the errors in the data and the observation operator, and so defines the accuracy of the measurements. We define (2.4) as the Measurement Model, which relates the data and the observed state.

2.1.1. Summary of Data Assimilation Model

Static and Dynamic states:

$$\alpha = \left[\ln\lambda\right], \quad \psi^i = \begin{bmatrix}\vec{Sw}^i \\ \vec{po}^i\end{bmatrix} (= \psi(t_i)) \tag{2.5a}$$

Nonlinear Physical Model:

$$\psi(x,t_i) = \mathscr{G}(\psi(x,t_{i-1}),\lambda) \tag{2.5b}$$

Measurement Model:

$$d_t = \mathscr{H}[\psi(t),\alpha] + \epsilon \qquad \epsilon \sim \mathscr{N}(0, C_{\epsilon\epsilon}) \tag{2.5c}$$

where the initial Static and Dynamic states are:

$$\lambda \sim Gamma(k,\theta), \quad \psi^0 = \begin{bmatrix}\vec{Sw}^0 = 0.29 \\ \vec{po}^0 = p^r = 10^5\end{bmatrix} \tag{2.5d}$$

where $k = \frac{(\lambda^0_{guess})^2}{\sigma^2_{\lambda\lambda}}, \theta = \frac{\lambda^0_{guess}}{\sigma^2_{\lambda\lambda}}$

[2] i.e. mistakes happen in the field.

2.2. Bayesian Formulation

Below we formulate the basic history matching problem in the Bayesian framework, and derive the associated cost function.

2.2.1. Sequentially

Using Bayes' theorem the estimation problem can be written as:

$$f(\psi, \alpha \mid d) \propto f(\psi, \alpha) f(d \mid \psi, \alpha)$$

Given that in the dynamical model, the prior information of the initial states could be formulated into distributions. We take advantage of these priors $f(\psi^0)$ and $f(\alpha)$, and thus we replace $f(\psi, \alpha)$ with $f(\psi, \alpha, \psi^0)$. However, since we assume to know the exact initial dynamic state in so that the prior for $f(\psi^0)$ is considered to be a degenerate deterministic distribution, then it marginalizes out immediately:

$$f(\psi, \alpha, \psi^0) = f(\psi, \alpha) = f(\psi \mid \alpha) f(\alpha)$$

As such, in this thesis, we will ignore $f(\psi^0)$ and not explicitly condition on ψ^0. Hence,

$$f(\psi, \alpha \mid d) \propto f(\psi \mid \alpha) f(\alpha) f(d \mid \psi, \alpha)$$

Here the probability density function (or pdf) $f(\psi \mid \alpha, \psi^0)$ is the prior density for the dynamic state.

2.2.2. Discrete Formulation (time-wise)

Here $\psi^i = \psi(x, t_i)$ with $t_{i-1} < t_i < t_{i+1},\ i = 1, \ldots, K$. We define the pdf for the model integration from t_{i-1} to t_i

$$f(\psi^i \mid \psi^{i-1}, \alpha)$$

which assumes the model is a first-order Markov process. The joint probability density function for the dynamic and static state is now:

$$f(\psi^1, \ldots, \psi^K, \alpha) \propto f(\alpha) \prod_{i=1}^{K} f(\psi^i \mid \psi^{i-1}, \alpha)$$

Now we assume $\mathbf{d} \in \mathbb{R}^M$ can be divided into subsets $d_j \in \mathbb{R}^{m_j}$ collected at times $t_{i(j)}$, with $j = 1, \ldots, J$ and $0 < i(1) < i(2) < \ldots < i(J) < K$. The subset d_j only depends on $\psi(t_{i(j)}) = \psi^{i(j)}$ and for α. Furthermore, ε_j and t_j are uncorrelated. Hence,

$$f(\mathbf{d} \mid \psi, \alpha) = \prod_{j=1}^{J} f(d_j \mid \psi^{i(j)}, \alpha)$$

and so it follows from Bayes Theorem, the model is a first order Markov Process:

$$f(\psi^1, \ldots, \psi^K, \alpha \mid \mathbf{d}) \propto f(\alpha) \prod_{i=1}^{K} f(\psi^i \mid \psi^{i-1}, \alpha) \prod_{i=1}^{J} f(d_j \mid \psi^{i(j)}, \alpha)$$

Recall that in this thesis, we assume that $m_j = m = 2$ and $i = j : 0 < i, j \leq J, K,\ K = J$. That is, we observe and assimilate the same amount of data at every time step. Furthermore, we assumed that there is no model error in the forecast. This reduces the model:

$$\begin{aligned} f(\psi^1, \ldots, \psi^K, \alpha \mid \mathbf{d}) &\propto f(\alpha) \prod_{i=1}^{K} f(\psi^i \mid \psi^{i-1}, \alpha) f(d_i \mid \psi^i, \alpha) \\ &\propto f(\alpha) \prod_{i=1}^{K} \delta(\psi^i - \mathcal{G}(\psi^{i-1}, \alpha)) f(d_i \mid \psi^i, \alpha) \\ &\propto f(\alpha) \prod_{i=1}^{K} f(d_i \mid \psi^i, \alpha) \end{aligned} \tag{2.6}$$

Taking $\boldsymbol{\psi} = \psi^1, \ldots, \psi^K$ such that:

$$f(\boldsymbol{\psi}, \alpha \mid \mathbf{d}) \propto f(\alpha) \prod_{i=1}^{K} f(\psi^i \mid \psi^{i-1}, \alpha) f(d_i \mid \psi^i, \alpha)$$

2.2.3. Likelihood and Cost Function formulation

We begin by recalling the initial sampling distribution of $\lambda \sim Gamma(k,\theta)$:

$$f(\lambda) = \frac{\lambda^{k-1} e^{-\frac{\lambda}{\theta}}}{\theta^k \Gamma(k)}$$

then with parameterization $\ln\lambda$, $\alpha \sim \log Gamma(k,\theta)$[3]:

$$f(\alpha) = \frac{e^{k\alpha - e^\alpha/\theta}}{\theta^k\Gamma(k)} = \frac{\lambda^{k-1}e^{-\frac{\lambda}{\theta}}}{\theta^k\Gamma(k)} \left(\frac{d}{d\lambda}\ln\lambda\right)^{-1} \tag{2.7}$$

Hence, the conditional joint density at data assimilation J for time step T:

$$\begin{aligned} f(\boldsymbol{\psi},\alpha \mid \mathbf{d}) &\propto f(\alpha) \prod_{i=1}^{K=T} f(\psi^i \mid \psi^{i-1},\alpha) \prod_{j=1}^{J=T} f(d_j \mid \psi^{i(j)},\alpha) \\ &\propto \left(\theta^k\Gamma(k)\right)^{-1} \exp\left(k\alpha - e^\alpha/\theta\right) \cdot \prod_{i=1}^{K=T} \delta\left(\psi^i - \mathscr{G}(\psi^{i-1})\middle|\alpha\right) \cdot \dots \\ &\quad \underbrace{\cdot \prod_{j=1}^{J} (2\pi)^{-\frac{m}{2}} \det\left(C_{\epsilon\epsilon_{i(j)}}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(d_j - \mathscr{H}\left[\psi^{i(j)},\alpha\right]\right)^{\mathsf{T}} C^{-1}_{\epsilon\epsilon_{i(j)}} \left(d_j - \mathscr{H}\left[\psi^{i(j)},\alpha\right]\right)\right)}_{L(\boldsymbol{\psi},\alpha|\boldsymbol{d})} \end{aligned} \tag{2.8}$$

where $L\left(\boldsymbol{\psi},\alpha \mid \boldsymbol{d}\right)$ is the likelihood. If we assume that the posterior is a member of the exponential family[16] such that:

$$f(\boldsymbol{\psi},\alpha \mid \mathbf{d}) \propto \exp\left(-\frac{1}{2}\tilde{\mathscr{J}}[\boldsymbol{\psi},\alpha]\right)$$

then the cost function $\tilde{\mathscr{J}}$ is

$$\begin{aligned} \tilde{\mathscr{J}}[\boldsymbol{\psi},\alpha] &= \sum_{j=1}^{J} \left\| d_j - \mathscr{H}\left[\psi^{i(j)},\alpha\right]\right\|^2_{C^{-1}_{\epsilon\epsilon_{i(j)}}} + \sum_{j=1}^{J} \ln\det\left(C_{\epsilon\epsilon_{i(j)}}\right) + \dots \\ &\quad + mJ\ln(2\pi) + 2\left[\ln\Gamma(k) + \frac{e^\alpha}{\theta} - k(\alpha - \ln\theta)\right] \end{aligned} \tag{2.9a}$$

equivalently in terms of λ

$$\begin{aligned} \mathscr{J}[\boldsymbol{\psi},\lambda] &= \sum_{j=1}^{J} \left\| d_j - \mathscr{H}\left[\psi^{i(j)},\ln\lambda\right]\right\|^2_{C^{-1}_{\epsilon\epsilon_{i(j)}}} + \sum_{j=1}^{J} \ln\det\left(C_{\epsilon\epsilon_{i(j)}}\right) + \dots \\ &\quad + mJ\ln(2\pi) + 2\left[\ln\Gamma(k) + \frac{\lambda}{\theta} - k\ln\frac{\lambda}{\theta}\right] \end{aligned} \tag{2.9b}$$

Given that the cost function $\tilde{\mathscr{J}}$ is equal to the negative log likelihood ($\times\frac{1}{2}$), it is clear that the minimum of $\tilde{\mathscr{J}}$ is the maximum likelihood estimate (MLE) for ψ and α as defined by the conditional joint probability density function in (2.8). Fortunately, the MLE is invariant under parameterization and hence minimizing $\tilde{\mathscr{J}}$ is equivalent to minimizing $\mathscr{J}$ which also yields the maximum likelihood estimates of ψ and λ.

[3] This is also sometimes known as the ExpGamma distribution.

2.3. Ensemble Kalman Filter

We finally formulate the basic Ensemble Kalman filter. Then we define it's computational costs. We also comment on some issues with the filter.

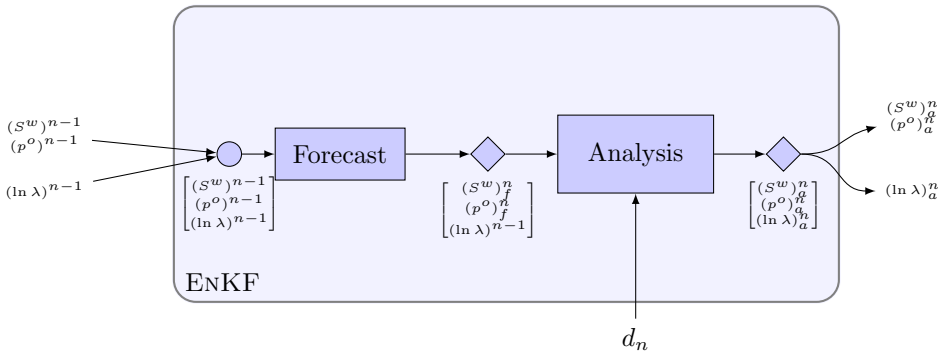


Figure 2.1: Ensemble Kalman Filter (EnKF) at nth Data Assimilation

2.3.1. Ensemble Formulation

This leads us to the Ensemble Kalman filter, where we define ensemble matrix $A \in \mathbb{R}^{n\times N}$ to be composed of an ensemble of model states $\psi(x,t_i) \in \mathbb{R}^{n_\psi}$, at a time t_i. Furthermore, in the case when we use the Ensemble Kalman filter to estimate parameters we augment the ensemble of states with the set of poorly known parameters $\alpha \in \mathbb{R}^{n_\psi}$. Thereby, our ensemble matrix A is a combination of dynamic variables $\psi(x,t_i)$ and static parameters $\alpha(t_i)$:

$$A_i = A(x,t_i) = \begin{bmatrix}\Psi^1 & \cdots & \Psi^N\end{bmatrix} = \begin{pmatrix}\psi^1(x,t_i) & \dots & \psi^N(x,t_i)\\ \alpha^1(t_i) & \dots & \alpha^N(t_i)\end{pmatrix} \in \mathbb{R}^{n\times N} \tag{2.10}$$

where $n = n_\psi + n_\alpha$ is the total dimension of the state vector and N is the number of realizations in the ensemble. Note that the parameters α are expected to be constant in time. However, in order to distinguish between the estimates at different update times, we allow the parameters to be dependent on time.
The ensemble covariance is defined as

$$C_{\psi\psi} = \overline{(\psi-\overline{\psi})(\psi-\overline{\psi})^\intercal}$$

where $\overline{\psi}$ is the ensemble mean, **is regarded as the best-guess estimate**, while the ensemble spread defines the error variance. The covariance is determined by the **smoothness of the ensemble members.** The ensemble mean is stored in each column of

$$\bar{A}(x,t_i) = A(x,t_i)\mathbf{1}_N$$

where $\mathbf{1}_N \in \mathbb{R}^{N\times N}$ where each element is $\frac{1}{N}$. We can define the ensemble perturbation matrix as

$$\begin{aligned}A'(x,t_i) &= A(x,t_i) - \bar{A}(x,t_i)\\ &= A(x,t_i)(I-\mathbf{1}_N)\end{aligned}$$

The ensemble covariance

$$C^e_{\psi\psi}(x_1,x_2,t_i) = \frac{A'(x,t_i)(A'(x,t_i))^\intercal}{N-1} \in \mathbb{R}^{n\times n}$$

Given a vector of measurements $d \in \mathbb{R}^m$, where m is the number of measurements, we define the N vectors of perturbed observation as

$$d_j = d + \epsilon_j, \qquad j = 1,\dots,N$$

which are stored in the columns of the matrix

$$D = (d_1, d_2, \dots, d_N) \in \mathbb{R}^{m\times N}$$

while the zero mean ensemble of perturbations are stored in the matrix

$$E = (\epsilon_1, \epsilon_2, \dots, \epsilon_N) \in \mathbb{R}^{m\times N}$$

from which we construct the ensemble representation of the measurement error covariance matrix

$$C^e_{\epsilon\epsilon} = \frac{1}{N-1}EE^\intercal \in \mathbb{R}^{m\times m}$$

The analysis equation can then be expressed in terms of the ensemble matrices:

$$A^a = A + \mathscr{H}^\intercal[C^e_{\epsilon\epsilon}]\left(\mathscr{H}^\intercal[\mathscr{H}[C^e_{\epsilon\epsilon}]] + C^e_{\epsilon\epsilon}\right)^{-1}(D - \mathscr{H}[A])$$

We define the ensemble of innovation vectors as

$$D' = D - \mathscr{H}[A]\,, \qquad D' \in \mathbb{R}^{m\times N}$$

Next the measurements of the ensemble perturbations is defined by $S \in \mathbb{R}^{m\times N}$.

$$S = \mathscr{H}[A']$$

Then matrix $C \in \mathbb{R}^{m\times m}$:

$$C = SS^\intercal + (N-1)C_{\epsilon\epsilon}$$

here we can use the full-rank, exact measurement error covariance matrix $C_{\epsilon\epsilon}$, or the low-rank representation $C^e_{\epsilon\epsilon}$. Using $D', S, C, C^e_{\epsilon\epsilon}, C^e_{\psi\psi}$, the analysis can be expressed as

$$\begin{aligned} A^a &= A + A'S^\intercal\left(SS^\intercal + EE^\intercal\right)^{-1}D' \\ &= A + A(I - 1_N)S^\intercal C^{-1}D' \\ &= A(I + (I - 1_N)S^\intercal C^{-1}D') \end{aligned}$$

where the matrix $X \in \mathbb{R}^{N\times N}$ is defined as $X = I + (I - 1_N)S^\intercal C^{-1}D'$. When we assume $1_N S^\intercal \equiv 0$ then clearly $X = I + S^\intercal C^{-1}D'$.

$$\begin{aligned} &= A(I + S^\intercal C^{-1}D') \\ &= AX \end{aligned} \tag{2.11}$$

2.3.2. Computational Costs

In this subsection we present the pseudo code of the Ensemble Kalman filter, as well as some computational costs. In this thesis we ran simulations in tensorflow which would automatically run operations on a Nvidia 6GB Geforce GTX 1660 with Nvidia's CuDNN libraries. Below we give the pseudo code for the EnKF update before defining the EnKF as a Data Assimilation loop:

Algorithm 1: A Single Ensemble Kalman Update. See [21] for computational complexity.

1 N :=Ensembles Size
2 $1_N := \frac{1}{N}J_N$ where J_N is a $N \times N$ matrix of ones
3 Pre Compute: $I - 1_N$ /* $\mathcal{O}\left(N^2\right)$ */
4 **EnKF Update** $\mathscr{E}(A, D_{obs})$**:**
5 $A' = A(I - 1_N)$ /* $\mathcal{O}\left(nN^2\right)$ */
6 Observation Model:
7 $S = \mathscr{H}(A')$ /* gather from nN to mN */
8 Generate Gaussian Noise:
9 $E \sim \mathcal{N}(0, C_{\epsilon\epsilon})$
10 Add Noise to Measurements:
11 $D = D_{obs} + E$ /* $\mathcal{O}(mN)$ */
12 Innovation Vectors:
13 $D' = D - S$ /* $\mathcal{O}(mN)$ */
14 Construct C:
15 $C = SS^\intercal + EE^\intercal$ /* $\mathcal{O}\left(m^2N\right), \mathcal{O}\left(m^2\right), \mathcal{O}\left(m^2N\right)$ */
16 $X = I + S^\intercal C^{-1}D'$ /* $\mathcal{O}\left(N^2\right), \mathcal{O}\left(mN^2\right), \mathcal{O}\left(m^3N\right)$ */
17 Analyzed Ensemble:
18 $A^a = AX$ /* $\mathcal{O}\left(nN^2\right)$ */
19 **return** A^a

Immediately, it is clear that much of the problem is delightfully parallelizable. Lines 11 and 13 do not require data from any other ensemble members. Meanwhile the remaining are composed of matrix multiplications, or some sort of scatter and or gather operation, which can be done in parallel across compute nodes. However, with larger ensemble sizes there may be an issue with loss of precision, due to the large number of reducing operations involved.

Algorithm 2: Ensemble Kalman Filter

N :=Ensembles Size
$\lambda \sim Gamma(k,\theta)$
$\alpha := \ln \lambda$
$A^a = \begin{bmatrix} \vec{S^w}^0 & \dots & \vec{S^w}^0 \\ \vec{p^o}^0 & \dots & \vec{p^o}^0 \\ \alpha^1 & \dots & \alpha^N \end{bmatrix}$
for *i timesteps* **do**
  Forecast Step:
  $A = \begin{bmatrix} \mathcal{G}(A^a;\alpha) \\ \alpha \end{bmatrix}$ /* $\mathcal{O}\left(\chi(N_g)\right)$ */
  Load measurement data:
  $D_{obs} = [d_1,\dots,d_N]$ /* scatter from m to mN */
  Analysis Step:
  $A^a = \mathcal{E}(A, D_{obs})$
end

2.3.3. Foreseeable Issues

Probabilistic Issues:
When the prior ensemble is non-Gaussian, the analyzed ensemble will inherit some of the non-Gaussian structures. It is also possible that the EnKF fails completely; e.g. if the weight on the prior is low and a multi-modal probability density function[16].

1. The issue with non Gaussian initial conditions is that they propagate into the next state and eventually become Gaussian provided that the nonlinearity of the physics is not too strong. This follows from the Likelihood function (2.9). In this thesis, we overcome this issue by having a high sampling rate for the data i.e. you can linearize almost anything if you are "fast" enough.

2. Another issue maybe non Gaussian posteriors; this could be an artifact from the initial sample, or an inherent outcome of the nonlinear physics. Why is this a problem? Firstly consider a posterior which is skewed to the point that a fat tail emerges, in which case the mean is drawn off towards the tail. Evidently then a Gaussian assumption may be able to capture the first 2 moments, i.e. the mean and variance, but fails to accurately represent the true density of an ensemble. This could result in a permanent bias in the estimate developing. For example, in this thesis, we use the gamma distribution over the normal distribution, wherein the mean of the gamma distribution is drawn towards the tail. Secondly, consider that the posterior maybe be multi-modal. For simplicity consider a posterior which is bimodal with thin tails and symmetric around the mean. In such a situation the mean would be accurately captured, but the probability of occurring would be vastly overestimated, whilst the probability of either mode occurring is underestimated. On the other hand, it is possible the ensemble gets stuck in the wrong mode and is unable to escape, falsely indicating a bias, and possibly resulting in degeneracy as the ensemble estimates attempt to escape.

Numerical Issues:
If measurements are nearly dependent or the quantity of measurements, $m \gg N$, is just too large then it is possible the required inversion of C (seen in (2.11)) will become numerically singular[17]. Hence, this implies that there is a minimum measurement noise required to be numerically stable. On the other hand, in the case of $m \ll N$, the needed correlations to update the ensemble may not be present, and spurious correlation may occur[17]. Alternatively, there may not be enough "information" in the data to provide an accurate estimate of the parameter λ. We will investigate this case in upcoming Subsection 3.4.1.

Issues with Updating the Physics:
Finally, it is clear that the multiplicative update of the Ensemble Kalman filter does not ensure conservation of mass in it's ensemble members. In [25], it is shown that in the mean sense there is conservation of mass. However, the example given does not augment the ensemble with static parameters. As a result it is unclear, even in the mean sense that the Ensemble Kalman filter is conservative when jointly estimating the dynamic state and static parameters. As a result we will further investigate a possible solution to this in Chapter 4.

2.4. Ensemble State Parameterizations

In the fourth section we analytically show the effect of the parameterizations of the static parameters analysis. Before progressing further, we first recall the ln-parameterization of the unknown parameter λ in the ensemble state. Firstly, recall that $\ln\lambda$ was taken to ensure that filter corrections would yield positive estimates of λ. Secondly, realize that the ln-parameterization is a variance-stabilizing transformation [9]. Thus by applying the delta method[20] we can approximate the immediate effect of the parameterization on the initial sample. Let λ be a random variable, with $\mathbb{E}[\lambda] = \lambda_{guess}$ and $\text{Var}[\lambda] = \sigma^2_{\lambda\lambda}$. Now given that $\alpha = \ln\lambda$, we take the 1st order Taylor approximation:

$$\ln\lambda \approx \ln\lambda_{guess} + \frac{\lambda - \lambda_{guess}}{\lambda_{guess}}$$

Thus we find that

$$\mathbb{E}[\ln\lambda] = \ln\lambda_{guess} \text{ and } \text{Var}\left[\ln\lambda_{guess}\right] \approx \frac{\sigma^2_{\lambda\lambda}}{\lambda^2_{guess}} \tag{2.12}$$

So immediately we see that taking the initial ln variance is scaled by the square of the mean initial guess. In the case $\lambda > 1$, the parameterized variance shrinks and may be far too small and lead to degeneracy. On the otherhand, if $\lambda < 1$ the parameterized variance "blows up" as $\lambda \to 0$. [4]

Thereafter, $\ln\lambda$ causes the multiplicative update of the EnKF to become a power update. As an illustrative example consider the single element case where a is multiplicatively updated by $1+\epsilon$ where in the EnKF ϵ is computed from the ensemble correlations and the data but for now we assume that the updates are the same regardless of the influence of the parameterization on the computation of the update (i.e. via the correlation computation) and that $\epsilon \to 0$. Applying the $\ln(a)$ parameterization we see that the update $a(1+\epsilon)$ becomes $a^{1+\epsilon}$. Now we wish to know for what values of a and ϵ is the update less potent, e.g. $a(1+\epsilon) > a^{1+\epsilon}$? Thus taking $a = e$ we see that when $\epsilon > 0$ then the update produced by the parameterization is greater than without, whilst when $\epsilon < 0$ the converse occurs. Simply, when a needs to be shrunk, $\epsilon < 0$, the parameterized parameter is shrunk less. Whilst when a needs to be enlarged, $\epsilon > 0$, it grows more. Note that since the update of the parameterized parameter is a power term, the size of a inherently affects the potency of the update. As such taking $a > e$, in the case $\epsilon > 0$ we see that the enlargement of the parameterized parameter scales with the size of a, to the extent that for $\epsilon \in (\sqrt[e]{1+\epsilon} = a, 0)$ the parameterized parameter is able to shrink more than when unparameterized. On the other hand for $0 < a < e$ the converse occurs, wherein for $\epsilon \in (-\infty, \sqrt[e]{1+\epsilon} = a)$ both the shrinkage and enlargement of the parameterized parameter is less potent than when unparameterized.

Granted, this illustrated example brings forth many combinations of the parameter a and the update ϵ, which is a little difficult to wrap ones head around. So a general (but a bit misleading) rule of thumb is that the $\ln(a)$ parameterization is gentler for sufficiently small updates of a, but steeper for sufficiently large updates of a. This is important to note as if conducted recursively then the unparameterized a could yield a more accurate final value.

However, returning to the EnKF the update is a $N \times N$ matrix and $A \in \mathbb{R}^{n\times N}$. On an element level there is a re-weighting and then a summation across the re-weighted ensemble elements:

$$AX = \left[\begin{array}{ccc} \vdots & \cdots & \vdots \\ \psi_1 & \cdots & \psi_N \\ \vdots & \cdots & \vdots \\ \hline \alpha_1 & \cdots & \alpha_N \end{array}\right] \left[\begin{array}{ccc} \vdots & \cdots & \vdots \\ x_1 & \cdots & x_N \\ \vdots & \cdots & \vdots \end{array}\right]$$

We denote the last row as $\vec{\alpha}$. Then examining only the update of the first element in this row, i.e. the update of α_1:

$$\langle\vec{\alpha}, x_1\rangle = \sum_{i=1}^{N} \alpha_i x_{1,i}$$

Now in the case of no parameterization, i.e. $\alpha = \lambda$:

$$= \sum_{i=1}^{N} \lambda_i x_{1,i}$$

[4] Furthermore, consider that we rewrote the Gamma(k,θ) sampling distribution in terms of only the initial guess (mean) λ^0 and scale θ, such that the shape $k = \frac{\lambda^0}{\theta}$. Then given that $\text{Var}[\ln\lambda] = \psi^{(1)}(k) = \psi^{(1)}\left(\frac{\lambda^0}{\theta}\right)$, where $\psi^{(1)}$ is the trigamma function. Later we use this trigamma formulation to compute $\text{Var}[\ln\lambda]$ directly.

However, with the $\ln(\cdot)$ parameterization, i.e. $\alpha = \ln\lambda$:

$$\langle \vec{\alpha}, x_1 \rangle = \sum_{i=1}^{N} \ln(\lambda_i) x_{1,i} = \sum_{i=1}^{N} \ln\left(\lambda_i^{x_{1,i}}\right) = \ln\left(\prod_{i=1}^{N} \lambda_i^{x_{1,i}}\right)$$

Now assuming that $x_{ij} > 0$, then it is clear by applying Jensen's inequality:

$$\ln\left(\sum_{i=1}^{N} \lambda_i x_{1,i}\right) > \ln\left(\prod_{i=1}^{N} \lambda_i^{x_{1,i}}\right)$$

This is an application of the arithmetic mean-geometric mean inequality which guarantees that the parameterized mean is bounded above by the unparameterized mean. On the other hand, when $x_{ij} \leq 0$ or a mixed case where $x_{ij} \leq 0$ and $x_{ik} > 0$ the inequality no longer holds. But, given an ensemble that encapsulates the true λ value and the physical requirement that $\lambda > 0$, then this bound will eventually hold! Thereby, we know that the skewness of the parameterization does not affect the ensemble mean but in fact improves it, though this is not necessarily the case for the spread.

2.5. Performance Metrics

The chapter concludes with the metrics that will be used to determine the effectiveness of subsequent ensemble methods as history matching procedures. Prior to running the twin experiments we define some traditionally used performance metrics which we focus on: Bias, Spread, the mean squared error, and the root mean squared error. We begin by defining the bias, given an estimator $\hat{\theta}$ of θ_{true}:

$$\left\| Bias(\hat{\theta}) \right\|_2^2 = \left\| \mathbb{E}\left[\hat{\theta}\right] - \theta_{true} \right\|_2^2 = \left\| \bar{\theta} - \theta_{true} \right\|_2^2$$

It is good to note that "with a few exceptions, implementing a simulation with a comfortably small bias is easier than with a comfortably small standard error"[20]. This quantifies any systematic errors incurred during the estimation process.

Another measure of the performance is the ensemble spread, which is the square root of the averaged variance of the ensemble, defined as

$$SPREAD = \sqrt{\mathbb{E}\left[\left\| \hat{\theta} - \mathbb{E}\left[\hat{\theta}\right] \right\|_2^2 \right]} = \sqrt{\mathrm{tr}(\mathrm{Var}\left[\hat{\theta}\right])}$$

This quantifies any statistical dispersion which in effect quantifies the uncertainty of the estimate.

The Mean Squared Error, or MSE, is calculated as the average squared difference between the actual and predicted values. Given that it is the second moment of the error, and thus incorporates both the variance of the estimator and its bias.

$$MSE = \mathbb{E}\left[\left\| \theta_{true} - \mathbb{E}\left[\hat{\theta}\right] \right\|_2^2 \right] = \mathrm{tr}(\mathrm{Var}\left[\hat{\theta}\right]) + \left\| Bias(\hat{\theta}, \theta_{true})) \right\|^2$$

Typically, choices involve trade-off between bias and variance, and as such it represents a combined metric of the two. Finally, the Root Mean Square Error, RMSE, is a measure of accuracy, to compare forecasting errors of different models for a particular dataset and not between datasets, as it is scale-dependent:

$$RMSE = \sqrt{\mathbb{E}\left[\left\| \theta_{true} - \mathbb{E}\left[\hat{\theta}\right] \right\|_2^2 \right]}$$

It is good to note that, the effect of each error on RMSE is proportional to the size of the squared error; thus larger errors have a disproportionately large effect on RMSE. Consequently, RMSE is sensitive to outliers.

2.6. Chapter Summary

In this chapter we began by stating that our dynamic state comprises of $\psi = [S^w, p^o]^\intercal$ from our nonlinear model $\mathscr{G}$. This dynamic state is augmented by the static parameters α, to form an ensemble member state, Ψ. However, the only available data, d, is of the oil pressure, p^o, located in the first and last control volume of the domain. This data is related to forecasts via an observation operator $\mathscr{H}$ which is perturbed by a measurement noise $\epsilon \sim \mathscr{N}\left(0, \sigma_{p^o} = 14000\right)$ present. Thereafter, in the second section, the basic history matching problem is formulated in the Bayesian framework and the cost function of the basic data assimilation scheme is derived as:

$$\mathscr{J}[\boldsymbol{\psi}, \lambda] = 2\left[\ln\Gamma(k) + \frac{\lambda}{\theta} - k\ln\frac{\lambda}{\theta}\right] + \sum_{j=1}^{J}\left[m\ln\left(2\pi\det\left(C_{\epsilon\epsilon_{i(j)}}\right)\right) + \left\|d_j - \mathscr{H}^o\left[\psi^{i(j)}, \ln\lambda\right]\right\|^2_{C^{-1}_{\epsilon\epsilon_{i(j)}}}\right]$$

Consequently, in the third section, we arrive to the formulation of the basic Ensemble Kalman filter, and are able to highlight how parallelizable the filter is. At the same time, we bring up some of the foreseeable issues: non Gaussian prior or posterior distributions which may cause the filter to fail, or the underlying Gaussian assumption may poorly estimate the true posterior. Numerically, C^{-1} may cause stability issues should there be too many measurements or the ensemble too accurate. On the other hand, too few measurements may not have enough information to accurately estimate λ. Finally, the update may not reflect the underlying physics and result in unphysical corrections. In the fourth section we showed analytically that although the ln-parameterization improves the ln estimate in the mean, it may blow up or overly shrink the uncertainty/variance of the estimate. The chapter concludes with the 4 metrics that will be used to determine the effectiveness of subsequent ensemble methods as history matching procedures; bias, spread, MSE, RMSE. The first two represent systematic error and statistical dispersion of the ensemble, whilst the 3rd combines them. Finally the 4th is for accuracy and used to compare estimators on a shared dataset.

3

Effectiveness of EnKF for Parameter Estimation

In chapter 1 we determined that due to the nonlinearity of the two-phase flow model,(1.17), parameter estimation was intractable by direct methods. Thus, in chapter 2, we are forced to resort to numerical methods such as the Ensemble Kalman filter (EnKF). In this chapter, we aim to determine the capability and effectiveness of the EnKF as a data assimilation method, in order to estimate an uncertain static parameter $\alpha = \ln\lambda$. In doing so, we endeavor to estimate unknown "true" λ. In our data assimilation model (2.5), we define our nonlinear physical model and measurement model in which we start the EnKF from fixed initial conditions for the dynamic state variables (S^w, p^o) and a fixed random sample of the unknown λ. The data d to be passed to the EnKF is synthetically generated ahead by (2.2) given a to be estimated "true" λ. We define the base twin experiment as; trying to estimate $\ln\lambda$, where $\lambda^{true} = 2.5$, given: an initial sample guess $\lambda_{guess} = 6.5$, the dynamic state variables (S^w, p^o) start from a fixed initial condition of $S^{w0} = 0.29$, $p^{o0} = p^r = 10^5$, and the number of control volumes(i.e. grid size) $N_g = 50$. Accompanying this, we generate synthetic data exactly for the $\lambda^{true} = 2.5$, which then yield measurements that only observe the oil pressure in the first and last grid cell and have measurement (/additive white Gaussian) noise of $\sigma_{p^o} = 14000$. From this base twin experiment we construct different experiments by varying one of these "settings" to understand the EnKF when applied to this nonlinear physical model.

Furthermore, we investigate if under the base filter settings the EnKF is capable of estimating λ. In the first section, we apply the EnKF to datasets created with different values of λ^{true}. In section 2, we investigate the effect of the initial sample where, first we investigate the effect of the mean initial guess and then the initial sample spread. In section 3, we investigate whether the EnKF is robust enough to filter the "true" λ^{true} when the fixed injection rate c is incorrectly set in the physical model (2.1). Having identified some effects of the nonlinearity on the EnKF fitting, we then aim to understand in section 4, how some filter "settings" affect the final estimate. Specifically we investigate: How does the ensemble size affect the final estimate? Does the parameterization affect the accuracy? What is the effect of the measurement noise? Does the grid size of the simulator, constructed with the IMPES scheme, (1.15), effect the final estimate?

3.1. Analysis of EnKF Parameter Estimates

We start the EnKF from fixed initial conditions for the dynamic state variables (S^w, p^o) and a fixed random sample of the unknown λ. We synthetically generate 4 datasets representing 4 values of λ each of which the EnKF needs to be able to estimate. The first of the 4 datasets is generated using the parameter base test value $\lambda_{base} = 2.5$, then denoted as d_{base}. The remaining datasets 3 are generated using parameter values selected from the initial sample of λs'. Specifically, we pick the smallest, $\lambda_{smallest} \approx 0.3$, and largest, $\lambda_{largest} \approx 20$, as well as the value closest to the base test value, $\lambda_{closest} \approx 2.47$. Having then applied the EnKF to each dataset then results in estimates of λ. By examining the ensemble spread, statistical bias and root mean squared error we can then determine whether the filter is capable of correctly estimating the parameter used to generate the respective dataset and under what certainty and given what (ensemble) bias. Together these metrics determine for which values of λ the EnKF is able to provide an accurate estimate given the nonlinearity of the physical model.

In Figure 3.1, in the top subplot we see the ensemble spread of λ being fitted to each dataset. Here we see using d_{base} the EnKF is able to reconstruct λ_{base}, which in fact yields the smallest root mean squared error and statistical bias, as seen in the lower row of subplots. Similarly, for $d_{closest}$ applying the EnKF performs well but renders a slightly less accurate result, under each metric. This could be the luck of the draw of the measurement noise or the initial sample favors λ_{base}. However, we can say that the overall behavior is very similar to λ_{base} but not enough to be mistaken for it. Concretely, the λ_{base} value does not fall into the ensemble spread of $d_{closest}$, but the dynamics are sufficiently similar that, provided a dataset generated from a value smaller than λ_{base}, a linear interpolation would be able to provide a sufficient estimate for λ_{base}.

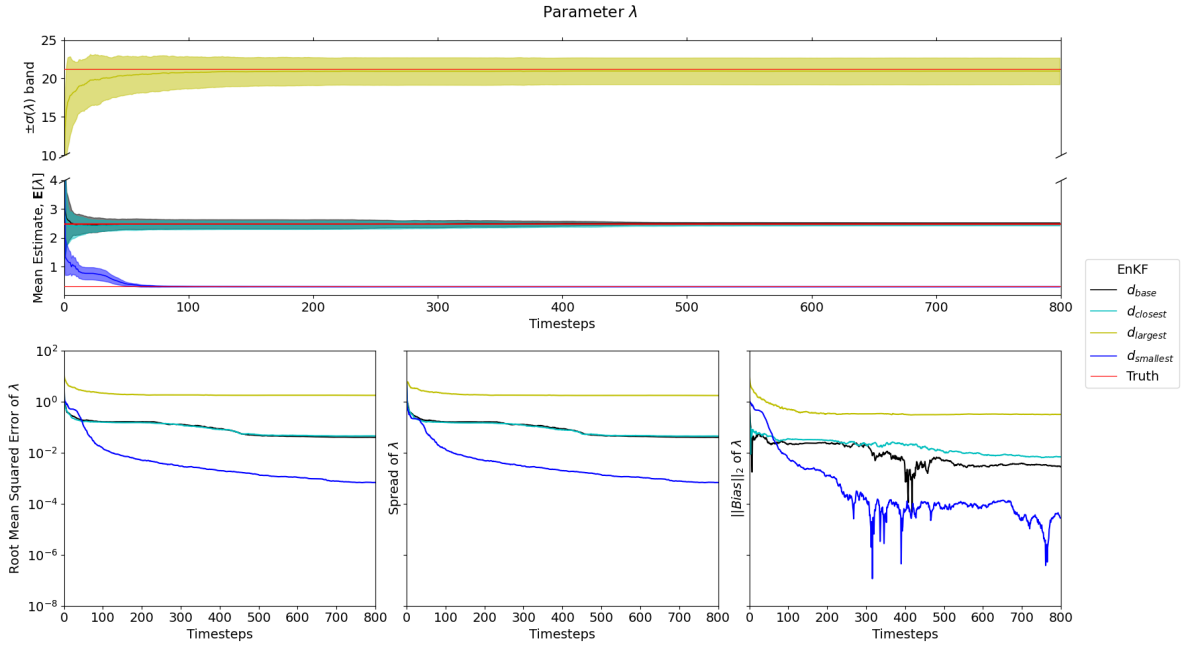


Figure 3.1: Synthetic datasets fitted by the EnKF where each dataset is generated using a different values of λ: (black) λ_{base}, (cyan) $\lambda_{closest}$ [closest to base], (yellow) $\lambda_{largest}$, (blue) $\lambda_{smallest}$, (red) True λ values

In the case of $d_{largest}$, generated by $\lambda_{largest}$, we see an enormous increase in the ensemble spread. This in part is due to the initial sample being very skewed against it, with the ensemble mean being far smaller. On the other hand, we also notice that the ensemble spread stabilizes much faster, indicating that due to the nonlinearity larger values of $\lambda_{largest}$ result in more uncertain predictions, i.e. larger spreads. Finally, in the case of $d_{smallest}$, generated by $\lambda_{smallest}$, the ensemble spread is much tighter than when estimating for larger values of λ. In this "lucky" sample the statistical bias continuously decreases almost monotonically. In general however, it has been observed that an ensemble bias, when the value of $\lambda_{smallest}$ is more than 1 standard deviation from the ensemble mean, is likely to form or the ensemble will become degenerate. Other choices for λ's sampling distribution and/or parameterization exist, for example a truncated normal distribution and/or clipping λ outside the range of the initial guess, e.g. $(0, 10)$. However, ancillary experiments did not produce reliable results when estimating the parameter, such as bias or filter failure, see appendix A.1.

3.2. The effect of the initial sample

In this section we wish to better understand the effect of the initial mean guess and spread on estimating the parameter λ. That is, firstly given a fixed initial spread how does the initial mean guess affect the accuracy of the estimate? Secondly, given a fixed initial mean guess, does a tighter initial spread yield a more certain estimate and thereby more accurate estimate? In either case would this differ given the value of the estimated parameter λ? To address these, we begin by generating two synthetic datasets, d_{guess} and d_{base} generated with λ_{guess} and λ_{base} respectively.

3.2.1. The effect of the initial mean guess

To answer the first question, we set a fixed initial sample spread, $\sqrt{\text{Var}[\lambda]} = 3.2$, and then sample 2 initial ensembles for each initial mean guess, $\mathbb{E}[\lambda] = \lambda_{base}$ and $\mathbb{E}[\lambda] = \lambda_{guess}$. Together with the 2 datasets, we construct a 2×2 experimental setup where the effect of fitting the EnKF given an accurate or inaccurate initial mean guess for 2 different parameter values is identified. In Figure 3.2, the top plot consists of the parameter mean estimate surrounded by the 1 standard deviation spread, with the bottom row of plots being the 3 metrics of interest: the statistical bias, the spread, and the root mean squared error. In the case of the estimating λ_{base} given an accurate initial mean guess of λ_{base}, denoted as $\lambda_{base}|\lambda_{base}$, the true value $\lambda = 2.5$ should already be the ensemble mean. Granted due to the ensemble sample size and numerical sampling size issues this is almost the case. Regardless we see that true value is always inside the ensemble spread and the mean ensemble value remains at most $1/4$ of the spread the true value.

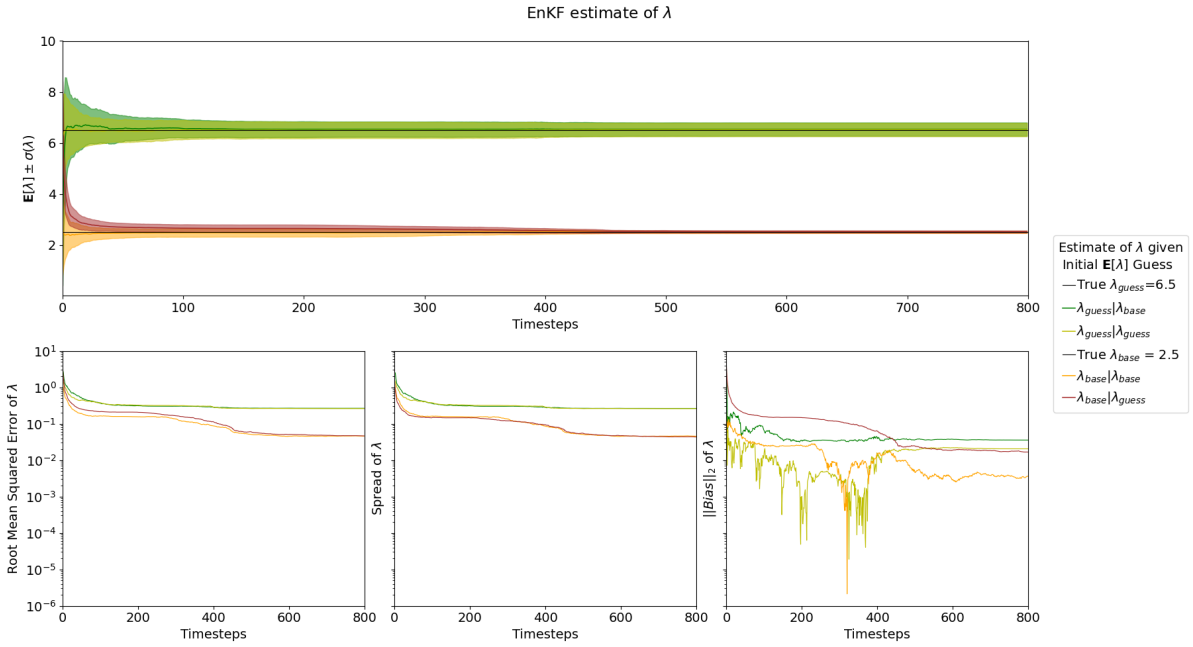


Figure 3.2: The effect of 2 initial mean guesses on the EnKF fitting of 2 datasets.

In the case of λ_{guess}, when the initial mean guess is accurate, i.e. $\lambda_{guess}|\lambda_{guess}$, and there we see a similar oscillatory behavior but instead the mean estimate levels out and remains rather large. In additions, with an inaccurate initial mean guess, i.e. $\lambda_{guess}|\lambda_{base}$, the ensemble is able to capture the true value within its spread much faster. This could be a result of the parameterization but may also be caused by the nonlinearity of the model. Recalling the physical interpretation of the Corey parameter λ relates the distribution index of pore sizes within the rock matrix, with large values of λ representing a very narrow distribution. As such, there is less variation between large values of λ which results in the EnKF update being less accurate as the correlation between ensembles computed from each λ is less clear. In contrast, the case of an inaccurate initial mean guess, i.e. $\lambda_{base}|\lambda_{guess}$, the spread shrinks too quickly and fails to encompass the true solution initially. This is most likely due to the fixed initial conditions of the dynamic state (I.C. S^w, p^o), which have yet to sufficiently have the parameter uncertainty propagate into them via the nonlinear model. If we consider

the initial ensemble state from a correlation standpoint then it is clear that at the first data assimilation there has been only a single forecast for the parameter uncertainty to propagate into ensemble state. With regard to the ensemble spread, there is almost unnoticeable difference when given an initial sample with an accurate or inaccurate mean guess. Indicating that there exists a minimum learnable lower bound on the spread. Overall, the root mean squared error indicates that the quality of the final estimate is not determined by how accurate your initial mean guess is but if there is sufficient data.

3.2.2. The effect of the initial sample spread

To answer the second question, we first return to Figure 3.1 where we reuse the λ_{base} estimate. In addition, we extract the initial sample and from this starting point in order to estimate λ_{guess} with the EnKF using dataset d_{guess}. Now recall that the spread on this initial sample was much larger than in the experiments of Figure 3.2, which was $\text{Var}[\lambda] = 27.5$ versus $\text{Var}[\lambda] = 3.2$ in Figure 3.1. Hence, we can now compare estimates of $\lambda_{base}|\lambda_{guess}$ and $\lambda_{guess}|\lambda_{guess}$ against the estimates of λ_{base} and λ_{guess}. Given that all these estimates have an initial mean guess of $\mathbb{E}[\lambda] = \lambda_{guess}$, without having done too much we can compare the effect of the initial spread on the final estimate. This reusing of experimental data allows for a clear consistent sequence of experiments and results whilst also being resource efficient.

In Figure 3.3, we have the same plot layout as in Figure 3.2. Clearly, when estimating λ_{guess} the initial advantage of having smaller initial spread quickly vanishes. Thereafter, we re encounter the lower bound of the spread. On the other hand, when estimating λ_{base}, we see that the larger initial spread is able to capture the true value within it the ensemble spread faster than when the initial spread is too small to capture the true value to begin with. In effect with a smaller initial spread the EnKF has difficulty locating the true value. This is observed in the statistical bias as the larger initial spread yields a smaller statistical bias, than a smaller initial spread leading to a larger statistical bias. Regardless of the initial spread, the final ensemble spread is the same when estimating a given λ.

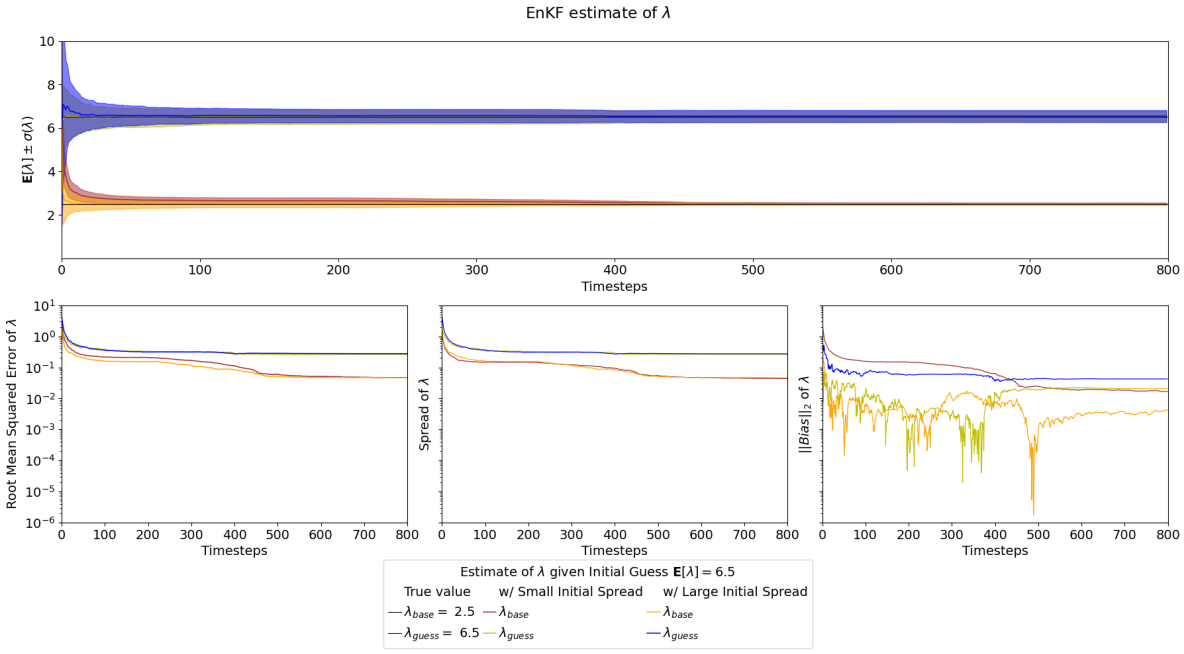


Figure 3.3: The effect of 2 initial spreads guesses on the EnKF fitting of 2 datasets.

This highlights that in situations where very little is known about the feasible range or variance of an unknown parameter, the EnKF is still able to identify up to a minimum degree of accuracy. Alternatively, the EnKF is also able to overcome overconfident inaccurate guesses, in which both the initial spread and mean guess fail to capture the unknown parameter. Overall, this spread lower bound we encounter renders the effect of the initial spread meaningless on the final accuracy of the ensemble as it dominates the root mean squared error.

3.3. Robustness interlude: mis-specified injection rate

In the previous experiment we remarked that the EnKF was capable of identifying an unknown parameter up to a minimum degree of accuracy. However, this is under the assumption that the rest of the model is entirely correct. It is entirely feasible that the underlying model is mis-specified. Hence, as a check of robustness of the EnKF, we mis-specify the value of a fixed parameter that we are able to control.

In this case, we start with the base experimental setup and add to the injection rate of the water phase a small fixed offset δc when applying the EnKF to estimate λ. We pick the injection rate specifically because it introduces additional epistemic uncertainty in the form of a model bias. This specific kind of model bias can easily be calibrated in a lab environment.

In contrast, choosing a parameter related to the rock, for example the absolute permeability K, would introduce a kind of model uncertainty that in a lab environment be considered part of the underlying irreducible aleatoric uncertainty of this model or in a larger problem an additional estimation objective. That is, specifically choosing a controllable parameter, like the injection rate, allows us to understand whether the EnKF is capable of estimating λ given the additional epistemic model bias caused by the additional δc? What effects does this have: would a corresponding fixed ensemble bias develop given a fixed offset? Otherwise, would the spread increase to incorporate the additional uncertainty?

Before proceeding to the experiment results, we recall the inlet Dirichlet and Neumann boundary conditions, i.e. (1.8b)&(1.8c), with injection rate c, which applied together at $x = 0$ reduce the total darcy velocity to:

$$c = -\lambda^w \cdot \nabla(p^w)$$

Had we replaced this at the inlet directly in (1.3a):

$$\varphi \frac{\partial S^w}{\partial t} = \nabla \cdot \left(\lambda^w \cdot \nabla p^w\right) = \varphi \frac{\partial S^w}{\partial t} = \nabla \cdot (c)$$
$$\implies S^w = \int \varphi^{-1} \nabla \cdot (c)\, dt$$

thus it is clear that S^w at the inlet is entirely determined by c. So when we assume a fixed offset is added to the injection rate, $c + \delta c = \hat{c}$:

$$\hat{S^w} = \int \varphi^{-1} \nabla \cdot (c + \delta c)\, dt = \int \varphi^{-1} \nabla \cdot (c)\, dt + \int \varphi^{-1} \nabla \cdot (\delta c)\, dt$$
$$\implies \hat{S^w} = S^w + \delta S^w$$

Alternatively, we could use $p^w = p^o - p^c$:

$$c = -\lambda^w \cdot \nabla(p^o - p^c)$$

Then assuming there exists an λ^{w-1} s.t.

$$-(\lambda^w)^{-1} c = \nabla(p^o - p^c)$$
$$p^o|_{x=0} \approx p^c(S^w|_{x=0}) - \int \left(\lambda^w(S^w|_{x=0})\right)^{-1} c\, dx$$

Recalling that p^c and λ^w are parameterized in terms of S^w which we just saw is in terms of c at $x = 0$. So when returning to the case of $c + \delta c = \hat{c}$, we can take a first order Taylor expansion in terms of $S^w + \delta S^w$ for effects caused by $c + \delta c$:

$$p^o(c + \delta c) \equiv p^o(S^w + \delta S^w) \approx p^c(S^w + \delta S^w) - \int (\lambda^w)^{-1}(S^w + \delta S^w)(c + \delta c)\, dx$$

$$p^o(c) + \mathcal{O}(\delta c) \equiv p^o|_{S^w} + \mathcal{O}\left(\delta S^w\right)$$
$$\approx p^c|_{S^w} + \mathcal{O}\left(\delta_{p^c} S^w\right) - \int \left[(\lambda^w)^{-1}|_{S^w} + \mathcal{O}\left(\delta_{(\lambda^w)^{-1}} S^w\right)\right](c + \delta c)\, dx$$
$$\approx \underbrace{\left[p^c|_{S^w} - \int (\lambda^w)^{-1}|_{S^w}\, c\, dx\right]}_{p^o(S^w)} + \left[\mathcal{O}\left(\delta_{p^c} S^w\right) - \int \mathcal{O}\left(\delta_{\lambda^w} S^w\right) c\, dx - \int \left[(\lambda^w)^{-1}|_{S^w} + \mathcal{O}\left(\delta_{\lambda^w} S^w\right)\right] \delta c\, dx\right]$$

Now clearly, there are effects that are directly related to the fixed offset δc, but there are also implicit effects through S^w. Propagating these into the data assimilation model, from the cost function in (2.9b) we consider

a single data assimilation. Herein, we only consider the inlet oil pressure, p^o_0, (i.e. exclude $p^o_{N_g}$):

$$\left\| d_j - \mathscr{H}\left[\psi^{i(j)}, \ln\lambda\right]\right\|^2_{C^{-1}_{\epsilon\epsilon_{i(j)}}} = \left\| d_j(p^o|_{x=0}) - (p^o + \mathscr{O}(\delta c)|^{i(j)}_{x=0}\right\|^2_{C^{-1}_{\epsilon\epsilon_{i(j)}}}$$

Applying the triangle inequality, we can bound the cost function from above:

$$\left\| d_j - \mathscr{H}\left[\psi^{i(j)}, \ln\lambda\right]\right\|^2_{C^{-1}_{\epsilon\epsilon_{i(j)}}} \leq \left\| d_j(p^o|_{x=0}) - p^o|^{i(j)}_{x=0}\right\|^2_{C^{-1}_{\epsilon\epsilon_{i(j)}}} + \left\|(\mathscr{O}(\delta c))^{i(j)}\right\|^2_{C^{-1}_{\epsilon\epsilon_{i(j)}}}$$

So it is clear that any offset implies the accuracy incurs an error at each data assimilation! However, it is not clear whether this will simply induce a bias or a larger spread. In general it does however highlight the necessity of having to ensure that all uncertainties are accounted for in the formulation of the data assimilation model or data collection needs to be well calibrated in order to then get an accurate fitting.

To start we run the base experiment as seen in Figure 3.4. Next, we rerun the EnKF with a small fixed positive offset of $\delta c = 2E-3cm/s$ (i.e. +4%) in the injection rate c. We see very clearly that the EnKF begins by underestimating the parameter with a very tight spread. Subsequently, it experiences massive filter divergence where the spread widens and the mean estimate starts to explode past the initial mean guess. As we begin to near breakthrough the estimate and spread start to shrink rapidly but just before breakthrough levels out with a small spread and a fixed ensemble bias which does not improve. It is interesting to note, although not clear from the statistical bias, that initially the bias is negative before flipping and becoming positive. Investigating the dynamic states we see that the offset causes the oil pressure to trail ahead below and to counter this the EnKF tries to pull back, in doing so dragging down the unknown parameter λ. This causes an over correction by the EnKF, where the "oil-wet" *behavior* requires the water pressure to be high to displace the oil. This implies the oil pressure remains higher which reduces the mismatch between the data and ensemble. However, since the truth has more "water-wet"-like *behavior* from the larger true value of λ, the water phase slips through the domain with greater ease. Whilst, the ensemble being "oil-wet"-like holds the wave front back causing the volume behind the front to fill to a greater extent and with the injection rate positive offset at a higher rate. So when the parameter estimate begins to increase, there is a cascade of built up water phase which floods into the remaining volumes as it becomes more "water-wet". As we get closer to breakthrough the production side oil pressure data point begins to inflate (a little), whilst up till now the majority of the update came from the injection side oil pressure data point.

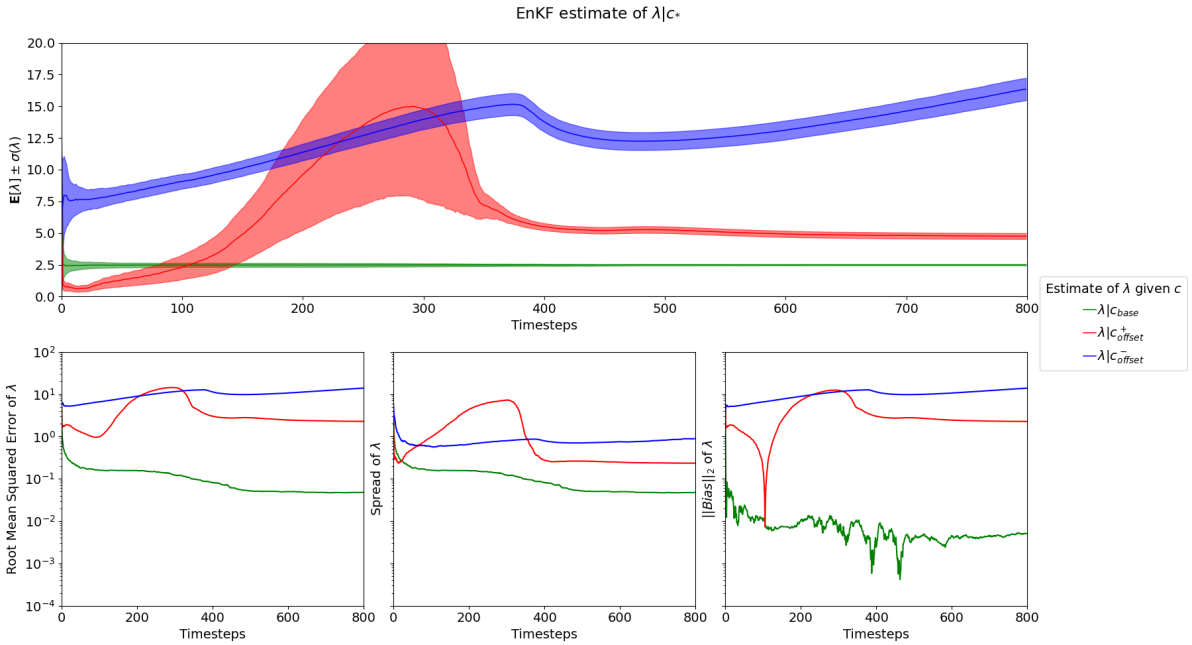


Figure 3.4: Comparing the effect of incorrect injection rate in the EnKF for the λ estimate of the base experiment

In contrast, running the EnKF with the equivalent negative offset, $\delta c = -2E{-}3cm/s$ (i.e. -4%), we see that the EnKF overestimates λ above the initial mean guess. Recalling Figure 1.6, we know that the oil pressure is more tolerant of high values of λ even though they lead below the true oil pressure. However, the negative offset slows down the front thereby increasing the oil pressure thus pumps up the parameter estimate even further. Interestingly, what we see is that an offset in the injection rate introduces an ensemble bias relative to the size of the offset, which clearly here are roughly of the same order. Moreover, we observe that transitioning towards breakthrough enforces this ensemble bias to be positive, regardless of the sign of the offset. Post breakthrough this positive ensemble bias forcing reduces, and the ensemble bias decreases. In the spread we see that the offset introduces additional unaccounted epistemic source of uncertainty and thereby increases the minimum spread lower bound. For this reason, we see that both offsets are about the same amount larger than without an offset.

Overall, in this case the accuracy of the estimate is dominated by the ensemble bias. As such, the positive offset yields a slightly smaller root mean squared error as it has a slightly smaller ensemble bias. This highlights the necessity of accounting for all sources of epistemic model uncertainty as they can yield an ensemble bias.

3.4. Effects of the EnKF Filter Settings

Having identified some affects of the nonlinearity on the EnKF fitting, we wish to understand how some filter "settings" may affect the final estimate? For example, how does the ensemble size affect the final estimate? Does the parameterization affect the accuracy? What is the effect of the measurement noise? Does the grid size of the physical model effect the final estimate?

3.4.1. The effect of the Ensemble Size

Hence, we wish to understand whether the ensemble size affects the uncertainty of the final estimate? Simply, we rerun the base experiment starting with a very small ensemble size of $N = 25$ and then again for increasing ensemble sizes up to $N = 400$. Furthermore, in order to overcome the effect of a "lucky draw" we rerun each experiment for 8 different seeds where if an experiment fails to produce a valid estimate (i.e. one without any NaNs) we rerun the experiment without updating the seed. This simple rerunning can be considered a "burn-in" of generator samples, see [20] for more details. Then by observing the mean squared error (where each seed MSE have been averaged together) vs the number of ensemble members we can determine whether an increase in ensemble members improves the estimate as well if a minimum number of ensemble members is required to achieve a reliable estimate.

In Figure 3.5 we investigate the effect on only the unknown parameter λ and not the entire ensemble. We make the distinction between the MSE with and without the ln parameterization of the parameter which uses the same fit but computes the MSE with and without the parameterization. In general, we observe that smaller ensemble sizes yield a slightly smaller mean squared error on average, but the "luck of the draw" plays a large role as seen by the error bars. Whilst ensembles larger than 100 tend to level out and be quite consistent. More generally, there does seem to be a slight bias to the initial sample for all the sample sizes as there is a consistent offset which does not match the variance exactly at time 0. This is similarly true for the parameterized $\ln\lambda$ which has shrunk as expected from(2.12). Thereafter, we see that the unparameterized λ always has a larger MSE, though this difference shrinks as more data is assimilated and the fit gets better. In fact, this difference can be roughly approximated by (2.12) where we use the current mean ensemble estimate which continues to improve as we get closer to the true value of λ. Finally we see that at the final timestep $i = 800$ both the unparameterized and parameterized MSE have reduced the bias sufficiently to be equivalent to the variance. That is, we hit again a lower bound but at least we know that the parameterization of λ has nothing to do with it.

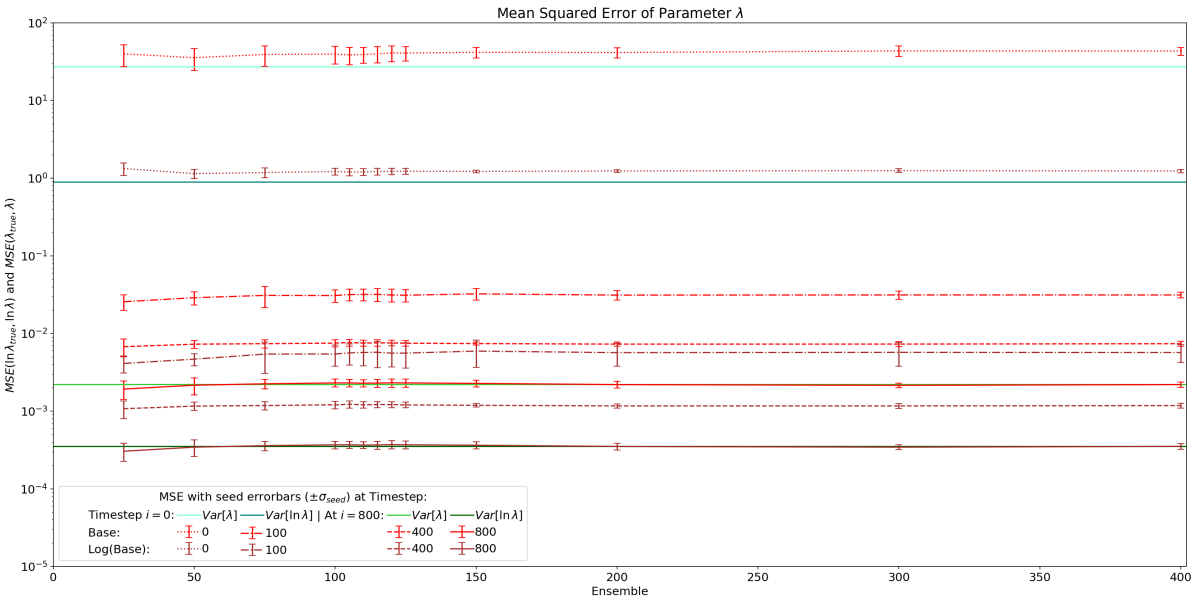


Figure 3.5: Comparing the effect of the Ensemble size and the parameterization on the λ estimate of the base experiment

Given that we start the simulation from a fixed exact initial condition for the dynamic state, S^w, p^o, where other than the initial uncertainty of λ the only other source of uncertainty is that of the measurements. Hence, by comparing the uncertainty of the augmented state against only the parameter λ we can discern if in fact the true uncertainty is reduced and not simply displaced. Here we do not make a distinction between λ and

parameterized $\ln\lambda$ as testing showed that the magnitude of the dynamic state variables dominate enough to drown out the λ whether it is parameterized or not. This is also true for the initial λ sample uncertainty which yields MSE estimates of $< \mathcal{O}\left(10^{-5}\right)$, as such we instead of starting at time step $i = 0$ we start at time step $i = 1$.

Instead of comparing the effect of the parameterization, we compare the effect of computing the MSE with the truth Ψ_{true} against computing the MSE with truth perturbed by the measurement noise, i.e. noisy truth, which we define as:

$$\Psi_{\text{noisy truth}} = \Psi_{truth} + H^{\intercal}\epsilon$$

In the real world it is never the case that we have perfect data, instead we have some noisy measurements from which we wish to estimate. Here we wish to know if uncertainty in these measurements, here noisy truth, influence whether we can detect error in our estimate. To be clear, the twin experiments are run as described above, only the computation of the MSE makes a distinction between $\Psi_{\text{noisy truth}}, \Psi_{truth}$.

In Figure 3.6, we observe at $i = 1$ the noisy truth simply adds to the forward uncertainty caused by λ. Examining the scale of the uncertainty, it is clear that uncertainty in the oil pressure dominates. Since λ is the only source of uncertainty, then the uncertainty in the oil pressure is determined by λ. Hence, the nonlinearity of a model can imply small changes in one component which can propagate into large changes in another.

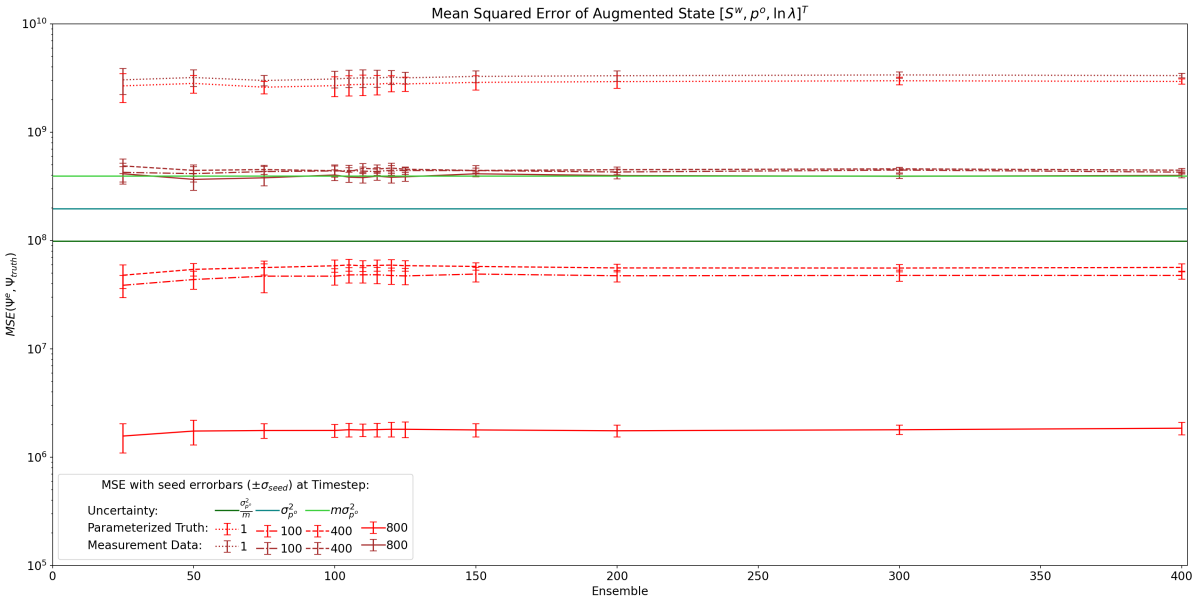


Figure 3.6: Comparing the effect of the Ensemble size and the MSE data on the augmented ensemble estimate of the base experiment

Furthermore, we observe that, when computing the MSE against noisy truth that it is not possible to overcome the measurement noise, which is self evident when we consider the MSE decomposition.

For example, consider the case X where X^* is the true state of X such that:

$$\mathbb{E}\left[\left\|X - X^*\right\|_2^2\right] = \mathbb{E}\left[\left\|X - \bar{X}\right\|_2^2\right] + \left\|\bar{X} - X^*\right\|_2^2$$

Now assume that instead we use a noisy true state, $X^* + \epsilon$, which has includes measurement error $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$:

$$\mathbb{E}\left[\left\|X - (X^* + \epsilon)\right\|_2^2\right] = \mathbb{E}\left[\left\|X - X^*\right\|_2^2\right] - 2\,\mathbb{E}\left[\langle X - X^*, \epsilon\rangle\right] + \mathbb{E}\left[\|0 - \epsilon\|_2^2\right]$$

The error ϵ is assumed to be independently and identically distributed, where independence includes from X. From the expectation of independent random variables, $\mathbb{E}\left[(X - X^*)^{\intercal}\epsilon\right] = \mathbb{E}\left[(X - X^*)^{\intercal}\right]\mathbb{E}\left[\epsilon\right]$:

$$\begin{aligned}
&= \mathbb{E}\left[\left\|X - X^*\right\|_2^2\right] - 2\langle\mathbb{E}\left[X - X^*\right], \mathbb{E}\left[\epsilon\right]\rangle + \mathbb{E}\left[\|0 - \epsilon\|_2^2\right] \\
&= \mathbb{E}\left[\left\|X - X^*\right\|_2^2\right] - 2\langle\mathbb{E}\left[X - X^*\right], 0\rangle + \sigma^2 \\
\implies \mathbb{E}\left[\left\|X - (X^* + \epsilon)\right\|_2^2\right] &= \mathbb{E}\left[\left\|X - X^*\right\|_2^2\right] + \sigma^2
\end{aligned}$$

where σ^2 is now part of the irreducible error. And so we see that measuring the MSE in an estimate against the true state, e.g. data without measurement noise, is bounded above by the MSE computed against a noisy true state, e.g. data with measurement noise:

$$\implies \mathbb{E}\left[\left\|X - X^*\right\|_2^2\right] \leq \mathbb{E}\left[\left\|X - (X^* + \epsilon)\right\|_2^2\right] = \mathbb{E}\left[\left\|X - X^*\right\|_2^2\right] + \sigma^2 \tag{3.1}$$

Hence, the uncertainty in the noisy true state is $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$ where $\sigma_i^2 = \sigma_{p^o}^2$ for the first and last control volume of the oil pressure component and $\sigma_i^2 = 0$ elsewhere. This implies that $\sigma^2 = m\sigma_{p^o}^2 = 2(14000^2)$ which is the same magnitude as the MSE computed with the noisy true state. Naturally the case computing the MSE against the truth doesn't suffer this problem, but it still encounters a lower bound. It is interesting to see how the EnKF is able to derive an update from noisy measurement data and approximate the truth. That is, when we compute the MSE using a noisy true state $X^* + \epsilon$ we see that the variance drowns out the actual variance and bias of the updated state. Where as, the MSE computed against the true state X^*, we see that the uncertainty of the noisy measurements data used in the filter update does not affect the estimate. Later in (3.10) we will see that the measurement uncertainty in the EnKF can only reduce the uncertainty in the estimate.

Regarding the lower bound encountered when computing the MSE against the truth, this is the Cramér-Rao bound:

Theorem 3.4.1 (Cramér-Rao bound[30])**.** *First we introduce a relation among symmetric positive definite matrices. Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be symmetric positive definite matrices. Then, we say* $\mathbf{A} \geq \mathbf{B}$ *exactly, when* $\mathbf{A} - \mathbf{B}$ *is a symmetric and positive semidefinite matrix. Let* x *be a vector and* $\hat{x}$ *be an unbiased estimate for* x*, where* z *is an observation of* x*. Let* $\nabla_x \ln L(x \mid z)$ *be the gradient and* $\nabla_x^2 \ln L(x \mid z)$ *be the Hessian of the likelihood function with respect to* x*.*

$$\begin{aligned}\operatorname{Cov}(\hat{x}, x) &\geq \left(\mathbb{E}\left[\left[\nabla_x \ln L(x \mid z)\right]\left[\nabla_x \ln L(x \mid z)\right]^\intercal\right]\right)^{-1} \\ &= -\left(\mathbb{E}\left[\nabla_x^2 \ln L(x \mid z)\right]\right)^{-1}\end{aligned}$$

Notice that there are two equivalent expressions for this lower bound - one involving only the first derivative and the other involving the second derivative of the likelihood function.

To illustrate this bound consider, the Cramér-Rao Lower bound for a linear Gaussian model, $z = Hx + w$ with observation matrix H and noise $v \sim \mathcal{N}(0, C)$ with $C = \sigma^2 I_{m\times m}$. Here the Gaussian model for $z \in \mathbb{R}^m$ with $v \in \mathbb{R}^m$ and $H \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$ is defined as:

$$p(z) = \frac{1}{(2\pi)^{m/2} \det C^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(z - Hx)^\intercal C^{-1}(z - Hx)\right]$$

then computing the Hessian of the likelihood with respect to x:

$$\nabla_x^2 \ln L(x \mid z) = -H^\intercal C^{-1} H$$

then to compute the expectation we further simplify and assume that $m = 2$ and $n = 3$ such that:

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ and } C = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$

$$-\mathbb{E}\left[\nabla_x^2 \ln L(x \mid z)\right] = \mathbb{E}\left[H^\intercal C^{-1} H\right] = \mathbb{E}\left[\begin{bmatrix} \frac{1}{\sigma^2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma^2} \end{bmatrix}\right] = \frac{2}{\sigma^2}$$

then using Theorem 3.4.1, the CRLB:

$$\implies \operatorname{Var}[x] \geq \frac{\sigma^2}{2}$$

or for more general m:

$$\implies \operatorname{Var}[x] \geq -\left(\mathbb{E}\left[H^\intercal C^{-1} H\right]\right)^{-1} = \frac{\sigma^2}{m}$$

However, here m may refer only to data at a single time step. Suppose we extended the linear system to include for all J sequential data assimilations at once. Then the Gaussian model for $\mathbf{z} \in \mathbb{R}^{mJ}$ with $\mathbf{v} \in \mathbb{R}^{mJ}$ and $\mathbf{H} \in \mathbb{R}^{mJ\times nJ}$, $\mathbf{x} \in \mathbb{R}^{nJ}$ is defined as:

$$p(\mathbf{z}) = \frac{1}{(2\pi)^{mJ/2} \det \mathbf{C}^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{H}\mathbf{x})^\intercal \mathbf{C}^{-1}(\mathbf{z} - \mathbf{H}\mathbf{x})\right]$$

Thus by the same steps:

$$\text{Var}[\mathbf{x}] \geq -\left(\mathbb{E}\left[\mathbf{H}^\intercal \mathbf{C}^{-1}\mathbf{H}\right]\right)^{-1} = -\left(\sum_{i=1}^{J} \mathbb{E}\left[H^\intercal C^{-1} H\right]\right)^{-1}$$

$$\implies \text{Var}[\mathbf{x}] \geq \frac{\sigma^2}{mT}$$

If we assume our problem to be sufficiently linear then we can similarly apply the same principles to our problem, wherein we have a data assimilation occurring at every timestep, e.g. $J = T$, and find that:

$$\implies \text{Var}[\Psi] \geq \left(\sum_{j=1}^{J=T} \mathbb{E}\left[H^\intercal C^{-1}_{\epsilon\epsilon_{i(j)}} H\right]\right)^{-1} = \frac{\sigma^2_{p^o}}{mT} \tag{3.2}$$

to be our Cramér-Rao Lower bound. However, the updated ensemble will still inherit many of the non-Gaussian properties from the forecast ensemble. That is, the EnKF "is approximate in the sense that it does not properly take into account non-Gaussian contributions in the prior for Ψ. In other words, it does not solve the Bayesian update equation for non-Gaussian pdfs. On the other hand, it is not a pure resampling of a Gaussian posterior distribution. Only the updates are linear and these are added to the prior non-Gaussian ensemble."[16] Thus as a result of the nonlinearity, this linear bound neglects to incorporate higher order terms and may be overcome or fail to reach the bound as a result of bias from non-Gaussian contributions.

Notice, in Figure 3.6, with each data assimilation the ensemble estimate continues to improve past this bound. However, not monotonically, as during the breakthrough transition the uncertainty of the ensemble increases rather than decreases past the injection flood uncertainty. This is most likely from the nonlinearity of the model as the ensemble members breakthrough at slightly different times.

To verify this, in Figure 3.7, we plot the MSE (computed against the truth) at every timestep for each ensemble size. We make a distinction between the MSE of the augmented and observed ensemble. Clearly, the convergence is not monotonic and is affected by the nonlinearity of the model. Most notably, the breakthrough transition wherein the effect of the ensemble size is present but very minor across the augmented ensemble. Small ensemble sizes slightly under estimate the MSE of the augmented ensemble as non-Gaussian contributions dominate and impose the nonlinear bias. Whereas larger ensemble sizes, although still affected, are better able to capture the uncertainty and thus avoid being dominated by the nonlinear bias.

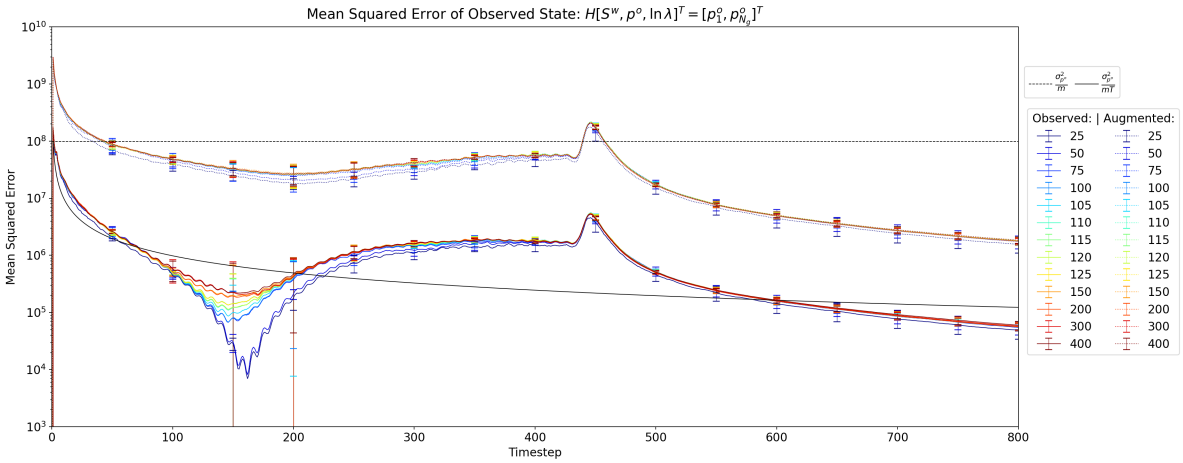


Figure 3.7: Comparing the effect of the Ensemble size and the MSE data on the observed & augmented ensemble estimates of the base experiment

Consider the observed ensemble when the MSE at $T = 1$ is just under $\sigma^2_{p^o} = \sigma^2_{p^o}/(m \times (T = 1))$. For the following 20 timesteps, the MSE follows almost exactly the bound described in (3.2), i.e. $\sigma^2_{p^o}/(mT)$. Thereafter, as the flood front develops and the nonlinearity of the model becomes more pronounced, i.e. less and less linear, the nonlinear bias from non-Gaussian contributions begin to dominate. In [16], it is shown that

as $N \to \infty$ the EnKF will converge to the Kalman Filter (KF), see [26]. That is the ensemble error statistics are an approximation of the true error statistics, which in the limit, as $N \to \infty$, they converge to. The EnKF then reduces to the KF which is the best linear unbiased estimator (BLUE) in the minimum mean-square-error sense. As such if the problem is sufficiently linear, then as $N \to \infty$ the EnKF should also result in the best linear unbiased estimator and provide the minimum mean squared error.

Notice in Figure 3.7 that the minima pre-breakthrough of the augmented and observed ensembles do not occur simultaneously, regardless of the ensemble size. This would indicate that changes in uncertainties at the observable points of the ensemble take time to propagate to the rest of the augmented ensemble, as non-Gaussian contributions.

Thus, if we were to assume that sufficient; linearity, ensemble size, and minimal non-Gaussian contributions, it is possible that the augmented ensemble also attains the lower bound. In reality we cannot force linearity, or enforce non-Gaussian contributions, or endlessly remeasure data to minimize the effect of "lucky draws". Hence, there must exist a relation that determines the achievable lower bound uncertainty. However, this lies outside the scope of this thesis. As a result, when it comes to determining the optimal ensemble size, we can only suggest that in order to sufficiently capture the model uncertainty the number of ensembles need be large enough. Evidently, the optimal choice is to have an infinitely large ensemble, however, this is not computationally feasible.

3.4.2. The effect of the Measurement Noise

Statistically this lower bound evidently is determined by the Cramér-Rao lower bound, but this lacks a physical interpretation. That is, does there exist an ensemble physics justification for a lower bound uncertainty on the estimate? Furthermore, what is the effect of the uncertainty in the measurement data? What does this imply physically? Finally, does less uncertain data improve the certainty of the final estimate?

As such, in order to better understand this bound we take a closer inspection at the covariance weighted updated that occurs in the EnKF. In doing so we are able to re-frame the elements of the sample covariance matrix as energy norms. Thereby providing a link establishing a link between statistical concepts and the physics. Furthermore, in this formalism we clearly see that the measurement noise is in effect an additional random noise potential, which proportionally shifts the accuracy of the estimate. By rerunning the base experiment with different measurement noise uncertainties, starting with a very small measurement noise $\sigma_{p^o} = 14$ and then increasing up to $\sigma_{p^o} = 1.4E6$, we can observe the effect of the measurement noise on the final estimate.

So to begin we recall our 2 very crucial points: Firstly, from statistical point of view the lower bound estimate is determined by the Cramér-Rao lower bound. Briefly, the uncertainty of the ensemble is bounded below by the uncertainty of the measurements, which in itself is somewhat informative but unclear. Hence at this juncture, it is useful to introduce the notion of Fisher information[30]:

Definition 3.4.1. *Fisher Information Matrix[30] Using the same notation as the Cramér-Rao Bound definition, the information matrix $\mathscr{I}(x)$ for the sample is defined by:*

$$\begin{aligned}\mathscr{I}(x) &= -\mathbb{E}\left[\nabla_x^2 \ln L(x \mid z)\right] \\ &= \mathbb{E}\left[\left[\nabla_x \ln L(x \mid z)\right]\left[\nabla_x \ln L(x \mid z)\right]^\top\right]\end{aligned}$$

which summarizes the amount of information in the observation.[1]

Furthermore, it can be shown that the Maximum Likelihood Estimate achieves this Cramér-Rao lower bound, at which point the smallest asymptotic variance of any unbiased estimator is said to be **asymptotically optimal**. Using this, we restate the Cramér-Rao bound [30] as

$$\mathrm{Cov}(\hat{x}, x) \geq \mathscr{I}^{-1}(x) \tag{3.3}$$

Contextualizing this with respect to our problem, we realize that the variance/spread of the parameter estimate is limited by how informative the measurable oil pressure is, and that the inverse of the error covariance matrix is exactly the inverse Fisher information matrix where the log likelihood is with respect to (λ)[2].

$$\mathrm{Cov}\left(\Psi^a, \Psi_{true}\right) \leq \mathscr{I}(\lambda) = -\mathbb{E}\left[\nabla_\lambda^2 \ln L(\Psi \mid d; \lambda)\right]$$

[1] There is a certain utility to the Fisher Information, in that it allows practitioners to set unknown values to have an initial information $\mathscr{I} = 0$ instead of an initial spread $= \infty$.

[2] Not $\ln \lambda$ as p^o is computed using λ.

In fact, the Fisher information matrix is the negative expected hessian of the log likelihood. Hence, it measures the sensitivity of the observed model output at each measurement, i.e. $(p^o)^n$, with respect to the unknown parameter $(\lambda)^n$. However, from a physical point of view the relevance is still not entirely clear with respect to what occurs in the fitting process to the physical system. That is until a closer inspection of the observed sample covariance matrix is taken, which becomes my second point.

We first examine the inner produce of the innovation S with itself:

$$\begin{aligned}\langle S,S\rangle &= \langle HA(I-1_N), HA(I-1_N)\rangle \\ &= \langle HA(I-1_N)(I-1_N), HA\rangle \\ &= \langle HA(I-1_N-1_N+1_N), HA\rangle \qquad \text{where } 1_N1_N = 1_N\end{aligned}$$

We realize the term $HA1_N = H\bar{A}$ is the observed ensemble component-wise (arithmetic) mean:

$$= \langle HA, HA\rangle - \langle HA1_N, HA1_N\rangle$$

Secondly, recall $< Y,Z >= \mathrm{tr}\,(YZ^\top)$ where $Y,Z \in \mathbb{R}^{p\times q}$.

$$\mathrm{tr}\left(SS^\top\right) = \mathrm{tr}\left(HA(HA)^\top\right) - \mathrm{tr}\left(H\bar{A}(H\bar{A})^\top\right)$$

Thirdly, realize that the resulting diagonal elements are the observable energy norms, $\|H\Psi_i\|_2$ of each ensemble member $\Psi_i \in [\Psi_0,\ldots,\Psi_N] = A$ and matrix observation operator $H \in \mathbb{R}^{m\times n}$ from (2.3). Thus the $\mathrm{tr}\,(HA(HA)^\top)$ represents the total observable energy across the ensemble. Whilst for the $\mathrm{tr}\left(H\bar{A}(H\bar{A})^\top\right)$ all diagonal elements are the observable (arithmetic) mean energy norm, $\left\|H\bar{\Psi}\right\|_2$, where $\bar{\Psi} = \frac{1}{N}\sum_{i=1}^{N}\Psi_i$. The resulting trace then represents the total observable mean energy across the ensemble which is just N times the observable mean energy norm:

$$\mathrm{tr}\left(SS^\top\right) = \sum_{i=1}^{N}\|H\Psi_i\|_2 - N\left\|H\bar{\Psi}\right\|_2 \tag{3.4}$$

This now very clearly highlights the connection between statistical uncertainty and physics defined concepts, i.e. the amount of energy across the observed ensemble. In that now we clearly see that observed sample covariance is the difference between the current total observed energy in the ensemble and N times the observed average energy in the ensemble.

Now let us assume that the ensemble is exact in every column. That is, every ensemble member reflects the true dynamic state and true parameter value, i.e. $\Psi_i = \Psi_{true} = [S^w_{true}, p^o_{true}, \lambda_{true}]^\top$. Evidently, the observed ensemble energy is equal to N times the observed ensemble average energy, and thus the variance of the observed ensemble is singular. Alternatively, in the context of Fisher information, the total observed information is infinite, which is to say everything that is learnable from the observation has bean learnt.

Now, let us relax this assumption and let only observed average energy be equal to the true observed energy, i.e. $\bar{\Psi} = \Psi_{true}$. Specifically, the ensemble mean is exactly the true solution and what's left to discern is if the variance can remain nonsingular when at an energetic minima equal to 0. Clearly it can, in fact the bounds of ensemble values are $[0, \sqrt{N\left\|H\bar{\Psi}\right\|_2}]$. Thus, energetically speaking, it is entirely possible to have an ensemble with an energy norm equivalent to the true observable energy and yet have an ensemble not equal to the true observable state. Notice that this energy shrinkage targets the entire ensemble and not ensemble members individually. Although, later we will see that the observation operator H in fact weights this ensemble by what is in fact observable. Regardless, the ensemble based shrinkage violates mass conservation laws of the individual ensemble members.

On a different note, recall that the error covariance matrix is the inverse of the Fisher information matrix and in so if the ensemble uncertainty decreases, the amount of information about the ensemble increases. As such realize that the ensemble energy shrinkage implies that the information content about the true state increases!

Subsequently, within this framework the sample measurement noise covariance is in effect an additional random noise potential:

$$\mathrm{tr}\left(EE^\top\right) \approxeq \sigma^2_{p^o}\,\mathrm{tr}\,(I) = m\sigma^2_{p^o}$$

then adding $\mathrm{tr}(SS^\top)$:

$$\mathrm{tr}\left(SS^\top + EE^\top\right) \approxeq \sum_{i=1}^{N}\|H\Psi_i\|_2 - N\left\|H\bar{\Psi}\right\|_2 + m\sigma^2_{p^o}$$

Now how does this actually relate to the entire ensemble and the fitting/analysis update of the EnKF? Well firstly, let us distinguish the forecast covariance ${C^e_{\Psi\Psi}}^f$ from the analyzed covariance ${C^e_{\Psi\Psi}}^a$ such that:

$$ {C^e_{\Psi\Psi}}^f = A(I-1_N)(I-1_N)^\intercal A^\intercal = A(I-1_N)A^\intercal \tag{3.5} $$

then from our analysis update (2.11) we can compute[3]:

$$ \begin{aligned} {C^e_{\Psi\Psi}}^a &= \left[I - A(I-1_N)S^\intercal C^{-1} H\right] {C^e_{\Psi\Psi}}^f \\ &= AA^\intercal - \bar{A}\bar{A}^\intercal - A' S^\intercal C^{-1} S A'^\intercal \quad \text{where } A' = A(I-1_N) \end{aligned} \tag{3.6} $$

Immediately we see that the analyzed covariance is a scaling of the forecast covariance. More importantly, the actual measurement data has no effect on the analyzed covariance, rather it is the uncertainty of the measurements that can shrink the analyzed ensemble covariance.

Evidently the trace of the forecast ensemble covariance:

$$ \mathrm{tr}\left({C^e_{\Psi\Psi}}^f\right) = \sum_{i=1}^{N} \|\Psi_i\|_2 - N\left\|\bar{\Psi}\right\|_2 = \sum_{i=1}^{N} \left(\|\Psi_i\|_2 - \left\|\bar{\Psi}\right\|_2\right) $$

Then by the reverse triangle inequality we know that

$$ \mathrm{tr}\left({C^e_{\Psi\Psi}}^f\right) \leq \sum_{i=1}^{N} \left(\left\|\Psi_i - \bar{\Psi}\right\|_2\right) = \mathrm{Var}\left[\Psi\right] \tag{3.7} $$

So the energy in the ensemble is smaller or equal to the ensemble variance, which is odd because this entails that there are ensemble members that deduct energetically whilst by construction we know that mass is still conserved. Therefore, the negative component must be caused by the momentum[4]of the forward propagation of the ensemble uncertainty. Moving on to the case of the trace of the analyzed ensemble covariance, we recall that $< Y, Z >= \mathrm{tr}\left(YZ^\intercal\right)$ where $Y, Z \in \mathbb{R}^{p\times q}$:

$$ \begin{aligned} \mathrm{tr}\left({C^e_{\Psi\Psi}}^a\right) &= \langle A, A\rangle - \langle \bar{A}, \bar{A}\rangle - \langle A(I-1_N), A(I-1_N)\rangle_{S^\intercal C^{-1} S} \\ &= \langle A, A\rangle - \langle \bar{A}, \bar{A}\rangle - \langle A(I-1_N), A(I-1_N)\rangle_{S^\intercal C^{-1} S} \\ &= \left[\langle A, A\rangle - \langle \bar{A}, \bar{A}\rangle\right] - \left[\langle A, A\rangle_{S^\intercal C^{-1} S} - 2\langle A, \bar{A}\rangle_{S^\intercal C^{-1} S} + \langle \bar{A}, \bar{A}\rangle_{S^\intercal C^{-1} S}\right] \end{aligned} $$

So first and foremost, we see that forecast ensemble covariance is additively (not multiplicatively) updated and that the update is the projection, $\Pi \in \mathbb{R}^{N\times N}$, given the observable ensemble innovation where the normalizing factor is perturbed by the measurement error sample covariance:

$$ \implies \Pi = S^\intercal C^{-1} S = S^\intercal (SS^\intercal + EE^\intercal)^{-1} S $$

Recall that $1_N = 1_N 1_N$ and $S1_N \equiv 0 \equiv 1_N S^\intercal$, and so the projections on averages dropout.

$$ \begin{aligned} \mathrm{tr}\left({C^e_{\Psi\Psi}}^a\right) &= \left[\langle A, A\rangle - \langle \bar{A}, \bar{A}\rangle\right] - \langle A, A\rangle_{S^\intercal C^{-1} S} \\ &= \langle A(I - S^\intercal C^{-1} S), A\rangle - \langle \bar{A}, \bar{A}\rangle \\ &= \langle A(I-\Pi), A\rangle - \langle \bar{A}, \bar{A}\rangle \end{aligned} \tag{3.8} $$

Now let us assume that $EE^\intercal = 0$, i.e. the measurements are perfect with no uncertainty, and make the heinous assumption that $S^\intercal (SS^\intercal)^{-1} S = I$[5]then

$$ \begin{aligned} &= \langle A(I - S^\intercal (SS^\intercal)^{-1}, S)A\rangle - \langle \bar{A}, \bar{A}\rangle \\ &= \langle 0, A\rangle - \langle \bar{A}, \bar{A}\rangle \\ &= -\langle \bar{A}, \bar{A}\rangle \end{aligned} $$

This is weird but valid. A negative variance? Impossible. Barbaric. But is it negative or simply non-negative? Consider what the variance of an ensemble is and what we are trying to do it. Statistically, the variance is how far the ensemble members are from the ensemble mean and what we want to do is shrink that variance,

[3] for a treatment in terms of the Kalman Gain consult [16].

[4] The mass of the momentum is the probabilistic mass of ensemble members, not the physical mass of the fluids! So this interpretation follows in by Newton's third law of motion

[5] Applying the Moore-Penrose inverse here is equally valid. This assumption shows that the projection is idempotent only when there is no measurement noise! Granted, the noise does not "de-center" the projection.

ideally to 0 (a perfect forecast). Remember we want the mean to converge to the true value and the variance to shrink around it. So here $EE^\intercal$ means that the ensemble is scaled w.r.t. measurement noise, which allows the ensemble mean to center itself towards the truth. Mathematically, firstly by definition of the inner product on real matrices $\text{tr}(A^\intercal A) \geq 0$ where equality is only when $A = 0$. Secondly, for random variables on the real line the $\text{Var}[\cdot] \geq 0$ by definition. Evidently either $\bar{A} = 0$ or $A = 0$ and so reusing $1_N = 1_N 1_N$ by way of $\langle \bar{A}, \bar{A} \rangle = \langle A, \bar{A} \rangle$:

$$\begin{aligned}
-\langle \bar{A}, \bar{A} \rangle &= -\frac{1}{2}\left[\langle \bar{A}, A \rangle + \langle \bar{A}, A \rangle\right] + \frac{1}{2}\left[\langle A, A \rangle - \langle A, A \rangle\right] \\
&= \frac{1}{2}\langle A - \bar{A}, A \rangle - \frac{1}{2}\langle \bar{A} + A, A \rangle - \frac{1}{2}\langle A - \bar{A}, \bar{A} \rangle + \frac{1}{2}\langle \bar{A}, A - \bar{A} \rangle \\
&= \frac{1}{2}\langle A - \bar{A}, A - \bar{A} \rangle - \frac{1}{2}\langle \bar{A} + A, A \rangle + \frac{1}{2}\langle \bar{A}, A \rangle - \frac{1}{2}\langle \bar{A}, \bar{A} \rangle \\
&= \frac{1}{2}\langle A - \bar{A}, A - \bar{A} \rangle - \frac{1}{2}\left[\langle A, A \rangle + \langle \bar{A}, \bar{A} \rangle\right]
\end{aligned}$$

Then if $A = 0$ by defintion $\bar{A} = A1_N = 0$, nonetheless we see that:

$$= \frac{1}{2}\langle \bar{A}, \bar{A} \rangle - \frac{1}{2}\left[\langle \bar{A}, \bar{A} \rangle\right] = 0$$

However, if $\bar{A} = 0$ then :

$$= \frac{1}{2}\langle A, A \rangle - \frac{1}{2}\left[\langle A, A \rangle + 0\right] = 0$$

Returning to (3.8), we can define the complementary projection $P = 1 - \Pi$. Then we assume there exists $P^{1/2} = (I - S^\intercal C^{-1} S)^{1/2}$:

$$\implies \text{tr}\left(C_{\Psi\Psi}^{e}{}^{a}\right) = \left\| AP^{\frac{1}{2}} \right\|_2 - N\left\|\bar{\Psi}\right\|_2$$

If we take $P^{1/2} = \left[p_1^{\frac{1}{2}}, \ldots, p_N^{\frac{1}{2}}\right]^\intercal \in \mathbb{R}^{N\times N}$. We find that $(AP^{\frac{1}{2}})_{i,j} = \sum_{k=1}^{N} \Psi_{i,k} p_{k,j}^{\frac{1}{2}}$ where $\Psi_k = [\Psi_{1,k}, \ldots, \Psi_{n,k}]$. For simplicity we denote this rescaled ensemble to have ensemble members; $AP^{\frac{1}{2}} = \left[\tilde{\Psi}_1, \ldots, \tilde{\Psi}_j, \ldots, \tilde{\Psi}_N\right] \in \mathbb{R}^{n\times N}$. Returning to the trace:

$$\implies \text{tr}\left(C_{\Psi\Psi}^{e}{}^{a}\right) = \sum_{j=1}^{N} \left\|\tilde{\Psi}_j\right\|_2 - N\left\|\bar{\Psi}\right\|_2 \tag{3.9}$$

So in essence, energetically, we rescale the ensemble energy with respect to the observation ensemble mean energy and a random energy potential, which must be present in order for the ensemble to be able to locate the true mean. Note however, this random energy potential is dependent on the measurement noise sample which as $N \to \infty \implies EE^\intercal \to \sigma_{p^o}^2 I$. As such consider the case where the measurement noise drowns out the observed ensemble covariance $C = SS^\intercal + EE^\intercal \approx EE^\intercal$, and assume that $EE^\intercal = \sigma_{p^o}^2 I$ s.t. $C^{-1} = \sigma_{p^o}^{-2} I$. Then,

$$\implies P^{1/2} = \sigma_{p^o}^{-1}(\sigma_{p^o}^2 I - S^\intercal S)^{1/2}$$

Hence, when $C \approx EE^\intercal$,

$$\text{tr}\left(C_{\Psi\Psi}^{e}{}^{a}\right) \approxeq \left\| A(I - \sigma_{p^o}^{-2} S^\intercal S)^{\frac{1}{2}} \right\|_2 - N\left\|\bar{\Psi}\right\|_2 \tag{3.10}$$

then as $\sigma_{p^o}^2 \to \infty$:

$$\begin{aligned}
\text{tr}\left(C_{\Psi\Psi}^{e}{}^{a}\right) &\approxeq \left\| A(I - 0)^{\frac{1}{2}} \right\|_2 - N\left\|\bar{\Psi}\right\|_2 \\
&= \sum_{i=1}^{N} \|\Psi_i\|_2 - N\left\|\bar{\Psi}\right\|_2
\end{aligned}$$

which we know from (3.7) is bounded above:

$$\text{tr}\left(C_{\Psi\Psi}^{e}{}^{a}\right) = \sum_{i=1}^{N} \|\Psi_i\|_2 - N\left\|\bar{\Psi}\right\|_2 \leq \text{Var}[\Psi]$$

Clearly, it the final spread is determined by the measurement noise. Wherein, the spread of the measurement noise, i.e. σ_{p^o}, determines the rate of shrinkage, and if too large the estimate will no longer improve, nor can it get worse!

In Figure 3.8, this is what exactly occurs as we increase the measurement noise from $\sigma_{p^o} = 14$ to $\sigma_{p^o} = 1.4E6$. We see that relatively $\sigma_{p^o} = 1.4E6$ is 10^5 times larger than $\sigma_{p^o} = 14$, and as a result the final spread is also 10^5 times smaller with $\sigma_{p^o} = 14$ than with $\sigma_{p^o} = 1.4E6$. This is true for all σ_{p^o} in between as well. However, for larger σ_{p^o} the estimate does not improve at all as the measurement noise dominates.

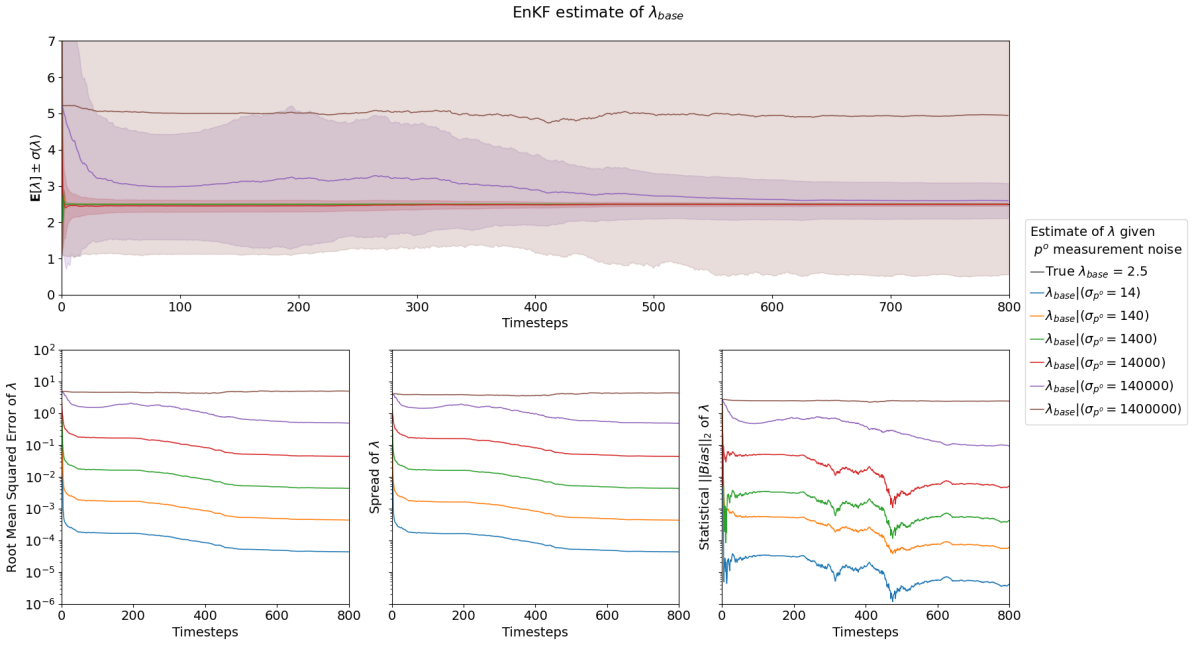


Figure 3.8: Comparing the effect of the measurement noise in the EnKF for the λ estimate of the base experiment

Originally, we began only with epistemic parametric uncertainty in λ which propagated into the dynamic states through the nonlinear model which was assumed to have no epistemic model uncertainty. Given noisy measurements, uncertain estimates for the augmented state Ψ are made, including λ. Although this analyzed λ estimate is more certain, it still has uncertainty. Whether the remaining irreducible uncertainty is the residual parametric uncertainty yet to be estimated or if uncertainty from the noisy measurements is "flowing back" into λ, is not clear. As a result, the remaining irreducible uncertainty is called aleatoric. The figure above clearly indicates however, that noisy measurements are one of the major reasons for the underlying irreducible uncertainty. In terms of Fisher information, noisier measurements have a lower information content than accurate measurements. As a result for a good estimate unbiased and precise measurements are needed.

3.4.3. The effect of the Grid Refinement

One way to improve the accuracy of the synthetic data and ensemble forecast is to increase the number of grid points/control volumes (or refine the mesh) of the nonlinear physical model. That is, we know from numerical analysis [5, 28] as we refine the grid the discretization error of the forecast decreases. Note the caveats of this being: the increased computational cost, dominating round off errors given a fine enough grid, and in hyperbolic problems the solver stability (e.g. CFL).

In-spite of these caveats, we wish to understand whether refining the grid, thereby increasing the ensemble forecast and synthetic data's accuracy, improves the final estimate? Does this have physical implications? Does this result in a lower mean squared error? Does this affect the information content of the ensemble?

Given how we discretized our spatial domain, see Figure 1.3, we know that as the grid is refined the extrapolation error of the Dirichlet boundary condition gets smaller with order $\mathcal{O}\left(\frac{d_{N_g}}{2}\right)$. However, refining the mesh increases the state size, $n = n_\psi + n_\alpha$ where $n_\alpha = 1$ and $n_\psi = N_g + N_g$, and as such is computationally costly, recall that a single forecast from the IMPES was defined to be $\mathcal{O}\left(\chi(N_g)\right)$. As a result we run an experiment, in which we fix the grid size of the ensemble forecasts to the base setting of $N_g = 50$, but then refine the synthetic dataset grid size. This is done to understand the effect of discretization errors of the synthetic

data on the spread of the estimate. That is, as the grid of the synthetic data is refined there is an $\mathcal{O}\left(\frac{d_{N_g}}{2}\right)$ introduced in the location of the measurements. Given, the data is measured from the first and last control volume, the center of the control volumes changes by $\mathcal{O}\left(\frac{d_{N_g}}{2}\right)$. In equation (3.3) we saw that the Cramér-Rao lower bound is the inverse of the Fisher information. This also implies that no other unbiased estimate can be more efficient than one that reaches this bound.

Definition 3.4.2 (Relative Efficiency[30]). *Let $\hat{A}_a$ and $\hat{A}_b$ be two estimates of an unknown parameter A. We say that the estimate $\hat{A}_a$ is more efficient relative to $\hat{A}_b$ if*

$$\mathrm{Var}\left[\hat{A}_a\right] \leq \mathrm{Var}\left[\hat{A}_b\right].$$

In Figure 3.9, initially we observe very little difference between estimates given data from a synthetic dataset generated with a very coarse mesh and a very fine mesh. However, by the end of the data assimilation, the experimental results clearly indicate $\mathcal{O}\left(\frac{d_{N_g}}{2}\right)$ trend that actually decreases at same rate as the uncertainty. Thus, indicating the impact of numerical error in the ensemble estimate. With respect to the informativeness of the grid, this would indicate that pre breakthrough synthetic data produced with finer meshes is as informative and efficient as coarse meshes. Whilst post breakthrough synthetic data produced by finer meshes is $\mathcal{O}\left(\frac{d_{N_g}}{2}\right)$ more efficient and informative that coarse meshes.

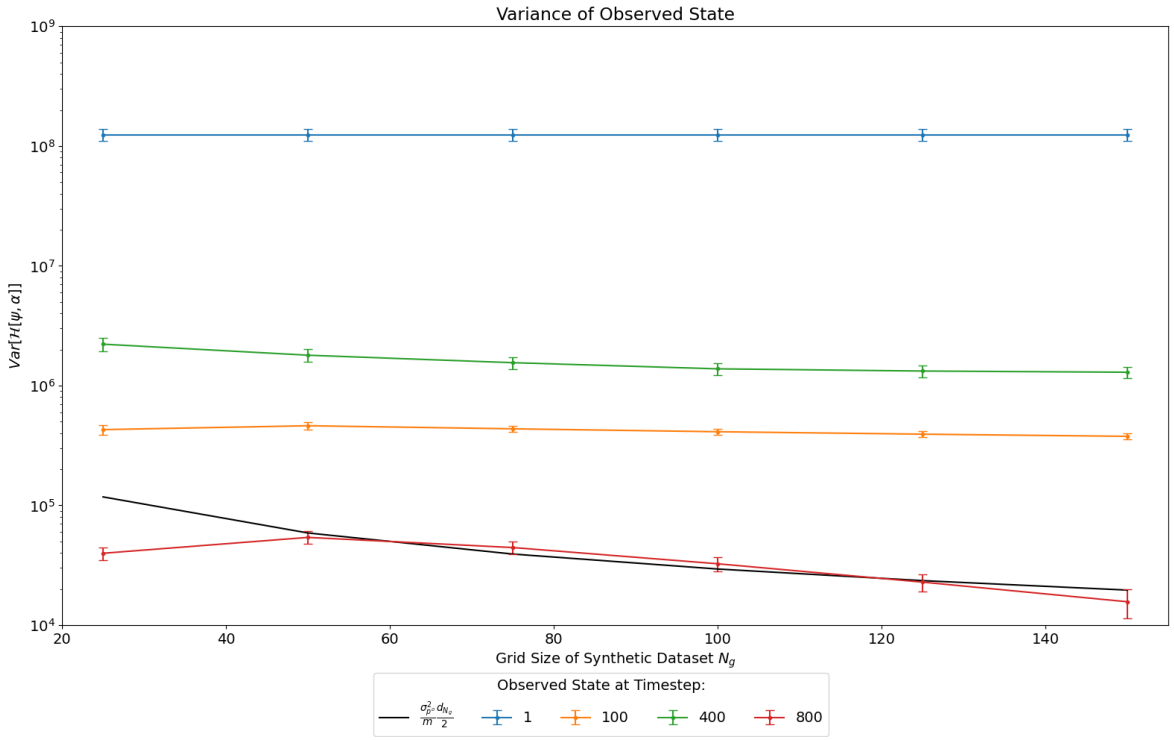


Figure 3.9: Comparing the effect of the grid refinement of the synthetic dataset on the variance of observed ensemble covariance at Timesteps : [1, 100, 400, 800]

However, ideally, we want to investigate whether the grid size of the forecasts impacts the quality of the EnKF estimate directly. Naturally, the trace of the full state covariance matrix will increase as the mesh size increases, rendering it meaningless. We could take the determinant of the covariance matrix, which is geometrically akin to computing the volume of the covariance ellipsoid. However, as the dimensionality increases, i.e. mesh is refined, the more difficult the determinant is to compute as the full state covariance gets larger and more self correlated. Recall that the degrees of freedom in the system is only 1 as only the parameter λ is unknown and the dynamic state is entirely dependent on it. This implies that the observed covariance matrix should only have a single eigenvalue and as a result should be singular. Thus, if we only consider the trace, the singularity will be hidden in the summation.

As a result, we should instead turn to the **condition number**[21]:

$$\kappa_2(C) = \frac{\sigma_{\max}(C)}{\sigma_{\min}(C)} = \frac{\lambda_{\max}(C)}{\lambda_{\min}(C)} \geq 1$$

where $\lambda_1(C) \geq \lambda_2(C) \geq \ldots \lambda_m(C) > 0$[30]. Matrices with small condition numbers are said to be *well-condition*, and κ_2 refers to the 2-norm condition of a matrix C which measures the elongation of the hyper-ellipsoid[21] $\{Cx : \|x\|_2 = 1\}$. In the 2-norm, orthogonal matrices are perfectly conditioned, then $\kappa_2 = 1, see$[21]. However, since there is only one degree of freedom, i.e. λ, the observed covariance matrix should become increasingly ill-conditioned as the parameter is estimate is improved.

Thus, we construct a twin experiment, in which using the finest mesh to generate the synthetic data ($N_g = 150$), we increase the grid size of the ensemble state, $N_g \in [25, 50, 75, 100, 125, (150)]$, and compute the observed covariance matrix condition number, $\kappa_2(H^\top C^e_{\Psi\Psi} H)$. In Figure 3.10, we observe that as the grid size increases, the problem becomes increasingly ill-conditioned. This follows as the ensemble state becomes increasingly correlated as the grid size increases. Hence, it is expected that the observed covariance matrix becomes ill-conditioned as the ensemble state grid is refined. However, this does have consequences, recall that if the condition number of a matrix is 10^d, then small errors (including the errors in the data and in the round-off) in the computation are magnified by the factor 10^d, [30]. Thus if we are dealing with finite precision arithmetic accurate up to d decimals, then small errors in time can wipe out the overall quality of the computations[30].

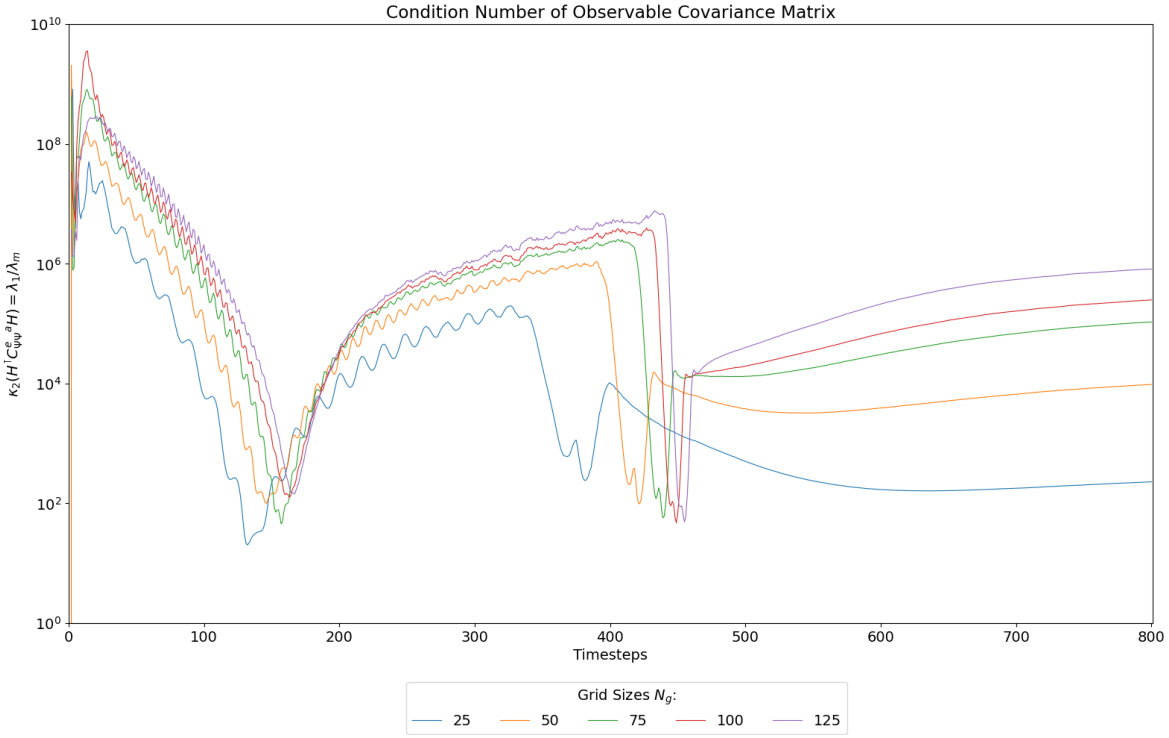


Figure 3.10: Comparing the effect of the grid refinement of the ensemble forecast on the condition number of the observed ensemble covariance given grid sizes N_g : [25, 50, 75, 100, 125, (150)]

To recap, if you refine the mesh of the synthetic dataset you can further reduce the uncertainty even though you incur a $\mathcal{O}\left(\frac{d_{Ng}}{2}\right)$ error due to the measurement location error. That is, although ensemble forecasts with coarse grids are just as informative as ensemble forecasts with fine grids, synthetic data with a finer grid is more informative than a coarse grid. This leads to the possibility of the EnKF being used with coarse forecasts to estimate the parameter then re-forecasted with a refined mesh to minimize discretization errors. However, the finer grids implies the observed ensemble covariance matrix becomes increasingly ill-conditioned. This in turn could magnify small spurious errors.

3.5. Chapter Summary

In this chapter we investigated if under the base filter settings the EnKF is capable of estimating λ.
In the first section, we found that the EnKF was able to assimilate each of the datasets created with different λ^{true}. That is, for each dataset/λ^{true}, the ensemble was able to capture λ^{true} within its spread, as well as accurately match to the mean estimate, $\mathbb{E}[\lambda] \approx \lambda^{true}$. However, for datasets using larger λ^{true} the final ensemble spread would end larger than datasets with smaller λ^{true}. This did mean that for extremely small values of λ^{true} there was a chance of an ensemble bias developing.
In section 2, we investigated the effect of the initial sample wherein we would change the initial mean or initial spread. Here, we used 2 datasets constructed from 2 different values of λ^{true}, $\lambda_{base} = 2.5$ and $\lambda_{guess} = 6.5$. This allowed for a 2x2 experimental setup when testing the sample, e.g. estimating λ_{base} given initial mean λ_{guess} (i.e. $\lambda_{base}|\lambda_{guess}$). Firstly, when testing the effect of the initial mean, we saw no discernible difference in the final accuracy when estimating for a given λ. Moreover, we noted the existence of a spread lower bound dependent on λ. Note however, with a poor initial guess it may take more data for the ensemble to capture the true value of λ.

Secondly, when testing the effect of the initial spread, we needed only half the 2x2 experiment setup, i.e. our initial mean $\mathbb{E}\left[\lambda^0\right] = \lambda_{guess}$, but then tested each dataset with a large and small variance. Similarly, we found no discernible difference in the final accuracy. As such, in situations where very little is known about variance of an unknown parameter, the EnKF is still able to identify up to a minimum degree of accuracy. Overall, this spread lower bound we encounter, rendered the effect of the initial spread meaningless on the final accuracy of the ensemble as it dominates the root mean squared error. That is, the EnKF is also able to overcome overconfident inaccurate guesses, in which both the initial spread and mean guess fail to capture the unknown parameter.

In section 3, we investigated whether the EnKF was robust enough to filter the "true" λ^{true} when the fixed injection rate c is incorrectly set ($\pm 4\%$) in the simulator (1.15). We found that the EnKF would utterly fail, developing large ensemble biases. In effect, this mis specification of a fixed variable is an additional source of epistemic model uncertainty. This highlights the necessity of accounting for all sources of uncertainty in order for the Ensemble Kalman filter to be successful.

In section 4, we investigated the effects of some of the EnKF settings; Ensemble size, Measurement noise, and Grid refinement. For each setting we create an experiment where we assimilate the base dataset for different values of that setting. In addition, we then reran each experiment using 8 different random seeds and averaged the end result to overcome "lucky draw" effects.

When investigating the Ensemble size, in the case of a linear Gaussian model, we showed that the Cramér-Rao lower bound should be $\frac{\sigma_{p^o}^2}{mT}$. Testing our problem for $25 \leq N \leq 400$, we initially found that this bound is followed closely however, as the model nonlinearity increases the ensemble deviates from this bound for any ensemble size. Larger ensemble sizes are better able to capture the true uncertainty and avoid being dominated by the nonlinear bias, but with diminishing returns.

Then in the case of the measurement noise, $14 \leq \sigma_{p^o} \leq 1400000$, we found that as we decreased the measurement noise we decreased the Cramér-Rao lower bound. i.e. the more accurate the measurements are the more accurate the estimate can be. In addition, we also show that the trace of the sample covariance matrix is the mean energy norm of the ensemble, where likewise the trace of the observable sample covariance matrix is observable (arithmetic) mean energy norm of the ensemble.

Finally, in the case of the grid refinement, $25 \leq N_g \leq 150$, we found that given what is observable, using synthetic data with a finer grid is more efficient and informative than a coarse grid, even though you incur a $\mathcal{O}\left(\frac{d_{N_g}}{2}\right)$ from measurement location error. Albeit, pre-breakthrough a finer grid is just as efficient and informative as a coarse grid. If we consider refining grid for the ensemble state, we find that fine grids are increasingly ill-conditioned, which follows as there is only one degree of freedom, i.e. λ. However, this ill-conditioning implies that fine grids are more susceptible to spurious numerical errors.

All in all, we have seen that the EnKF can efficiently estimate λ. Although, the EnKF does suffer from many failings; mass conservation in ensemble state is not conserved, small unaccounted errors can compound and might cause filter divergence and ensemble bias. However, there is the possibility of the EnKF being used with coarse forecasts to estimate the parameter then re-forecasting with a refined mesh. In the next chapter we investigate whether this is a viable in the context of history matching.

4

Confirmation Step

In the previous chapter we saw that the EnKF could efficiently provide an accurate parameter estimate of λ. However, the EnKF often produces unphysical results given that during the data assimilation no constraints regarding the physics are directly imposed. For example, mass conservation is not taken into account and the analyzed state may be physically inconsistent. Although the EnKF has been shown to guarantee mass conservation in the mean sense[1]. Whether this is sufficient in practice is unclear, it does follow by the law of large numbers, in the long run the ensemble will converge to the mean, and unless there is an external bias, the mean is the true value. However, any roulette player knows that just because an event has very seldomly occurred before does not imply it is less likely to occur again. So, in a finite time horizon, it is not necessarily the case that the mean is the truth. Furthermore, in the physical domain, one should always remember that even the flap of an Amazonian butterfly's wing can cause tornadoes in Kansas[36]. That is, due to the non-linearity of systems, even small unlikely changes in a system's state can cause large changes in the outcome. In the last chapter this was clear in Section 3.3, in which a 4% change in the injection rate caused the filter to fail. In addition to the connection we uncovered in the last chapter between the ensemble uncertainty and the energy in the system, wherein it became clear that mass conservation is violated by the data assimilation, we now introduce a confirmation step! Thus, we will term the EnKF with a confirmation step as the EnKC.

This will allow for mass conservation between time steps to be retained by the dynamic state, which also implies we are able to retain a little bit of the non-linearity of the physics. In this chapter we will be investigating the potential of this method for history matching by comparing the EnKF against the EnKC. To reiterate, the EnKC uses the EnKF to assimilate data and then re-forecasts (i.e. the confirmation step) the ensemble between data assimilations but with the new parameter estimate:

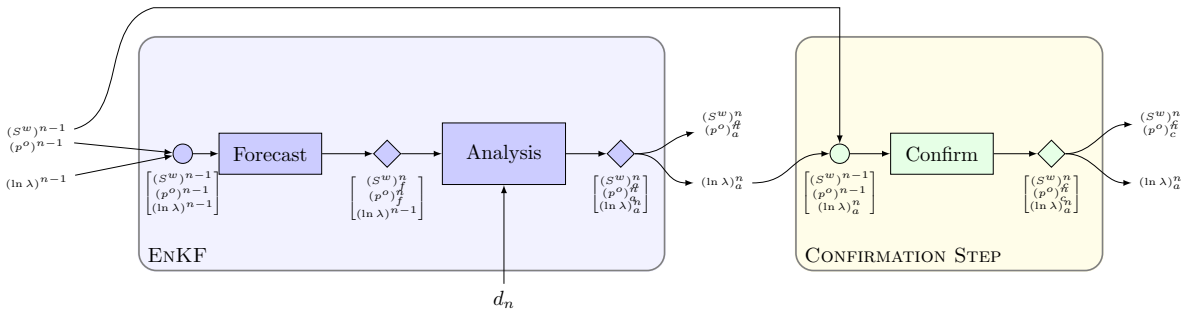


Figure 4.1: Ensemble Kalman Filter with Confirmation Step (EnKC) at nth Data Assimilation

However, it is not clear how this will affect the parameter estimate. Does this reduce the uncertainty of the parameter estimate or does it introduce a bias? Will the confirmation step introduce more of the nonlinearity present in the physics? How is the fitting process affected? Is it more informative? How does it compare with the EnKF without a confirmation step? Are there any computational implications? Does it handle a reduced dataset well?

[1] The example shown in [25] does not augment the state with a static parameter.

In order to investigate the above, we first review some of the computational aspects of the EnKC. Then, we run the base twin experiment with a confirmation step to verify if it is indeed capable of providing a parameter estimate, which we compare against the EnKF. We also check for a final ensemble bias and spread, as well as any other peculiarities that might arise from the nonlinear forecast. Thereafter, we compute the difference between the observable energy norms before and after the confirmation step. This will tell us if energy is conserved across the observed ensemble and how this affects the fitting. Moreover, by extending our information formalism, we can quantify how informative the confirmation step is compared to without. Thus, comparing the EnKF fit against the EnKC we review what the confirmation step implies for the final parameter estimate.

Thereafter, we investigate whether the EnKC is capable given a reduced dataset. By running the base experiment with subsampled data, i.e. oil pressure data is only assimilated every $[1,2,4,8,16,50,100]$ forecasts/timesteps, we then can investigate if it is still capable of providing a parameter estimate, and reasonable water saturation profile by examining the bias and spread. In addition, by examining the traces of the observed confirmed EnKC ensemble covariances we can observe the informativeness of oil pressure data assimilations. Similarly, we also investigate the case with an augmented dataset, in which water saturation snapshots are added to the base experiment's oil pressure dataset. We conclude with a chapter summary.

4.1. Computational Aspects of EnKC

First, we briefly review some of the computational aspects of the EnKC. Since the first step is simply to run the EnKF, the EnKC inherits all the Numerical issues discussed in Section 2.3. However, by reforecasting/confirming the ensemble we allow for conservation of mass within each ensemble member. This implies that each ensemble member is a possible truth, and not simply a fitting. Evidently, this does imply that the simulator needs to be called twice as often, i.e. an additional computational cost of $\mathcal{O}\left(\chi(N_g)\right)$, whilst also requiring the prior ensemble to be retained in memory. On the other hand, in the EnKF update, there is no need to update the entire ensemble state as only α^a needs to be passed to the Confirmation step. Recall however, this does reintroduce more of the inherent nonlinearity into the ensemble, thus reinforcing non-gaussian posteriors.

Algorithm 3: Ensemble Kalman Filter with Confirmation (EnKC)

N :=Ensembles Size
$\lambda \sim Gamma(k,\theta)$
$\alpha := \ln \lambda$
$A^c = \begin{bmatrix} \vec{S^w}^0 \\ \vec{p^o}^0 \\ \alpha \end{bmatrix}$
for *i timesteps* **do**
  Forecast Step:
  $A = \begin{bmatrix} \mathcal{G}(A^c;\alpha) \\ \alpha \end{bmatrix}$ /* $\mathcal{O}\left(\chi(N_g)\right)$ */
  Load measurement data:
  $D_{obs} = [d_1,\ldots,d_N]$ /* scatter from m to mN */
  Analysis Step:
  $\alpha^a = \mathcal{E}(A, D_{obs})$
  /* wherein the update step $\alpha^a = \alpha X$ only costs $\mathcal{O}\left(N^2\right)$ */
  Confirm Step:
  $A^c = \begin{bmatrix} \mathcal{G}(A^c;\alpha^a) \\ \alpha^a \end{bmatrix}$ /* $\mathcal{O}\left(\chi(N_g)\right)$ */
end

Note however, in the previous chapter it was remarked that during the parameter estimation, a coarse grid could be used for the parameter estimation procedure and a fine grid used for the confirmed history match. As such, reducing the computational cost of the additional confirmation step without compromising the parameter estimation and history matching procedures.

4.2. The effects of Confirmation step on the Base Experiment

Below in Figure 4.2 we run the base experiment for the EnKC and then repeat with the EnKF to compare the effect of the confirmation step. We see that the addition of the confirmation step yields a smaller statistical bias but a larger spread. Granted in both cases the bias is well within the spread and thus not relevant. However, the larger spread is interesting and consistent with what we would expect. That is, by applying a confirmation step, i.e. re-forecasting the dynamic state with the filter fitted parameters, we in effect apply a one-step smoothener on the parameter λ which implicitly updates the dynamic state forecast.

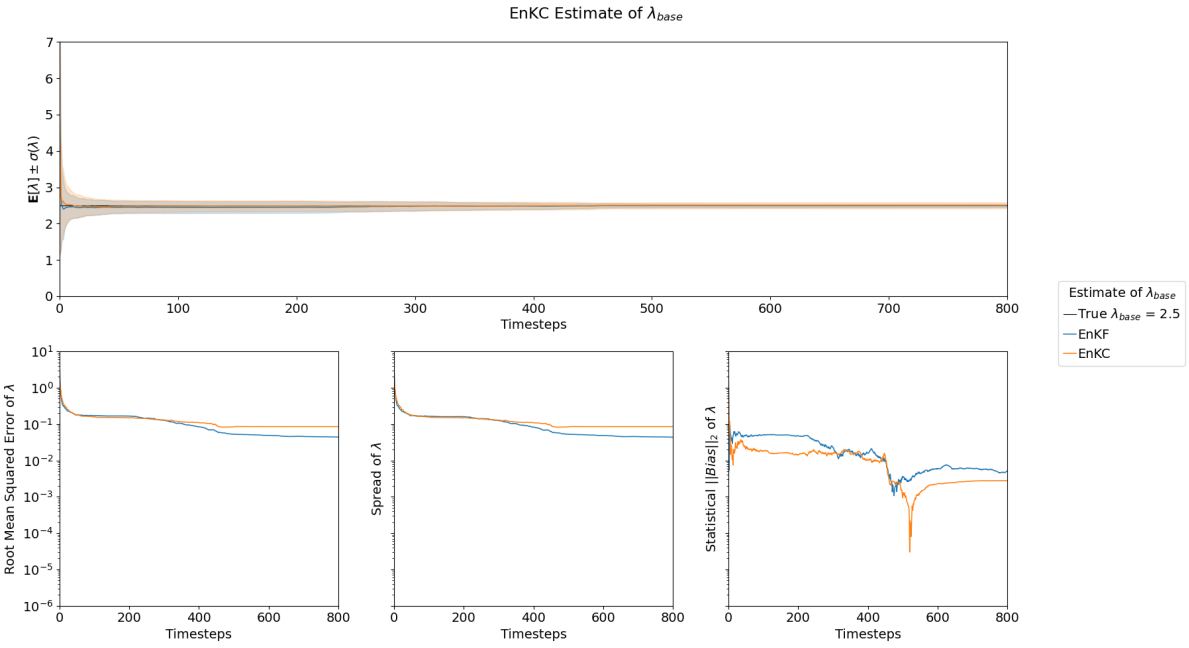


Figure 4.2: Base Experiment Estimate of λ with EnKF (blue), and EnKC (orange)

In Figure 4.3 we examine these effects on the dynamic state variables S^w, p^o and additionally we include the initial forecast and intermediary analysis of the EnKC. What we clearly see is that the confirmation step acts as a one-step smoothener on the analyzed ensemble whereby any abrupt updates are smoothened out by the confirming forecast. This is most noticeable prior breakthrough wherein the parameter uncertainty continues to infiltrate the dynamic state variables, S^w, p^o. In the case of the water saturation, S^w, as the filter slows down the uncertainty propagation caused by the parameter, the rate oscillates smoothly rather than jagged over corrections. This does then cause the water saturation spread of the EnKC to lag behind the EnKF.

However, in terms of the oil pressure EnKC does very well during the Injection Flood formation, but as we get closer to breakthrough the Bias becomes more volatile and a layer between the EnKF and EnKC spreads starts to form. Then during breakthrough, the additional confirmation step does suffer some mild divergence in the oil pressure. This could be caused by the compounding of forecast errors from the initial uncertainty in the parameter. Alternatively, the available data becomes less and less relevant to the fitting of λ. Fortunately, this only slows the rate at which the water saturation is fitted, and after some time the divergent episode of the oil pressure is partially recovered.

Unfortunately, it is not clear why exactly the filter diverges during breakthrough. Whether the compounding forecast errors or the uninformativeness of the data are the dominating cause or even an issue is not clear. Alternatively, this divergence could be inherent to the problem, and a result of a critical point situated on the saturation wave front crossing the boundary. This would imply that the EnKC successfully predicted the uncertainty of a physically driven event that the EnKF would over-correct for.

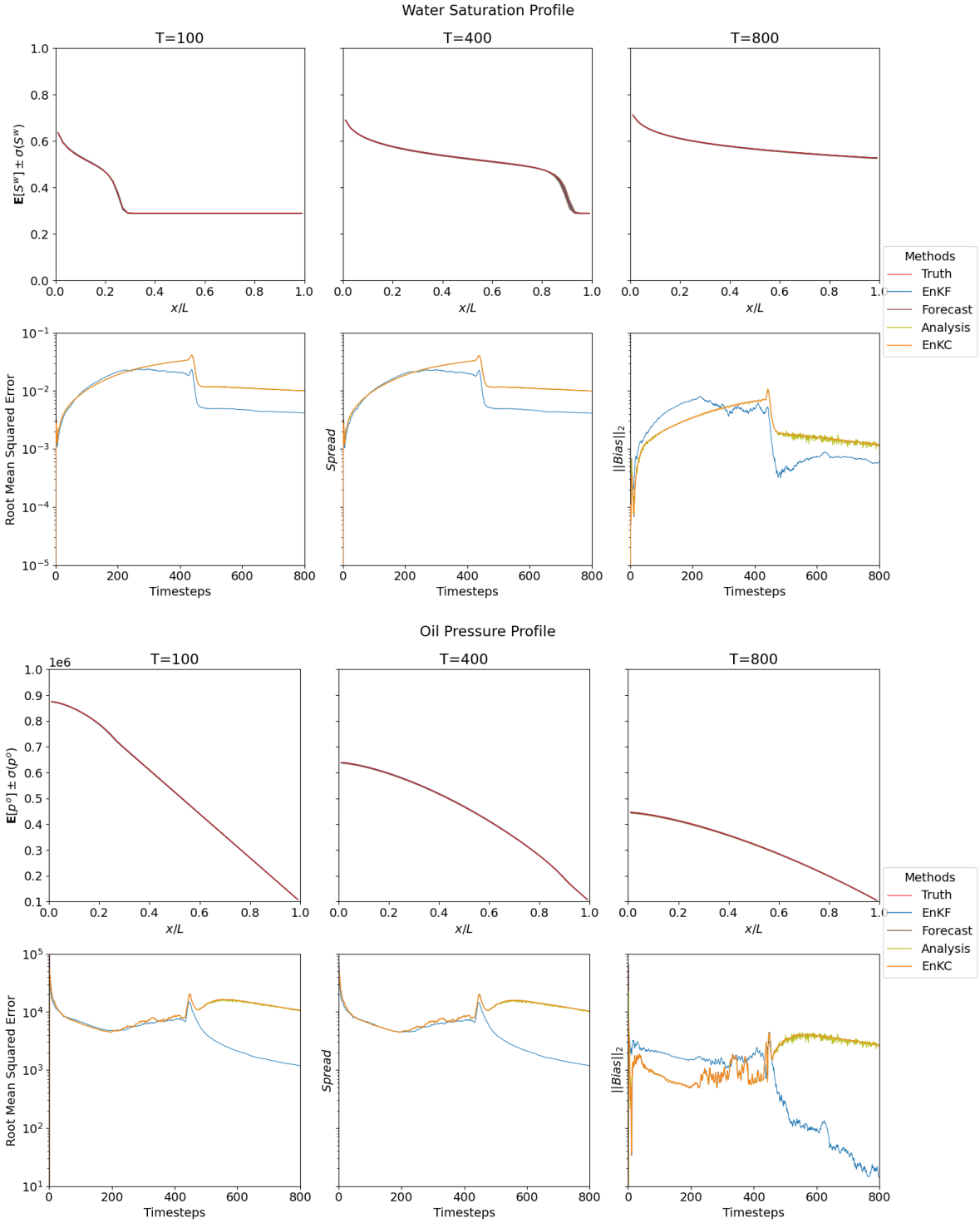


Figure 4.3: Base Experiment Estimate of S^w, p^o with EnKF (blue), and EnKC (Forecast: brown, Analysis: olive, Confirmation: orange)

4.3. Comparing the Analyzed and Confirmed Ensembles

Now to compare the effect of the Confirmation step we first return to (3.9) and recall we found that the covariance of the analyzed ensemble ${C^e_{\Psi\Psi}}^a$ was:

$$\mathrm{tr}\left({C^e_{\Psi\Psi}}^a\right) = \langle A^{f_a}P^a, A^{f_a}\rangle - \langle \bar{A^{f_a}}, \bar{A^{f_a}}\rangle$$

given a forecast A^{f_a} to be analyzed. Then without loss of generality we assume the covariance of the confirmed ensemble ${C^e_{\Psi\Psi}}^c$ is of similar form:

$$\mathrm{tr}\left({C^e_{\Psi\Psi}}^c\right) = \langle A^{f_c}P^c, A^{f_c}\rangle - \langle \bar{A^{f_c}}, \bar{A^{f_c}}\rangle$$

where similarly a forecast A^{f_c} is to be analyzed but then also confirmed. Now then taking the difference, and then assume that $A^{f_c} - A^{f_a} = \delta A$ s.t. $\bar{A^{f_c}} - \bar{A^{f_a}} = \bar{\delta A}$:

$$\begin{aligned}
\mathrm{tr}\left({C^e_{\Psi\Psi}}^c\right) - \mathrm{tr}\left({C^e_{\Psi\Psi}}^a\right) &= \langle (A^{f_a}+\delta A)P^c, A^{f_a}+\delta A\rangle - \langle A^{f_a}P^a, A^{f_a}\rangle \\
&\quad - \langle \bar{A^{f_a}}+\bar{\delta A}, \bar{A^{f_a}}+\bar{\delta A}\rangle - \langle \bar{A^{f_a}}, \bar{A^{f_a}}\rangle \\
&= \langle A^{f_a}(P^c - P^a), A^{f_a}\rangle + 2\langle A^{f_a}P^c, \delta A\rangle + \langle \delta A P^c, \delta A\rangle \\
&\quad - 2\langle \bar{A^{f_a}}, \bar{\delta A}\rangle - \langle \bar{\delta A}, \bar{\delta A}\rangle
\end{aligned}$$

If we now assume that both the confirmed and analyzed covariances are derived from the same initial forecast, i.e. $A^{f_c} = A^{f_a} \implies \delta A = 0 \implies \bar{\delta A} = 0$, then the resulting ensemble of α is the same, i.e. $\alpha^c = \alpha^a$. So equivalently, we can state the following:

$$\mathrm{tr}\left({C^e_{\Psi\Psi}}^c\right) - \mathrm{tr}\left({C^e_{\Psi\Psi}}^a\right) = \mathrm{tr}\left({C^e_{\psi\psi}}^c\right) - \mathrm{tr}\left({C^e_{\psi\psi}}^a\right)$$

then denoting $\psi^f_i = [\vec{Sw}^f_i, \vec{po}^f_i]^\intercal \in \mathbb{R}^{2N_g}$, and assume $[\psi^f] = \left[\psi^f_1, \ldots, \psi^f_N\right]$:

$$= \langle [\psi^f](P^c - P^a), [\psi^f]\rangle$$

where then $[\psi^f](P^c - P^a) = \delta\psi$ is the difference in the analyzed ensemble and the confirmed ensemble.

$$= \langle \delta\psi, [\psi^f]\rangle \tag{4.1}$$

We also can interpret it energetically as the difference in traces represents the difference in ensemble energies of the dynamic state variables, S^w, p^o, from each update. Reusing our initial assumptions:

$$\mathrm{tr}\left({C^e_{\psi\psi}}^c\right) - \mathrm{tr}\left({C^e_{\psi\psi}}^a\right) = \sum_{j=1}^{N} \left\| \left([\psi^f]P^{c\,\frac{1}{2}}\right)_j \right\|_2 - \left\| \left([\psi^f]P^{a\,\frac{1}{2}}\right)_j \right\|_2 \tag{4.2}$$

Notice that this energetic variation is the amount of dispersion incurred to the dynamic state variables by the EnKF analysis. That is, given the same point of origin an analyzed ensemble should be equivalent to a confirmed ensemble, but due to the nonlinearity of the model physics the uncertainty propagates at a nonlinear rate which the EnKF's Gaussian Assumed Update cannot account for accurately.

In Figure 4.4, we plot the trace of the observed ensemble covariance for the Analyzed and Confirmed ensembles of the EnKC, as well as the EnKF ensemble for later comparison. As remarked in Figure 4.3 the confirmation step is a one-step smoothener on the analyzed ensemble. This is now supported by the resulting trace trajectory results in smooth transitions, whilst the analyzed ensemble alone is far more jagged. On the other hand, this smoothening creates an artificial lag, which accumulates with each confirmation step. This is most notable post-breakthrough where the confirmed ensemble has a consistently larger trace above the analyzed ensemble.

If we compare against the EnKF we see that, without the lag of the confirmation step, it is better able to take advantage of the available information and minimize the uncertainty in the first 160 timesteps (data assimilations/forecasts). However, this could be a sign of severe overfitting to the data as thereafter the uncertainty increases almost reaching the measurement noise level. Only at breakthrough does it cross the measurement noise threshold only to after sharply decrease again. Whilst the EnKC (analyzed and confirmed ensembles) needs time to correct course before decreasing linearly.

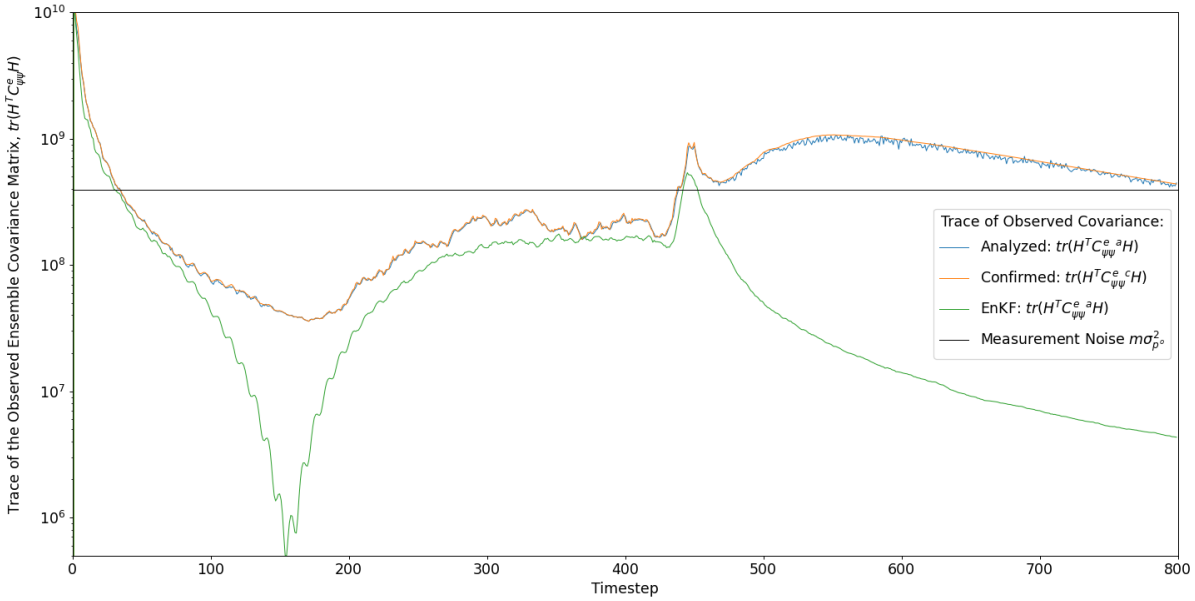


Figure 4.4: Traces of the observed Analyzed and Confirmed EnKC ensemble covariances. (For comparison the trace of the observed EnKF ensemble covariance)

Next, we instead compute the sample covariance matrices from the perspective of the posterior states. That is we denote Ψ^a as the analyzed ensemble, whilst Ψ^c is the confirmed ensemble:

$$C^e_{\Psi\Psi}{}^a = \Psi^a(I-1_N)(I-1_N)\Psi^{a\intercal}$$
$$C^e_{\Psi\Psi}{}^c = \Psi^c(I-1_N)(I-1_N)\Psi^{c\intercal}$$

where we assume that $\alpha^c = \alpha^a$, then denoting dynamic state variables $\psi^* = [S^{w*}, p^{o*}]^\intercal$

$$\implies C^e_{\Psi\Psi}{}^c - C^e_{\Psi\Psi}{}^a = \begin{bmatrix} \psi^c\psi^{c\intercal} - \psi^a\psi^{a\intercal} & (\psi^c - \psi^a)\alpha^{a\intercal} \\ \alpha^a(\psi^c - \psi^a)^\intercal & 0 \end{bmatrix} - \begin{bmatrix} \bar{\psi}^c(\bar{\psi}^c)^\intercal - \bar{\psi}^a(\bar{\psi}^a)^\intercal & (\bar{\psi}^c - \bar{\psi}^a)\alpha^{a\intercal} \\ \alpha^a(\bar{\psi}^c - \bar{\psi}^a)^\intercal & 0 \end{bmatrix}$$

Then the resulting trace is only in terms of the dynamic state variables p^o, S^w:

$$\begin{aligned} \mathrm{tr}\left(C^e_{\Psi\Psi}{}^c - C^e_{\Psi\Psi}{}^a\right) &= \mathrm{tr}\left(\psi^c\psi^{c\intercal} - \psi^a\psi^{a\intercal} - \bar{\psi}^c(\bar{\psi}^c)^\intercal + \bar{\psi}^a(\bar{\psi}^a)^\intercal\right) \\ &= \langle\psi^c,\psi^c\rangle - \langle\psi^a,\psi^a\rangle - \langle\bar{\psi}^c,\bar{\psi}^c\rangle + \langle\bar{\psi}^a,\bar{\psi}^a\rangle \end{aligned}$$

On a side note, realize that because of $\alpha^c = \alpha^a$, then the trace of (augmented) ensemble covariances are equivalent to the dynamic state covariances:

$$\mathrm{tr}\left(C^e_{\Psi\Psi}{}^c - C^e_{\Psi\Psi}{}^a\right) = \mathrm{tr}\left(C^e_{\psi\psi}{}^c - C^e_{\psi\psi}{}^a\right)$$

Now let $\psi^c - \psi^a = \delta\psi$ s.t. $\bar{\psi}^c - \bar{\psi}^a = \bar{\delta\psi}$

$$\begin{aligned} &= \langle\psi^a + \delta\psi, \psi^a + \delta\psi\rangle - \langle\psi^a,\psi^a\rangle - \langle\bar{\psi}^a + \bar{\delta\psi}, \bar{\psi}^a + \bar{\delta\psi}\rangle + \langle\bar{\psi}^a,\bar{\psi}^a\rangle \\ &\vdots \\ &= 2\left[\langle\delta\psi,\psi^a\rangle - \langle\bar{\delta\psi},\bar{\psi}^a\rangle\right] + \langle\delta\psi,\delta\psi\rangle - \langle\bar{\delta\psi},\bar{\delta\psi}\rangle \end{aligned}$$

where $\psi 1_N = \bar{\psi}$ and $\delta\psi 1_N = \bar{\delta\psi}$

$$= 2\langle\delta\psi,\psi^a\rangle_{(I-1_N)} + \langle\delta\psi,\delta\psi\rangle_{(I-1_N)} \tag{4.3}$$

Taking instead $\sqrt{2}(I-1_N)$ as the vector space, then clearly we see it is a quadratic form:

$$= \frac{1}{2}\langle\delta\psi,\delta\psi\rangle_{\sqrt{2}(I-1_N)} - \langle\delta\psi,-\psi^a\rangle_{\sqrt{2}(I-1_N)}$$

which is the potential energy in each estimate, and implies that the mean deviation of the analyzed ensemble is half the difference between confirmed and analyzed mean deviations:

$$\frac{1}{2}\delta\psi(1_N - I) = \psi^a(I-1_N) \tag{4.4}$$
$$\implies \psi^c(I-1_N) = \psi^a(1_N - I) \tag{4.5}$$

Unfortunately $(I - 1_N)$ is singular, with no inverse the problem cannot be solved by direct inversion alone. However, notice that ψ^c is the $(I - 1_N)$ has rank $N-1$ also implies there are $N-1$ degrees of freedom for residuals[2]. Thus we have shown that a relation between the residuals $r^* = \psi^*(I - 1_N)$ of the analysis and confirm ensembles exists. For now we will alternatively consider $\delta\psi$ as a differential δ_a with respect to ψ^a and in so we can apply the dot product rule to the left-hand term :

$$\implies \text{tr}\left(C^e_{\psi\psi}{}^c - C^e_{\psi\psi}{}^a\right) = \sum_{i=1}^{N} \delta_a \left\|\psi^a_i\right\|_{(I-1_N)} + \left\|\delta_a \psi^a_i\right\|_{(I-1_N)} \tag{4.6}$$

This again gives us a relation of the energy potential between the analyzed and confirmed dynamic state variables, only this time over the vector space of residuals (mean deviations) and with respect to ψ^a. Optimally, we want the residuals to vanish (i.e. $= 0$), and so long as the data does encompass the truth then estimation will be perfect. So as $\psi^a \to \bar{\psi}^a \iff r^a \to 0$ the left hand term in (4.6) does "disappear", regardless of what occurs to ψ^c. However, the right-hand term will not, in fact it may continue to grow. That is, if the analysis produces the true parameter value given a forecast from an uncertain initial ensemble with uncertain dynamic states, then the confirmed ensemble will have propagated that initial uncertainty of the dynamic states. Ensemble divergence will then occur in the dynamic state variables of the confirmed ensemble. On the other hand, suppose that the confirmed ensemble is "perfect" but the analyzed ensemble is not[3], then from (4.3) we simply retain the analyzed covariance which is bounded below by the measurement noise.

In Figure 4.5 we plot the Energy potentials (i.e. difference between traces of the observed ensemble covariances) between Analyzed and Confirmed Ensembles of the EnKC. Given that the energy potential can be positive or negative, we take the $\|\cdot\|_2^2$ to stay within the positive reals. However, in (4.1) we notice that it is only the $2 < \delta\psi, \psi^a >_{(I-1_N)}$ can turn negative whilst the $< \delta\psi, \delta\psi >_{(I-1_N)}$ always remains positive. Recall that the ensemble is always positive in the dynamic state variables, and that the static parameter is the same between the analyzed and confirmed ensembles. As such any change in sign in the potential is a result of the difference between the confirmed and analyzed ensemble, wherein by definition the confirmed ensemble conserves mass, thus any changes indicate the analyzed ensemble has either gained or lost mass. Specifically, when the potential is positive then $\mathbb{E}_\Omega\left[\psi^c - \psi^a\right] > 0 \implies \mathbb{E}_\Omega\left[\psi^c\right] > \mathbb{E}_\Omega\left[\psi^a\right]$. Hence, we can say the analyzed ensemble has lost "mass" due to the analysis step. Whereas when the potential is negative, $\mathbb{E}_\Omega\left[\psi^c - \psi^a\right] < 0 \implies \mathbb{E}_\Omega\left[\psi^c\right] < \mathbb{E}_\Omega\left[\psi^a\right]$, the analyzed ensemble has artificially gained "mass". Only when $\mathbb{E}_\Omega\left[\psi^c - \psi^a\right] = 0 \implies \mathbb{E}_\Omega\left[\psi^c\right] = \mathbb{E}_\Omega\left[\psi^a\right]$ has mass been conserved across the entire domain.

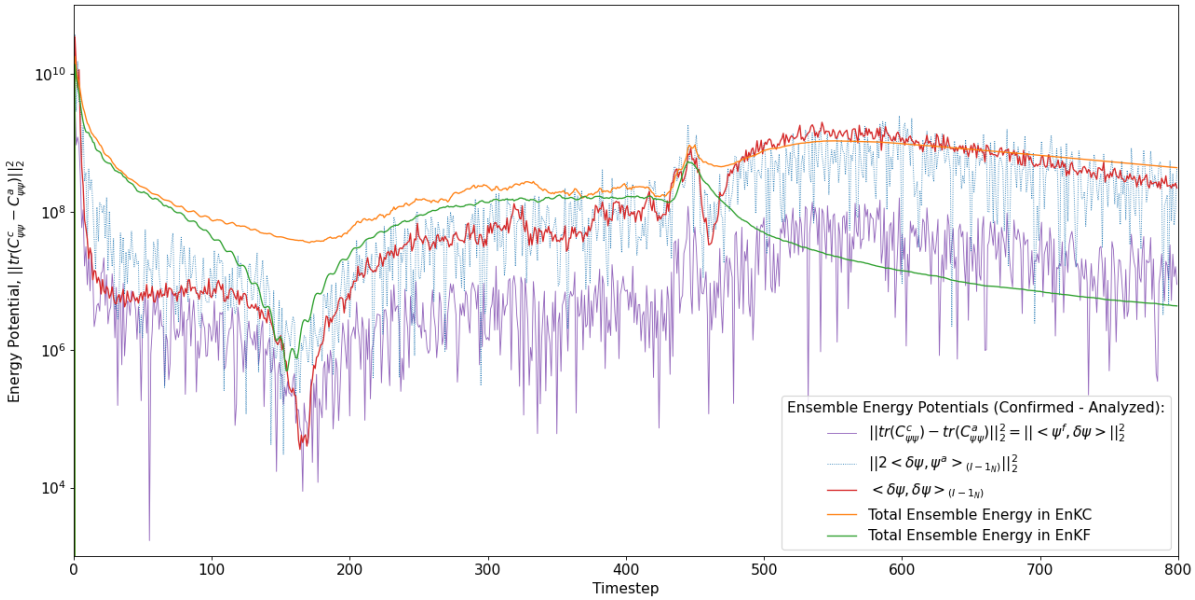


Figure 4.5: Energy Potentials between the Analyzed and Confirmed Ensemble, as well as the EnKC and EnKF Total Ensemble Energies

[2]This follows as there is only 1 degree of freedom, i.e. λ.
[3]for example consider that during the first data assimilation we fit perfectly the parameter and as such the uncertainty does not have the chance to propagate into the dynamic state variables of the confirmed ensemble.

Note however, that under this interpretation mass may still fluctuation across the domain unnaturally even though it is "conserved". Then again this isn't an issue for the confirmed ensemble as this only occurs in the analyzed ensemble. That being said, it is now clear why $2 < \delta\psi, \psi^a >_{(I-1_N)}$ is the oscillatory component and can turn negative. What is interesting is that $< \delta\psi, \delta\psi >_{(I-1_N)}$ retains the extrema of the EnKC and EnKF and thus may be optimized.

Now given that the trace is a linear mapping, the prior and posterior formulations of the energy potential between the analyzed and confirmed states should be equal:

$$\begin{aligned} & \operatorname{tr}\left(C^e_{\Psi\Psi}{}^c\right) - \operatorname{tr}\left(C^e_{\Psi\Psi}{}^a\right) = \operatorname{tr}\left(C^e_{\Psi\Psi}{}^c - C^e_{\Psi\Psi}{}^a\right) \\ \iff & \operatorname{tr}\left(C^e_{\psi\psi}{}^c\right) - \operatorname{tr}\left(C^e_{\psi\psi}{}^a\right) = \operatorname{tr}\left(C^e_{\psi\psi}{}^c - C^e_{\psi\psi}{}^a\right) \end{aligned}$$

Now we take the prior formulation from (4.1) and the posterior formulation from (4.3):

$$\langle \psi^f, \delta\psi \rangle = 2 \langle \delta\psi, \psi^a \rangle_{(I-1_N)} + \langle \delta\psi, \delta\psi \rangle_{(I-1_N)} \tag{4.7}$$

$$\implies 0 = \langle \psi^f, \delta\psi \rangle - 2 \langle \delta\psi, \psi^a \rangle_{(I-1_N)} - \langle \delta\psi, \delta\psi \rangle_{(I-1_N)} \tag{4.8}$$

Clearly, we see that it is not possible to directly reconcile the two formulations within vector space of residuals. However, if we are able to minimize $\delta\psi$ on the space of residuals then the two formulations both converge to 0! This could lead to a further minimization in order to improve the estimate further. Notice that if we take the left hand side of (4.7):

$$\langle \psi^f, \delta\psi \rangle = \langle \psi^f, \psi^c - \psi^a + (\psi^a - \psi^a)1_N \rangle = \langle \psi^f, \psi^c - \psi^a 1_N + \psi^a (1_N - I) \rangle$$

we can then apply (4.5):

$$= \langle \psi^f, \psi^c - \psi^a 1_N \rangle + \langle \psi^f, \psi^c \rangle_{(I-1_N)}$$

where if we reuse that $\bar{\psi}^c - \bar{\psi}^a = \bar{\delta\psi}$ and assume that $\bar{\delta\psi} \to 0$:

$$\begin{aligned} &= 2 \langle \psi^f, \psi^c \rangle_{(I-1_N)} + \langle \psi^f, \bar{\delta\psi} \rangle \\ &= 2 \langle \psi^f, \psi^c \rangle_{(I-1_N)} \end{aligned}$$

Returning this to (4.8) we find another quadratic form within the space of residuals:

$$\langle \psi^f, \psi^c \rangle_{(I-1_N)} = \langle \delta\psi, \psi^a \rangle_{(I-1_N)} + \frac{1}{2} \langle \delta\psi, \delta\psi \rangle_{(I-1_N)} \tag{4.9}$$

Notice that $\langle \psi^f, \psi^c \rangle_{(I-1_N)}$ is the cross-covariance between the ensemble forecast and the confirmed ensemble and that minimizing this is equivalent to minimizing the error brought by the uncertain parameter. Recall, that we assumed the model forecasts are assumed to be deterministic with no stochastic uncertainty, and so any uncertainty between forecasts is a result of the uncertain parameter. Examining the right hand terms, we see that the right most term $\langle \delta\psi, \delta\psi \rangle_{(I-1_N)}$ is the Forecast Error Covariance between the analysis and confirmation. Meanwhile, recall that the analysis can be rewritten in terms of the data and forecast, we thus assume the remaining term is composed of the additional contribution in terms of the data and confirmed ensemble. However, given that developing a further minimization is outside the scope of this thesis, this was not further investigated. Nonetheless, in appendix C, an attempt was made to reconcile these additional contributions in the Data Error Covariance, and together with the Forecast Error Covariance a Confirmed Online EnKC-Maximization algorithm (EnKCMC) proposed.

We have shown through twin experiments that the confirmation step acts as a one-step smoother on the dynamic state variables. Analytically we show that the difference in trace between the analyzed and confirmed ensemble covariances is the amount of dispersion incurred to the dynamic state variables by the EnKF analysis. In addition, by analytically decomposing the trace we show which components cause the mass to "fluctuate" unnaturally across the domain as a result of the analysis, and why the EnKF may fail to be mass conservative across the domain. Finally, we show that there is a quadratic form in terms of Forecast Error Covariance between the analysis and confirmation, plus an additional data contribution equal to the cross-covariance between the ensemble forecast and confirmed ensemble, which is equivalent to uncertainty from the uncertain parameter. As previously stated, it is still not fully understood what the additional data contribution is, but it lies outside the scope of this thesis. In the next section we will investigate the effect that data has on the history matching and parameter estimation procedures

4.4. Data Effects on the EnKC

In the previous section we noted that the effects of the data on the parameter estimation and history matching were still unclear. Thus, in this section, we investigate whether the EnKC is still capable given a reduced oil pressure dataset or if enhancing the dataset with water saturation snapshots is beneficial.

In the case of a reduced dataset, we run the base experiment with subsampled data, i.e. oil pressure data is only assimilated every $[1,2,4,8,16,50,100,400]$ forecasts (or timesteps). We then can investigate if the EnKC is still capable of providing a parameter estimate and physically reasonable water saturation profile, through examining the bias and spread of the ensemble but also of the propagated uncertainty of the initial ensemble (i.e. no data). In addition, by examining the traces of the observed covariances we determine the informativeness of oil pressure data assimilations.

Thereafter, in the case of an enhanced dataset, the oil pressure dataset is extended at certain instances to include a water saturation snapshot of the entire spatial domain. Again, by observing the bias and spread of the parameter estimate and water saturation profile we determine the effects of additional snapshots. In addition, by examining the traces of the observed covariances of the confirmed ensemble we determine the informativeness of each water saturation snapshot.

4.4.1. Reduced Dataset: Oil Pressure

What we intend to determine is how the quantity and timing of oil pressure data effects the parameter estimate, as well as the history match of the water saturation profile. We run the base experiment with reduced (sub-sampled) datasets. To clarify, the base setting used in the rest of the thesis is 1, i.e. for every forecast there is oil pressure data to be assimilated. Whilst a reduced dataset j implies that only after the jth forecast is there data available for assimilation. Alternatively put, out of 800 forecasts, there are respectively $J = \frac{T}{j} - 1$ data assimilations in total. Below we investigate the cases when data is only assimilated every $j \in [1,2,4,8,16,50,100,400]$, which can be seen as $J \in [799,399,199,99,49,15,7,1]$ data assimilations in total.

In Figure 4.6 we have the parameter λ estimate plot and metrics. Evidently, the λ estimate only improves with each data assimilation. Given that fewer data assimilations occur, the estimate of λ does not improve. That is, the spread and bias only change during data assimilation. Interestingly, this does not imply that data always improves the estimate! In fact, there is some slight filter divergence between the flood front formation ($T \approx 100$) and prior to breakthrough ($T \approx 400$). Moreover, the peaks of the divergence seem to coincide for numerous datasets $[16,50,100]$. In the case of dataset of 16 forecasts per data assimilation (fpd), we observe that the peak is between time step 160 and 176, which seems to coincide with the minima of $\langle \delta\psi, \delta\psi \rangle_{(I-1_N)}$ in Figure 4.5 (which occurs at 163).

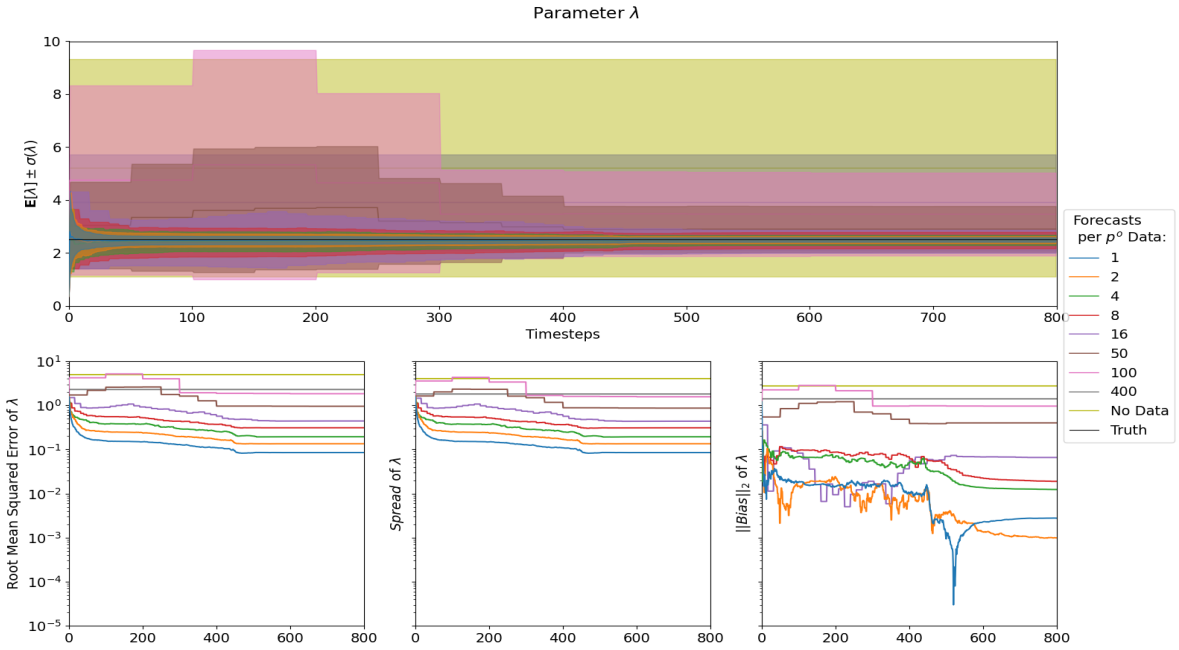


Figure 4.6: Comparing the effect of reducing the oil pressure dataset in the EnKC for the λ estimate of the base experiment.

Keep in mind that the EnKC is a one-step smoothener, and thus the enlarged spread is a result of the next data assimilation and not the data assimilation at the start of the divergence in the spread. For example, take $100fpd$, the data at 100 and 300 time steps cause the reduction in spread, but the data near 200 time steps is the cause of the divergence in spread. Then comparing this point of divergence (for $100fpd$) with the original ensemble sample (i.e. no data), the parameter estimate is worse in every way, resulting a larger spread and bias. This shows that incorrectly timed data can produce a worse estimate than no data at all. This can be no clearer than when comparing both $50fpd$ and $100fpd$ against $400fpd$ which has only 1 (well timed) data assimilation for the entire simulation and does not seem affect by this spread divergence.

On the other hand, notice that if the rate of data assimilation is high enough, e.g. $8fpd$, then this filter divergence is negligible or even nonexistent. This reinforces the idea, put forward in Subsection 2.3.3, that if you sample "fast" enough then you can avoid filter divergence caused by the nonlinearity of the physics. On the other hand, one could argue that this results in overfitting to the data and under-represents the true uncertainty in the physics.

From Figure 4.7 we very clearly see that the case with "no data" provides an uncertainty upper bound to the spread, as the initial parametric uncertainty of λ dominates the water saturation profile. That is, given fewer corrections of the parameter estimate of λ, the initial uncertainty continues to propagate the ensemble further apart, resulting in larger spreads respectively. Due to the Confirmation step these spreads are irrecoverable, interestingly however, there is a negligible difference in the final spread for $50 \leq fpd \leq 400$. This indicates that although just one data assimilation can be as good as multiple poorly placed data assimilations, there is a minimum sampling rate required to further improve the history match.

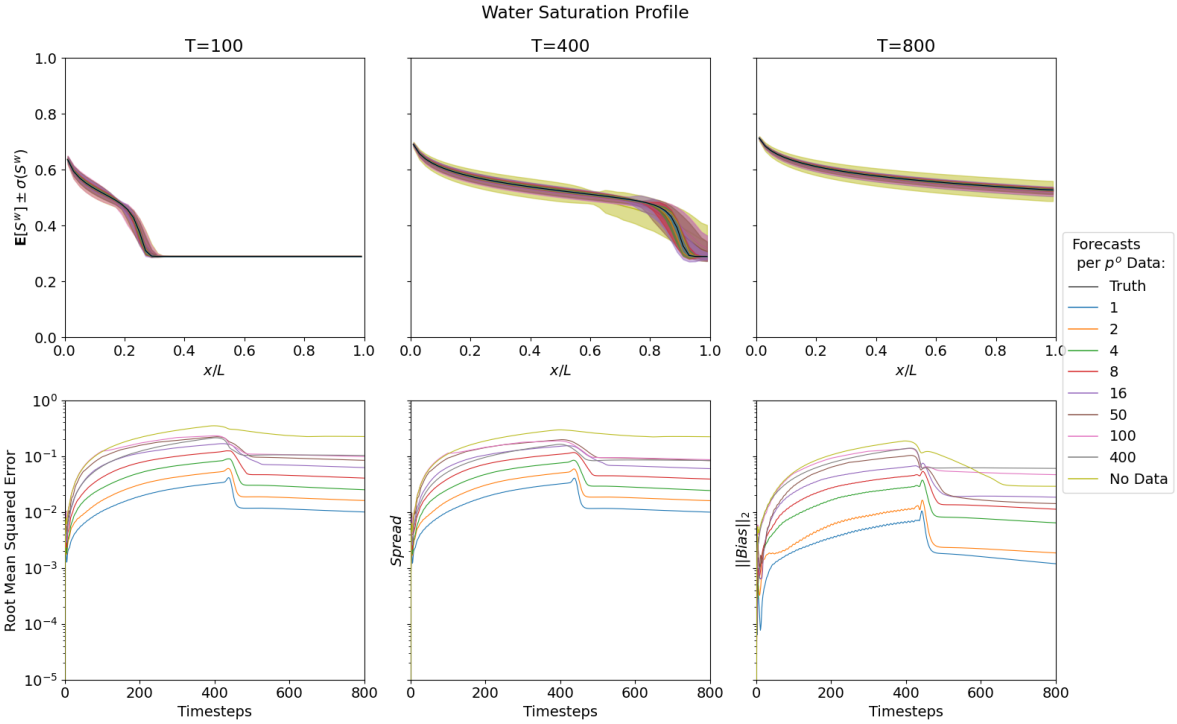


Figure 4.7: Comparing the effect of reducing the oil pressure dataset in the EnKC for the S^w profile history match of the base experiment.

On a side note, we would also like to mention the misleading "water-wet" behavior of the water saturation profile. That is, although the band is computed from the spread of the ensemble, it may not represent the inherent skewness in the ensemble. As a result, the profile may appear to be far more "water-wet" than any confirmed ensemble member is, let alone the true profile. For example in the cases with the little data (50 or 100), an unaware practitioner may mistakenly conclude the core to be very "water-wet". Should a practitioner midway through an experiment want to use data that exploits the exact moment of breakthrough, they will undershoot the true moment given their mistaken "water-wet" belief.

In Figure 4.8, we have the trace of the observed ensemble covariance for the EnKC, wherein again the case with "no data" provides an upper bound. Note here, we assume that we are able to observe the ensemble at every timestep in-spite the actual availability of the data. As remarked above, more available data results in a tighter spread. As the covariance trace is lower, thus the total information in the ensemble is greater.

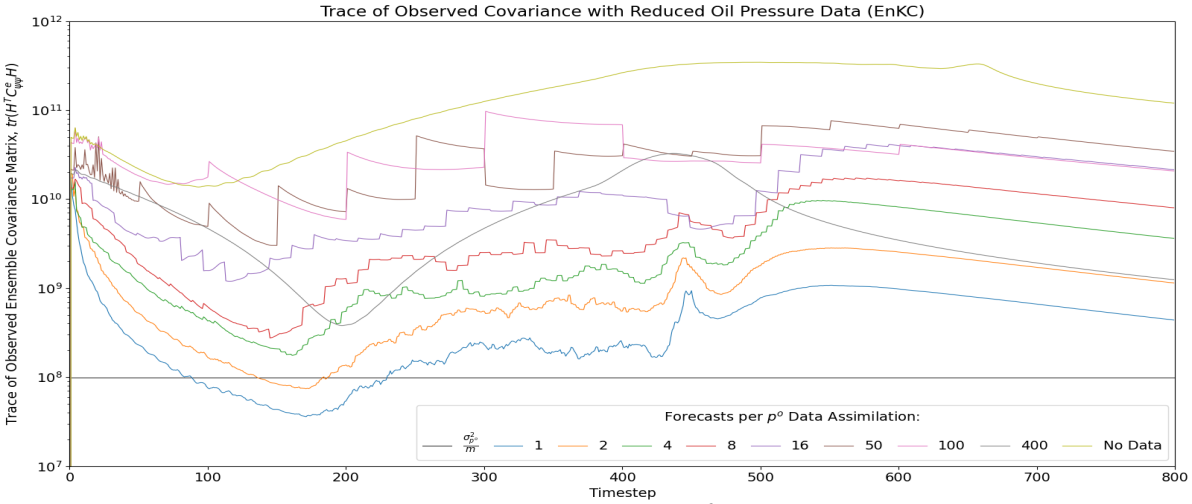


Figure 4.8: Traces of the observed EnKC ensemble covariances given reduced p^o datasets with increasing time between data assimilations

However, not all data is equally as informative, whereby some are in fact mis-informative. Such as, in $100fpd$ where the first data assimilation causes the ensemble to be less informative than the forecast with "no data". On the other hand, notice the case of $400fpd$, where one well-timed data assimilation has a massive "perceived" impact on the informativeness of the ensemble. To such an extent that at end time $400fpd$ appears to be nearly as informative as $2fpd$! However, from Figure 4.7, we know this is not the case with respect to S^w, and so this trace does not allow us to perceive this informational skewness.

Having seen how a single of data point can have a very large impact on the informativeness of the ensemble, we now endeavor to understand the effect of the timing of a single data assimilation. Are some points in time more informative than others?

In Figure 4.9 below, we run the base experiment using a dataset consisting of a single of data point at different timesteps, i.e. $[100, 200, 300, 400, 500, 600, 700]$, which we compare against the case with "no data". Clearly data points occurring between 400 and 500 time steps are most informative. This happens during breakthrough which coincides with the critical points on the wave front passing through the observable domain $(p^o(\Omega_1), p^o(\Omega_{N_g}))$. As such, the further away a critical point is from an observable location the less informative a data point will appear to be. This clearly seems to be the case at 100, during the flood front formation near the inlet of the core, which renders the ensemble more uncertain than with "no data".

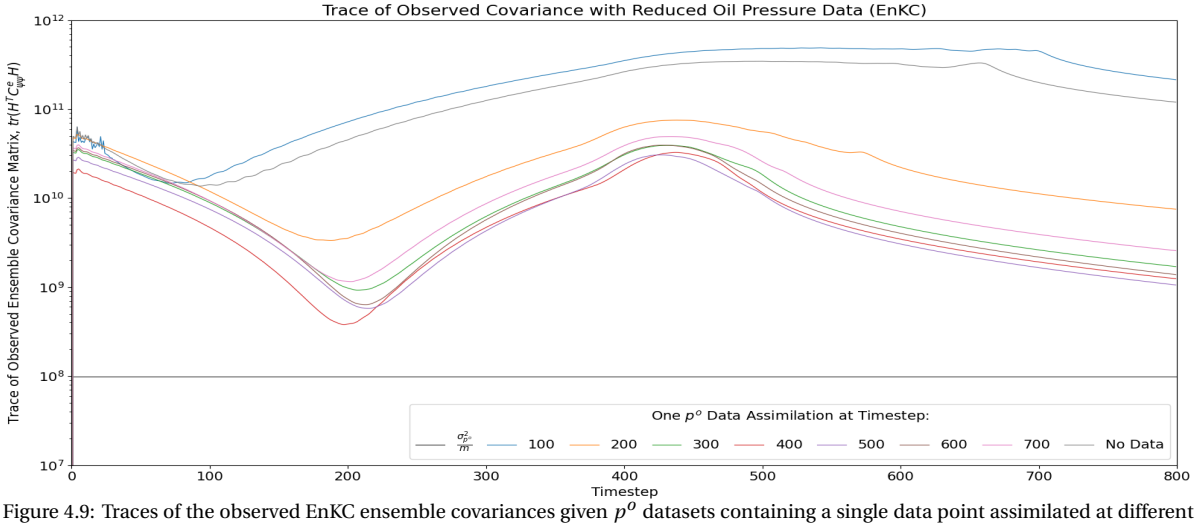


Figure 4.9: Traces of the observed EnKC ensemble covariances given p^o datasets containing a single data point assimilated at different points in time.

Notice, this relation is not symmetric about the "optimal" time step, observe the case at 700 which is far more certain than with a data point at 200. This has several crucial implications for technicians: First, if you sample whilst the flood front is forming or approaching the middle of the core, then you are more likely to introduce uncertainty than reduce it. Secondly, it is effective to reduce the uncertainty after break through but as you move further away from break through the less informative data is. As a result, estimation of the time of breakthrough is critical, as it is here when data is the most informative time, wherein it is better to be a little late than too early.
In our investigations of the effect of reduced oil pressure data we found timing the observation of the flood front to be critical, i.e. breakthrough point. Although locating alternative spatial measurement locations was not investigated, ancillary experiments show that the inlet oil pressure is far more informative than the end oil pressure, see appendix A.2, and similarly to the timing, the closer data is measured to the source of distortion the more accurate the parameter estimate is.

4.4.2. Enhanced Dataset: Water Saturation Snapshots

Next, we intend to determine how the quantity and timing of water saturation snapshots effects the parameter estimate, as well as the history match of the water saturation profile. We begin by running the case with no additional snapshots, denoted as []. We denote each scenario by a list with the iterations at which a snapshot is taken, i.e. snapshots at $[k,l,\ldots]$ for $1 \leq k,l \leq T$. In order to determine whether the timing of a snapshot impacts the final result, we investigate if a single water saturation snapshot is better placed at [100], [400], or [700]. Then we increase the number of snapshots in the base experiment to determine how the number of snapshots used impacts the final result. Here we take, [200,400,600], [100,300,500,700], and [100,200,300,400,500,600,700]. Notice that we will also be able to infer if the timing between snapshots has an impact.
In Figure 4.10 it is clear that each snapshot causes a sharp increase in the accuracy of the λ estimate. Fortunately, there is no filter divergence, given that we have supplemented the base dataset with oil pressure data available for every forecast, as it is possible that the p^o data is not sufficiently informative to prevent filter divergence. Thus, adding a snapshot only improves the estimate. We observe that when only a single snapshot is available, the earlier the better for an accurate parameter estimate. This follows as the compounding of forecast errors from the confirmation step is not possible. However, there is very little difference between 100 and 400, compared to 700 which suggests that snapshots after breakthrough are far less effective.

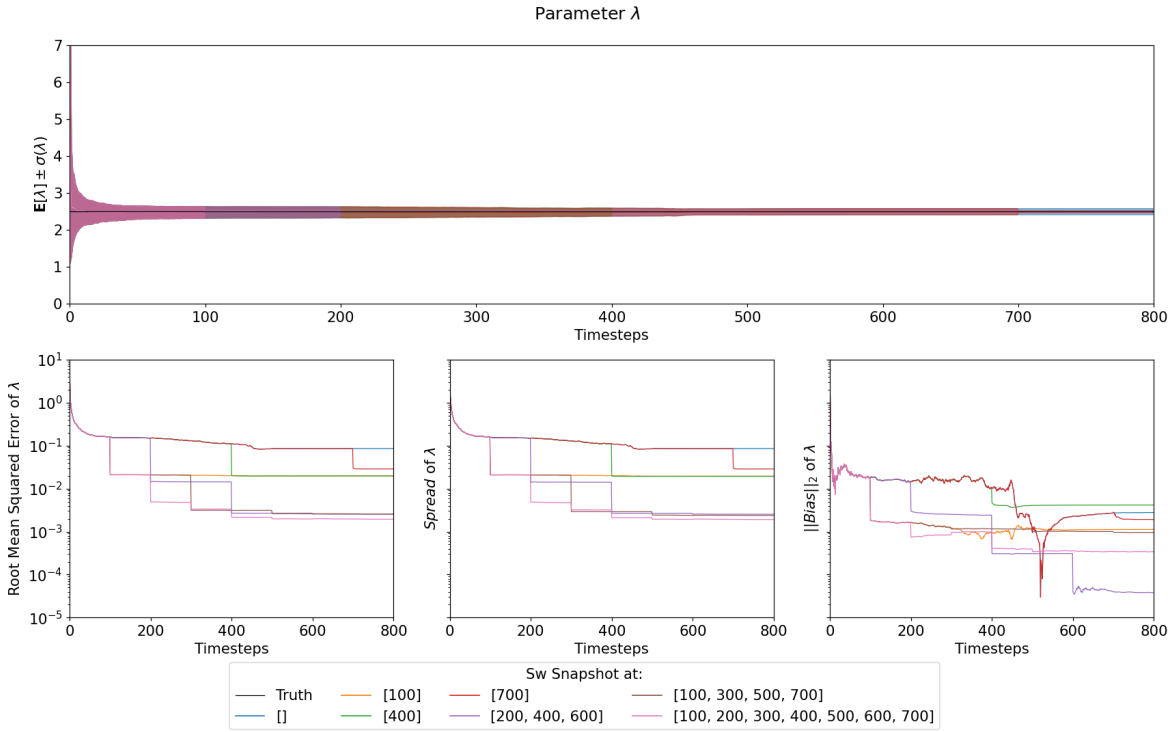


Figure 4.10: Comparing the effect of datasets augmented with S^w snapshots in the EnKC for the λ estimate of the base experiment.

One would assume that more snapshots imply a more accurate parameter estimate, but this is not strictly the case. Firstly, the reduction from each subsequent available snapshot is smaller than previous, which is seen from the step-like convergence in [100,200,300,400,500,600,700]. Both [200,400,600] and [100,300,500,700] are able to attain nearly the same accuracy with just (under and over) half as many snapshots. Hence, more snapshots do not strictly imply a more accurate estimate for λ.
Now, it is interesting to observe the effect of snapshots on the water saturation profile given the compounding of forecast errors from the confirmation step. In Figure 4.11, pre-breakthrough, as soon as a snapshot is used the propagation of the initial parametric uncertainty stops, only then does the oil pressure start to fit the water saturation profile towards the truth. Without such a snapshot we would have needed to wait for breakthrough to occur. At which point, the compounding of forecast errors would be irrecoverable without forgoing mass conservation (i.e. no confirmation step). For example, observe the difference in accuracy between using a single snapshot at either 100 and/or 700. It is clear that the earlier the first snapshot is used, the greater the final accuracy of the water saturation profile history match. Snapshots that occur after are much less informative, and have an almost negligible effect on the water saturation history match. Meanwhile, snapshots that occur during breakthrough also have very little effect unlike p^o data. This is due to the measurement functional including the water saturation which observes the entire water saturation profile, and as a result all the critical points of the water saturation profile. Thus, post breakthrough, snapshots also seem to have very little effect on the final accuracy, as the critical point on the flood front has vanished.

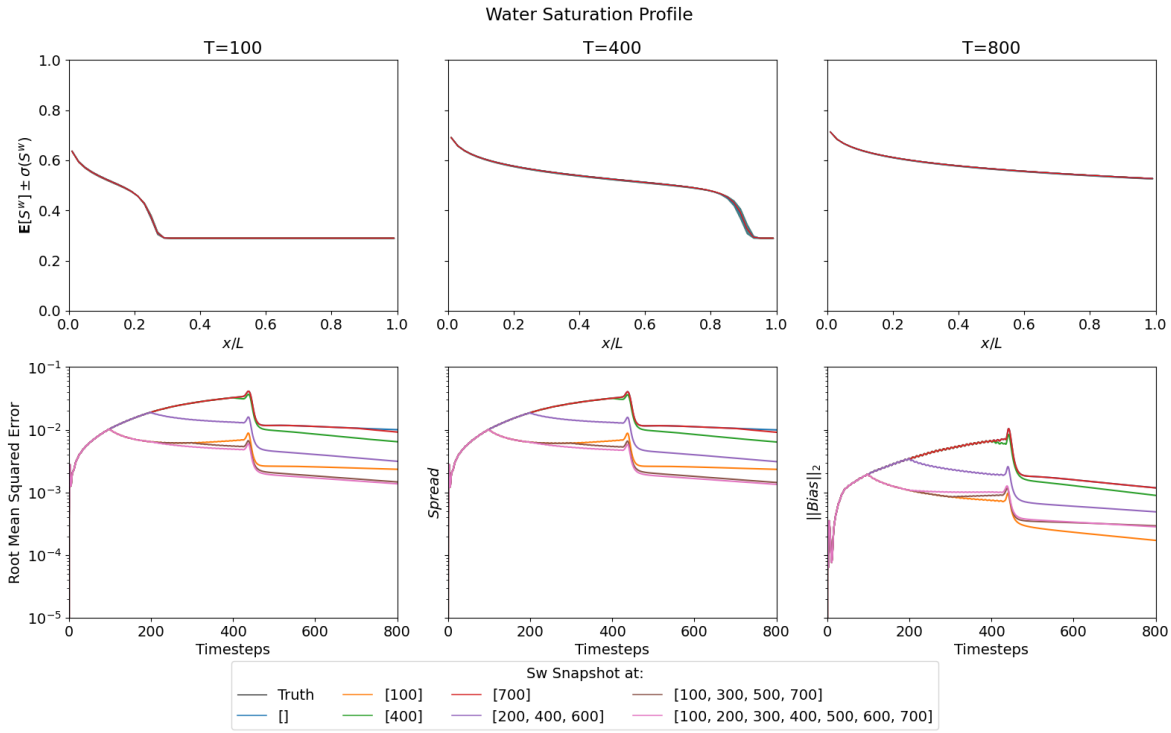


Figure 4.11: Comparing the effect of datasets augmented with S^w snapshots in the EnKC for the S^w profile history match of the base experiment.

Up till now it seems when snapshots are assimilated the uncertainty in ensemble "vanishes". In order to better understand what happens to the uncertainty, in Figure 4.8, we have the trace of the observed ensemble covariance for the EnKC. We observe that in fact the snapshots introduce uncertainty into the ensemble, whereby the first snapshot introduces the most uncertainty. This implies that the analyzed λ estimates which then used in the confirmation step causes the large increase in the uncertainty in the oil pressure. This could indicate that correction that incurs as a result of the analyzed λ disrupts the physics. For example, consider an ensemble with a λ forecasting "water-wet" core, then introducing a snapshot produces an analyzed λ which forecasts a more "oil-wet" core. That is, the core requires the oil pressure to be much higher in order for the water flood to proceed, evidently the inverse scenario can occur. As a result, jumps in the oil pressure uncertainty occur, and with it the entire ensemble. It could be the case that if a larger confirmation window were to be used then such large corrections would not occur. Note however, this phenomenon is purely a result of

the mathematical construction of the nonlinear physical model. The couplings in the model determine how such corrections occur.

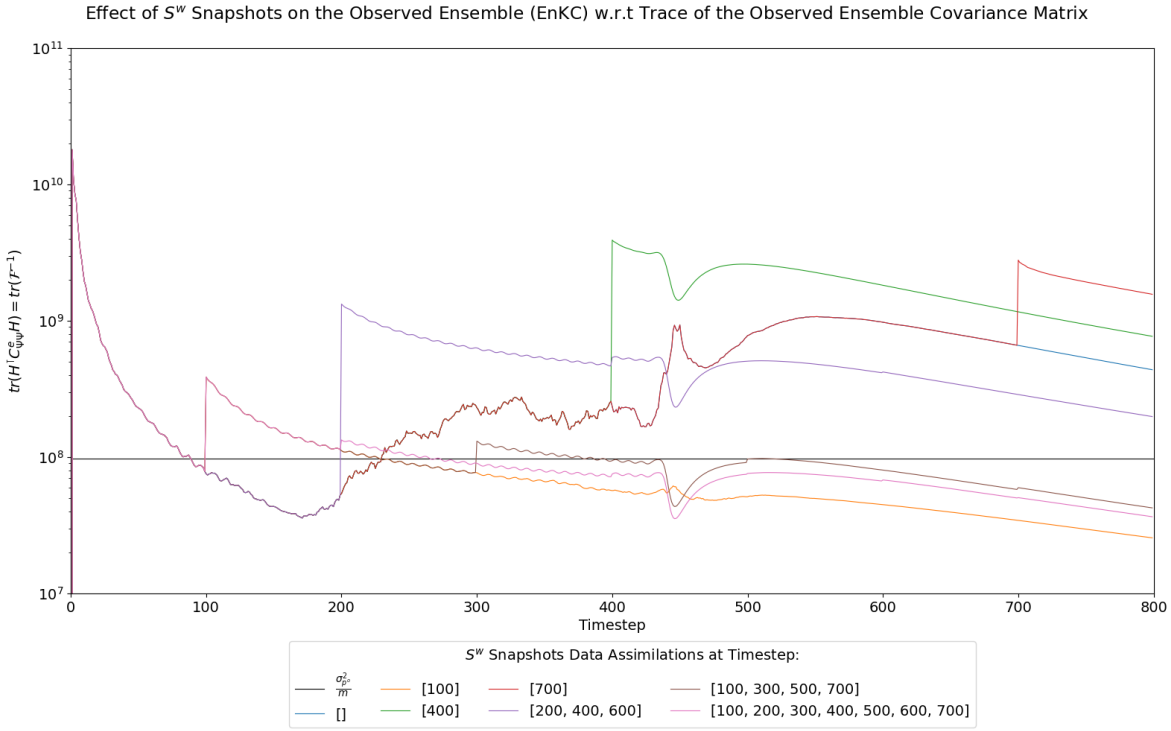


Figure 4.12: Traces of the observed EnKC ensemble covariances given datasets augmented with S^w snapshots

Hence, in our investigations of the effect of snapshot data we found timing to be critical. Although, multiple snapshots yield better estimates, the earlier the snapshot data is assimilated the better the final estimate is. Note that, the planning of the optimal schedule of snapshots is not investigated. However, ancillary experiments show promise wrapping the genetic algorithm around the current framework to find the optimal schedule of snapshots. Albeit this approach being exceedingly parallel, it is also exceedingly computationally expensive, as in order to gauge the effectiveness of a schedule the entire data assimilation needs to be run for each set of snapshots.
Although the effects of data with the confirmed ensemble are not fully understood, we can underscore the importance of timing, location and type of data has on the parameter estimation and history match. Specifically, in the case of the oil pressure data observing the point of breakthrough was critical to a good estimate. Whilst in the case of water saturation snapshots, the earlier snapshot data is applied the better the estimate. In neither case was having lots of data crucial, although it typically did not hurt, as was the case with snapshots albeit with diminishing additional returns. However, although snapshots did improve estimates, they cause large jumps in uncertainty. Furthermore, oil pressure data assimilated as the flood front passed into the middle of the core actually increased the uncertainty. These highlight the need for additional predictor-corrector methods to control such corrections and deviation. One such method could be the implementation of a Maximization step. This lies outside the scope of this thesis, see appendix [**?**] for an attempt. Nonetheless, the relationship between uncertainty and data still needs to be further developed, see appendix [**?**] for an attempt at further developing.

4.5. Chapter Summary

In chapter 4, we investigated whether EnKF with a Confirmation step (EnKC) was an adequate application to deal with the conservation of mass issues of the EnKF, whilst at the same time still able to solve the history matching problem. In the first section, we compared the parameter estimate when using the EnKC compared to when just using the EnKF. We saw that although EnKC was able to estimate the parameter accurately, it proved to be less accurate and less efficient than the EnKF. In that, by only fitting the parameter and then using a confirmation step, we continue to accumulate the parameter uncertainty into the dynamic state. This in turn would feedback into the parameter estimate. Initially, this meant that the uncertainty in dynamic state and static parameter was as good as the EnKF. However, this compounding of parametric uncertainty in the confirmed forecasts (i.e. as forecast errors) implies that the EnKC will always under-perform against the EnKF. Furthermore, during breakthrough the EnKC experienced filter divergence in the oil pressure which further hindered the history match. On the other hand, the confirmation step reintroduced the non-linearity of the physics, and consequently represents a more accurate portrayal of the physics. Herein, we are presented with this trade-off between an efficient and "accurate" data assimilation technique via the EnKF, or an additional confirmation step, respecting the non-linearity of the physics which results in a larger spread in return for mass conservation within ensemble members. All in all, in the case of the base experiment, the EnKC is a viable history matching procedure.

In the second section, we compared the analyzed and confirmed ensembles of the EnKC. Using what was found in Section 3.4, we analytically examined the traces of the observed analyzed and confirmed EnKC ensemble covariances and found that the difference in these traces represented the amount of dispersion incurred to the dynamic state variables by the EnKF analysis. From the results of the base experiment we confirmed this, and realized that the confirmation step acts as a single step smoother on the analyzed ensemble. We then examined this in the context of Energy Potentials between the analyzed and confirmed ensembles and by analytically decomposing this difference we show which terms cause the mass to "fluctuate" unnaturally across the domain as a result of the EnKF analysis. Hence, answering why EnKF may fail to be mass conservative across the domain. Finally, we show that there is a quadratic form in terms of a Forecast Error covariance between the analysis and confirmation plus an additional data contribution, this is equal to the cross-covariance between the ensemble forecast and confirmed ensemble which is equivalent to uncertainty from the uncertain parameter. However, the additional data contribution is still not fully understood, and constructing a further minimization lies outside the scope of this thesis. Although, in appendix C an attempt was made.

In the third section, we examined the effects of data on the parameter estimation and history matching procedure. Specifically, in the first subsection, we investigated whether the EnKC could effectively provide a parameter estimate and history match given a reduced oil pressure data set. As expected, we found that with less data available, the greater the uncertainty in the final data assimilation, for both parameter estimate and history match. Poorly timed data caused filter divergence, whilst well timed data could outperform multiple data assimilations. In the case of the parameter estimate, data assimilated as the profile front approached in the middle of the core caused filter divergence, unless the data sampling rate was sufficiently high. That is, if you sample data "fast" enough then you can avoid filter divergence caused by the nonlinearity of the physics, but then again this might just be overfitting. In terms of the S^w history match, we observed that the "no-data" case served as an upper bound on the uncertainty, and although even a little data improved the history match but to really minimize the uncertainty a minimum data sampling rate was required. Interestingly, although more data typically implied greater accuracy, we found that the timing of the data was most critical. Observing the covariance trace, we clearly saw how each data assimilation could introduce or remove uncertainty in the ensemble. Meanwhile, we also witnessed how a single data point can outperform far larger data sets. Specifically, data assimilated at the point of breakthrough had the greatest reduction of the uncertainty in the history match. Given the effectiveness of well-placed data, we further investigated the timing of a single data assimilation and found that the point of breakthrough produced the smallest uncertainty. Whilst assimilating data as the flood front is forming can actually cause filter divergence and produce a worse estimate than no data at all.

Subsequently, in the second subsection, we investigated the effect of augmenting the data set with water saturation snapshots. Again, we found timing to be critical, the earlier the better as each snapshot mitigates the compounding of forecast errors. Wherein the closer we got to breakthrough the less effective snapshots were at reducing the uncertainty in both the parameter estimate and the S^w history matching. In the case of the parameter estimation multiple data snapshots were shown to be effective in reducing the uncertainty. Whilst for the S^w history matching, a single early snapshot could outperform 3 snapshots and be nearly as

good 4 or 7 snapshots. However, the uncertainty does not simply “vanish” instead large jumps in the oil pressure occur as the large corrections in λ cause discontinuities in the physical state that are then accounted for in the confirmed oil pressure. Observing the covariance trace, we clearly saw how each snapshot results in a large jump in the observable uncertainty of the ensemble, whereby the first snapshot introduces the greatest amount of uncertainty.

5

Conclusions and Recommendations

5.1. Conclusion

As technology has advanced, oil analysis has become more important as predictions regarding well output is probably the most cost-effective technique on the market today, especially in the field of secondary extraction. The research objective of this thesis was to develop a tool that can quantify the uncertainty of a core flood model and a parameter estimation routine. Based on the results of this study, it is concluded that the Ensemble Kalman Filter is capable of estimating the Corey parameter λ and with an additional confirmation step can be effective in history matching. As mathematical modeling is increasingly relied upon, this could be a good first choice amongst those techniques available due to its flexibility and relative robustness which allows it to be applied to many classes of geophysical problems with ease. Although some issues still need to be ironed out and the filter developed further. The most important achievement here, is that the developed model has shown that it is possible to estimate the Corey Parameter and minimize its uncertainty as well as the parametric uncertainties propagated into the model's water saturation and oil pressure.

A successful parameter estimation and history matching procedure for core flooding requires a model that accurately incorporates the features and physical phenomena which influence the physics be derived. It should be obvious that history matching is heavily dependent on the quality of present parameters used and the past historical data it is matched with: the more accurate the initial knowledge the more accurate the outcome. The physics of two-phase flow in porous media is influenced by geophysical features and phenomena (e.g. k_{r*}, p^c), as well as constraints, parameterizations, and other conditions on the dynamic variables (e.g. S^w, p^o) and are then used to construct a mathematical model. An implicit (oil) pressure and explicit (water) saturation (IMPES) scheme is then discretized in order to produce predictions that can be used in a history matching procedure, with the aim of estimating the parameter that produces the prediction matching closest to the measured data. Although, estimating λ is intractable and consequently ill-conditioned which therefore implies it is an ill-posed problem.

Traditionally, this would require technicians to manually calibrate λ to match simulations to data which was a time-consuming task. As a result, much research has been invested in automated history matching procedures, wherein Monte Carlo-like approaches show the potential to be accurate and time efficient in estimating unknown parameters and history matching. These methods are known to be easy, flexible and are able to incorporate all available measurable numerical data. However, such approaches are typically limited to small-scale problems and are prone to incur enormous computational costs, as they marginalize over increasing number of components in a problem as well as a sufficiently large number of samples/ensembles, $N_{samples} > \mathcal{O}\left(10^6\right)$. In addition, many history matching methods only operate on batched datasets and cannot assimilate data points iteratively, implying that there is the problem of continuous real-time model updating.

In order to overcome the obstacles above, we implemented the sub-optimal Kalman filtering method; the ensemble Kalman filter (EnKF). The EnKF employed was able to estimate the unknown parameter λ accurately and computationally efficiently. However, the EnKF assumes Gaussian measurement noise, and as a result we foresaw that estimating λ would yield physically infeasible estimates which extend outside the parameter's domain $\lambda > 0$. To solve this, we implemented an ln-parameterization to λ which we showed to improve convergence in the mean at the possible expense of the uncertainty.

Using the base twin experiment, the final accuracy of the parameter estimate produced by the EnKF was unaffected by the initial mean guess and/or spread of the unknown parameter. Only the true value to be estimated was affected by the final accuracy of the estimate.

As a robustness check we introduced an unaccounted fixed offset $\delta c = 4\%$ in the injection rate, as typically this would be a fixed known laboratory setting. What resulted was unphysical EnKF estimates which compounded and caused filter divergence and ensemble bias. In addition, we showed that there would be a $\|\mathcal{O}(\delta c)\|^2_{C^{-1}_{ee_{i(j)}}}$ error. Hence, for a successful data assimilation using the EnKF, all uncertainties must be incorporated into the data assimilation model, otherwise unphysical errors will compound.

Therefore, for the EnKF to successfully estimate the parameter, all sources of epistemic uncertainty must be accounted for.

Through twin experiments, we showed that the parameter uncertainty was bound below by the Cramér-Rao Lower bound. In the case of a linear model, given data $d \in \mathbb{R}^m$ the variance must be greater than $\frac{\sigma_{p^o}}{m}$, then including T additional observations, the variance must be greater than $\frac{\sigma_{p^o}}{mT}$. As a result, the CRLB is determined by the number of measurements, the uncertainty from the measurement noise or inversely the Fisher informativeness of the unknown parameter. Thus, we showed, as measurements became more certain the uncertainty of the acheivable λ estimate decreased, I.e. $\sigma_{p^o} \to 0 \implies \mathrm{Var}[\lambda] \to 0$. However, the nonlinearity model induced a bias dominating which would cause the ensemble error to overcome or fail to reach this bound. We found larger ensemble sizes could better capture the true uncertainty and as a result were less prone to being dominated by the nonlinear bias, albeit with diminishing returns. In addition, in spite of measurement location error of $\mathcal{O}\left(\frac{d_{Ng}}{2}\right)$, we showed that estimating with synthetic datasets with finer grids sizes were more informative and efficient estimators than with a coarser grids. This implies that the EnKF could be used with coarse forecasts to estimate the parameter then re-forecasted with a refined mesh. However, increasing the grid of the ensemble state was shown to increase the condition number, i.e. more ill-conditioned. Although this was expected, as there is only one degree of freedom (i.e. λ), it also implies that the risk of introducing spurious numerical errors increases.

We also found analytically that the shrinkage in the EnKF update violated the mass conservation of individual ensemble members at each data assimilation in addition to the aforementioned nonlinearity.

In order to tackle the problem of mass conservation between data assimilations, we introduced the same confirmed ensemble Kalman filter (EnKC), or an additional 'confirmation' step at the end of each data assimilation. This re-forecasts the dynamic state with the more certain parameter estimates, and so allows for a conserved 'fitted' forecast. This 'fitted' forecast acts as a one-step smoothener with respect to the parameter estimate. However, it is not computationally efficient as the EnKF as it doubles the forecast's computational costs and, forecast errors continue to compound as the dynamic state is never corrected. Still, the confirmation step is needed in order to produce a physically realistic history match.

Using a smoothener, the confirmed ensemble Kalman filter (EnKC)further improves the EnKF's ability to handle nonlinearity although it yields a smaller statistical bias.

By analytically showing that the trace of the sample covariance matrix is the mean energy norm of the ensemble, which we use to study the difference in traces of the observed analyzed and confirmed EnKC ensemble covariances and found this difference represented the amount of dispersion incurred to the dynamic state variables by the EnKF analysis. In addition, by analytically decomposing the trace we show which components cause the mass to "fluctuate" unnaturally across the domain as a result of the analysis, and why the EnKF may fail to be mass conservative across the domain. Furthermore, given that we know the ensemble energy is equivalent to the uncertainty in the observable state, and this is defined by the uncertainty in the parameter estimate then inversely indicates the informativeness on the estimate. Both the EnKC and EnKF can estimate unknown parameters. However, the EnKC does not perform as well as the EnKF when estimating the unknown parameter or minimizing the uncertainty in the history matching procedure, and as a result is not as informative. On the other hand, the history matches provided by the EnKC respect the nonlinearity of the physics, whilst the EnKF only does so in the mean sense. Finally, we show that there is a quadratic form in terms of Forecast Error Covariance between the analysis and confirmation, plus an additional data contribution equal to the cross-covariance between the ensemble forecast and confirmed ensemble, which is equivalent to uncertainty from the uncertain parameter.

Quantity, timing, location and type of data affected the parameter estimate and history match produced by the EnKC greatly. Typically, less data implies more uncertainty in the final parameter estimate and history match. However, a single well-timed and placed point of data could nearly outperform a dataset of 400 data points. That is, the timing of data was found to be absolutely critical! Whereby we showed that oil pressure

data measured nearer to the time of breakthrough produced the best history matches. Whereas oil pressure data measured whilst the flood front approached the middle of the core would lead to filter divergence. Augmenting datasets with water saturation snapshots proved to be effective in improving the λ estimate and history match for the water saturation but would introduce uncertainty in the oil pressure. Herein the earlier the snapshot the greater the benefit, where-after each subsequent snapshot contributed diminishing returns to the history matching. Specifically, a single snapshot at the 100th iteration was more informative than snapshots occurring at the [200, 400, 600] iterations. This clearly underscores the importance of timing, location and type of data.

5.2. Recommendations

Below are a series of recommendations:

- To deduce if this study can be used for multiple parameters it would be interesting to study a variety of parameters to estimate the phenomena of parameterization in depth. We have only investigated the case wherein only one parameter is uncertain. However, in real world applications there are often multiple parameters that are uncertain. Moreover, many of the couplings that describe physical phenomena are often derived through statistical means.

- To see what the effects are on the uncertainty given such a coupling is removed. It could be interesting to study the Brooks-Corey Model for relative permeabilities and capillary pressure which under certain conditions may accurately represent reality. One such condition is the assumption that the Corey parameter for relative permeabilities and capillary pressure are the same, where in reality the values that best represent the two phenomena may be distinct, i.e. $\lambda_{k_r} \neq \lambda_{p^c}$. Moreover, whether the EnK(F/C) is still capable of estimating either/both parameters and provide an adequate history match. Preliminary experiments indicate this is the case (see appendix A.3). However, why this is the case, what are the effects of the missing constraint, and why they occur is very unclear.

- To understand some phenomena that may be the result of the simulator, EnK(F/C), or the combination thereof, for example, the explicit effects of the IMPES formulation, it would be interesting to compare against the simultaneous solution (SS) method, which is much more stable albeit computationally costly. Moreover, it would increase the size of the dynamic state in exchange for correlations between all the components of the model explicitly. Perhaps one simulator is better at capturing the true correlations during certain periods of the core flooding, but thereafter suffers from filter divergence. It would thus be interesting to investigate if there is a way to combine the two simulators, running the "best" simulator(/expert) for the time it delivers the "least regrettable" results. However, this is a very "meta" approach, and without fully understanding the effects of the EnKF may still suffer from it.

- To study the effects of generated "randomness" and "lucky draws" which are ever present, and typically solved with increased ensemble sizes and/or repeated experiments, it would be interesting to apply confidence regions and other common statistical techniques, as more samples could always be generated if a certain confidence is needed. However, as the Gamma distributed sampling distribution is incomplete when computing quantiles, a more workable sampling distribution would need to be selected.

- To implement quasi monte Carlo sampling techniques. Certain techniques are constructed on the bases that the total number of samples generated ensure that the space on which the monte Carlo integration is integrated guarantees the variation to be bounded. Thereby, minimizing the variation caused by "generated" randomness.

- To study the whole coupled uncertain forecast and analyzed ensembles in terms of a thermodynamic system, in which the "temperature" is the probability of an ensemble occurring. Here one would formulate the goal as reaching an adiabatic net state where the total uncertainty across the ensemble is ergodic. Such a formulation would reinforce the energy information formulation and would allow technicians more familiar with thermodynamics greater intuition for information-theoretic concepts of uncertainty. The beginnings of such a study can be seen in appendix B in which the Bayesian framework is extended with information-theoretic concepts and a broader measure of information, i.e. Kullback-Leibler divergence but were found to be beyond the scope of the present thesis, although the preliminary investigation shows promise.

- To study Lyapunov exponents as a measure of the rate of dispersion of ensemble members. This could be used to calibrate; simulators (e.g. grid refinement), or the number of forecasts possible before data is needed to prevent filter divergence. That is, Lyapunov exponents can be used to construct a forecast window. In addition, given that we used the confirmation step as a "smoother" with respect to the static parameters, it would be interesting to investigate how Lyapunov exponents could be used to minimize the propagation of forecasting errors. On the other hand, even with the confirmation step, estimates would still suffer from the effects of the EnKF fitting of the ensemble given observable correlations. As the dimensionality of the ensemble increases, either through more uncertain static parameters or a

larger dynamic state (e.g. via grid refinement), the chance of spurious correlations or filter degeneracy increases.

- To study whether Model reduction can be applied to EnKC, and whether this would improve the parameter estimate, history match, and/or prevent spurious correlation or filter degeneracy. One such method of interest is proper orthogonal decomposition. For example, applying it to the physical dynamic state, the problem can be reduced with respect to unknown static parameter(s) of interest. This reduced problem is in effect solved in an orthogonal subspace. In doing so, we would be able to avoid spurious spatial correlations whilst also avoiding filter degeneracy. However, to construct such an orthogonal subspace several "snapshots" may be needed at different time instances, i.e. a trailing window, to ensure that the variance of the resulting de-correlated components is maximized, i.e. the decomposition is well calibrated. However, such reductions are often limited to removing first order correlations and may not address the effects of the confirmation step on the parameter estimate.

- To extend the EnKC into the Expectation-Maximization framework in order to improve the parameter estimate is of great interest. In appendix C, an attempt was to build on the work of [11], but with a twist. Specifically, replacing the EnKS smoothed state with the EnKC confirmed state, in order to estimate the covariance matrix of the forecast error and "re-measurement" error. That is, by then augmenting the data assimilation model, as in [16, pg.142], with the confirmed state (which now has an additive sampled forecast error) as a prediction of measurements given the analyzed unknown parameter, we re-analyze the unknown parameter of the forecast. Preliminary investigations show the method to be very effective, however the justification and theoretical build up is incomplete, see appendices B, C. Specifically, how to cast the confirmed state as "measurement predictions" with Gaussian errors using the EM covariance result, and similarly the "re-measurement" data errors.

Evidently such techniques would isolate the effects caused by the EnKF and allow clearer investigation into the nonlinear model effects. Nonlinear system identification is of much interest with many different approaches being investigated, we encourage the reader to be curious.

A

Appendix

A.1. Effects of Initial Parameter Sampling Distribution and Parameterization of Static components of Ensemble State

Using the EnKF under the base experiment, we show the effect of using a clipped parameterization $clip(\lambda)$, as well as the effect of a Truncated Normal initial sampling distribution for λ with the $\ln\lambda$ parameterization or with a clipped parameterization $clip(\lambda)$:

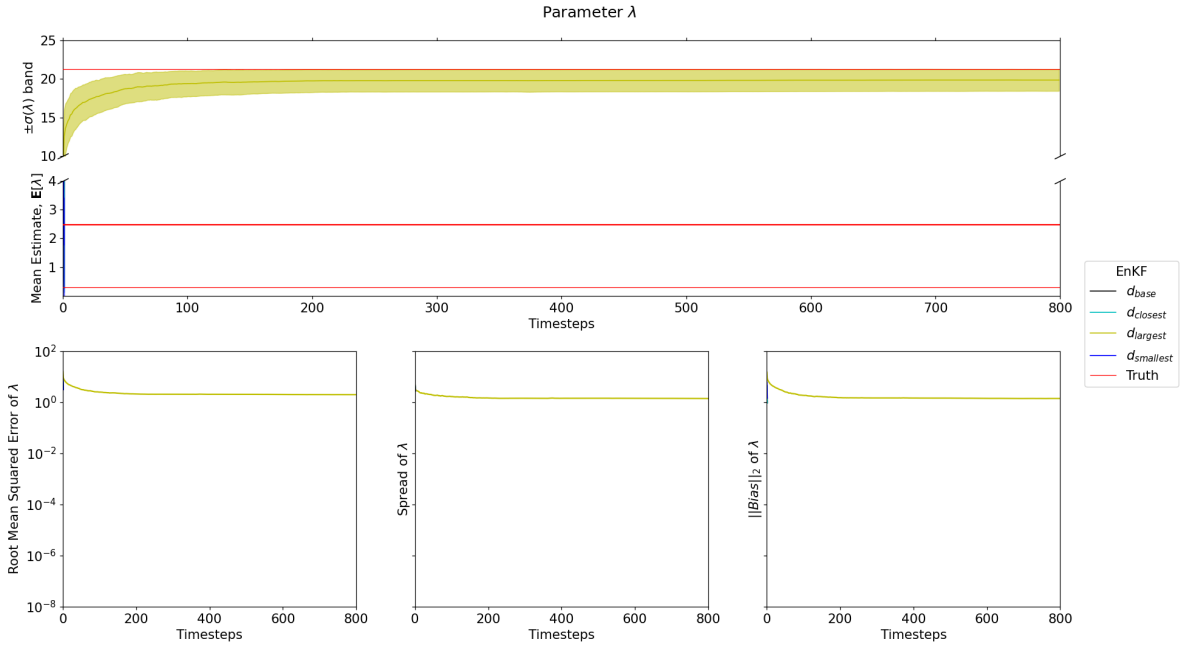


Figure A.1: $\alpha = clip(\lambda)$, where $clip$ applies $\max\lambda = 10$ and $\min\lambda = 0$

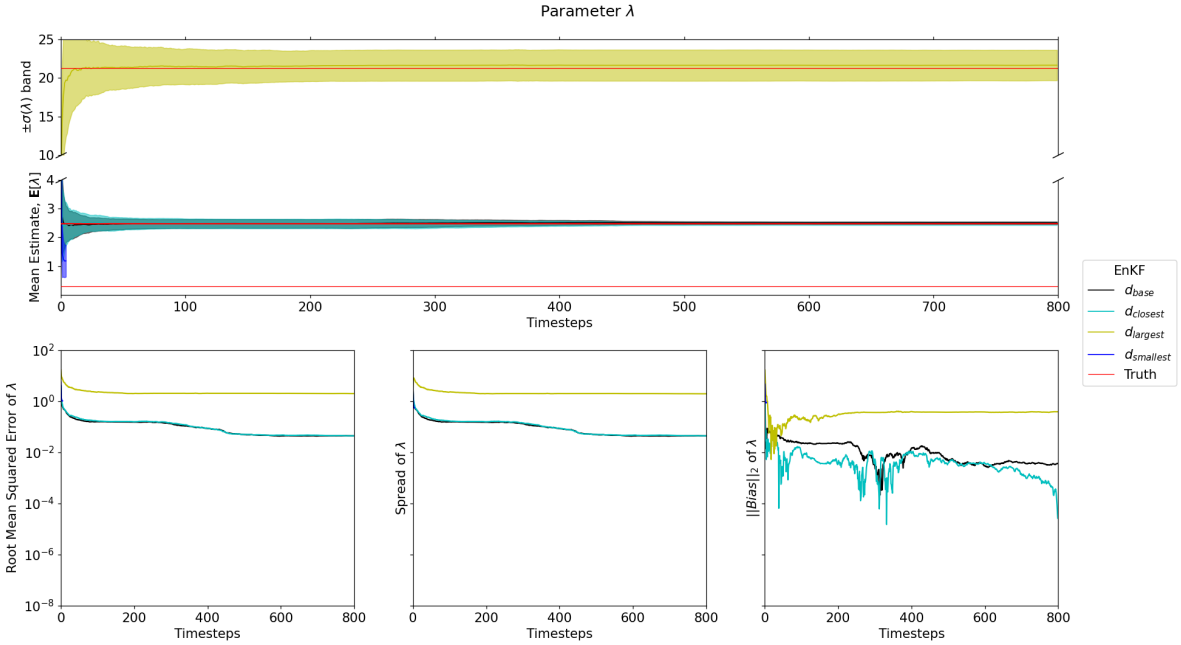


Figure A.2: λ ~TruncatedNormal($\lambda_{guess}, \sigma^2_{\lambda\lambda}$)

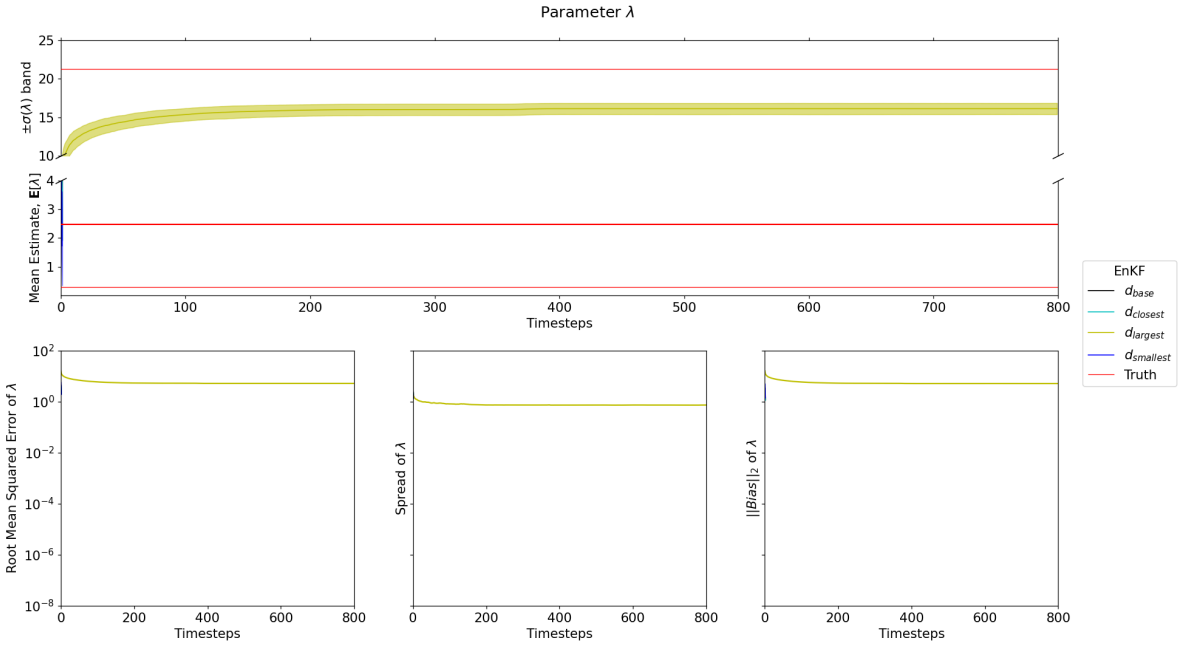


Figure A.3: $\alpha = clip(\lambda)$ and λ ~TruncatedNormal($\lambda_{guess}, \sigma^2_{\lambda\lambda}$), where $clip$ applies $\max \lambda = 10$ and $\min \lambda = 0$

A.2. Effects of Location of Oil Pressure Data

Using the EnKC under the base experiment (i.e. assuming data at every forecast), we show the effect of placing the oil pressure sensor at different locations:

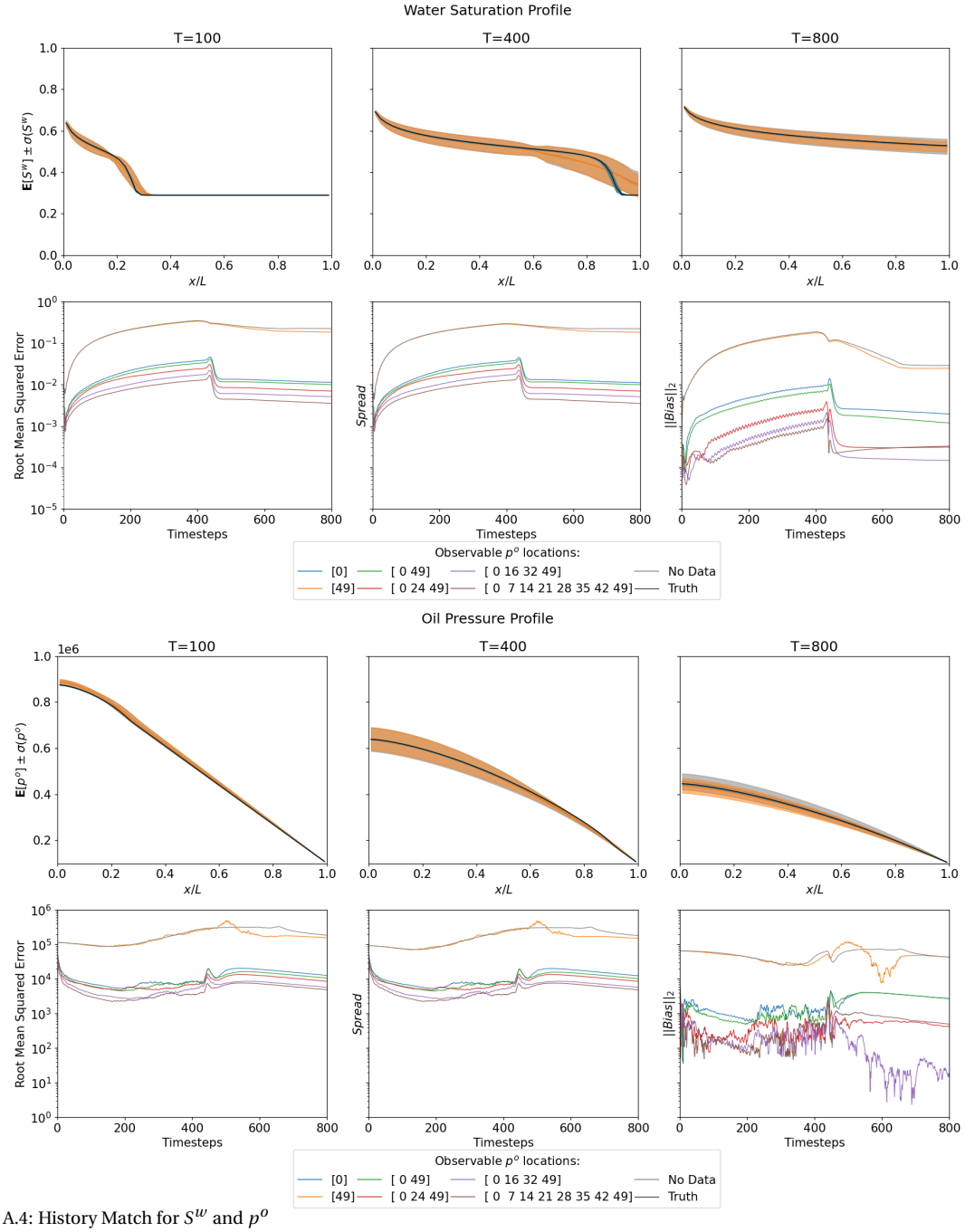


Figure A.4: History Match for S^w and p^o

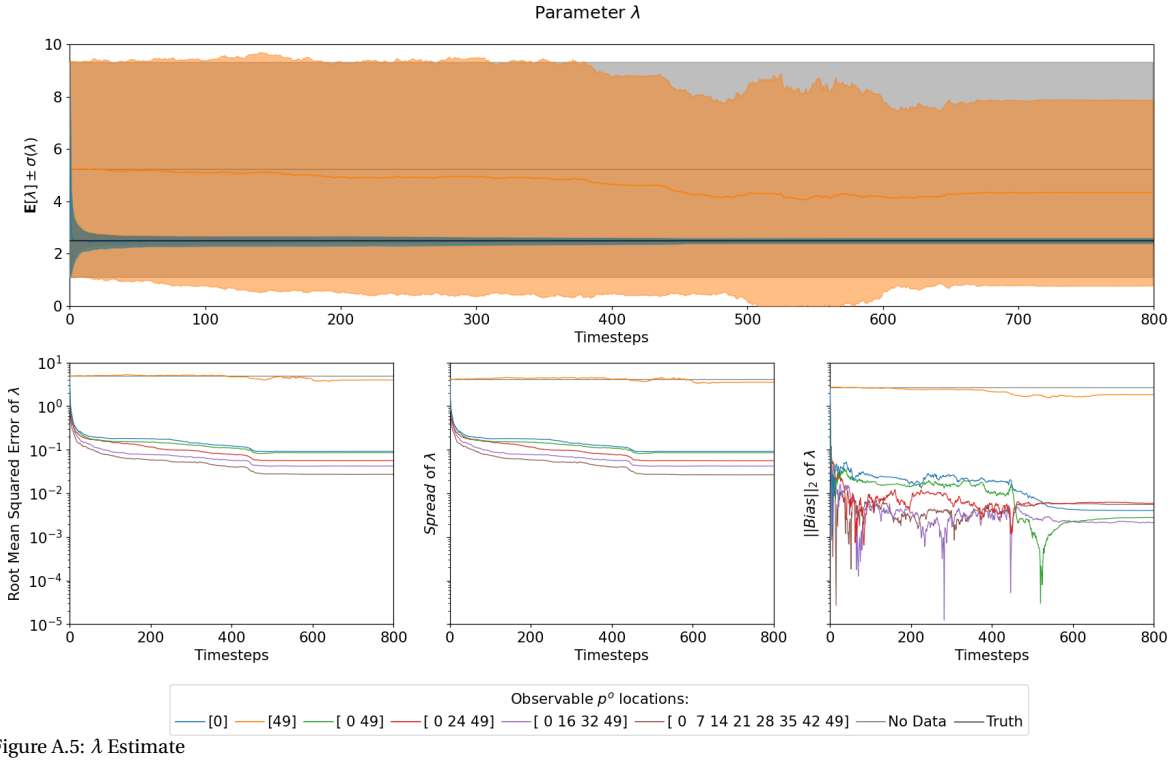


Figure A.5: λ Estimate

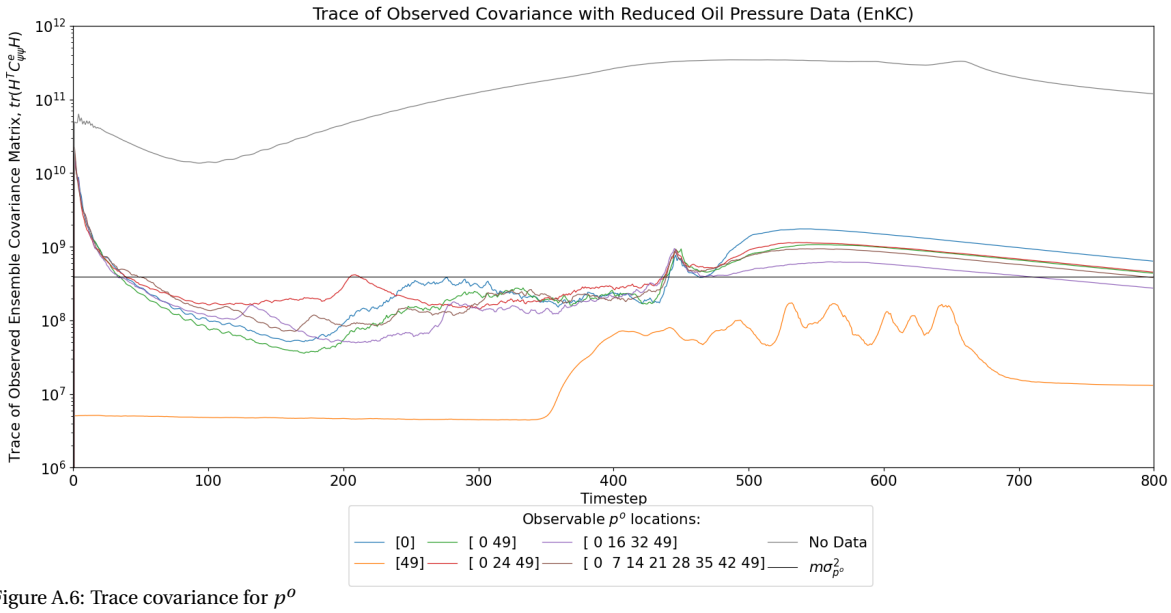


Figure A.6: Trace covariance for p^o

A.3. Splitting λ into λ_{k_r} and λ_{p^c}

Using the EnKC under the base experiment (i.e. assuming that $\lambda = \lambda_{k_r} = \lambda_{p^c}$), we attempt the simultaneous parameter estimation proceedure of $(\lambda_{k_r}, \lambda_{p^c})$, and provide a history match for S^w:

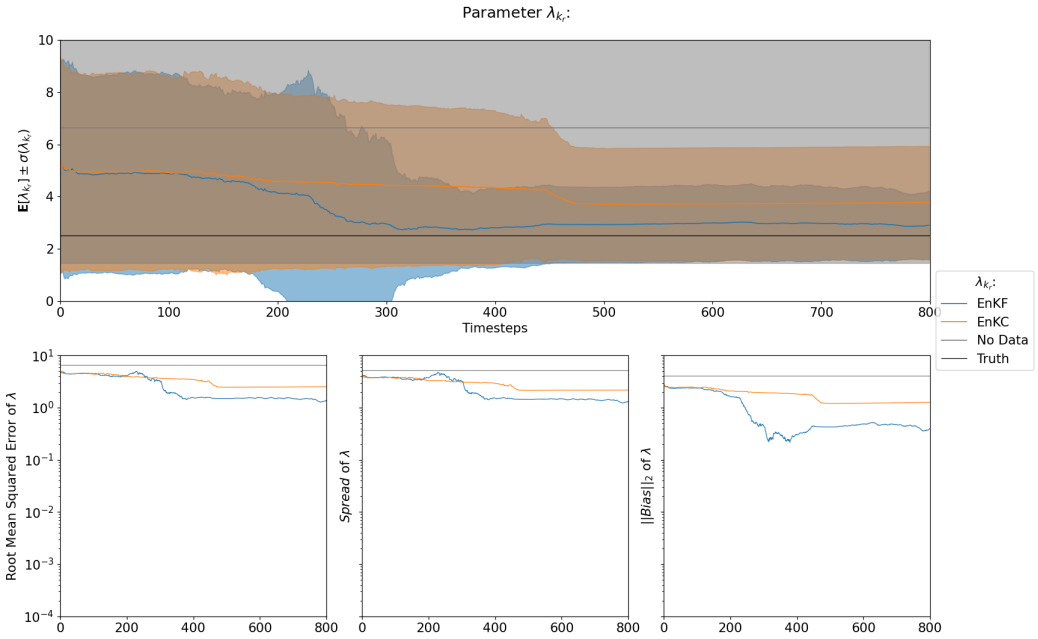


Figure A.7: Estimating λ_{k_r} assuming base settings for λ

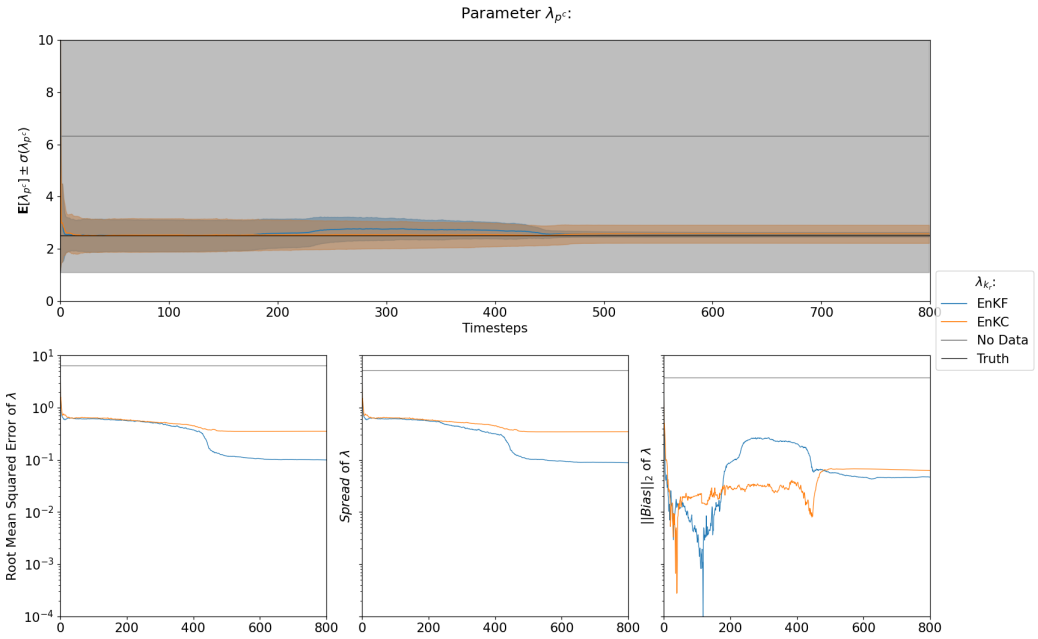


Figure A.8: Estimating λ_{p^c} assuming base settings for λ

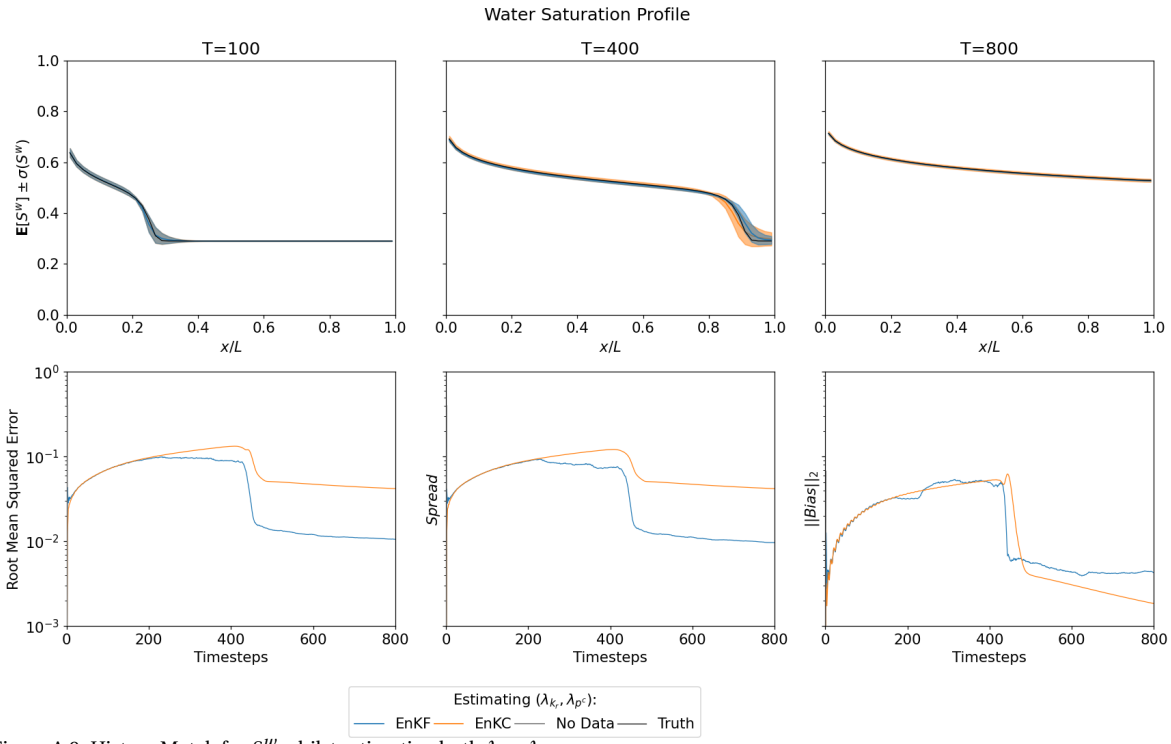


Figure A.9: History Match for S^w whilst estimating both $\lambda_{p^c}, \lambda_{k_r}$

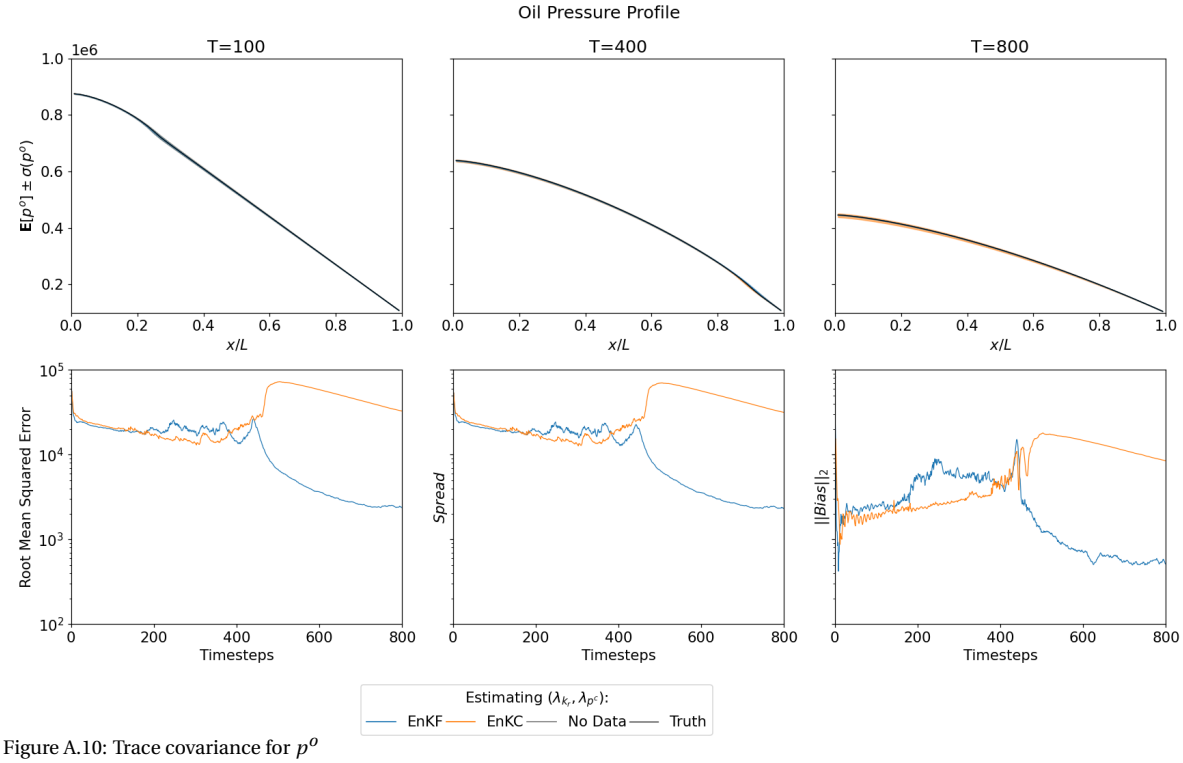


Figure A.10: Trace covariance for p^o

B

Information Energy

In the preceeding chapters we have defined the concept of information as the inverse of uncertainty, but this Fisher formulation is a frequentist's conclusion and so is only reactive of observations. As such, the current information formulation is incomplete. Is there a more constructed link between the ensemble energy and information? If so, can we then concretely define how informative improving the parameter estimate is to the dynamic state uncertainty? Moreover, how does the data come into play in this framework? Is there a more concrete construction between the ensemble uncertainty and the data?

Keeping in line with the EnKF's Bayesian formulation to estimate our unknown parameter λ, we must be proactive given that we update our uncertain beliefs with incoming data. That being said in order to differentiate between prior and posterior distributions, or analyzed and confirmed ensemble, we must define some kind of measure of similarity. In this chapter, we begin by defining the Kullback-Leibler (KL) Divergence wherein the fisher information defines the Hessian of KL-divergence between two distributions. Thereafter, we show that minimizing the KL-Divergence is equivalent to minimizing the maximum likelihood estimate. Given this broader notion of information we show that there is a connection to the ensemble energy as a function of the static parameters. In order to build some intuition for this broader notion of information we follow Särkkä's interpretation of an "Energy Function", for an unknown α, as the unnormalized negative log-posterior[37]. We derive the energetic effect of the parameterization on this "energy function". There after, we apply the base experiment results when using the EnKF, and in order to overcome the increasingly high dimensionality of data, we make some approximations/heuristics.

B.1. Information Interpretation:

This must also be able to incorporate our fisher formulation and be consistent with the likelihood. Only then can quantify how informative the confirmation step is compared to without. In order to do this, we first take the following preliminary steps:

1. We begin by defining the Kullback-Leibler (KL) Divergence, wherein the Fisher Information Matrix defines the local curvature in distribution space for which KL-divergence is the metric.

2. In addition, we realize that maximizing the likelihood is equivalent to minimizing KL-Divergence.

By combining the following steps we are able to establish a link between the energy across the ensemble and the informativeness of the ensemble.

B.1.1. Kullback-Leibler divergence and its' Hessian, the Fisher information

Having defined our likelihood function we now must define some kind of measure of similarity between sets of ensembles. That is, a directed 'distance' between any two distributions in which information is used as the measure. Only then can we determine what ensembles are most similar to each other, or the true posterior distribution. One such measure, is the Kullback-Leibler divergence:

Definition B.1.1 (Kullback-Leibler information[39])**.** *Suppose P and Q are two probability measures with densities p and q with respect to a common measure ν. The Kullback-Leibler (KL) information of Q from P is defined by*

$$D_{KL}[P \,||\, Q] = \int \log \frac{p(y)}{q(y)} p(y) d\nu(y)$$

In case of parametric families, where $Y \sim P_\vartheta$ and $P_\vartheta \ll \nu$ for all ϑ with densities $f_Y(\cdot \mid \vartheta)$, we write

$$D_{KL}\left[\vartheta \,||\, \vartheta^*\right] = \mathbb{E}_\vartheta \left[\log \frac{f_Y(Y \mid \vartheta)}{f_Y(Y \mid \vartheta^*)}\right], \quad \vartheta, \vartheta^* \in \Omega$$

This is the KL information of ϑ^* from ϑ based on the data Y.

Given this, what exactly is the connection between Fisher Information Matrix and KL-divergence? It turns out, Fisher Information Matrix $\mathscr{I}(\vartheta)$ defines the Hessian of KL-divergence between two distributions $f(y \mid \vartheta)$ and $f(y \mid \vartheta^*)$, with respect to ϑ^*, evaluated at $\vartheta^* = \vartheta$.

KL-divergence can be decomposed into entropy and cross-entropy term, i.e.:

$$D_{KL}\left[f(y \mid \vartheta) \,||\, f(y \mid \vartheta^*)\right] = \mathbb{E}_{f(y|\vartheta)}\left[\log f(y \mid \vartheta)\right] - \mathbb{E}_{f(y|\vartheta)}\left[\log f(y \mid \vartheta^*)\right]$$

The first derivative with respect to ϑ^*:

$$\begin{aligned}
\nabla_{\vartheta^*} D_{KL}\left[f(y \mid \vartheta) \,||\, f(y \mid \vartheta^*)\right] &= \nabla_{\vartheta^*} \mathbb{E}_{f(y|\vartheta)}\left[\log f(y \mid \vartheta)\right] - \nabla_{\vartheta^*} \mathbb{E}_{f(y|\vartheta)}\left[\log f(y \mid \vartheta^*)\right] \\
&= -\mathbb{E}_{f(y|\vartheta)}\left[\nabla_{\vartheta^*} \log f(y \mid \vartheta^*)\right] \\
&= -\int \nabla_{\vartheta^*} \log f(y \mid \vartheta^*)\, f(y \mid \vartheta) dy
\end{aligned}$$

The second derivative with respect to ϑ^*:

$$\nabla^2_{\vartheta^*} D_{KL}\left[f(y \mid \vartheta) \,||\, f(y \mid \vartheta^*)\right] = -\int \nabla^2_{\vartheta^*} \log f(y \mid \vartheta^*)\, f(y \mid \vartheta) dy$$

Thus, the Hessian with respect to ϑ^* evaluated at $\vartheta^* = \vartheta$ is:

$$\begin{aligned}
H_{D_{KL}[f(y|\vartheta)||f(y|\vartheta^*)]} &= -\int \nabla^2_{\vartheta^*} \log f(y \mid \vartheta^*)\big|_{\vartheta^*=\vartheta}\, f(y \mid \vartheta) dy \\
&= -\mathbb{E}_{f(y|\vartheta)}\left[\nabla^2_{\vartheta^*} \log f(y \mid \vartheta^*)\big|_{\vartheta^*=\vartheta}\right] \\
&= \mathscr{I}(\vartheta)
\end{aligned}$$

More generally, $\mathscr{I}(\vartheta)$ defines the local curvature in distribution space for which $D_{KL}\left[f(y \mid \vartheta) \,||\, f(y \mid \vartheta^*)\right]$ is the metric.

B.1.2. MLE minimizes the KL-divergence

Now to show that the MLE minimizes the KL-divergence, we take independent and identically distributed (i.i.d) data from $f(y \mid \vartheta^*)$:

$$\begin{aligned}\hat{\vartheta} &= \arg\max_{\vartheta} \prod_{i=1}^{N} f(y_i \mid \vartheta) = \arg\max_{\vartheta} \sum_{i=1}^{N} \log f(y_i \mid \vartheta) \\ &= \arg\max_{\vartheta} \frac{1}{N} \sum_{i=1}^{N} \log f(y_i \mid \vartheta) - \frac{1}{N} \sum_{i=1}^{N} \log f(y_i \mid \vartheta^*) \\ &= \arg\max_{\vartheta} \frac{1}{N} \sum_{i=1}^{N} \log \frac{f(y_i \mid \vartheta)}{f(y_i \mid \vartheta^*)}\end{aligned}$$

then by the law of large numbers

$$\begin{aligned}&= \arg\min_{\vartheta} \int \log \frac{f(y_i \mid \vartheta^*)}{f(y_i \mid \vartheta)} f(y_i \mid \vartheta^*) \, dy \\ \implies \hat{\vartheta} &= \arg\min_{\vartheta} D_{KL}\left[f(y_i \mid \vartheta^*) \,||\, f(y_i \mid \vartheta) \right]\end{aligned}$$

where the integral term is the KL-Divergence. Hence, it is clear that maximizing the MLE is equivalent to minimizing the Kullback-Leibler Divergence. [32]

In order to verify this is indeed what occurs, we will compute the KL-Divergence between the data with noisy measurements and the observed ensemble state which we assume to be Gaussian. This assumption implies we can use the closed form of the KL-Divergence between 2 multivariate normal distributions:

$$D_{KL}[\mathcal{N}_0 \,||\, \mathcal{N}_1] = \frac{1}{2}\left[\operatorname{tr}\left(\Sigma_1^{-1}\Sigma_0\right) + (\mu_1 - \mu_0)^\intercal \Sigma_1^{-1}(\mu_1 - \mu_0) - k + \ln \frac{\det \Sigma_0}{\det \Sigma_1} \right]$$

here k is the dimension of the observed data, i.e. $k = m$ or $k = N_g + m$. The inversion of Σ_1 implies we are only able to compute the reverse KL-Divergence:

$$D_{KL}\left[\mathcal{N}_e \,||\, \mathcal{N}_\psi\right] = \frac{1}{2}\left[\operatorname{tr}\left(C_{\epsilon\epsilon}^{-1} S S^\intercal\right) + (d_{obs} - \mathcal{H}\bar{\psi})^\intercal C_{\epsilon\epsilon}^{-1}(d_{obs} - \mathcal{H}\bar{\psi}) - k + \ln \frac{\det C_{\epsilon\epsilon}}{\det S S^\intercal} \right]$$

Note that in the case that $\mu_1 = \mu_0$:

$$D_{KL}[\mathcal{N}_0 \,||\, \mathcal{N}_1] = \frac{1}{2}\left[\operatorname{tr}\left(\Sigma_1^{-1}\Sigma_0\right) - k + \ln \frac{\det \Sigma_1}{\det \Sigma_0} \right].$$

Hence, $\bar{d} = \mathbb{E}[d] = \mathbb{E}[d_{obs} + \epsilon] = d_{obs}$, and so we see when comparing Sample Measurements against Asymptotic Measurements:

$$D_{KL}[\mathcal{N}_e \,||\, \mathcal{N}_d] = \frac{1}{2}\left[\operatorname{tr}\left(C_{\epsilon\epsilon}^{-1} E E^\intercal\right) - k + \ln \frac{\det C_{\epsilon\epsilon}}{\det E E^\intercal} \right]$$

then assuming that $EE^\intercal \to C_{\epsilon\epsilon}$ as the ensemble size $N \to \infty$:

$$= \frac{1}{2}\left[k - k + \ln 1\right] = 0$$

In Figure B.1, we plot the Kullback-Leibler Divergence between the data with noisy measurements and the results of the EnKF and EnKC, before and after the confirmation step.

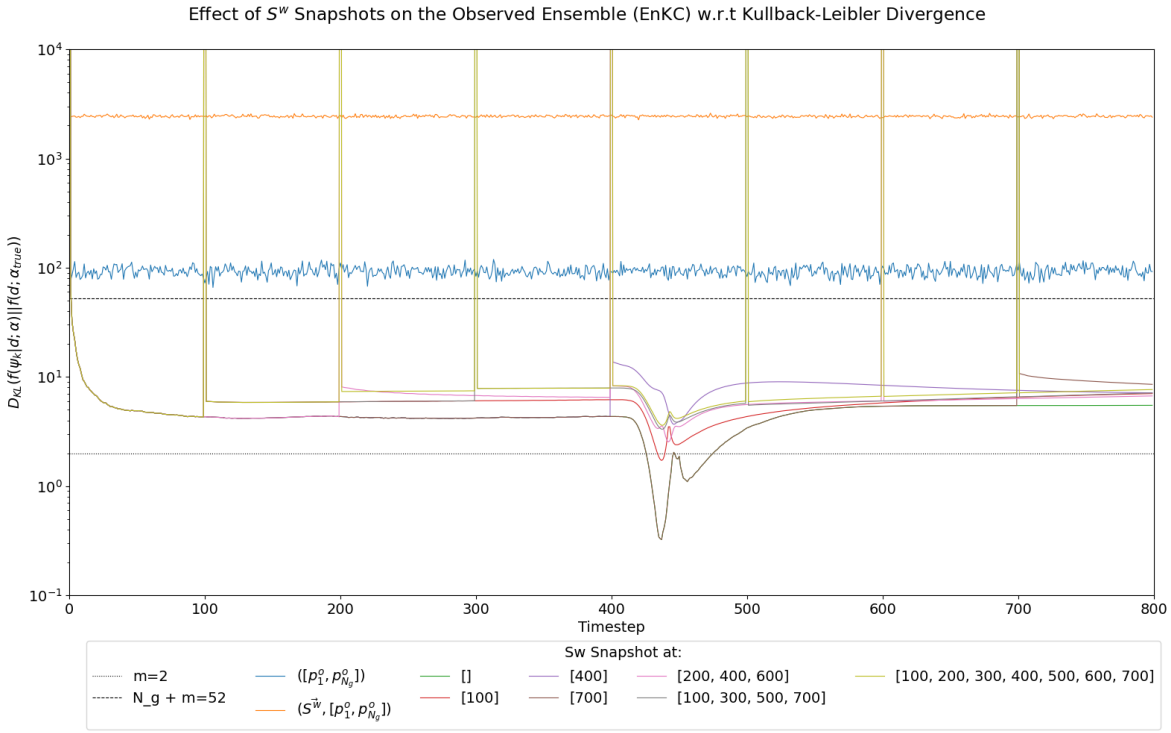


Figure B.1: Kullback-Divergence between the data with noisy measurements and the observed ensemble state which we assume to be Gaussian.

B.2. The Energy Connection

In what follows, we want to incorporate the information theoretic Kullback-Divergence into our Bayesian formulation in order to define energetic components that can then be used to refine the estimate.

Suppose $p(d \mid \alpha)$ is the true intractable distribution of the observed data, onto which we apply Bayes Theorem:

$$p(d \mid \alpha) = \frac{p(d, \psi \mid \alpha)}{p(\psi \mid d, \alpha)}$$
$$\implies \ln p(d \mid \alpha) = \ln p(d, \psi \mid \alpha) - \ln p(\psi \mid d, \alpha)$$

Now suppose there is a tractable approximation $q(\psi) \approx p(\psi \mid d, \alpha')$:

$$\begin{aligned}\ln p(d \mid \alpha) &= \ln p(d, \psi \mid \alpha) - \ln p(\psi \mid d, \alpha) + \ln q(\psi) - \ln q(\psi) \\ &= \left(\ln \frac{p(d, \psi \mid \alpha)}{q(\psi)}\right) + \left(-\ln \frac{p(\psi \mid d, \alpha)}{q(\psi)}\right)\end{aligned}$$

Since expectation with respect to p are assumed to be intractable we take the expectation with respect to $q(\psi)$:

$$\mathbb{E}_q\left[\ln p(d \mid \alpha)\right] = \left(\int q(\psi) \ln \frac{p(d, \psi \mid \alpha)}{q(\psi)} d\psi\right) + \left(-\int q(\psi) \ln \frac{p(\psi \mid d, \alpha)}{q(\psi)} d\psi\right)$$

Here since the loglikelihood $p(d|\alpha)$ is independent of ψ and $\int q(\psi) d\psi = 1$ we have $\mathbb{E}_q\left[p(d|\alpha)\right] = p(d|\alpha)$. Whilst on the RHS the right hand term is $D_{KL}\left[q(\psi) \,||\, p(\psi \mid d, \alpha)\right]$, and the left hand term we denote evidence lower bound as $\mathscr{L}(q, \alpha)$:

$$\ln p(d \mid \alpha) = \mathscr{L}(q, \alpha) + D_{KL}\left[q(\psi) \,||\, p(\psi \mid d, \alpha)\right]$$

Firstly, notice that the KL-Divergence is not symmetric, and so the right hand term is in fact the reverse KL-Divergence[1]. If one were to minimize this reverse KL, q will lock on to one of multiple possible modes. As a result q will typically under-estimate the support of p. That is, $D_{KL}\left[q \,||\, p\right]$ is infinite if $p = 0$ and $q > 0$.

[1] also known as an I-projection or information projection

Hence, if $p = 0$ we must ensure $q = 0$. Thus the reverse KL is known to be zero forcing for q[32]. Secondly, the KL divergence is always non-negative, and so we see that $\mathscr{L}(q,\alpha)$ is a lower bound of the log-likelihood, and thus is known as the Evidence Lower BOund (ELBO)[11]. This in turn can be interpreted as an energy functional[27], and thus is also known as the "Free-Energy" or "Helmholtz Energy". That is, if we separate $\mathscr{L}(q,\alpha)$ into 2 terms:

$$\begin{aligned}\int q(\psi)\ln\frac{p(d,\psi\mid\alpha)}{q(\psi)}d\psi &= \int q(\psi)\ln p(d,\psi\mid\alpha)d\psi - \int q(\psi)\ln q(\psi)d\psi \\ \implies \mathscr{L}(q,\alpha) &= \underbrace{\mathbb{E}_q\left[\ln p(d,\psi\mid\alpha)\right]}_{\text{Internal Energy Term}} - \underbrace{\mathbb{E}_q\left[\ln q(\psi)\right]}_{\text{Entropy Term}}\end{aligned} \tag{B.1}$$

Clearly the right term, is the entropy of q. This entropy term can be interpreted as a measure of spontaneous change of the ensemble given the data, in which changes requires work done (i.e. energy). Alternatively, this spontaneous change can be considered a measure of the inherent irreducible uncertainty (i.e. aleatoric) or the amount of missing information. Whilst the left term, is the internal energy in the ensemble state and data, which is composed partially of kinetic energy and potential energy. Notice that if we continue following our Gaussian assumption, in that we expect the mean to be the truth, then by taking the trace[2] of these expectations we recover the same energetic interpretations from subsection 3.4.2, only now with respect to the tractable approximation q. However, if q is approximately exact, i.e.

$$q(\psi) \approxeq p(\psi|d,\alpha') = p(d|\alpha) = p(\psi,d|\alpha),$$

then (B.1) is equivalent to the total energy in the mean state which by the LLN is the true state. Evidently, q is not trivial to deduce, but crucial as different forms of q may yield different results.

[2] In following the Gaussian assumption, optimizing for the trace implies minimizing the variance in favor for the mean. We could have chosen an alternative criteria in our experimental design. For example the determinant, which would have maximized all sources of fisher information (including cross terms) and thus taken into account free energy as well as the mean energy.

B.3. Computing the "Energy" of a parameter

So we have shown that within this broader notion of information there is a connection to the energy, at least in the Bayesian probabilistic sense. Is this even useful? In that, can the energy of the parameter even be computed? Evidently it is a function of α, but is there a closed form or just an approximation? Is there one that flows naturally from the EnKF? What kind of approximation does it yield?

In order to answer these questions, we first go to Särkkä [37], where they define the unnormalized negative log-posterior as an "Energy Function":

Definition B.3.1 (Energy function [37])**.**

$$\phi_T(\alpha) = -\log p(d_{1:T} \mid \alpha) - \log p(\alpha) \tag{B.2}$$

where $\phi_T(\alpha)$ is the energy after time step T.

This leads to the following remark:

Remark. The definition of the energy function thus implies that the posterior probability density of α is:

$$p(\alpha \mid d_{1:T}) \propto \exp(-\phi_T(\alpha)) \tag{B.3}$$

Now if we could marginalize (2.8) over the dynamic state ψ to get a marginal posterior in terms of α:

$$p(\alpha \mid d_{1:T}) = \int p(\psi_{1:T}, \alpha \mid d_{1:T}) d\psi_{1:T}$$

then we could easily compute the total energy in (B.3). Unfortunately, computation of this high-dimensional integral is hard thus becomes even harder as we obtain more measurements.

In which case only an approximation based on recursion is possible:

Theorem B.3.1 (Recursion for energy function[37])**.** *The energy function defined in* (B.2) *can be evaluated recursively as follows:*

- *Start from $\psi_0 = -\log p(\alpha)$*
- *At each step $k = 1, 2, \ldots, T$ compute the following:*

$$\phi_k(\alpha) = \phi_{k-1}(\alpha) - \log p(d_k \mid d_{k-1}, \alpha)$$

Särkkä builds on this for nonlinear gaussian models, which having gone through the derivations in his appendix, are applicable to our model which does not have model noise Q:

Theorem B.3.2 (Gaussian filtering based energy function[37])**.** *The recursion for the energy function is given as*

$$\phi_k(\alpha) = \phi_{k-1}(\alpha) + \frac{1}{2}\log(|2\pi| \det C_k(\alpha)) + \frac{1}{2} V_k^\intercal(\alpha) C_k^{-1}(\alpha) V_k(\alpha) \tag{B.4}$$

where the terms $V_k(\alpha)$ and $C_k(\alpha)$ are terms given by the Kalman/Gaussian filter with the parameters fixed to α.

Thus if we fix α and run the above algorithm from $\phi_0(\alpha) = -\log p(\alpha)$ at $k = 0$ to the step $k = T$, then we can compute the full energy function $\phi_T(\alpha)$. Notice in (B.4); the right most term $\frac{1}{2} V_k^\intercal(\alpha) C_k^{-1}(\alpha) V_k(\alpha)$ is the data-fit, the middle term $\frac{1}{2}\log(|2\pi| \det C_k(\alpha))$ is the complexity of the data.

Remark. However, the EnKF operates on an ensemble, and to achieve the same estimate we assume: $V_k(\alpha) = D(\alpha^{true}) - \mathscr{H}\big(\bar{A}_k(\alpha)\big)$ and $C_k(\alpha) = S_k(\alpha) S_k^\intercal(\alpha) + E_k E_k^\intercal$.

Now we can clarify what is meant by the data-fit and the complexity. Firstly, the complexity doesn't just define the volume of the covariance matrix of observed ensemble and measurements, i.e. $S_k S_k^\intercal + E_k E_k^\intercal$, but the total feasible ensemble energy which includes the free-energy between ensemble members and the added noise potential of the measurements.

Then, to compute the "energy" of each ensemble member's α and decouple the "energy" of other ensemble members we only use the diagonal of the data-fit:

$$\phi_k(\alpha) = \phi_{k-1}(\alpha) + \frac{1}{2}\log(|2\pi| \det C_k(\alpha)) + \frac{1}{2} V_k^\intercal(\alpha) C_k^{-1}(\alpha) V_k(\alpha))$$

Furthermore, the original formulation assumes α to be fixed, whilst in our case we have augmented the state to include α, so there is an energy mismatch due to updating $\alpha^{(k-1)} \to \alpha^{(k)}$ from each data assimilation, and eventually the covariance will become degenerate as the accuracy of the parameter exceeds computer precision. As a result we realize that the data-fit highlights that minimizing C may in fact lead to the data-fit blowing up as $\|V_k\|_{C^{-1}}$ does not converge sufficiently fast enough.

C

Expectation-Maximization

The original intent was to provide a supplementary procedure to λ, whilst further investigating the physical implications. Unfortunately, this would have required the explicit computation of the gradient with respect to the parameter. That is, $\frac{d\psi}{d\alpha}$, or alternatively $\frac{dS^w}{d\lambda}$ and $\frac{dp^o}{d\lambda}$, are not always computable in practice, unless the intention is to explicitly fit the ensemble to a model or approximation.

In what follows is an attempt at implementing an online Expectation-Maximization algorithm by building onto the work of [11], but instead using the EnKC as the E-step. Then in the M-step, we instead maximize the error statistics using the results of EnKC, from which we then construct an augmented data assimilation model to re-analyse the data and estimate the unknown parameter λ.

Recall, in equation (4.9) we showed that there exists a quadratic form that when minimized, minimizes the cross-covariance between the ensemble forecast and confirmation. Moreover, given that the model is assumed to be deterministic, we know that this uncertainty originates from the parameter uncertainty. Furthermore, (4.9) is in terms of the Forecast Error Covariance between the analyzed and confirmed ensembles, plus an additional data contribution term. Here we will assume that the Data Error Covariance encompasses this additional data contribution term:

$$\langle \psi^f, \psi^c \rangle_{(I-1_N)} = \underbrace{\langle \delta\psi, \psi^a \rangle_{(I-1_N)}}_{\sim R(\lambda)} + \frac{1}{2} \underbrace{\langle \delta\psi, \delta\psi \rangle_{(I-1_N)}}_{\sim Q(\lambda)} \tag{C.1}$$

As it is not possible to compute $\frac{\partial\psi}{\partial\lambda}$, it is not possible to minimize this directly. Therefore, in what follows we construct an expectation-maximization algorithm using the EnKC as the E-step and maximizing $\vartheta = \{Q, R\}$ in the M-step. As a result, we then construct an augmented data assimilation model, where the confirmed ensemble is perturbed by $q \sim \mathcal{N}(0, Q)$ and taken to be model predicted measurements, whilst the data is now assumed have measurement noise $\nu \sim \mathcal{N}(0, R)$. Finally, using this augmented data assimilation model we recompute the estimate for λ.

C.0.1. MAP estimates

The maximum a posteriori (MAP) estimate is obtained by determining the location of the maximum of the posterior distribution and using it as the point estimate:

$$\hat{\alpha}^{MAP} = \arg\max_{\alpha} \left[p(\alpha \mid d_{1:T}) \right]$$

The MAP estimate can be equivalently computed as the minimum of the error function defined in (B.2):

$$\hat{\alpha}^{MAP} = \arg\min_{\alpha} \left[\phi_T(\alpha) \right]$$

which is usually numerically more stable and easier to compute. It is also possible to use a Laplace approximation which uses the second derivative (Hessian) of the energy function to form a Gaussian approximation to the posterior distribution:

$$p(\alpha \mid d_{1:T}) \approxeq \mathcal{N}\left(\alpha \mid \hat{\alpha}^{MAP}, [H(\hat{\alpha}^{MAP})]^{-1}\right)$$

where $H(\hat{\alpha}^{MAP})$ is the Hessian matrix evaluated at the MAP estimate. However, to implement the Laplace approximation, we need to have a method to compute (or approximate) the second order derivatives of the energy function. This implies that the derivatives of the dynamic states S^w, p^o with respect to α, or via the chain rule λ, would need to be known. Unfortunately, $\frac{d\psi}{d\alpha}$, or alternatively $\frac{dS^w}{d\lambda}$ and $\frac{dp^o}{d\lambda}$, are not always computable. Hence, we will take several intermediary steps of using the EM algorithm to maximize Q, R, and then estimate λ using the augmented data assimilation model.

C.1. EM-algorithm

The expectation-maximization (EM) algorithm is a method to iteratively find an ML estimate of the parameters when the direct optimization of the posterior distribution (or equivalently, energy function) is not feasible. However, we show how it can also be easily modified for computation of MAP estimates.

The EM algorithm is based on the result that even when we cannot evaluate the marginal likelihood as such, we are still often able to compute a lower bound for it as follows.

Definition C.1.1 (Expectation Maximization algorithm)**.** *Given an initial guess α_0, the i-th iteration of the EM algorithm encompasses*

Expectation (E) - step: *Maximize the Free Energy as a function of q given a set of parameters α_{i-1}:*

$$q_{i-1} = \arg\max_q \mathscr{L}(q, \alpha_{i-1})$$

The optimal density is :

$$q_{i-1}(\psi_{0:j}) = p(\psi_{0:j} \mid d_{0:j}; \alpha_{i-1})$$

Maximization (M) - step: *Maximize the Free Energy w.r.t. α given a fixed q_{i-1}:*

$$\alpha_i = \arg\max_\alpha \mathscr{L}(q_{i-1}, \alpha)$$

As long as we find an α_i that increases the ELBO in the maximization step (this family of methods are called generalized EM algorithms) and a function q_{i-1} that increases the ELBO in the expectation step (called as incremental EM algorithms), the convergence is guaranteed.

C.1.1. Derivation of the Batch EM-Algorithm

In this subsection we follow Coccuci et al.'s derivation of the Batch EM-algorithm, see [11]. The aim of the traditional batch EM algorithm applied to a hidden Markov model is to iteratively find the statistical parameters ϑ that maximize the complete likelihood function $f(d_{1:K}; \vartheta)$, given a batch of observations distributed in a time interval, $d_{1:K} = \{d_1, \ldots, d_K\}$.

Using the Markov property of the state, and that observations only depend on the current state, the joint probability density in a time interval $0:K$ results in

$$\begin{aligned} f(\Psi_{0:K}, d_{1:K}; \vartheta) &= f(\Psi_0; \vartheta) \prod_{k=1}^{K} f(\Psi_k \mid \Psi_{k-1}; \vartheta) \prod_{k=1}^{K} f(d_k \mid \Psi_k; \vartheta) \\ &= f(\Psi_0; \vartheta) \prod_{k=1}^{K} f(\Psi_k, d_k | \Psi_{k-1}; \vartheta). \end{aligned} \tag{C.2}$$

Using the product form of the joint density, in (C.2), the resulting evidence lower bound, ELBO, function for hidden Markov models is

$$\begin{aligned} \mathscr{L}(f(\Psi_{0:K} \mid d_{1:K}; \vartheta_{i-1}), \vartheta) &= \sum_{k=1}^{K} \int f(\Psi_{k-1:k} \mid d_{1:K}; \vartheta_{i-1}) \log f(\Psi_k, d_k \mid \Psi_{k-1}; \vartheta) d\Psi_{k-1:k} + C \\ &\triangleq \sum_{k=1}^{K} \mathscr{E}\left[\log f(\Psi_k, d_k \mid \Psi_{k-1}) \mid d_{1:K}; \vartheta_{i-1}\right] + C \end{aligned} \tag{C.3}$$

where all the constant terms w.r.t. ϑ are included in C and dropped from $\mathscr{L}$ in what follows. Furthermore, note that

$$\begin{aligned} \mathscr{E}\left[g(\Psi_k, \Psi_{k-1}) \mid d_{1:K}; \vartheta\right] &= \int g(\Psi_{k-1}, \Psi_k) f(\Psi_{0:K} \mid d_{1:K}; \vartheta) d\Psi_{k-1:k} \\ &= \int g(\Psi_{k-1}, \Psi_k) f(\Psi_{k-1}, \Psi_k \mid d_{1:K}; \vartheta) d\Psi_{k-1:k}. \end{aligned}$$

In the general case, the parameters that maximize the ELBO, (C.3), in a hidden Markov model in the maximization step need to be determined numerically. However, if $f(\Psi_k, d_k \mid \Psi_{k-1})$ belongs to an exponential family (condition satisfied when both observational and model errors belong to the exponential family), it is possible to derive an analytical expression for the parameters that maximize the ELBO. The joint density of state and observation given the previous state required in (C.2) is in this case expressed as

$$f(\Psi_k, d_k \mid \Psi_{k-1}; \vartheta) = h(\Psi_k, d_k) \exp\left[\Phi(\vartheta) \cdot s(\Psi_{k-1}, \Psi_k, d_k) - A(\vartheta)\right] \tag{C.4}$$

where $s(\Psi_{k-1}, \Psi_k, d_k)$ is the natural sufficient statistic, $\Phi(\vartheta)$ is called the natural parameter and h and A are well-defined functions.
The gradient of the ELBO w.r.t. ϑ by introducing (C.4) into (C.3) is

$$\nabla_\vartheta \mathscr{L}[f(\Psi_{0:K} \mid y_{1:K}; \vartheta_{i-1}), \vartheta] = \nabla_\vartheta \Phi(\vartheta) \cdot \sum_{k=1}^{K} \mathscr{E}\left[s(\Psi_{k-1}, \Psi_k, d_k) \mid d_{1:k}; \vartheta_{i-1}\right] - K \nabla_\vartheta A(\vartheta) \tag{C.5}$$

Assuming the expression in (C.5) has one root, corresponding to the maximum of $\mathscr{L}$, the resulting E-step of the ith iteration becomes

$$S_{i-1} = \frac{1}{K} \sum_{k=1}^{K} \mathscr{E}\left[s(\Psi_{K-1}, \Psi_K, d_K)\right] \tag{C.6}$$

while the M-step is

$$\vartheta_i = \hat{\vartheta}(S_{i-1}) \tag{C.7}$$

where we define $\hat{\vartheta}(S_{i-1})$ as the solution for ϑ of the equation:

$$\nabla_\vartheta \Phi(\vartheta) \cdot S_{i-1} - \nabla_\vartheta A(\vartheta) = 0 \tag{C.8}$$

The critical value for ϑ which sets this gradient to zero, will indeed maximize the ELBO. This is due to properties of the Hessian of the likelihood of exponential families.

C.1.2. Derivation of the Online EM-Algorithm

Again, in this subsection we follow Coccuci et al.'s derivation of the Online EM-algorithm, see [11].

The major drawback of batch EM is that for each new observation, and a given value of the parameters ϑ, the ELBO has to be recomputed using the whole sequence of observations from 1 to K. Our goal is to update the parameter with each new observation. Thus, in what follows the EM iteration index i is dropped and will be replaced with the index corresponding to the assimilation cycle. With this notation, S_K is an analogue for the quantity in (C.6) but which is indexed by the last observation taken into account. We will approximate S_K (from(C.6) in terms of S_{K-1} and the information provided by observation d_K . If new observations are received sequentially, the value of K is in principle changing with time. Therefore, this leads to a recursive formula valid for every $k \in \{1, \ldots, L\}$, meaning we can write an approximation of S_k in terms of S_{k-1} and a term involving observation d_k with k ranging from 1 to K.

We begin by writing:

$$S_K = \frac{1}{K}\left(\sum_{k=1}^{K-1} \int s(\Psi_{k-1}, \Psi_k, d_k) f(\Psi_{k-1}, \Psi_k \mid d_{1:K}; \vartheta_{K-1}) d\Psi_{k-1:k} + \int s(\Psi_{k-1}, \Psi_k \mid d_{1:K}; \vartheta_{K-1}) d\Psi_{K-1:K}\right) \tag{C.9}$$

We can recognize that the first $K-1$ terms in (C.9) correspond to $(K-1)S_{K-1}$ but incorporating information of the newly available observation d_K, so that the posterior density corresponds to smoothing. We make the approximation that d_k does not significantly influence the previous state estimates but only provides information to the last term, which corresponds to the sufficient statistics at K. This results in:

$$\tilde{S}_K = (1-\gamma_K)\tilde{S}_{K-1} + \gamma_K \int s(\Psi_{k-1}, \Psi_K, d_K) f(\Psi_{k-1}, \Psi_K \mid d_{1:K}; \vartheta_{K-1}) d\Psi_{K-1:K} \tag{C.10}$$

$$= (1-\gamma_K)\tilde{S}_{K-1} + \gamma_K \mathscr{E}\left[s(\Psi_{K-1}, \Psi_K, d_K) \mid d_{1:K}; \vartheta_{K-1}\right], \tag{C.11}$$

where we introduced $\gamma_k \in (0,1)$ as a step-size (instead of $\frac{1}{K}$). In essence, (C.11) is a form of exponential moving average, where γ_k controls the memory of the statistics, determining the importance of the old statistics relative to the contribution of the current observation. The initialization parameter $\tilde{S}_0$ has to be provided as a first guess. In the Gaussian case, it coincides with the first guess for the parameter ϑ_0.

C.1.3. Explicit Implementation

Below we explicitly illustrate how the previous subsection is used to compute to estimate the Forecast Error Covariance Q and Data Error Covariance R, which in the previous section would be $\vartheta = \{Q, R\}$. Here we begin by assuming that $\tilde{Q}_0 = 0_{N_g \times N_g}$ and $\tilde{R}_0 = C_{\epsilon\epsilon}$. Then at an arbitrary kth data assimilation, with data d_k, we run the EnKC to estimate the analyzed dynamic state ψ^a, and confirmed dynamic state ψ^c. Then using (C.8) we compute the proposal $\mathscr{E}[Q]$, $\mathscr{E}[R]$ using the root for the gradient of the ELBO for the exponential family representation of a multivariate Gaussian density.

In the case of the covariance of the multivariate Gaussian, $S_{i-1} = \Sigma = (x-\mu)(x-\mu)^\top$:

$$\mathscr{E}[Q] = (\psi^a - \psi^c)(\psi^a - \psi^c)^\top$$
$$\mathscr{E}[R] = (d - \psi^c)(d - \psi^c)^\top$$

Then using (C.11) we compute

$$\tilde{Q}_K = (1-\gamma_K)\tilde{Q}_{K-1} + \gamma_K \mathscr{E}[Q]$$
$$\tilde{R}_K = (1-\gamma_K)\tilde{R}_{K-1} + \gamma_K \mathscr{E}[R]$$

We also assume $\gamma_k = 0.6$ and as far has not been changed.

C.2. Using the Confirmed Ensemble for Model Predicted Measurements

Now in order to for this to be useful we need to some use these Error Statistics to compute (re)estimate λ. As a result, we will define an augmented data assimilation model in which the confirmed ensemble will be perturbed and used a predicted measurements, and the data will be (re)measured using the data error ν instead of the measurement error ϵ. Recall that the original data assimilation model, (2.5), with Static and Dynamic states:

$$\alpha = \left[\ln\lambda\right], \quad \psi^i = \begin{bmatrix} \vec{S^w}^i \\ \vec{p^o}^i \end{bmatrix} (= \psi(t_i)) \tag{C.12a}$$

Nonlinear Physical Model:

$$\psi(x, t_i) = \mathscr{G}(\psi(x, t_{i-1}), \lambda) \tag{C.12b}$$

Measurement Model:

$$d_t = \mathscr{H}[\psi(t), \alpha] + \epsilon \qquad \epsilon \sim \mathscr{N}(0, C_{\epsilon\epsilon}) \tag{C.12c}$$

where the initial Static and Dynamic states are:

$$\lambda \sim Gamma(k,\theta), \quad \psi^0 = \begin{bmatrix} \vec{S^w}^0 = 0.29 \\ \vec{p^o}^0 = p^r = 10^5 \end{bmatrix} \tag{C.12d}$$

where $k = \frac{(\lambda^0_{guess})^2}{\sigma^2_{\lambda\lambda}}, \theta = \frac{\lambda^0_{guess}}{\sigma^2_{\lambda\lambda}}$. Then we introduced a Confirmation step which was stated to be a one-step smoothened forecast:

$$\psi^c(x, t_i) = \mathscr{G}(\psi(x, t_{i-1}), \lambda^a) \tag{C.12e}$$

C.2.1. The augmented Data assimilation model

We now can define the error prediction-correction model. In order to do this, we assume that the unknown parameter λ is forecasted, which in fact results in the analyzed parameter λ^a. Then the confirmation step is the deterministic forecast using this forecasted parameter, however the model predicted data has model uncertainty $q \sim \mathscr{N}(0, \tilde{Q})$ and the maximization measurement model has data error $\nu \sim \mathscr{N}(0, \tilde{R})$:

Parameter Forecast:

$$\lambda^a = \lambda + \lambda'$$

Nonlinear Model taking into account the incorrect forecast:

$$\begin{aligned} \mathscr{H}[\psi^*(x, t_i)] &= \mathscr{H}[\mathscr{G}(\psi(x, t_{i-1}), \lambda^a) + q] \\ &= \mathscr{H}[\psi^c(x, t_i) + q] \end{aligned} \tag{C.13}$$

Maximization Measurement Model taking into account that we corrected for affected by uncertain λ to get λ^a:

$$d_t = \hat{\mathscr{H}}[\psi(t), \alpha, \mathscr{H}[\psi^*(x,t)]] + \nu \tag{C.14}$$

wherein the Maximization Observation Operator $\hat{\mathscr{H}} : \mathbb{R}^{n+m} \to \mathbb{R}^m$ for each ensemble member. In this way we are able to provide the variance minimized estimate for $\ln\lambda$ (which was an issue see Section 2.4).

C.2.2. How is this actually implemented

And just like in [16, pg142], we can treat $\psi^*(x,t_i)$ as the prediction of the measurements given λ^a. Augmenting further the ensemble state to include these prediction of the measurements, now we apply the variance minimizing solution found in [16, pg.144].

We augment the system:

$$\Psi = \begin{bmatrix} \psi \\ \alpha \\ \mathscr{H}[\psi^*] = \mathscr{H}[\psi^c + q] \end{bmatrix}$$

We define the Maximization Measurement Observer:

$$\hat{\mathscr{H}}_i[\psi, \alpha, \mathscr{H}[\psi^*]] = \begin{bmatrix} p^{oc}(x_1, t_i) + q(p^o(x_1, t_i)) \\ p^{oc}(x_{N_g}, t_i) + q(p^o(x_{N_g}, t_i)) \end{bmatrix} \in \mathbb{R}^m \quad \text{where } m = 2$$

Then it follows from the general EnK framework:

$$\hat{S} := \mathscr{H}[\psi^*](I - 1_N)$$
$$d := d_{obs} + \nu$$
$$\hat{E} := [\nu_1, \ldots, \nu_N] \in \mathbb{R}^{m \times N}$$

As such:

$$\alpha^{cm} = \alpha^f \left(I + \hat{S}^\intercal \left(\hat{S}\hat{S}^\intercal + \hat{E}\hat{E}^\intercal \right)^{-1} \left(d - \psi^f \right) \right)$$

Hereafter, another confirmation step can optionally be run. When and why to do this is yet to be investigated rigorously.

C.3. The 2 Online EM methods: EnKCM and EnKCMC

In order to ensure that energy is conserved and the first order markov property still holds in the E-step analysis we re-confirm the augmented data assimilation mode. Thereby we create a distinction between the EnKCM and EnKCMC:

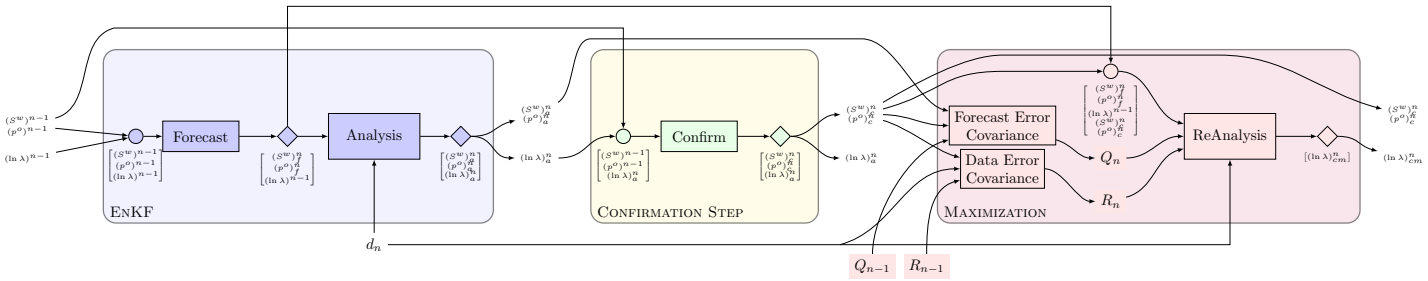


Figure C.1: Confirmed Ensemble Kalman Filter with One-Step Maximization (EnKCM) at nth Data Assimilation

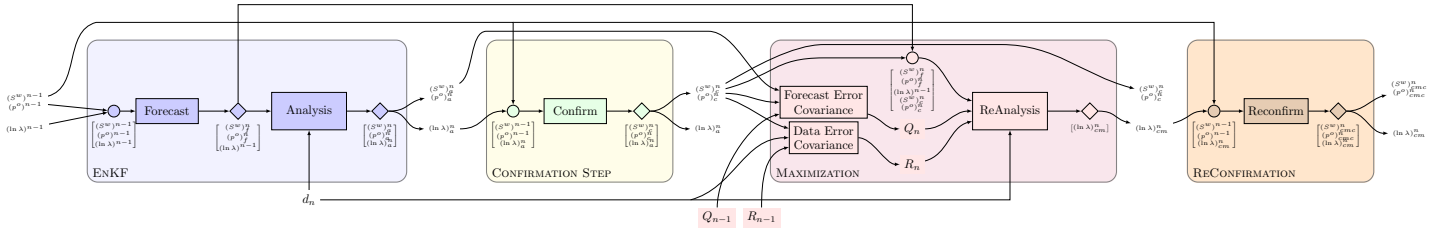


Figure C.2: ReConfirmed Ensemble Kalman Filter with One-Step Maximization (EnKCMC) at nth Data Assimilation

In the figures below we run under the base experiment settings, some preliminary tests that show promise.

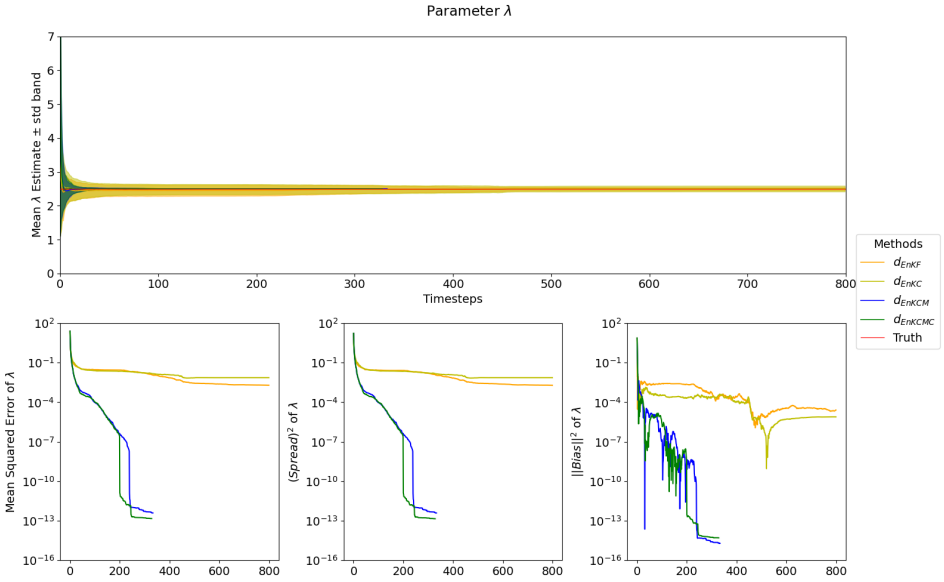


Figure C.3: Comparing the fit of λ estimate

Note that EnKFP and EnKCP refer to EnKF-Potential and EnKC-Potential, which simply means that Q, R are computed but not used. e.g. EnKF=EnKFP, only we also compute Q, R for demonstration purposes.

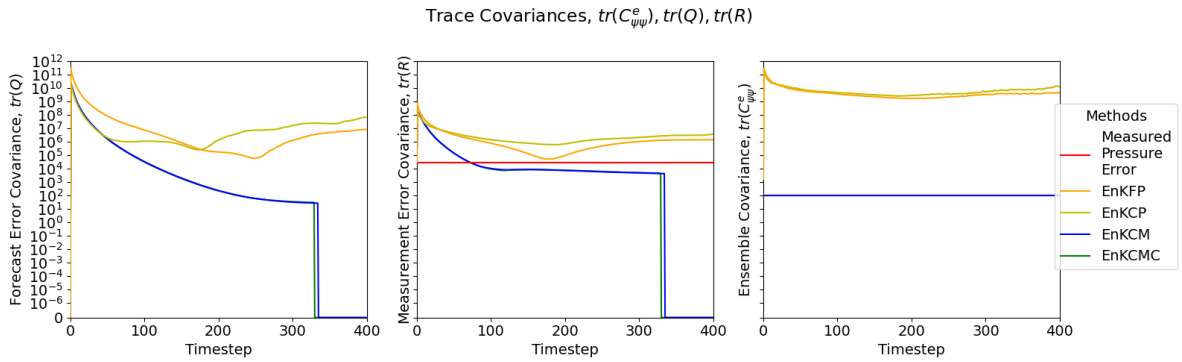


Figure C.4: Trace of Q, R, $C^e_{\psi\psi}$

Figure C.4 shows the trace of the following covariances:

Forecast Error, Q continues to decrease, i.e. forecasts look more and more like each other.

Data Error, R continues to decrease, i.e. forecast look more and more like the data.

Ensemble, $C^e_{\psi\psi}$ remains constant at a fixed level(just above 10^2), i.e. ensemble members actually do look like each other, but there is an initial uncertainty that cannot be removed.

Bibliography

[1] *Real-Time Reservoir Model Updating Using Ensemble Kalman Filter*, volume All Days of *SPE Reservoir Simulation Conference*, 01 2005. doi: 10.2118/92991-MS. URL https://doi.org/10.2118/92991-MS. SPE-92991-MS.

[2] Shell Energy Transition Report. Technical report, Royal Dutch Shell, apr 2018. URL https://www.shell.com/energy-and-innovation/the-energy-future/shell-energy-transition-report.html.

[3] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

[4] Frank E. Ashford. Computed relative permeability drainage and imbibition. In *Fall Meeting of the Society of Petroleum Engineers of AIME*. Society of Petroleum Engineers, 1969. doi: 10.2118/2582-ms. URL https://doi.org/10.2118%2F2582-ms.

[5] K. Aziz and A. Settari. *Petroleum Reservoir Simulation*. Springer Netherlands, 1979. ISBN 9780853347873.

[6] Richard O. Baker, Harvey W. Yarranton, and Jerry L. Jensen. 8 - special core analysis—rock–fluid interactions. In Richard O. Baker, Harvey W. Yarranton, and Jerry L. Jensen, editors, *Practical Reservoir Engineering and Characterization*, pages 239–295. Gulf Professional Publishing, Boston, 2015. ISBN 978-0-12-801811-8. doi: https://doi.org/10.1016/B978-0-12-801811-8.00008-0. URL https://www.sciencedirect.com/science/article/pii/B9780128018118000080.

[7] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, apr 2011. ISBN 9780387310732. URL https://www.xarg.org/ref/a/0387310738/.

[8] Royal Harvard Brooks. *Hydraulic properties of porous media*. Colorado State University, 1965.

[9] George Casella and Roger Berger. *Statistical Inference*. Duxbury Resource Center, June 2001. ISBN 0534243126.

[10] Zhangxin Chen, Guanren Huan, and Yuanle Ma. *Computational methods for multiphase flows in porous media*. Computational science & engineering. Society for Industrial and Applied Mathematics, Philadelphia, 2006. ISBN 9780898716061. OCLC: ocm62804700.

[11] Tadeo Cocucci, Manuel Pulido, Magdalena Lucini, and Pierre Tandeo. Model error covariance estimation in particle and ensemble kalman filters using an online expectation-maximization algorithm. *Quarterly Journal of the Royal Meteorological Society*, 147, 10 2020. doi: 10.1002/qj.3931.

[12] A.T. Corey. *Mechanics of Immiscible Fluids in Porous Media*. Water Resources Publications, 1994. ISBN 9780918334831.

[13] L. P. Dake. *Fundamentals of reservoir engineering*. Elsevier, Amsterdam Boston, 1978. ISBN 044441830X.

[14] Erle Donaldson. *Wettability*. Gulf Pub. Co, Houston, Tex, 2008. ISBN 9781933762296.

[15] T. Ertekin, J.H. Abou-Kassem, and G.R. King. *Basic Applied Reservoir Simulation*. SPE textbook series. Society of Petroleum Engineers, 2001. ISBN 9781555630898.

[16] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer Berlin Heidelberg, 2009. ISBN 9783642037108.

[17] Geir Evensen. The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, Nov 2003. ISSN 1616-7228. doi: 10.1007/s10236-003-0036-9. URL https://doi.org/10.1007/s10236-003-0036-9.

[18] Reinhard Furrer and Thomas Bengtsson. Estimation of high-dimensional prior and posterior covariance matrices in kalman filter variants. *Journal of Multivariate Analysis*, 98(2):227 – 255, 2007. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2006.08.003. URL http://www.sciencedirect.com/science/article/pii/S0047259X06001187.

[19] Tarek Ganat. *Pumping System of Heavy Oil Production*. 12 2019. ISBN 978-1-83968-409-8. doi: 10.5772/intechopen.87077.

[20] Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Stochastic Modelling and Applied Probability v. 53. Springer, 1 edition, 2003. ISBN 0387004513,9780387004518.

[21] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins studies in the mathematical sciences. The Johns Hopkins University Press, fourth edition edition, 2013. ISBN 9781421407944.

[22] J. Hadamard. Sur les problèmes aux dérivés partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902.

[23] P.L. Houtekamer and Herschel L. Mitchell. Ensemble kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3269–3289, oct 2005. doi: 10.1256/qj.05.135. URL https://doi.org/10.1256%2Fqj.05.135.

[24] C.M.. M. Istchenko and I.D.. D. Gates. Well/wormhole model of cold heavy-oil production with sand. *SPE Journal*, 19(02):260–269, June 2013. doi: 10.2118/150633-pa. URL https://doi.org/10.2118/150633-pa.

[25] Tijana Janji?, Dennis McLaughlin, Stephen E. Cohn, and Martin Verlaan. Conservation of mass and preservation of positivity with ensemble-type kalman filter algorithms. *Monthly Weather Review*, 142 (2):755 – 773, 2014. doi: 10.1175/MWR-D-13-00056.1. URL https://journals.ametsoc.org/view/journals/mwre/142/2/mwr-d-13-00056.1.xml.

[26] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Dover Books on Electrical Engineering Series. Dover Publications, 2007. ISBN 9780486462745.

[27] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. Adaptive computation and machine learning. MIT Press, 2009. ISBN 9780262013192. URL https://books.google.nl/books?id=7dzpHCHzNQ4C.

[28] R.J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2002. ISBN 9781139434188.

[29] M.C. Leverett. Capillary Behavior in Porous Solids. *Transactions of the AIME*, 142(01):152–169, 12 1941. ISSN 0081-1696. doi: 10.2118/941152-G. URL https://doi.org/10.2118/941152-G.

[30] J.M. Lewis, S. Lakshmivarahan, S. Dhall, and Cambridge University Press. *Dynamic Data Assimilation: A Least Squares Approach*. Number v. 13 in Dynamic Data Assimilation: A Least Squares Approach. Cambridge University Press, 2006. ISBN 9780521851558.

[31] Daniel Loeve, F. Wilschut, R.H. Hanea, Jos Maas, P.M.E. Hooff, Paul Van den Hoek, S.G. Douma, and J.F.M. Doren. Simultaneous determination of relative permeability and capillary pressure curves by assisted history matching several scal experiments. 09 2011.

[32] Kevin P. Murphy. *Machine learning : a probabilistic perspective*. MIT Press, Cambridge, Mass. [u.a.], 2013. ISBN 9780262018029 0262018020. URL https://www.amazon.com/Machine-Learning-Probabilistic-Perspective-Computation/dp/0262018020/ref=sr_1_2?ie=UTF8&qid=1336857747&sr=8-2.

[33] NVIDIA. Cuda, release: 11.5, 2021. URL https://developer.nvidia.com/cuda-toolkit.

[34] Society of Petroleum. *The SI Metric System of Units and SPE Metric Standard, Second Edition*. 1982. ISBN 9995538814,9789995538811.

[35] Alfio Quarteroni, Riccardo Sacco, and Fausto Saleri. *Numerical Mathematics (Texts in Applied Mathematics)*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 3540346589. doi: 10.5555/1212166.

[36] S.S.M.W.H. Robert L. Devaney, M.W. Hirsch, S. Smale, and R.L. Devaney. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Pure and Applied Mathematics - Academic Press. Elsevier Science, 2004. ISBN 9780123497031. URL https://books.google.nl/books?id=INYJuKGmgd0C.

[37] S. Särkkä. *Bayesian Filtering and Smoothing*. Bayesian Filtering and Smoothing. Cambridge University Press, 2013. ISBN 9781107030657. URL https://books.google.de/books?id=5VlsAAAAQBAJ.

[38] T. J. Sullivan. *Introduction to uncertainty quantification*. Springer, Cham, 2015. ISBN 978-3-319-23394-9.

[39] Frank van der Meulen. *Lecture Notes Statistical Inference*. Course Code WI4455. TU Delft, Delft, The Netherlands, 2018.

[40] J. J. I. M. van Kan, A. Segal, and Fredericus Johannes Vermolen. *Numerical Methods in Scientific Computing*. Delft Academic Press, Delft, 2014. ISBN 9789065623638.

[41] Peter Jan van Leeuwen. Comment on "data assimilation using an ensemble kalman filter technique". *Monthly Weather Review*, 127(6):1374 – 1377, 01 Jun. 1999. doi: 10.1175/1520-0493(1999)127<1374:CODAUA>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/mwre/127/6/1520-0493_1999_127_1374_codaua_2.0.co_2.xml.