NTNU
Innovation and Creativity

Hierarchical Ensemble Kalman Filter

for Observations of Production and 4-D Seismic Data

Jon Sætrom

Master of Science in Physics and Mathematics
Submission date: July 2007
Supervisor: Karl Henning Omre, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

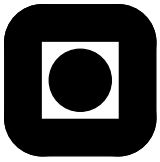
Problem Description

Purpose of the work: To define a linear Gaussian model connecting reservoir properties to 3-D seismic AVO data and incorporate an hierarchical model in the Ensemble Kalman Filter.

Assignment given: 07. February 2007
Supervisor: Karl Henning Omre, MATH

NTNU
Norwegian University of Science and Technology

Faculty of Information Technology, Mathematics and Electrical Engineering



MASTER'S THESIS

FOR

STUD.TECHN. Jon Sætrom

FACULTY OF INFORMATION TECHNOLOGY, MATHEMATICS AND ELECTRICAL ENGINEERING

NTNU

Date due: July 5, 2007

Discipline: Statistics

Title: "Hierarchical Ensemble Kalman Filter for observations of production and 4-D seismic data."

Purpose of the work: To define a linear Gaussian model connecting reservoir properties to 3-D seismic AVO data and incorporate an hierarchical model in the Ensemble Kalman Filter.
Steps:
- *Define a linear model, and an algorithm that connects seismic data to reservoir properties.*
- *Define a model and implement an algorithm for Hierarchical Ensemble Kalman Filtering based on seismic data and production history.*
- *Evaluate the model on a synthetic case study.*
- *Discuss the results.*

This Master's Thesis is to be carried out at the Department of Mathematical Sciences under guidance of Professor Henning Omre.

Trondheim, February 5, 2007.

Trond Digernes Instituttleder Dept. of Mathematical Sciences	Henning Omre Professor Dept. of Mathematical Sciences

Preface

This Master's Thesis is written as the final part of the Master of Technology degree within the Industrial Mathematics Program at the Norwegian University of Science and Technology. The work is to be conduced over a period of 20 weeks.

The idea of this Thesis has been to combine the Ensemble Kalman Filter with Bayesian Kriging, to develop a hierarchical Bayesian method, suitable for sequential reservoir history matching based on observed production and 4-D seismic. The focus in this Thesis has been both the theoretical model formulation, and implementation of the methology on a synthetic test case inspired by the Troll field offshore of Norway.

I would like to thank my supervisor Henning Omre, for his helpful advice and contributions to this Thesis. Further, I would like to thank Jo Eidsvik at NTNU, for providing me with the computer code used as a basis in the seismic forward modelling, Ole Petter Lødøen at Statoil for providing me with the Eclipse data files, and the reference reservoir used in his PhD. Thesis and Per Kristian Hove at NTNU for assisting me with some UNIX and MATLAB related issues. I would also like to thank the Department of Petroleum engineering and applied Geophysics at NTNU, for granting me access to their reservoir fluid flow simulator Eclipse. Finally I would like to thank Hege Bjørnsen for all her help, love and support.

Jon Sætrom Trondheim, July 5th, 2007

Abstract

Finding unknown reservoir properties based on observations of well-logs, production and 4-D seismic, known in the petroleum literature as reservoir characterisation and production history matching, is a very complicated task. In recent years, technological advancements has led to an increase in the amount of available data regarding a reservoir under production. Thus, development of methods that solves the history matching problem, both automatically and sequentially as new data becomes available, has been given a lot of attention.

The Ensemble Kalman Filter is a Bayesian method for performing reservoir history matching which has shown promising results. However, this method requires multiple realisations of the reservoir, obtained through fluid flow simulation. When considering very large reservoir models represented on a fine scale, repeated fluid flow simulation will be computationally too expensive. Therefore, the reservoir has to be represented on a coarser scale before fluid flow simulation can be performed. This approximation is, however, known to introduce bias in the history matched reservoir, which should be accounted for in the model formulation. Moreover, inclusion of 4-D seismic data, requires a fine scale representation of the reservoir.

In this Thesis, we present a Hierarchical Bayesian solution to the challenges outlined above, requiring only fluid flow simulations on a coarse scale. Further, we outline a linearised Bayesian seismic AVO inversion technique, that fits into the Ensemble Kalman Filter methology. The complete method, which we have called the Hierarchical Scale-Corrected Ensemble Kalman Filter, is demonstrated on a case study inspired by the Troll field in the North Sea, where we have included both observations of production, and 4-D synthetic seismic data.

Contents

1 Introduction

Evaluation and prediction of the properties in a petroleum reservoir, such as permeability, hydrocarbon saturation and porosity, based on production history, well-logs and seismic data, is a very important task when the goal is to extract as much hydrocarbons as possible. This task, however, requires solutions to a complex, ill-posed and non-linear inverse problem. Traditional methods for performing this task has been based on a deterministic approach.

The statistical approach to solving this problem is Bayesian inversion, which in recent years has gained a strong popularity when solving geophysical inverse problems (Scales & Snieder 1997). In Bayesian inversion *a priori* knowledge of the model, before any observations have been made, is reflected through a probability distribution. By further assigning a probabilistic term, the likelihood, connecting observations to the model, *a posteriori* knowledge of the model can be obtained through a probability distribution.

Technological advancements has in recent years made permanent monitoring of a reservoir possible through production data and 4-D seismic. The use of a traditional history matching method will in such a case be too time consuming, since the whole process has to be repeated as new observations become available. Thus, methods that are able to assimilate the model to observed data, both continuously and sequentially, has in recent years been given much attention.

The Kalman Filter (Kalman 1960) is a Bayesian method that can solve space-time inversion problems sequentially. The key assumption in this method, however, is that the model connecting previous observations to future observations is a linear Gaussian function. Such an assumption will not be valid for the fluid flow of a petroleum reservoir, which is a highly non-linear, non-Gaussian process.

The Ensemble Kalman Filter (EnKF), introduced by Evensen (1994) and later modified by Evensen et al. (1998), is a method inspired by the Kalman Filter that has shown promising results within fields such as weather forecasting, oceanography, ground water hydrology and petroleum engineering (Evensen 2003). When the function connecting the future observations to past observations is non-linear, the model will be analytically intractable in a stochastic setting. The idea of the EnKF is therefore to use a set of realisations called an ensemble, such that relevant statistics such as the mean and the covariance can be estimated. As new observations are available, the EnKF also provides a sequential method for updating the ensemble members.

Evaluation of the fluid flow in a reservoir involves solving a system of partial differential equations, refered to as fluid flow simulation. This is known to be a computational demanding process when the dimension of the problem increases. In very large reservoir evaluation problems, containing up to 10^9 grid blocks, repeated fluid flow simulation may take weeks or months to complete. Approximate models, that reduces the dimension of the problem is therefore required. Usually, this involves representing the very fine scale reservoir on a coarser scale such that repeated fluid flow simulation can be performed. This approximation is, however, known to introduce bias, which should be accounted for (Omre & Lødøen 2004).

Lødøen & Omre (2005) presented an extension to the EnKF, the Scale-Corrected EnKF, where the bias of coarse scale fluid flow simulation was corrected for. This approach, however, required a calibration step involving fluid flow simulations on both fine and coarse scale. Moreover, their approach required inversion of empirical covariance matrices, which in most practical settings would be rank deficient.

In this Thesis, we have combined the idea of Lødøen & Omre (2005) with Bayesian Kriging (Lee & Zidek 1992), into a hierarchical method we have called the Hierarchical Scale-Corrected EnKF (HScCEnKF). The main advantages of this approach is that we only require fluid flow simulations on a coarse scale, while being able to predict the outcome on a fine scale, preserving the rank of all covariance matrices. This feature is important when we include observations of 4-D seismic, as these usually appear on a much finer grid than the coarse scale grid used for fluid flow simulation.

The inclusion of 4-D seismic in the EnKF methology requires that there exists a linear Gaussian relationship between the reservoir and seismic observations. Combining the work of Buland & Omre (2003), connecting seismic amplitude versus offset (AVO) data to rock physics, and the work of Bachrach (2006) and Batzel & Wang (1992), connecting rock physics to the reservoir properties, we have developed a linear Gaussian model connecting seismic AVO data to the reservoir properties.

The Thesis proceeds as follows: In Section 2, we introduce some of the notation used throughout this Thesis, followed be a short review to the physical equations governing fluid flow in a reservoir in its simplest form in Section 3. The linear Bayesian seismic AVO inversion method is presented in Section 4, and the concept of Bayesian reservoir history matching is introduced in Section 5. Section 6 presents the model assumptions used in the EnKF together with an Algorithm in a setting including both observations of production history and 4-D seismic. Approximate fluid flow simulation is presented in Section 7, followed by the model description and an Algorithm for the HScCEnKF in Section 8. The case study used in this Thesis is presented in Section 9, and Section 10 presents and discusses the results. Finally we draw some conclusion, and outline further work in Section 11.

2 Notation

Throughout this Master's Thesis all vectors, meaning scalar vectors, stochastic vectors and realisations, will be denoted by boldface small letters. That is $\boldsymbol{a} \in \mathbb{R}^{n\times 1}$ means that $\boldsymbol{a}$ is a column vector with n elements and $\boldsymbol{a}^T \in \mathbb{R}^{1\times n}$ its transpose. Capital boldface letters will similarly denote both stochastic matrices and realisations. The notation $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ means that $\boldsymbol{A}$ is a matrix with n rows and m columns, where each entry, a_{ij}, is a real number. Further, note that for two numbers t_1 and t_2, $\Delta t_1 = t_2 - t_1$, and the notation $\bar{x}$ for $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$, will represent the mean value of the vector given as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Functions $f(\bullet)$ will denote probability density functions (pdfs). Given the pdf $f(\boldsymbol{x})$, then a realisation from this distribution will be denoted by $\boldsymbol{x}^i \sim f(\boldsymbol{x})$, where the symbol $\sim$ implies that the realisation follows this distribution. Conditional probability density function of a multivariate random variable, $\boldsymbol{z}$, given another multivariate random variable, $\boldsymbol{y}$ will be denoted by $f(\boldsymbol{z}|\boldsymbol{y})$.

Using Bayes' rule (Casella & Berger 2002), we can express this in terms of their joint distribution as follows

$$\begin{aligned} f(\boldsymbol{z}|\boldsymbol{y}) &= \frac{f(\boldsymbol{z},\boldsymbol{y})}{f(\boldsymbol{y})} \\ &\propto f(\boldsymbol{z},\boldsymbol{y}). \end{aligned} \tag{1}$$

That is, we know the structure of the pdf up to a possibly unknown normalisation constant. The multivariate expectation and covariance of a stochastical vector $\boldsymbol{x}$, as given in Definition 1 and 2, Appendix A, will be denoted by $\boldsymbol{\mu_x}$ and $\boldsymbol{\Sigma_x}$ respectively. Further, will the multivariate covariance between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, as given in Definition 3, Appendix A, be denoted by $\boldsymbol{\Sigma_{xy}}$.

Consider a reservoir domain $\mathcal{D} \subset \mathbb{R}^3$, which is discretised into a lattice $\mathcal{L}_\mathcal{D} = \mathcal{L}^{\boldsymbol{z}} \times \mathcal{L}^{\boldsymbol{xy}}$ containing n grid nodes as shown in Figure 1. Here $\mathcal{L}^{\boldsymbol{z}} \subset \mathbb{R}^1$ corresponds to the n_z grid nodes in the vertical direction, and $\mathcal{L}^{\boldsymbol{xy}} \subset \mathbb{R}^2$ corresponds to the $n_{xy} = n_x \cdot n_y$ grid nodes in the horizontal plane, giving a total of $n = n_x \cdot n_y \cdot n_z$ lattice nodes.

Since each lattice node in the discretised reservoir $\mathcal{L}_\mathcal{D}$ contains certain reservoir properties, meaning porosity ϕ, log permeability κ, saturation s, and pressure p, we let the state of the reservoir at a discrete timestep $t \in \{0,\ldots,T\}$, be described by the vector

$$\boldsymbol{r}_t = \left[\boldsymbol{\kappa}^T, \boldsymbol{\phi}^T, \boldsymbol{s}_t^T, \boldsymbol{p}_t^T\right]^T \in \mathbb{R}^{n_r\times 1},$$

where n_r denotes the number of lattice nodes times the number of reservoir properties considered. These properties are considered to be unknown throughout the reservoir, with log-permeability and saturation being static variables, and saturation and pressure being spatio-temporal variables.

It should be noted that the variable $\boldsymbol{s}_t$ includes water, gas and oil saturation. Since the sum of water, oil and gas saturation is equal to one, it is therefore

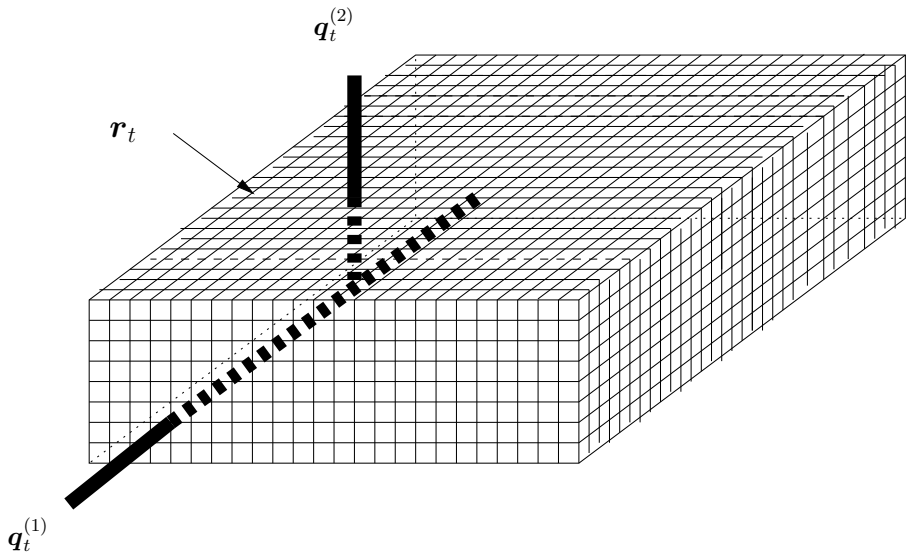


Figure 1: The Figure shows an oil reservoir discretised into a three-dimensional lattice. Here $\boldsymbol{r}_t$ denotes the vector containing the reservoir properties at each lattice node at timestep $t \in \{0, \ldots, T\}$, and $\boldsymbol{q}_t^{(k)},\ k \in \{1, 2\}$, denote the vectors containing the production properties for the production and injection well respectively.

sufficient to only consider the gas and water saturation as the unknown variables for saturation. In this Thesis we will further assume that the water saturation is a known constant, thus the notation $\boldsymbol{s}_t$ will refer to the gas saturation in the reservoir at timestep t.

The reservoir production properties at timestep t, denoted by $\boldsymbol{q}_t \in \mathbb{R}^{n_q \times 1}$, is the second uncertain variable. Here n_q is the number of injection and productions wells, times the number of production properties considered. Combining $\boldsymbol{r}_t$ and $\boldsymbol{q}_t$, we get a vector $\boldsymbol{x}_t \in \mathbb{R}^{(n_r+n_q)\times 1}$, containing the state of the reservoir at given timestep t. For notational convenience let

$$\boldsymbol{x}_t = \begin{bmatrix} \boldsymbol{r}_t \\ \boldsymbol{q}_t \end{bmatrix},$$

refer to the reservoir state at timestep t.

Observed values of the reservoir at timestep t will be denoted by

$$\boldsymbol{x}_t^o = \begin{bmatrix} \boldsymbol{r}_t^o \\ \boldsymbol{q}_t^o \end{bmatrix} = \begin{bmatrix} \boldsymbol{D}_{\boldsymbol{r}t}\boldsymbol{r}_t + \boldsymbol{\epsilon}_{\boldsymbol{r}t} \\ \boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{q}_t + \boldsymbol{\epsilon}_{\boldsymbol{q}_t} \end{bmatrix} = \boldsymbol{D}_{\boldsymbol{x}t}\boldsymbol{x}_t + \boldsymbol{\epsilon}_{\boldsymbol{x}t}, \tag{2}$$

where $\boldsymbol{D}_{\boldsymbol{r}t}$ and $\boldsymbol{D}_{\boldsymbol{q}_t}$ are shift matrices. Here the two vectors $\boldsymbol{\epsilon}_{\boldsymbol{r}t}$ and $\boldsymbol{\epsilon}_{\boldsymbol{q}_t}$ are included to model the uncertainties in the observations, and they are assumed to follow some probability distribution. A model such as the one above will in this Thesis be refered to as the likelihood model, or forward model. Note that $\boldsymbol{r}_t^o$ can include both observations through well-logs and seismic surveys. For notational convenience we will let $\boldsymbol{w}_t^o$, $\forall t \in T_{w^o}$, denote the complete set of observed well-logs, while $\boldsymbol{d}_t^o$, $\forall t \in T_{d^o}$, will refer to the complete set of seismic data. Here T_{w^o} and T_{d^o} are the sets of timesteps where observations of well-log and seismic data are available respectively.

3 Fluid Flow Simulation

In this section a simple version of the system of equations used in fluid flow simulation is presented. Further, we define the term fluid flow simulator.

Reservoir fluid flow simulation is generally the process of applying numerical methods to solve a system of partial differential equations that describe flow in a porous medium. We now give a brief introduction to this system in its simplest form. More thorough descriptions, with more complex models can be found in e.g. Azis & Settari (1979), Ewing (1983), Keoderitz (2005) or Peaceman (1977).

If we consider the flow of a single fluid in an isotropic medium, that is, properties of the fluid is independent of the direction, then Darcy's law (Darcy 1856) is given as

$$\boldsymbol{u} = -\frac{\boldsymbol{K}}{\mu}\left(\nabla p - \frac{\rho}{g}\nabla z\right), \tag{3}$$

where $\boldsymbol{u} = u\boldsymbol{i} + v\boldsymbol{j} + w\boldsymbol{k}$, with $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ being unity vectors in x, y and z direction, and u, v, w are flux components. Moreover, $\boldsymbol{K}$, p, g and μ are the permeability tensor, pressure, gravity constant and viscosity respectively. According to Azis & Settari (1979), the permeability tensor can in most practical problems be reduced to a diagonal tensor, and if we assume an isotropic medium, then $K = K_x = K_y = K_z$. Combining this equation with an equation stating that the mass should be preserved, namely (Durlofsky 2003)

$$\frac{\partial}{\partial t}(\phi\rho) + \nabla \cdot (\rho \boldsymbol{u}) + \xi = 0, \tag{4}$$

we get

$$\frac{\partial}{\partial t}(\phi\rho) - \nabla \cdot \left(\frac{\rho}{\mu} K \left(\nabla p - \frac{\rho}{g}\nabla z\right)\right) + \xi = 0, \tag{5}$$

where ρ is the density, ϕ is the porosity and ξ is a source/sink term representing a mass flow rate per unit volume (injection/production). This last equation is referred to as the pressure equation in the literature. If we ignore the gravity and assume that the porosity and density are independent of time, Expression 5 is reduced to

$$\nabla \cdot \left(\frac{1}{\mu} K \cdot \nabla p\right) = \frac{\xi}{\rho}. \tag{6}$$

To be able to solve this system, according to Peaceman (1977), we also need a set of boundary conditions, and functions connecting porosity ϕ and density ρ to the pressure:

$$\phi = \phi(p), \qquad \rho = \rho(p).$$

When these conditions are set, then the system can be solved using a numerical solver meaning a finite difference or finite element method (Iserles 1996, White 1986). Regardless of the complexity of the underlying system, we will from now on denote the solver of this system of equations as an unknown, possibly highly complex function $\omega : \mathbb{R}^{n_r+n_q} \to \mathbb{R}^{n_r+n_q}$. This will further be referred to as the fluid flow simulator.

Assume that we are given the reservoir timestep $t-1$, $\boldsymbol{x}_{t-1}$. Then the state of the reservoir at a later timestep t, can be written as

$$\begin{bmatrix} \boldsymbol{r}_t \\ \boldsymbol{q}_t \end{bmatrix} = \omega_{\Delta_{t-1}} \left(\begin{bmatrix} \boldsymbol{r}_{t-1} \\ \boldsymbol{q}_{t-1} \end{bmatrix} \right) + \boldsymbol{\epsilon}_{\Delta_{t-1}}, \tag{7}$$

where the subscript in the fluid flow simulator expresses the difference between two consecutive timesteps, and $\boldsymbol{\epsilon}_{\Delta_{t-1}} \in \mathbb{R}^{(n_r+n_q)\times 1}$ is a vector representing numerical-, and model simplification errors.

4 Seismic Inversion

Figure 2 shows the outline of a marine seismic survey. An airgun array behind the seismic vessel is fired, generating waves that propagate in the water until they reach a subsurface. Here the properties of the medium change, and portions of the waves are reflected to the surface where hydrophones on the seismic streamers behind the vessel registers the amplitudes of the reflected waves. The Result is a set of reflected wave amplitudes $\boldsymbol{d}^o_{s_k,j,\,\theta_l}$, at different locations $j \in \mathcal{L}^{\boldsymbol{xy}}$, as a function of discretised seismic reflection travel time $s_k,\ k \in \{1,\dots,S\}$ and reflection angles $\theta_l,\ l \in \{1,\dots,n_\theta\}$.

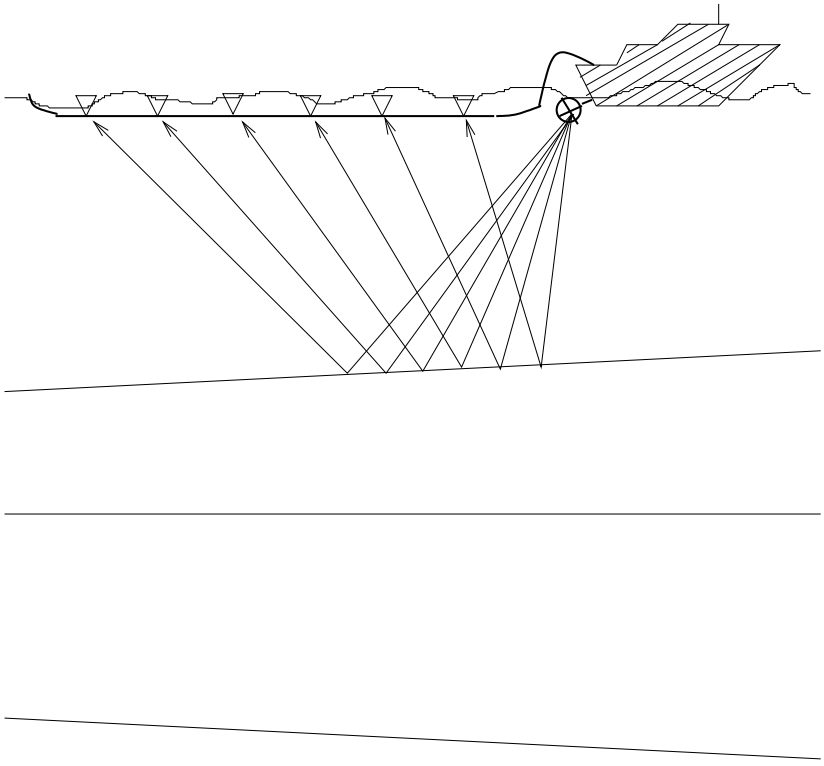


Figure 2: Outline of marine seismic acquisition.

The seismic trace, meaning the set of reflected wave amplitudes along a vertical profile $\boldsymbol{z}_j \in \mathcal{L}^{\boldsymbol{z}}$, for different reflection angles $\theta \in \{\theta_1,\dots,\theta_{n_\theta}\}$, at lattice node j will further be represented by the vector

$$\boldsymbol{d}^o_j = \left[d^o_{j,s_1,\,\theta_1},\dots,d^o_{j,s_S,\,\theta_1},\dots,d^o_{j,s_1,\,\theta_{n_\theta}},\dots,d^o_{j,s_S,\,\theta_{n_\theta}}\right]^T \in \mathbb{R}^{n_\theta \cdot S \times 1}. \tag{8}$$

In this section we present a forward model connecting the reservoir $\boldsymbol{r}_t$ to the observed seismic data $\boldsymbol{d}^o_t$ at certain timesteps $t \in T_{d^o}$. Further, we present a Bayesian solution to the inverse problem of finding the reservoir properties, given pre-stack amplitude versus offset (AVO) seismic data.

Please note that in this Thesis we are not concerned with the complicated task of preprocessing the seismic data. We therefore refer the reader to e.g.

Avseth et al. (2005) and Sheriff & Geldart (1995) for further reading on this topic. Also, in order to get an analytically tractable model such that the seismic inversion can be performed fast, a simplified linear model, where we invert using seismic traces separately, is developed.

4.1 Seismic Forward Model

Consider a position $\boldsymbol{y} \in \mathbb{R}^{3\times 1}$ in an isotropic, elastic reservoir domain $\mathcal{D} \subset \mathbb{R}^3$. Then, according to Sheriff & Geldart (1995), the material properties at $\boldsymbol{y}$ can be completely described by the pressure wave (P-wave) velocity $\alpha(\boldsymbol{y})$, shear wave (S-wave) velocity $\beta(\boldsymbol{y})$, and density $\rho(\boldsymbol{y})$.

Given

$$[\alpha(\boldsymbol{y}), \beta(\boldsymbol{y}), \rho(\boldsymbol{y})], \quad \forall \boldsymbol{y} \in \mathcal{D},$$

a synthetic seismogram can then be found by solving the wave equation describing the wave propagation (Sheriff & Geldart 1995). In most non-trivial cases, this must be done by applying a numerical partial differential equation solver such as the finite difference or finite element methods (Iserles 1996, White 1986).

4.1.1 Seismic Likelihood

Instead of solving the complete wave equation, another approach is instead to use ray tracing methods (Cerveny et al. 1977). Central in these methods is Snell's law which for the transition between layer l and $l+1$ is given as

$$\frac{\sin(\theta_{P,l})}{\alpha_l} = \frac{\sin(\theta_{P,\,l+1})}{\alpha_{l+1}} = \frac{\sin(\theta_{S,l})}{\beta_l} = \frac{\sin(\theta_{S,\,l+1})}{\beta_{l+1}} = \text{const.}, \tag{9}$$

and the Zoeppritz' equations (Zoeppritz 1919). Snell's law simply states that when a P-wave travelling in a solid layer l with a non-normal angle of incidence $\theta_{P,l}$ and velocity α_l, hits the boundary of a solid layer $l+1$, the P-wave is converted into four reflected and transmitted P- and S-waves. This process is shown in Figure 3.

An expression for the distribution of the energy in the four converted waves as a nonlinear function of the material properties and the angle of incidence, $\theta_{P,l}$, is given through the reflection and transmission coefficients in the Zoeppritz' equation. For small angles of incidence Aki & Richards (1980) showed that the reflection coefficient for a reflected P-wave can be approximated by

$$c_{PP}(\theta_P) = a_\alpha(\theta_P, \bar{\alpha}, \bar{\beta})\frac{\Delta\alpha}{\bar{\alpha}\Delta z} + a_\beta(\theta_P, \bar{\alpha}, \bar{\beta})\frac{\Delta\beta}{\bar{\beta}\Delta z} + a_\rho(\theta_P, \bar{\alpha}, \bar{\beta})\frac{\Delta\rho}{\bar{\rho}\Delta z}, \tag{10}$$

where

$$\begin{aligned}
a_\alpha(\theta_P, \bar{\alpha}, \bar{\beta}) &= \frac{1}{2}(1 + \tan^2\theta_P) \\
a_\beta(\theta_P, \bar{\alpha}, \bar{\beta}) &= -4\frac{\bar{\beta}^2}{\bar{\alpha}^2}\sin^2\theta_P \\
a_\rho(\theta_P, \bar{\alpha}, \bar{\beta}) &= \frac{1}{2}(1 - 4\frac{\bar{\beta}^2}{\bar{\alpha}^2}\sin^2\theta_P),
\end{aligned}$$

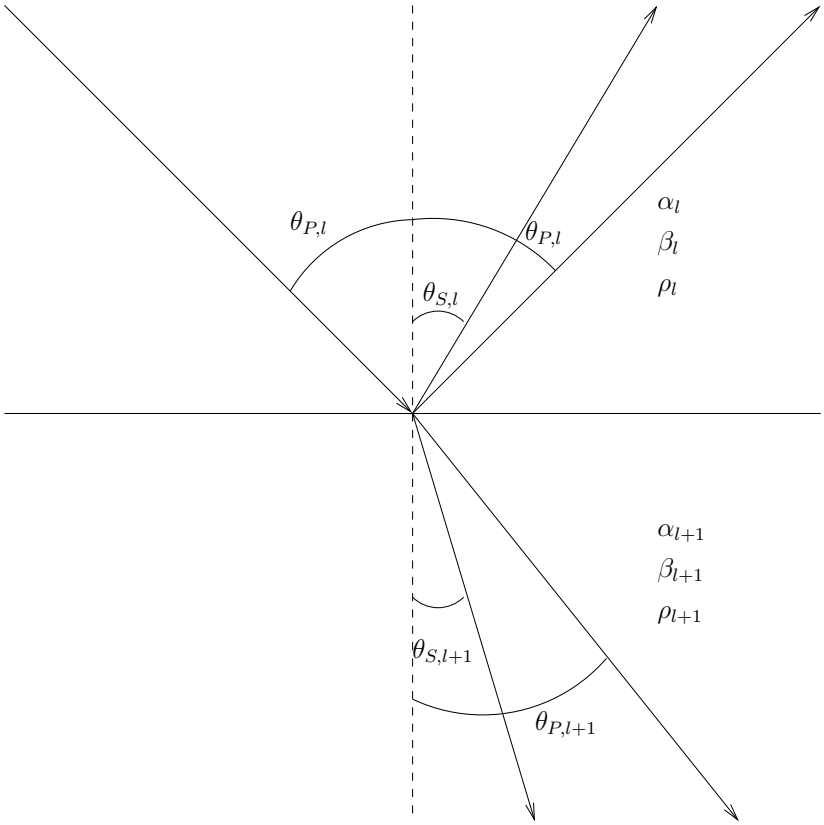


Figure 3: Converted waves generated as an incident P-wave passes the interface between the solid layers l and $l+1$.

and Δz is the thickness of the geological layer.

According to (Stolt & Weglein 1985), Expression (10) can be given as a continuous function of seismic reflection time by letting $\Delta s \to 0$ such that

$$\begin{aligned}
\frac{\Delta\alpha}{\bar{\alpha}\Delta s} &\to \frac{1}{\alpha(s)} \lim_{\Delta s \to 0} \frac{\alpha(s+\Delta s) - \alpha(s)}{\Delta s} = \frac{\mathrm{d}}{\mathrm{d}s} \ln \alpha(s), \\
\frac{\Delta\beta}{\bar{\beta}\Delta s} &\to \frac{\mathrm{d}}{\mathrm{d}s} \ln \beta(s), \\
\frac{\Delta\rho}{\bar{\rho}\Delta s} &\to \frac{\mathrm{d}}{\mathrm{d}s} \ln \rho(s).
\end{aligned}$$

As in Buland & Omre (2003), we can generalise the coefficients in Expression (10), to time dependent coefficients by letting the average velocities also be dependent of time. If we further assume that for each seismic trace

$$\frac{\bar{\beta}(s)}{\bar{\alpha}(s)} = \frac{\mu_\beta}{\mu_\alpha},$$

we can write

$$c_{PP}(\theta_{P,\,k}, s) = \boldsymbol{a}(\theta_{P,\,k})^T \boldsymbol{m}'(s), \tag{11}$$

where

$$\boldsymbol{a}(\theta_{P,\,k}) = \begin{bmatrix} \frac{1}{2}(1 + \tan^2 \theta_{P,\,k}) \\ -4\frac{\mu_\beta^2}{\mu_\alpha^2} \sin^2 \theta_{P,\,k} \\ \frac{1}{2}(1 - 4\frac{\mu_\beta^2}{\mu_\alpha^2} \sin^2 \theta_{P,\,k}) \end{bmatrix},$$

and

$$\boldsymbol{m}'(s) = \left[\frac{\mathrm{d}}{\mathrm{d}s} \ln \alpha(s), \frac{\mathrm{d}}{\mathrm{d}s} \ln \beta(s), \frac{\mathrm{d}}{\mathrm{d}s} \ln \rho(s) \right]^T.$$

From now on we will drop the subscript P in the angles of incidence, and write θ_k. For each discretised seismic trace $\boldsymbol{d}_j$, $j \in \{1, \ldots, n_{xy}\}$, we then get

$$\boldsymbol{c}_{PP,j} = \boldsymbol{A}_j \boldsymbol{m}'_j, \tag{12}$$

where

$$\begin{aligned}
\boldsymbol{c}_{PP,j} &= [c_{PP,j}(\theta_1, s_1), \ldots, c_{PP,j}(\theta_1, s_S), \ldots, c_{PP,j}(\theta_{n_\theta}, s_1), \ldots, c_{PP,j}(\theta_{n_\theta}, s_S)]^T \in \mathbb{R}^{n_\theta \cdot S \times 1} \\
\boldsymbol{m}'_j &= \left[\boldsymbol{m}'_j(s_1)^T, \ldots, \boldsymbol{m}'_j(s_S)^T \right]^T \in \mathbb{R}^{3S \times 1}
\end{aligned}$$

and

$$\boldsymbol{A}_j = \left[\boldsymbol{A}_{j,1}^T, \ldots, \boldsymbol{A}_{j,n_\theta}^T \right]^T, \tag{13}$$

with $\boldsymbol{A}_{j,k} \in \mathbb{R}^{S \times 3 \cdot S}$ being a block matrix, with $\boldsymbol{a}_j(\theta_k)^T$ on each block diagonal element and zero otherwise.

According to Dobrin & Savit (1988), the seismic observations are connected to the reflection coefficients $\boldsymbol{c}_{PP}$ through the convolution model

$$\boldsymbol{d}(\theta, s) = \int \omega_\theta(s - u) c_{PP}(\theta, u) \mathrm{d}u + \epsilon(\theta, s), \tag{14}$$

where $\omega_\theta(s-u)$, is a seismic wavelet dependent on the angle θ and $\epsilon(\theta, s)$ is an error term. Similarly as in Buland & Omre (2003), Expression (14) can further be discretised such that

$$\boldsymbol{d}_j = \boldsymbol{W}\boldsymbol{c}_{PP,j} + \boldsymbol{\epsilon}_\omega, \tag{15}$$

where

$$\boldsymbol{\epsilon}_\omega = [\epsilon(\theta_1, s_1), \ldots, \epsilon(\theta_1, s_S), \ldots, \epsilon(\theta_{n_\theta}, s_1), \ldots, \epsilon(\theta_{n_\theta}, s_S)]^T \in \mathbb{R}^{n_\theta \cdot S \times 1},$$

and $\boldsymbol{W} \in \mathbb{R}^{n_\theta \cdot S \times n_\theta \cdot S}$ is a block-diagonal matrix with diagonal elements

$$\boldsymbol{W}_k = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{0} & \ldots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{w}_w & \ddots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \ldots & \boldsymbol{0} & \boldsymbol{w}_S \end{bmatrix} \in \mathbb{R}^{S \times S}, \quad k \in \{1, \ldots, n_\theta\}.$$

Here

$$\boldsymbol{w}_m = \begin{cases} \left[\omega_{-(m-1)}, \ldots, \omega_0, \ldots, \omega_{n_\omega}\right], & m < n_\omega + 1 \\ \left[\omega_{-(n_\omega)}, \ldots, \omega_0, \ldots, \omega_{n_\omega}\right], & n_\omega + 1 \leq m \leq n_S - n_\omega \\ \left[\omega_{-(n_\omega)}, \ldots, \omega_0, \ldots, \omega_{n_\omega - (n_S - m)}\right], & m > n_S - n_\omega \end{cases}$$

represents the discretised wavelet along a seismic trace, with length $\in \{n_\omega + 1, \ldots, 2n_\omega + 1\}$.

Since differentiation is a linear operation, we can write

$$\boldsymbol{m}'_j = \boldsymbol{D}\boldsymbol{m}_j, \tag{16}$$

where

$$\boldsymbol{D} = \frac{1}{2\Delta s} \begin{bmatrix} 0 & -\sqrt{2} & \sqrt{2} & 0 & \ldots & 0 \\ -1 & 0 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & -1 & 0 & 1 \\ 0 & 0 & \ldots & 0 & -\sqrt{2} & \sqrt{2} \end{bmatrix} \in \mathbb{R}^{3S \times 3S}.$$

Combining Expressions (12), (15) and (16), yields

$$\boldsymbol{d}_j = \boldsymbol{G}_j\boldsymbol{m}_j + \boldsymbol{\epsilon}_\omega, \tag{17}$$

with $\boldsymbol{G}_j = \boldsymbol{W}\boldsymbol{A}_j\boldsymbol{D} \in \mathbb{R}^{n_\theta S \times 3S}$.

4.1.2 Rock Physics Likelihood

Using a similar approach as in Bachrach (2006), we can link the elastic parameters α, β and ρ to the reservoir properties ϕ, s and p through a non-linear function, $g : \mathbb{R}^3 \to \mathbb{R}^3$, by using the Biot-Gassmann theory (Gassmann 1951, Biot 1962), and the empirical results of Batzel & Wang (1992).

Gassmann's equation is given on the following form:

$$\frac{K_{\text{sat}}}{K_0 - K_{\text{sat}}} = \frac{K_{\text{dry}}}{K_0 - K_{\text{dry}}} + \frac{K_{\text{fl}}}{\phi(K_0 - K_{\text{fl}})}, \quad G_{\text{sat}} = G_{\text{dry}}, \tag{18}$$

where

$$\begin{aligned} K_{\text{dry}} &= \text{effective bulk modulus of dry rock,} \\ K_{\text{sat}} &= \text{effective bulk modulus of the rock with pore fluid} \\ K_0 &= \text{matrix bulk modulus of the material} \\ K_{\text{fl}} &= \text{effective bulk modulus of the pore fluid} \\ G_{\text{dry}} &= \text{effective shear modulus of dry rock} \\ G_{\text{sat}} &= \text{effective shear modulus of rock with pore fluid} \end{aligned}$$

As shown in Mavko et al. (1998), Expression (18) can be equally written as

$$\begin{gathered} \frac{K_{\text{sat1}}}{K_0 - K_{\text{sat1}}} - \frac{K_{\text{fl1}}}{\phi(K_0 - K_{\text{fl1}})} = \frac{K_{\text{sat2}}}{K_0 - K_{\text{sat2}}} - \frac{K_{\text{fl2}}}{\phi(K_0 - K_{\text{fl2}})}, \\ G_{\text{sat1}} = G_{\text{sat}} \end{gathered} \tag{19}$$

in a situation with two different fluids. For partially saturated rocks we can, according to Mavko et al. (1998), model the effective fluid modulus as a function of saturation as follows:

$$K_{\text{fl1}}(s) = \left(\frac{s_w}{K_{\text{b}}} + \frac{s}{K_{\text{gas}}} + \frac{1 - s_w - s}{K_{\text{oil}}} \right)^{-1}, \tag{20}$$

where s_w, K_{b}, K_{gas} and K_{oil} are the water saturation, brine, gas and oil bulk modulus respectively. We further assume that

$$\begin{aligned} K_{\text{sat2}} &= K_{\text{sat2}}(\phi, \boldsymbol{\theta}_{K_\phi}) + \epsilon_{K_\phi} \\ G_{\text{sat}} &= G_{\text{sat}}(\phi, \boldsymbol{\theta}_{G_\phi}) + \epsilon_{G_\phi}, \end{aligned} \tag{21}$$

where $K_{\text{sat}}(\phi, \boldsymbol{\theta}_{K_\phi})$ and $G_{\text{sat}}(\phi, \boldsymbol{\theta}_{G_\phi})$ are functions found through e.g. regression based on well-logs, with $\theta_{(\bullet)}$ being the corresponding model parameters such as regression coefficients in the case where the model is found through regression, and $\epsilon_{(\bullet)}$ describes the model errors.

The bulk modulus K of a fluid or gas is according to Mavko et al. (1998) given as

$$K = \rho \cdot v^2 \tag{22}$$

where ρ is the density and v is the acoustic velocity in the corresponding fluid. The results of Batzel & Wang (1992), can therefore be used to connect the bulk modulus of brine, oil and gas to pressure, through non-linear functions

$$\begin{aligned} K_{\text{brine}} &= K_{\text{brine}}(p, \boldsymbol{\theta}_{K_{\text{brine}}}) \\ K_{\text{oil}} &= K_{\text{oil}}(p, \boldsymbol{\theta}_{K_{\text{oil}}}) \\ K_{\text{gas}} &= K_{\text{gas}}(p, \boldsymbol{\theta}_{K_{\text{gas}}}) \end{aligned} \tag{23}$$

Here $\boldsymbol{\theta}_{K_{(\bullet)}}$ are model parameters, such as temperature, brine salinity, gas gravity and API gravity (American Petroleum Institute oil gravity), which are assumed to be known throughout the reservoir. A summary of the these functions can be found in Mavko et al. (1998).

Combining Expressions (20), (21) and (23), and assuming that $K_{\text{fl2}} = K_{\text{brine}}$ in Expression (19), we can express K_{sat1} as a function of porosity, saturation and pressure, namely

$$K_{\text{sat1}} = K_{\text{sat1}}(\phi, s, p, \boldsymbol{\theta}_{K_{\text{sat1}}}). \tag{24}$$

$\boldsymbol{\theta}_{K_{\text{sat1}}}$ is here a set of physical constants, such as the ones described above, or regression coefficients, assumed to be known throughout the reservoir through well-logs and core samples.

Finally, by using the Expressions in Bachrach (2006), we can establish a connection between the elastic parameters ρ, α and β and ϕ, s and p through the non-linear functions

$$\begin{aligned}\alpha &= g_\alpha(\phi, s, p, \boldsymbol{\theta}_\alpha),\\ \beta &= g_\beta(\phi, s, p, \boldsymbol{\theta}_\beta),\\ \rho &= g_\rho(\phi, s, p, \boldsymbol{\theta}_\rho).\end{aligned}$$

Again, the notation $\boldsymbol{\theta}_{\alpha,\beta,\rho}$ corresponds to known model parameters such as regression coefficients, or physical properties of the reservoir.

For notational convenience, we from now on express the relation between the elastic properties of the reservoir, and the reservoir properties for a seismic trace through the non-linear function $g: \mathbb{R}^{3n_z} \to \mathbb{R}^{3S}$, such that

$$\boldsymbol{m}_j = g(\boldsymbol{r}_{-\boldsymbol{\kappa}_j}), \tag{25}$$

where $\boldsymbol{r}_{-\boldsymbol{\kappa}_j} = \left[\boldsymbol{\phi}_j^T, \boldsymbol{s}_j^T, \boldsymbol{p}_j^T\right]^T \in \mathbb{R}^{3n_z \times 1}$.

4.2 Bayesian Linearised Seismic Amplitude vs. Offset Inversion

We will now outline a linear Bayesian inversion method of finding the reservoir properties along a seismic trace. Note that in this section we have omitted the subscript indicating which seismic trace we are examining.

Consider the Graph shown in Figure 4, where single arrows mean that there is a stochastic relation between the variables and thick arrows imply a deterministic relationship.

Using the seismic forward model in the previous section, we further assume that

$$\begin{aligned}\boldsymbol{d}^o|\boldsymbol{d} &= \boldsymbol{D}_{\boldsymbol{d}^o}\boldsymbol{d} + \boldsymbol{\epsilon}_{\boldsymbol{d}^o},\\ \boldsymbol{d}|\boldsymbol{m} &= \boldsymbol{G}\boldsymbol{m} + \boldsymbol{\epsilon}_\omega,\\ \boldsymbol{m} &= g(\boldsymbol{r}_{-\boldsymbol{\kappa}}),\end{aligned} \tag{26}$$

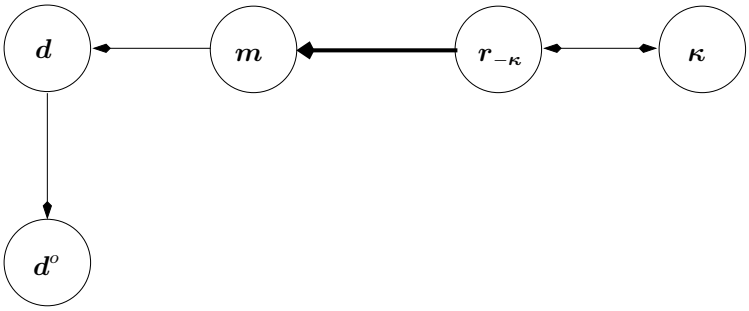


Figure 4: The figure shows the graph of the stochastic model considered in the seismic inversion. Thin arrows imply a stochastic relationship, while thick arrows imply a deterministic relationship.

where $\boldsymbol{D}_{\boldsymbol{d}^o} \in \mathbb{R}^{n_\theta \cdot S_{d^o} \times n_\theta \cdot S}$ is an indicator matrix that specifies which locations seismic observations are available, with $S_{d^o} \leq S$, and

$$\begin{aligned} \boldsymbol{\epsilon}_{\boldsymbol{d}^o} &\sim \text{Gauss}_{n_\theta \cdot S_{d^o}}(\boldsymbol{0}, \sigma_{d^o}^2 \boldsymbol{I}) \\ \boldsymbol{\epsilon}_\omega &\sim \text{Gauss}_{n_\theta \cdot S_{d^o}}(\boldsymbol{0}, \sigma_\omega^2 \boldsymbol{I}). \end{aligned}$$

We also assign the following prior distribution to $\boldsymbol{r}$:

$$\boldsymbol{r} \sim \text{Gauss}_{4 \cdot n_z}(\boldsymbol{\mu}_{\boldsymbol{r}}, \boldsymbol{\Sigma}_{\boldsymbol{r}}), \tag{27}$$

where

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{r}} &= [\boldsymbol{\mu}_{\boldsymbol{\kappa}}{}^T, \boldsymbol{\mu}_{\boldsymbol{\phi}}{}^T, \boldsymbol{\mu}_{\boldsymbol{s}}{}^T, \boldsymbol{\mu}_{\boldsymbol{p}}{}^T]^T \in \mathbb{R}^{4n_z \times 1} \\ \boldsymbol{\Sigma}_{\boldsymbol{r}} &= \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\kappa}} & \boldsymbol{\Sigma}_{\boldsymbol{\kappa\phi}} & \boldsymbol{\Sigma}_{\boldsymbol{\kappa s}} & \boldsymbol{\Sigma}_{\boldsymbol{\kappa p}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\phi\kappa}} & \boldsymbol{\Sigma}_{\boldsymbol{\phi}} & \boldsymbol{\Sigma}_{\boldsymbol{\phi s}} & \boldsymbol{\Sigma}_{\boldsymbol{\phi p}} \\ \boldsymbol{\Sigma}_{\boldsymbol{s\kappa}} & \boldsymbol{\Sigma}_{\boldsymbol{s\phi}} & \boldsymbol{\Sigma}_{\boldsymbol{s}} & \boldsymbol{\Sigma}_{\boldsymbol{sp}} \\ \boldsymbol{\Sigma}_{\boldsymbol{p\kappa}} & \boldsymbol{\Sigma}_{\boldsymbol{p\phi}} & \boldsymbol{\Sigma}_{\boldsymbol{ps}} & \boldsymbol{\Sigma}_{\boldsymbol{p}} \end{bmatrix} \in \mathbb{R}^{4n_z \times 4n_z}. \end{aligned}$$

Here the notation $\boldsymbol{x} \sim \text{Gauss}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that $\boldsymbol{x}$ is a p-dimensional vector having the multivariate Gaussian distribution, as described in Appendix A.1, with expectation $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma}$.

Using the standard linear Gaussian theory in Appendix A.2, we see that

$$\boldsymbol{r}_{-\boldsymbol{\kappa}} \sim \text{Gauss}_{3 \cdot n_z}(\boldsymbol{\mu}_{\boldsymbol{r}_{-\boldsymbol{\kappa}}}, \boldsymbol{\Sigma}_{\boldsymbol{r}_{-\boldsymbol{\kappa}}}), \tag{28}$$

where

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{r}_{-\boldsymbol{\kappa}}} &= \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{\phi}} \\ \boldsymbol{\mu}_{\boldsymbol{s}} \\ \boldsymbol{\mu}_{\boldsymbol{p}} \end{bmatrix} \\ \boldsymbol{\Sigma}_{\boldsymbol{r}_{-\boldsymbol{\kappa}}} &= \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\phi}} & \boldsymbol{\Sigma}_{\boldsymbol{\phi s}} & \boldsymbol{\Sigma}_{\boldsymbol{\phi p}} \\ \boldsymbol{\Sigma}_{\boldsymbol{s\phi}} & \boldsymbol{\Sigma}_{\boldsymbol{s}} & \boldsymbol{\Sigma}_{\boldsymbol{sp}} \\ \boldsymbol{\Sigma}_{\boldsymbol{p\phi}} & \boldsymbol{\Sigma}_{\boldsymbol{ps}} & \boldsymbol{\Sigma}_{\boldsymbol{p}} \end{bmatrix}. \end{aligned}$$

As we can see from the stochastic model defined above, the only non-linear term is the function $g(\bullet)$, connecting the elastic parameters $\boldsymbol{m}$, to $\boldsymbol{r}_{-\boldsymbol{\kappa}}$. If we assume that the function g has second order continuous partial derivatives in a neighbourhood of the expected value $\boldsymbol{\mu}_{\boldsymbol{r}_{-\boldsymbol{\kappa}}}$, and the diagonal elements of the covariance matrix are sufficiently small, then we can linearise the function g by Taylor expanding it around the expected value (Høyland 1988). That is,

$$\boldsymbol{m} = g(\boldsymbol{r}_{-\boldsymbol{\kappa}}) \approx g(\boldsymbol{\mu}_{\boldsymbol{r}_{-\boldsymbol{\kappa}}}) + \boldsymbol{J}_g(\boldsymbol{r}_{-\boldsymbol{\kappa}} - \boldsymbol{\mu}_{\boldsymbol{r}_{-\boldsymbol{\kappa}}}), \tag{29}$$

where $\boldsymbol{J}_g \in \mathbb{R}^{3\cdot S \times 3\cdot n_z}$ is the Jacobi matrix of g given as

$$\boldsymbol{J}_g = \left[\begin{array}{ccc} \frac{\partial g_1(\boldsymbol{\phi},\boldsymbol{s},\boldsymbol{p})}{\partial \boldsymbol{\phi}} & \frac{\partial g_1(\boldsymbol{\phi},\boldsymbol{s},\boldsymbol{p})}{\partial \boldsymbol{s}} & \frac{\partial g_1(\boldsymbol{\phi},\boldsymbol{s},\boldsymbol{p})}{\partial \boldsymbol{p}} \\ \frac{\partial g_2(\boldsymbol{\phi},\boldsymbol{s},\boldsymbol{p})}{\partial \boldsymbol{\phi}} & \frac{\partial g_2(\boldsymbol{\phi},\boldsymbol{s},\boldsymbol{p})}{\partial \boldsymbol{s}} & \frac{\partial g_2(\boldsymbol{\phi},\boldsymbol{s},\boldsymbol{p})}{\partial \boldsymbol{p}} \\ \frac{\partial g_3(\boldsymbol{\phi},\boldsymbol{s},\boldsymbol{p})}{\partial \boldsymbol{\phi}} & \frac{\partial g_3(\boldsymbol{\phi},\boldsymbol{s},\boldsymbol{p})}{\partial \boldsymbol{s}} & \frac{\partial g_3(\boldsymbol{\phi},\boldsymbol{s},\boldsymbol{p})}{\partial \boldsymbol{p}} \end{array} \right]_{\boldsymbol{r}_{-\boldsymbol{\kappa}}=\boldsymbol{\mu}_{\boldsymbol{r}_{-\boldsymbol{\kappa}}}}.$$

This again implies that

$$\boldsymbol{m} \sim \text{Gauss}_{3S}(\boldsymbol{\mu}_{\boldsymbol{m}}, \boldsymbol{\Sigma}_{\boldsymbol{m}}), \tag{30}$$

approximately, with

$$\boldsymbol{\mu}_{\boldsymbol{m}} = g(\boldsymbol{\mu}_{\boldsymbol{r}_{-\boldsymbol{\kappa}}})$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{m}} = \boldsymbol{J}_g \boldsymbol{\Sigma}_{\boldsymbol{r}_{-\boldsymbol{\kappa}}} \boldsymbol{J}_g^T.$$

The computation of the matrix $\boldsymbol{A}$ in Expression (13), requires that $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$, and $\boldsymbol{\mu}_{\beta}$ are known. By standard Gaussian theory,

$$\boldsymbol{m}_{-\boldsymbol{\rho}} = \left[\begin{array}{c} \ln \boldsymbol{\alpha} \\ \ln \boldsymbol{\beta} \end{array} \right] \sim \text{Gauss}_{2S} \left(\left[\begin{array}{c} \boldsymbol{\mu}_{\boldsymbol{m}1} \\ \boldsymbol{\mu}_{\boldsymbol{m}2} \end{array} \right], \left[\begin{array}{cc} \boldsymbol{\Sigma}_{\boldsymbol{m}1} & \boldsymbol{\Sigma}_{\boldsymbol{m}12} \\ \boldsymbol{\Sigma}_{\boldsymbol{m}21} & \boldsymbol{\Sigma}_{\boldsymbol{m}2} \end{array} \right] \right).$$

Thus, by Definition 8 and Result 6 in Appendix A.1, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have the log-Gaussian distribution. That is:

$$\begin{aligned} \boldsymbol{\alpha} &\sim \Lambda_S(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}), \\ \boldsymbol{\beta} &\sim \Lambda_S(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}), \end{aligned} \tag{31}$$

with

$$\begin{aligned} \boldsymbol{\mu}_{\alpha} &= \exp\left\{ \boldsymbol{\mu}_{\boldsymbol{m}1} + \frac{\text{diag}(\boldsymbol{\Sigma}_{\boldsymbol{m}1})}{2} \right\}, \\ \boldsymbol{\mu}_{\beta} &= \exp\left\{ \boldsymbol{\mu}_{\boldsymbol{m}2} + \frac{\text{diag}(\boldsymbol{\Sigma}_{\boldsymbol{m}2})}{2} \right\}, \\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}(i,j)} &= \exp\left\{ \boldsymbol{\mu}_{\boldsymbol{m}1,i} + \boldsymbol{\mu}_{\boldsymbol{m}1,j} + \frac{1}{2}(\boldsymbol{\Sigma}_{\boldsymbol{m}1,(i,i)} + \boldsymbol{\Sigma}_{\boldsymbol{m}1,(j,j)}) \right\} \left[\exp\{\boldsymbol{\Sigma}_{\boldsymbol{m}1,(i,j)}\} - 1 \right], \\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}(i,j)} &= \exp\left\{ \boldsymbol{\mu}_{\boldsymbol{m}2,i} + \boldsymbol{\mu}_{\boldsymbol{m}1,j} + \frac{1}{2}(\boldsymbol{\Sigma}_{\boldsymbol{m}2,(i,i)} + \boldsymbol{\Sigma}_{\boldsymbol{m}2,(j,j)}) \right\} \left[\exp\{\boldsymbol{\Sigma}_{\boldsymbol{m}2,(i,j)}\} - 1 \right], \\ & \qquad i,j \in \{1,\ldots,S\}. \end{aligned}$$

Using the expressions above, we see that the likelihood model connecting the reservoir to the seismic observations can be written as:

$$\boldsymbol{d}^o = \boldsymbol{D_r}\boldsymbol{r} + \boldsymbol{\epsilon_d}, \tag{32}$$

where

$$\boldsymbol{D}_r = \begin{bmatrix} \mathbf{0} & \boldsymbol{D_{d^o}}\boldsymbol{G}\boldsymbol{J}_g \end{bmatrix} \in \mathbb{R}^{n_\theta \cdot S_{d^o} \times 4 \cdot n_z}, \tag{33}$$

$$\boldsymbol{\epsilon_d} = \boldsymbol{D_{d^o}}\left(\boldsymbol{\mu_{\epsilon_{rd^o}}} + \boldsymbol{\epsilon_\omega}\right) + \boldsymbol{\epsilon_{d^o}},$$

and

$$\boldsymbol{\mu_{\epsilon_{rd^o}}} = \boldsymbol{G}\left(g(\boldsymbol{\mu_{r_{-\kappa}}}) - \boldsymbol{J}_g\boldsymbol{\mu_{r_{-\kappa}}}\right).$$

Finally, since we have assumed a linear Gaussian relationship in all likelihood models in the graph shown in Figure 4, and have assigned a Gaussian prior distribution, the results in Ripley (1996) and Appendix A.2, gives

$$\begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{d}^o \end{bmatrix} \sim \text{Gauss}_{4n_z+n_\theta S}\left(\begin{bmatrix} \boldsymbol{\mu_r} \\ \boldsymbol{\mu_{d^o}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma_r} & \boldsymbol{\Sigma_{rd^o}} \\ \boldsymbol{\Sigma_{rd^o}}^T & \boldsymbol{\Sigma_{d^o}} \end{bmatrix}\right),$$

where

$$\boldsymbol{\mu_{d^o}} = \boldsymbol{D_{d^o}}\boldsymbol{G}\boldsymbol{\mu_m}, \tag{34}$$

$$\boldsymbol{\Sigma_{rd^o}} = \boldsymbol{\Sigma_r}\boldsymbol{D_r}^T,$$

$$\boldsymbol{\Sigma_{d^o}} = \boldsymbol{D_r}\boldsymbol{\Sigma_r}\boldsymbol{D_r}^T + \sigma_\omega^2\boldsymbol{D_{d^o}}\boldsymbol{D}_{\boldsymbol{d}^o}^T + \sigma_{d^o}^2\boldsymbol{I}. \tag{35}$$

Thus, by Result 2 in Appendix A.2

$$\boldsymbol{r}|\boldsymbol{d}^o \sim \text{Gauss}_{4n_z}(\boldsymbol{\mu_{r|d^o}}, \boldsymbol{\Sigma_{r|d^o}}), \tag{36}$$

where

$$\begin{aligned} \boldsymbol{\mu_{r|d^o}} &= \boldsymbol{\mu_r} + \boldsymbol{\Sigma_r}\boldsymbol{D_r}^T\boldsymbol{\Sigma}_{\boldsymbol{d}^o}^{-1}(\boldsymbol{d}^o - \boldsymbol{D_{d^o}}\boldsymbol{G}\boldsymbol{\mu_m}) \\ \boldsymbol{\Sigma_{r|d^o}} &= \boldsymbol{\Sigma_r} - \boldsymbol{\Sigma_r}\boldsymbol{D_r}^T\boldsymbol{\Sigma}_{\boldsymbol{d}^o}^{-1}\boldsymbol{D_r}\boldsymbol{\Sigma_r}. \end{aligned} \tag{37}$$

We complete this section by outlining the procedure for seismic inversion in Algorithm 1:

Algorithm 1: Linearised Bayesian Seismic AVO Inversion

for $j = 1$ *to* n_{xy} **do**

Generate $\boldsymbol{\epsilon_{r}}_j \sim \text{Gauss}_{4n_z}(\mathbf{0}, \boldsymbol{\Sigma_{r|d^o}}_{,j})$

Evaluate $\boldsymbol{r}_j = \boldsymbol{\mu_r}_j + \boldsymbol{\Sigma}_{r,j}\boldsymbol{D_r}_j^T\boldsymbol{\Sigma}_{\boldsymbol{d}^o,j}^{-1}(\boldsymbol{d}_j^o - \boldsymbol{D_{d^o}}_{,j}\boldsymbol{G}_j\boldsymbol{\mu_m}_j) + \boldsymbol{\epsilon_r}_j$

5 Reservoir History Matching

In this section we present the problem of assimilating unknown reservoir properties to observations made through reservoir production and seismic surveys, known as reservoir history matching. We will further outline a statistical approach to this problem, and finally a sequential Bayesian algorithm for reservoir history matching is given.

5.1 History Matching

Consider the stochastic model described in Figure 5. The goal here is to find

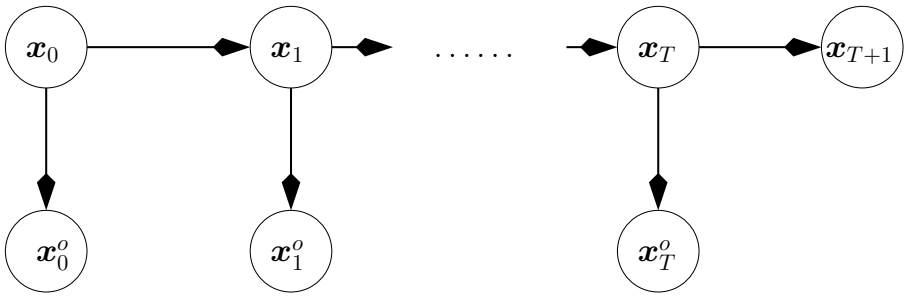


Figure 5: The figure shows the Directed Acyclic Graph (DAG) of the model considered. Single arrows mean that there is a stochastic relationship between the variables.

the unknown properties of the reservoir $\boldsymbol{x}_t$, given observations of well-logs $\boldsymbol{w}_t^o$, acquired at discrete timesteps $t \in T_{w^o}$, observed production data from the reservoir $\boldsymbol{q}_t^o$, acquired at discrete timesteps $t \in \{1, 2, \ldots, T\}$, and 4-D seismic, $\boldsymbol{d}_t^o$, acquired at discrete timesteps $t \in T_{d^o}$. We also want to predict the future production properties based on previous data, $\boldsymbol{x}_{T+1}$. This problem can then be described as a non-linear, ill-posed inverse problem (Omre & Tjelmeland 1996).

Various deterministic methods was earlier used to solve this problem (Keoderitz 2005, Ewing & Lin 1991, Bissel et al. 1992, e.g). These methods however make it difficult to quantify any uncertainty linked to the forecast of the reservoir (Lødøen & Omre 2005). In this report we therefore focus on the Bayesian solution to this inverse problem, where the uncertainty in the reservoir forecast is represented by the posterior distribution.

As in Section 2 we assume that

$$\boldsymbol{x}_t^o = \boldsymbol{D}_{\boldsymbol{x}t}\boldsymbol{x}_t + \boldsymbol{\epsilon}_{\boldsymbol{x}t},$$

where $\boldsymbol{\epsilon}_{\boldsymbol{x}t}$ follow some type of probability distribution. It is therefore natural to define a stochastic model for the state of the reservoir.

The model in Figure 5 describes a Directed Acyclic Graph (DAG). By Ripley (1996), this implies that the state of the reservoir appears as a Markov Process in time. That is the forward model can equally be written as

$$\begin{aligned} f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \ldots, \boldsymbol{x}_0) &= f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) \\ &= \omega_{\Delta_{t-1}}(\boldsymbol{x}_{t-1}) + \boldsymbol{\epsilon}_{\Delta_{t-1}}, \quad \forall t \geq 1. \end{aligned}$$

By sequential decomposition, the prior model for all evaluated timesteps then becomes:

$$f(\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{T+1}) = f(\boldsymbol{x}_0) \prod_{t=1}^{T+1} f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}).$$

Again, using the Markov properties of a DAG, we see that the posterior model

$$f(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{T+1}|\boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_{T-1}^o, \boldsymbol{x}_T^o) = \frac{f(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{m+1}, \boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_{T-1}^o, \boldsymbol{x}_T^o)}{f(\boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_{T-1}^o \boldsymbol{x}_T^o)}$$

$$\propto f(\boldsymbol{x}_0^o|\boldsymbol{x}_0) f(\boldsymbol{x}_0) f(\boldsymbol{x}_{T+1}|\boldsymbol{x}_T) \prod_{t=1}^{T} f(\boldsymbol{x}_t^o|\boldsymbol{x}_t) f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}).$$

The prediction of the future properties of the reservoir is of special interest, that is, the marginal posterior distribution expressed as

$$f(\boldsymbol{x}_{T+1}|\boldsymbol{x}_0^o, \ldots, \boldsymbol{x}_T^o) = \int \ldots \int f(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{T+1}|\boldsymbol{x}_0^o, \ldots, \boldsymbol{x}_T^o) \mathrm{d}\boldsymbol{x}_0 \ldots \mathrm{d}\boldsymbol{x}_T. \quad (38)$$

Using the Law of Total Probability (Walpole et al. 2002) and again the Markov property of the DAG, we see that a recursive expression for the state of the reservoir at timestep t is given as

$$\boldsymbol{x}_t^i \sim f(\boldsymbol{x}_t|\boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_{t-1}^o) = \int f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) f(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_{t-1}^o) \mathrm{d}\boldsymbol{x}_{t-1}. \quad (39)$$

Moreover, realisations honouring the observed value at timestep t, $\boldsymbol{x}_t^{c,i}$, is given as:

$$\begin{aligned} \boldsymbol{x}_t^{c,i} &\sim f(\boldsymbol{x}_t|\boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_t^o) \\ &\propto f(\boldsymbol{x}_t^o|\boldsymbol{x}_t) f(\boldsymbol{x}_t|\boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_{t-1}^o). \end{aligned} \quad (40)$$

Algorithm 2 describes a general sequential method for the history matching inverse problem, based on the general formulation in Tarantola (2005). This algorithm require that we generate realisations from the distribution described in Expression (40), which we know up to an unknown normalisation constant. The problem here is that this constant will in most practical applications be too computer demanding to determine. A solution to this problem is to use a Markov chain Monte Carlo (MCMC) (Barndorff-Nielsen et al. 2001), rejection sampling (RS) or sampling importance sampling (SIR) algorithm (Coles et al. 2002, Omre 2000), which again will be a computer demanding process.

However, if $f(\boldsymbol{x}_t|\boldsymbol{x}_0^o, \ldots, \boldsymbol{x}_{t-1}^o)$ are assumed to be Gaussian, and the likelihood models are assumed to be Gauss linear, then, as shown in Appendix B, $f(\boldsymbol{x}_t|\boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_t^o)$ will be Gaussian, thus making the normalisation constant analytically tractable. Due to the complexity of the fluid flow simulator $\omega_{\Delta_t}(\bullet)$, this assumption will unfortunately not be valid. An approximation is therefore that $f(\boldsymbol{x}_t|\boldsymbol{x}_0^o, \ldots, \boldsymbol{x}_{t-1}^o)$ at an arbitrary timestep follow the Gaussian distribution with empirically estimated parameters, hence making the generation of $\boldsymbol{x}_t^{c,i}$ very computer efficient. This method corresponds to the Ensemble Kalman Filter introduced by Evensen et al. (1998).

Algorithm 2: Sequential Bayesian History Matching

- Assign some *a prior* distribution to the model at the initial state, $f(\boldsymbol{x}_0)$.
- Introduce the physical theory connecting $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t-1}, \ldots, \boldsymbol{x}_0$, namely the forward model $f(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \ldots, \boldsymbol{x}_0)$.
- Assign some likelihood model, $f(\boldsymbol{x}_t^o|\boldsymbol{x}_t)$, connecting observations of the field, to the true field.
- Generate a realisation, from the *a posteriori* distribution for the initial state, $\boldsymbol{x}_0^c \sim f(\boldsymbol{x}_0|\boldsymbol{x}_0^o)$.

for $t = 1$ to T **do**

- Advance model to the next timestep, honouring previous observations, using the forward model. That is, generate $\boldsymbol{x}_t \sim f(\boldsymbol{x}_t|\boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_{t-1}^o)$
- Adjust model to honour the observed field by generating a realisation $\boldsymbol{x}_t^c \sim f(\boldsymbol{x}_t|\boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_{t-1}^o, \boldsymbol{x}_t^o)$

- Predict state of the reservoir in the future, using previous observations: $\boldsymbol{x}_{T+1} \sim f(\boldsymbol{x}_{T+1}|\boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_{T-1}^o, \boldsymbol{x}_T^o)$.

6 Ensemble Kalman Filter

This section presents the theory and algorithm of the traditional Ensemble Kalman Filter (EnKF), when applied to a reservoir history matching setting with observed production and 4-D seismic data.

Consider the model given in Figure 6, where the stochastic variables are assumed to have the following distributions:

$$\boldsymbol{x}_t|\boldsymbol{x}_0^o, \boldsymbol{x}_1^o, \ldots, \boldsymbol{x}_{t-1}^o \sim \text{Gauss}_{n_r+n_q}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t),$$

with

$$\boldsymbol{\mu}_t = \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{r}t} \\ \boldsymbol{\mu}_{\boldsymbol{q}_t} \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_t = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{r}t} & \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t} \\ \boldsymbol{\Sigma}_{\boldsymbol{q}\boldsymbol{r}_t} & \boldsymbol{\Sigma}_{\boldsymbol{q}_t} \end{bmatrix}.$$

As in Expression (32), Section 4.2, we assume that

$$\boldsymbol{d}_{j,t}^o = \boldsymbol{D}_{\boldsymbol{r}j,t}\boldsymbol{r}_{j,t} + \boldsymbol{\epsilon}_{\boldsymbol{d}j,t}, \quad \in \mathbb{R}^{n_\theta \cdot S_{d^o,j} \times 1}, \ \forall j \in \{1 \ldots n_{xy}\}. \tag{41}$$

Moreover, the observed production is given as

$$\boldsymbol{q}_t^o = \boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{q}_t + \boldsymbol{\epsilon}_{\boldsymbol{q}_t}, \quad \in \mathbb{R}^{n_{q^o} \times 1}, \quad \forall t > 0, \tag{42}$$

where $\boldsymbol{D}_{\boldsymbol{q}_t} \in \mathbb{R}^{n_q^o \times n_q}$, indicates which observations we have available at timestep t, $\boldsymbol{\epsilon}_{\boldsymbol{d}^o,j,t}$ follow the distribution in Expression (32), independent of $\boldsymbol{x}_t$, $\boldsymbol{\epsilon}_{\boldsymbol{d}^o,k,s}$, and $\boldsymbol{\epsilon}_{\boldsymbol{q}_t}$, $\forall k \neq j$, $\forall s \neq t$, $\forall t$, and

$$\boldsymbol{\epsilon}_{\boldsymbol{q}_t} \sim \text{Gauss}_{n_{q^o}}(0, \boldsymbol{\Sigma}_{\boldsymbol{q}_t^o}^o),$$

independent of $\boldsymbol{x}_t$ and $\boldsymbol{\epsilon}_{\boldsymbol{r}j,t} \quad \forall j, \forall t$. Note that in this Section we have omitted observations of well-logs, meaning that $\boldsymbol{x}_t^o$ only consists of production and seismic data.

Result 1, Appendix A.2 then gives

$$\begin{aligned} \boldsymbol{q}_t^o &\sim \text{Gauss}_{n_{q^o}}(\boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\mu}_{\boldsymbol{q}_t}, \boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T + \boldsymbol{\Sigma}_{\boldsymbol{q}_t^o}^o), \quad \forall t \geq 1, \\ \boldsymbol{d}_{j,t}^o &\sim \text{Gauss}_{n_\theta \cdot S_{d_j^o}}(\boldsymbol{\mu}_{\boldsymbol{d}^o\,j,t}, \boldsymbol{\Sigma}_{\boldsymbol{d}^o\,j,t}), \quad \forall j, \forall t \in T_{d^o}, \end{aligned} \tag{43}$$

where $\boldsymbol{\mu}_{\boldsymbol{d}^o\,j,t}$ and $\boldsymbol{\Sigma}_{\boldsymbol{d}^o\,j,t}$ are given in Expressions (34) and (35), Section 4.2. Moreover, Result 2, Appendix A.2, gives

$$\boldsymbol{r}_t^q = \left[\boldsymbol{r}_t|\boldsymbol{x}_0^o, \ldots, \boldsymbol{x}_{t-1}^o, \boldsymbol{q}_t^o\right] \sim \text{Gauss}_{n_r}(\boldsymbol{\mu}_{\boldsymbol{r}|\boldsymbol{q}^o\,t}, \boldsymbol{\Sigma}_{\boldsymbol{r}|\boldsymbol{q}^o\,t}), \tag{44}$$

$$\boldsymbol{q}_t^q = \left[\boldsymbol{q}_t|\boldsymbol{x}_0^o, \ldots, \boldsymbol{x}_{t-1}^o, \boldsymbol{q}_t^o\right] \sim \text{Gauss}_{n_q}(\boldsymbol{\mu}_{\boldsymbol{q}|\boldsymbol{q}^o\,t}, \boldsymbol{\Sigma}_{\boldsymbol{q}|\boldsymbol{q}^o\,t}), \tag{45}$$

$$\boldsymbol{r}_{j,t}^{qd} = \left[\boldsymbol{r}_{j,t}^q|\boldsymbol{d}_{j,t}^o\right] \sim \text{Gauss}_{4 \cdot n_z}(\boldsymbol{\mu}_{\boldsymbol{r}^q|\boldsymbol{d}^o\,j,t}, \boldsymbol{\Sigma}_{\boldsymbol{r}^q|\boldsymbol{d}^o\,j,t}), \tag{46}$$

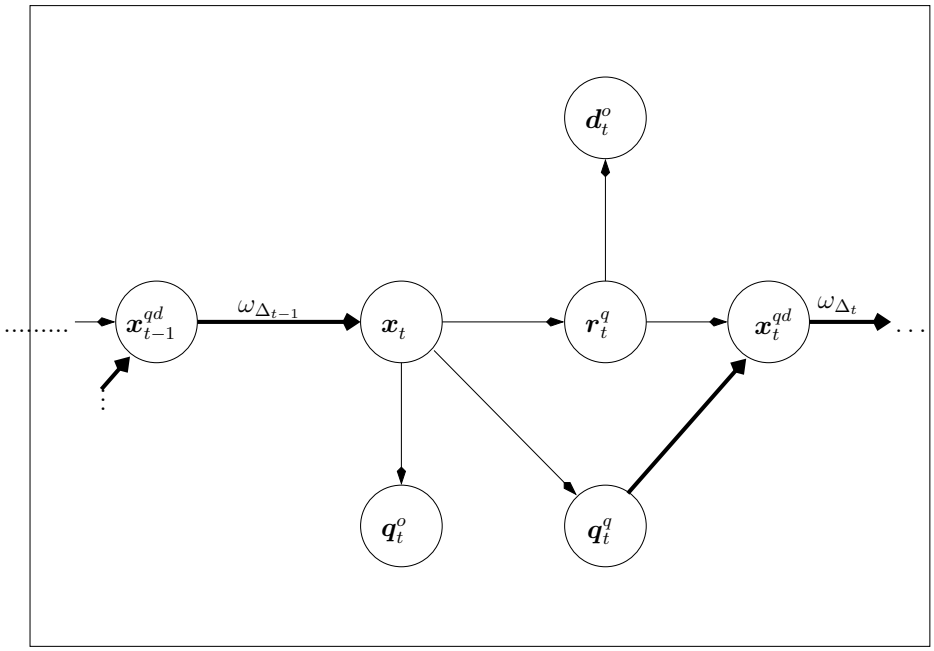


Figure 6: The figure shows the Directed Acyclic Graph (DAG) of the the assimilation part in the EnKF. Single arrows mean that there is a stochastic relationship between the variables, while thick arrows imply that there is a deterministic relationship.

where

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{r}|\boldsymbol{q^o}\,t} &= \boldsymbol{\mu}_{\boldsymbol{r}t} + \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T\left(\boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}^o\right)^{-1}\left(\boldsymbol{q}_t^o - \boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\mu}_{\boldsymbol{q}_t}\right), \qquad (47)\\
\boldsymbol{\mu}_{\boldsymbol{q}|\boldsymbol{q^o}\,t} &= \boldsymbol{\mu}_{\boldsymbol{q}_t} + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T\left(\boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}^o\right)^{-1}\left(\boldsymbol{q}_t^o - \boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\mu}_{\boldsymbol{q}_t}\right),\\
\boldsymbol{\mu}_{\boldsymbol{r}^q|\boldsymbol{d}^o,j,t} &= \boldsymbol{\mu}_{\boldsymbol{r}|\boldsymbol{q^o}\,,j,t} + \boldsymbol{\Sigma}_{\boldsymbol{r}|\boldsymbol{q}^o,j,t}\boldsymbol{D}_{\boldsymbol{r},j,t}^T\boldsymbol{\Sigma}_{\boldsymbol{d}^o,j,t}^{-1}(\boldsymbol{d}_{j,t}^o - \boldsymbol{\mu}_{\boldsymbol{d}^o,j,t}),
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{r}|\boldsymbol{q^o}\,t} &= \boldsymbol{\Sigma}_{\boldsymbol{r}t} - \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T\left(\boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}^o\right)^{-1}\boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}^T, \qquad (48)\\
\boldsymbol{\Sigma}_{\boldsymbol{q}|\boldsymbol{q^o}\,t} &= \boldsymbol{\Sigma}_{\boldsymbol{q}_t} - \boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T\left(\boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}^o\right)^{-1}\boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\Sigma}_{\boldsymbol{q}_t},\\
\boldsymbol{\Sigma}_{\boldsymbol{r}^q|\boldsymbol{d}^o,j,t} &= \boldsymbol{\Sigma}_{\boldsymbol{r}|\boldsymbol{q}^o,j,t} - \boldsymbol{\Sigma}_{\boldsymbol{r}|\boldsymbol{q}^o,j,t}\boldsymbol{D}_{\boldsymbol{r},j,t}^T\boldsymbol{\Sigma}_{\boldsymbol{d}^o,j,t}^{-1}\boldsymbol{D}_{\boldsymbol{r},j,t}\boldsymbol{\Sigma}_{\boldsymbol{r}|\boldsymbol{q}^o,j,t},
\end{aligned}$$

$\boldsymbol{\mu}_{\boldsymbol{d}^o,j,t}$ is given in Expression (34), and $\boldsymbol{\mu}_{\boldsymbol{r}^q|\boldsymbol{d^o}\,j,t}$, $\boldsymbol{\Sigma}_{\boldsymbol{r}^q|\boldsymbol{d^o}\,j,t}$ are given in Expression (37), replacing $\boldsymbol{\mu}_{\boldsymbol{r}j,t}$ with $\boldsymbol{\mu}_{\boldsymbol{r}|\boldsymbol{q^o}\,j,t}$ and $\boldsymbol{\Sigma}_{\boldsymbol{r}j,t}$ with $\boldsymbol{\Sigma}_{\boldsymbol{r}|\boldsymbol{q^o}\,j,t}$.

As shown in Appendix B, realisations from the posterior distributions $\boldsymbol{r}_t^q \sim f(\boldsymbol{r}_t|\boldsymbol{r}_0^o,\ldots,\boldsymbol{q}_t^o)$, $\boldsymbol{q}_t^q \sim f(\boldsymbol{r}_t|\boldsymbol{r}_0^o,\ldots,\boldsymbol{q}_t^o)$ and $\boldsymbol{r}_t^{qd} \sim f(\boldsymbol{r}_{j,t}^q|\boldsymbol{d}_{j,t}^o)$ in Expressions (44) to (46) can then be generated as follows:

$$\begin{aligned}
\boldsymbol{r}_t^{q,i} &= \boldsymbol{r}_t^i + \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T\left(\boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}^o\right)^{-1}\left(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_{\boldsymbol{q}_t}^{\,i} - \boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{q}_t^t\right),\\
\boldsymbol{q}_t^{q,i} &= \boldsymbol{q}_t^i + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T\left(\boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}^o\right)^{-1}\left(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_{\boldsymbol{q}_t}^{\,i} - \boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{q}_t^i\right),\\
\boldsymbol{r}_{j,t}^{qd,i} &= \boldsymbol{r}_{j,t}^{q,i} + \boldsymbol{\Sigma}_{\boldsymbol{r}^q,j,t}\boldsymbol{D}_{\boldsymbol{r}j,t}^T\boldsymbol{\Sigma}_{\boldsymbol{d}^o j,t}^{-1}\left(\boldsymbol{d}_{j,t}^o + \boldsymbol{\epsilon}_{\boldsymbol{r}j,t} - \boldsymbol{D}_{\boldsymbol{r}j,t}\boldsymbol{r}_{j,t}^{q,i}\right),
\end{aligned}$$

with $\boldsymbol{\epsilon}_{\boldsymbol{q}_t}^1, \ldots, \boldsymbol{\epsilon}_{\boldsymbol{q}_t}^{n_s} \overset{\text{i.i.d}}{\sim} \text{Gauss}_{n_q}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{q}_t}^o)$, independent of

$$\boldsymbol{r}_t^j,\ \boldsymbol{q}_t^j,\ , \boldsymbol{\epsilon}_{\boldsymbol{r}_t}^j,\ \forall t, i, j,$$

and $\boldsymbol{\epsilon}_{\boldsymbol{r}_{j,t}}^i$ following the distribution in Expression (32), independent of

$$\boldsymbol{r}_t^j,\ \boldsymbol{q}_t^j,\ , \boldsymbol{\epsilon}_{\boldsymbol{q}_t},\ \boldsymbol{\epsilon}_{\boldsymbol{r}_{k,t}}^i\ \forall t, i, j, k \neq j.$$

As we can see from the expressions above, $\boldsymbol{\Sigma}_t$ is the only unknown parameter required when we want to generate realisations $\boldsymbol{r}_t^q$, $\boldsymbol{q}_t^q$ and $\boldsymbol{r}_{j,t}^{qd}$. The idea of the EnKF Algorithm is to use a set of realisations $\{\boldsymbol{x}_t^1, \ldots, \boldsymbol{x}_t^{n_s}\}$, referred to as an ensemble, to estimate this unknown covariance matrix $\boldsymbol{\Sigma}_t$. A typical estimator used will be

$$\hat{\boldsymbol{\Sigma}}_t = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}t} & \hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}\boldsymbol{q}t} \\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}\boldsymbol{r}t} & \hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}t} \end{bmatrix} = \frac{1}{n_s - 1} \sum_{i=1}^{n_s} (\boldsymbol{x}_t^i - \bar{\boldsymbol{x}}_t)(\boldsymbol{x}_t^i - \bar{\boldsymbol{x}}_t)^T,$$

which is both an unbiased and consistent estimator of $\boldsymbol{\Sigma}_t$ (Casella & Berger 2002).

Approximate realisations from the posterior distribution can then be generated by substituting the covariance matrices in the expressions above, by their empirically estimated equivalents.

Algorithm 3 summaries the EnKF procedure. Note that the only difference from the EnKF algorithm proposed by Evensen et al. (1998) is that here we also have included seismic observations at each timestep $t \in T_{d^o}$.

Remark. When the number of ensemble members is less than the dimension of the reservoir, the estimated covariance matrix will become rank deficient (Johnson & Wichern 2002). As long as the prior covariance matrix $\boldsymbol{\Sigma}_{(\bullet)}^o$ is positive definite, the inversion step of the EnKF can in most cases be performed. However, a low rank representation of the covariance matrix can lead to divergence when the ensemble members are updated (Evensen 2007).

Algorithm 3: EnKF

for $i = 1$ *to* n_s **do**
  Generate: $\boldsymbol{x}_0^{qd,i} \sim f(\boldsymbol{x}_0|\boldsymbol{x}_0^o)$
for $t = 1$ *to* T **do**
  for $i = 1$ *to* n_s **do**
    Generate: $\boldsymbol{x}_t^i = \omega_{\Delta_{t-1}}(\boldsymbol{x}_{t-1}^{qd,i})$
  $\hat{\boldsymbol{\Sigma}}_t = \frac{1}{n-1}\sum_i (\boldsymbol{x}_t^i - \bar{\boldsymbol{x}}_t)(\boldsymbol{x}_t^i - \bar{\boldsymbol{x}}_t)^T$
  for $i = 1$ *to* n_s **do**
    Generate: $\boldsymbol{\epsilon}_{\boldsymbol{q}_t}^i \sim \text{Gauss}_{n_q}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{q}_t}^o\right)$
    $\boldsymbol{r}_t^{q,i} = \boldsymbol{r}_t^i + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T\left(\boldsymbol{D}_{\boldsymbol{q}_t}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}^o\right)^{-1}\left(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_{\boldsymbol{q}_t}^i - \boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{q}_t^t\right)$
    $\boldsymbol{q}_t^{q,i} = \boldsymbol{q}_t^i + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T\left(\boldsymbol{D}_{\boldsymbol{q}_t}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}_t}\boldsymbol{D}_{\boldsymbol{q}_t}^T + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}^o\right)^{-1}\left(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_{\boldsymbol{q}_t}^i - \boldsymbol{D}_{\boldsymbol{q}_t}\boldsymbol{q}_t^i\right)$
  if $t \in T_{d^o}$ **then**
    for $i = 1$ *to* n_s **do**
      for $j = 1$ *to* n_{xy} **do**
        Generate: $\boldsymbol{\epsilon}_{\boldsymbol{\omega}}{}_j^i \sim \text{Gauss}_{n_\theta \cdot S}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}t}\right)$
        Generate: $\boldsymbol{\epsilon}_{\boldsymbol{d}^o}{}_j^i \sim \text{Gauss}_{n_\theta \cdot S}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{d}^o t}\right)$
        $\boldsymbol{\epsilon}_{\boldsymbol{r}}{}_{j,t}^i = \boldsymbol{D}_{\boldsymbol{d}^o,j,t}\left(\boldsymbol{\mu}_{\boldsymbol{\epsilon}_{\boldsymbol{r}\boldsymbol{d}^o}}{}_{j,t}^i + \boldsymbol{\epsilon}_{\boldsymbol{\omega}}{}_j^i\right) + \boldsymbol{\epsilon}_{\boldsymbol{d}^o}{}_j^i$
        $\boldsymbol{r}_{j,t}^{q,i} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}^q}{}_{j,t}\boldsymbol{D}_{\boldsymbol{r}}{}_{j,t}^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{d}^o}^{-1}{}_{j,t}\left(\boldsymbol{d}_{j,t}^o + \boldsymbol{\epsilon}_{\boldsymbol{r}}{}_{j,t}^i - \boldsymbol{D}_{\boldsymbol{r}j,t}\boldsymbol{r}_{j,t}^{q,i}\right)$
  else
    $\boldsymbol{r}_t^{qd,i} = \boldsymbol{r}_t^{q,i}, \forall i$
  $\boldsymbol{x}_t^{qd,i} = \left[\boldsymbol{r}_t^{qd,i^T}, \boldsymbol{q}_t^{q,i^T}\right]^T, \ \forall i$

7 Approximate Fluid Flow Simulation

In this section we motivate the use of approximate fluid flow simulators, present various upscaling techniques, and outline two downscaling methods that can incorporate the bias introduced by using a coarse scale fluid flow simulator.

When the dimension of the reservoir n_r, is large, fluid flow simulation on the original fine scale reservoir may be computationally prohibitive in a practical setting. Since both manual and automatic approaches to history matching require multiple realisations of the reservoir properties, meaning repeated fluid flow simulations, a high number of reservoir grid blocks would restrict the usefulness of any history matching procedure.

A solution to this problem is to replace the fluid flow simulator given in Expression (7) by an approximate simulator, as shown in Figure 7,

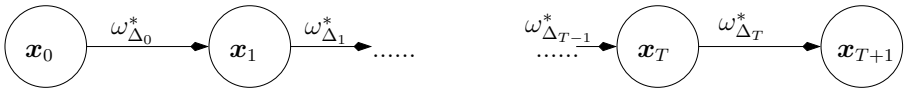


Figure 7: The figure shows the Directed Acyclic Graph (DAG) of the approximate fluid flow simulator. Thin arrows imply that there is a stochastic relationship.

$$\boldsymbol{x}_t = \omega^*_{\Delta_{t-1}}(\boldsymbol{x}_{t-1}) + \boldsymbol{\epsilon}^*_{\Delta_{t-1}}, \tag{49}$$

where $\omega^*_{\Delta t} : \mathbb{R}^{(n_r+n_q)\times 1} \rightarrow \mathbb{R}^{(n_r+n_q)\times 1}$, and $\boldsymbol{\epsilon}^*_{\Delta t} \in \mathbb{R}^{n_r+n_q \times 1}$ is a centred Gaussian approximation error with covariance matrix $\boldsymbol{\Sigma}^*_{\Delta t}$. That is

$$\boldsymbol{\epsilon}^*_{\Delta t} \sim \text{Gauss}_{n_r+n_q}(\boldsymbol{0}, \boldsymbol{\Sigma}^*_{\Delta t}).$$

The covariance matrix $\boldsymbol{\Sigma}^*_{\Delta t}$ will contain the the covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{r}|\boldsymbol{r}^*\,t}$ and $\boldsymbol{\Sigma}_{\boldsymbol{q}|\boldsymbol{q}^*\,t}$ outlined in Sections 7.3.1 and 7.3.2.

The approximate fluid flow simulator $\omega^*_{\Delta t}(\bullet)$ will normally be divided into three parts:

1. Upscaling,
2. Coarse scale fluid flow simulation,
3. Downscaling.

7.1 Upscaling

According to Farmer (2002), upscaling generally is the approximation of a system of partial differential equations by another, often of the same form, in such a way that the new system can be solved with fewer computing resources. In the reservoir setting this implies that we apply a function, possibly highly complex, $\nu^* : \mathbb{R}^{n_r \times 1} \rightarrow \mathbb{R}^{n^*_r \times 1}$, such that

$$\begin{bmatrix} \boldsymbol{r}^*_t \\ \boldsymbol{q}^*_t \end{bmatrix} = \begin{bmatrix} \nu^*(\boldsymbol{r}_t) \\ \boldsymbol{q}_t \end{bmatrix} + \boldsymbol{\epsilon}^*_t = \nu^*(\boldsymbol{x}_t) + \boldsymbol{\epsilon}^*_t, \tag{50}$$

where $\boldsymbol{\epsilon}_t^* \in \mathbb{R}^{n_r+n_q\times 1}$ is a centred Gaussian variable with covariance matrix $\boldsymbol{\Sigma}_t^*$ reflecting the error introduced by moving the reservoir from a fine scale to a coarse scale (Deutsch 2002, Omre & Lødøen 2004, Lødøen et al. 2004). An illustration of this process is shown in Figure 8. The number of lattice nodes

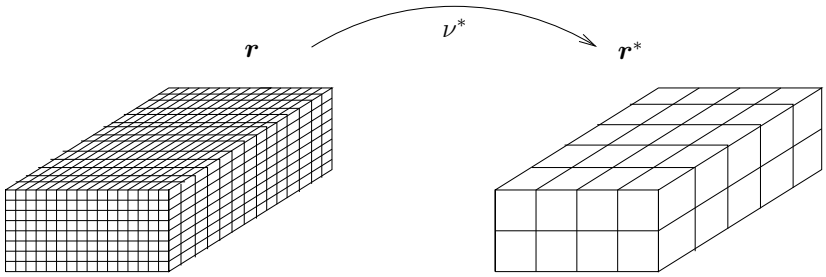


Figure 8: The figure shows the process of upscaling a reservoir, reducing the number of lattice nodes.

in the reservoir can therefore be considerably reduced, making the fluid flow simulation computationally affordable.

We now give a brief review of some of the methods used in the upscaling of a reservoir. A more thorough description can be found in Durlofsky (2003), Farmer (2002), or Chen et al. (2003).

Reservoir upscaling techniques can, according to Farmer (2002), be classified as either

- Local-Local,
- Local-Global,
- Global-Local,
- Global-Global.

In a Local-Local method, only the fine scale reservoir within the corresponding coarse scale block is considered. An example is taking the arithmetic mean of the fine grid blocks within the domain of the coarse grid block. Porosity and saturation are both volumetric concentrations, and arithmetic averaging will therefore in theory be correct. As noted by Deutsch (2002), upscaling of the permeability will be more problematic, where non-linear operation must be applied. The simplest two methods are the harmonic mean

$$K_h^* = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{K_i}\right]^{-1}, \tag{51}$$

and the geometric mean

$$\begin{aligned} K_g^* &= \left(\prod_{i=1}^{n} K_i\right)^{\frac{1}{n}} \\ &= \exp\left\{\frac{1}{n}\sum_{i=1}^{n}\log(K_i)\right\}. \end{aligned} \tag{52}$$

As we can see from Expression (52), the geometric mean is reduced to the arithmetic mean when we consider the log permeability $\boldsymbol{\kappa}$. Thus

$$\log(K_g^*) = \kappa^* = \frac{1}{n}\sum_{i=1}^{n}\kappa_i. \tag{53}$$

Since Local-Local methods by far are the easiest to implement, they are also the most popular methods used.

Global-Global methods require both fine, and coarse scale fluid flow simulation outputs, therefore they will possibly be the most computationally expensive method to use. According to Farmer (2002), these methods try to minimise a global functional measuring the difference between the fine and coarse grid solutions. For an introduction to the Global-Local and Local-Global methods, we refer to Farmer (2002).

7.2 Coarse Scale Fluid Flow Simulation

By coarse scale fluid flow simulation we simply mean that we replace the fine scale simulator ω by a coarse scale equivalent $\bar{\omega} : \mathbb{R}^{n_{q^*}+n_{r^*}} \rightarrow \mathbb{R}^{n_{q^*}+n_{r^*}}$, that requires less computer resources. This implies that we solve a set differential equations in a lower dimensional space. The result of this approximation is that the number of fluid flow simulations that can be afforded, in most cases will be considerably increased.

7.3 Downscaling

Downscaling of the reservoir properties and correcting the production properties, that is representing the reservoir and production properties on a fine scale given a coarse scale representation, can in general be done as follows:

$$\begin{bmatrix} \boldsymbol{r}_t \\ \boldsymbol{q}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_{\boldsymbol{r}t}^{\boldsymbol{s}} + \boldsymbol{A}_{\boldsymbol{r}t}\boldsymbol{r}_t^* \\ \boldsymbol{a}_{\boldsymbol{q}_t}^{\boldsymbol{s}} + \boldsymbol{A}_{\boldsymbol{q}_t}\boldsymbol{q}_t^* \end{bmatrix}, \tag{54}$$

where $\boldsymbol{a}_{\boldsymbol{r}t}^{\boldsymbol{s}} \in \mathbb{R}^{n_r \times 1}$ and $\boldsymbol{a}_{\boldsymbol{q}_t}^{\boldsymbol{s}} \in \mathbb{R}^{n_q \times 1}$ are shift vectors, and $\boldsymbol{A}_{\boldsymbol{r}t} \in \mathbb{R}^{n_r \times n_{r^*}}$ and $\boldsymbol{A}_{\boldsymbol{q}_t} \in \mathbb{R}^{n_q \times n_{q^*}}$ are correction matrices (Lødøen & Omre 2005). A more detailed description will be given in Sections 7.3.1 and 7.3.2. Note that both of the downscaling procedures presented in this Thesis minimises the Mean Squared Prediction Error (MSPE) under the Gaussian approximation used in the EnKF. This is shown in Result 13, Appendix A.3.

7.3.1 Downscaling in The Scale-Corrected Ensemble Kalman Filter

As noted above, the dimension of the reservoir often prohibits fluid flow simulation on a fine scale. Upscaling of the reservoir is therefore done, in order to perform fluid flow simulation on a coarser scale. The process of upscaling is known to introduce bias in the fluid flow simulation (Omre & Lødøen 2004). Based on this result, Lødøen & Omre (2005) proposed an empirical statistical method, where the relation between the reservoir on fine-, and coarse scale can be estimated.

Let

$$\boldsymbol{\mathfrak{x}}_t = \begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{x}_t^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_t \\ \boldsymbol{q}_t \\ \boldsymbol{r}_t^* \\ \boldsymbol{q}_t^* \end{bmatrix},$$

where $\boldsymbol{x}_t^* \in \mathbb{R}^{n_r^* + n_q}$ denotes the reservoir on a coarse scale. As in the EnKF algorithm, the assumption here was that $\boldsymbol{\mathfrak{x}}_t \in \mathbb{R}^{p \times 1}$, where $p = (n_r + n_q) + (n_{r^*} + n_{q^*}) = u + g$, was approximately Gaussian distributed with mean and covariance matrix given as

$$\boldsymbol{\mu}_{\boldsymbol{\mathfrak{x}}t} = \begin{bmatrix} \boldsymbol{\mu}_t \\ \boldsymbol{\mu}_t^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{r}t} \\ \boldsymbol{\mu}_{\boldsymbol{q}t} \\ \boldsymbol{\mu}_{\boldsymbol{r}^*t} \\ \boldsymbol{\mu}_{\boldsymbol{q}^*t} \end{bmatrix}, \tag{55}$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{x}}t} = \left[\begin{array}{c|c} \boldsymbol{\Sigma}_t & \boldsymbol{\Sigma}_{1*t} \\ \hline \boldsymbol{\Sigma}_{*1t} & \boldsymbol{\Sigma}_{*t} \end{array}\right] = \left[\begin{array}{cc|cc} \boldsymbol{\Sigma}_{\boldsymbol{r}t} & \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}t} & \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{r}^*t} & \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}^*t} \\ \boldsymbol{\Sigma}_{\boldsymbol{q}\boldsymbol{r}t} & \boldsymbol{\Sigma}_{\boldsymbol{q}t} & \boldsymbol{\Sigma}_{\boldsymbol{q}\boldsymbol{r}^*t} & \boldsymbol{\Sigma}_{\boldsymbol{q}\boldsymbol{q}^*t} \\ \hline \boldsymbol{\Sigma}_{\boldsymbol{r}^*\boldsymbol{r}t} & \boldsymbol{\Sigma}_{\boldsymbol{r}^*\boldsymbol{q}t} & \boldsymbol{\Sigma}_{\boldsymbol{r}^*t} & \boldsymbol{\Sigma}_{\boldsymbol{r}^*\boldsymbol{q}^*t} \\ \boldsymbol{\Sigma}_{\boldsymbol{q}^*\boldsymbol{r}t} & \boldsymbol{\Sigma}_{\boldsymbol{q}^*\boldsymbol{q}t} & \boldsymbol{\Sigma}_{\boldsymbol{q}^*\boldsymbol{r}^*t} & \boldsymbol{\Sigma}_{\boldsymbol{q}^*t} \end{array}\right]. \tag{56}$$

By performing fluid flow simulation on both coarse-, and fine scale, on a calibration ensemble $\boldsymbol{\mathfrak{x}}_t^1, \ldots, \boldsymbol{\mathfrak{x}}_t^{n_c}$, empirical estimates of the mean and covariance matrix could be obtained. This was referred to the calibration step of the algorithm.

A larger set of realisations, $\boldsymbol{x}_t^{*,1}, \ldots, \boldsymbol{x}_t^{*,n_s}$, was then used in what was referred to as the simulation step. Here all fluid flow simulation was performed on a coarse scale, hence the number of realisations in the simulation set n_s, was larger than the number of realisations in the calibration set n_c. At each timestep, fine scale forecasts could be generated using the coarse scale realisations and the empirical estimates from the calibration step. By Result 2, Appendix A.2, downscaling of the coarse scale reservoir was then done by

$$\boldsymbol{r}_t^i = \hat{\boldsymbol{\mu}}_{\boldsymbol{r}t} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}\boldsymbol{r}^*t}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}^*t}^{-1}(\boldsymbol{r}_t^{*,i} - \hat{\boldsymbol{\mu}}_{\boldsymbol{r}^*t}) + \hat{\boldsymbol{\epsilon}}_t^i \tag{57}$$

$$\boldsymbol{q}_t^i = \hat{\boldsymbol{\mu}}_{\boldsymbol{q}t} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}\boldsymbol{q}^*t}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}^*t}^{-1}(\boldsymbol{q}_t^{*,i} - \hat{\boldsymbol{\mu}}_{\boldsymbol{q}^*t}) + \tilde{\boldsymbol{\epsilon}}_t^i, \tag{58}$$

where

$$\hat{\boldsymbol{\epsilon}}_t^1, \ldots, \hat{\boldsymbol{\epsilon}}_t^{n_s} \overset{\text{i.i.d}}{\sim} \text{Gauss}_{n_r}(\boldsymbol{0}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}t} - \hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}\boldsymbol{r}^*t}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}^*t}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}^*\boldsymbol{r}t})$$
$$\tilde{\boldsymbol{\epsilon}}_t^1, \ldots, \tilde{\boldsymbol{\epsilon}}_t^{n_s} \overset{\text{i.i.d}}{\sim} \text{Gauss}_{n_q}(\boldsymbol{0}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}t} - \hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}\boldsymbol{q}^*t}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}^*t}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}^*\boldsymbol{q}t}).$$

As we can see from the expressions above, this downscaling method is similar to the one described in Expression (54), where the bias introduced by performing the simulation on a coarser grid is corrected for. The downscaled ensemble could then be assimilated using the EnKF equations described in Section 6.

It should be noted that if we only consider observations of the production properties, downscaling of the reservoir properties as in Expression (57) is only

required during prediction of a future reservoir $\boldsymbol{r}_{m+1}$ (see Lødøen & Omre (2005) for details). This further omitted the need to compute, and invert $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}^*t}$, which will be singular when the number of ensemble members in the calibration set is smaller than the dimension of the coarse scale reservoir, n_r^* (Johnson & Wichern 2002, Result 3.3). This will usually be the case when the dimension of the fine scale reservoir is large, since the calibration step of the Scale-Corrected EnKF also requires repeated fluid flow simulation on a fine scale grid.

In order to predict the future reservoir, an *ad hoc* solution using Bayesian regularisation was therefore proposed by Lødøen & Omre (2005).

7.3.2 Downscaling in a Hierarchical Bayesian Setting

As noted above, the empirical coarse scale covariance matrix $\hat{\boldsymbol{\Sigma}}_{*t}$, will be singular when the number of realisations in the calibration set is less than the dimension of the covariance matrix, which will typically be the case when fluid flow simulation has to be performed on both fine and coarse scale. It also requires inversion of the the matrix $\boldsymbol{\Sigma}_{\boldsymbol{q}_t^*}^*$ at each timestep, which will be rank deficient when we observe no variability in at least one of the production properties considered.

The idea in this Thesis is to instead use a hierarchical Bayesian approach, where we assign prior distributions to the expectation and covariance matrix, calibrate using only coarse scale realisations, and downscale using the posterior distributions. We will now give a detailed description of the model considered, some theoretical results, and finally the algorithm used in our proposed approximate fluid flow simulator used in the Hierarchical Scale-Corrected Ensemble Kalman Filter (HScCEnKF) outlined in Section 8.

Note that in the current study we only consider a linear transformation in the upscaling of the reservoir, $\nu^*(\boldsymbol{r}_t) = \boldsymbol{A}\boldsymbol{r}_t$, where $\boldsymbol{A} \in \mathbb{R}^{n_r^* \times n_r}$. Hence,

$$\begin{bmatrix} \boldsymbol{r}_t^* \\ \boldsymbol{q}_t^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}\boldsymbol{r}_t \\ \boldsymbol{q}_t \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{\boldsymbol{r},t}^* \\ \boldsymbol{\epsilon}_{\boldsymbol{q},t}^* \end{bmatrix}, \tag{59}$$

where $\boldsymbol{\epsilon}_{\boldsymbol{r},t}^* \sim \text{Gauss}_{n_{r^*}}(\boldsymbol{\mu}_{\boldsymbol{\epsilon}_{\boldsymbol{r}}t}^*, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_{\boldsymbol{r}}t}^*)$ independent of $\boldsymbol{x}_t$, $\boldsymbol{\epsilon}_{\boldsymbol{q}_t^*}^*$, $\forall t$, and $\boldsymbol{\epsilon}_{\boldsymbol{q}_t^*}^* \sim \text{Gauss}_{n_q}(\boldsymbol{\mu}_{\boldsymbol{\epsilon}_{\boldsymbol{q}}t}^*, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_{\boldsymbol{q}}t}^*)$ independent of $\boldsymbol{x}_t$, $\boldsymbol{\epsilon}_{\boldsymbol{r}_t^*}^*$, $\forall t$.

Omitting the subscript t and using Definitions 1 and 2 in Appendix A, we can now write the expectation and covariance of $\boldsymbol{\mathfrak{X}}$ as

$$\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}} = \begin{bmatrix} E[\boldsymbol{x}] \\ E[\boldsymbol{x}^*] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{r}} \\ \boldsymbol{\mu}_{\boldsymbol{q}} \\ \boldsymbol{A}\boldsymbol{\mu}_{\boldsymbol{r}} + \boldsymbol{\mu}_{\boldsymbol{\epsilon}_{\boldsymbol{r}}}^* \\ \boldsymbol{\mu}_{\boldsymbol{q}} + \boldsymbol{\mu}_{\boldsymbol{\epsilon}_{\boldsymbol{q}}}^* \end{bmatrix}.$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} = \left[\begin{array}{cc|cc} \boldsymbol{\Sigma}_{\boldsymbol{r}} & \boldsymbol{\Sigma}_{\boldsymbol{rq}} & \boldsymbol{\Sigma}_{\boldsymbol{r}}\boldsymbol{A}^T & \boldsymbol{\Sigma}_{\boldsymbol{rq}} \\ \boldsymbol{\Sigma}_{\boldsymbol{qr}} & \boldsymbol{\Sigma}_{\boldsymbol{q}} & \boldsymbol{\Sigma}_{\boldsymbol{qr}}\boldsymbol{A}^T & \boldsymbol{\Sigma}_{\boldsymbol{q}} \\ \hline \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{r}} & \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{rq}} & \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{r}}\boldsymbol{A}^T + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_{\boldsymbol{r}}}^* & \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{rq}} \\ \boldsymbol{\Sigma}_{\boldsymbol{qr}} & \boldsymbol{\Sigma}_{\boldsymbol{q}} & \boldsymbol{\Sigma}_{\boldsymbol{qr}}\boldsymbol{A}^T & \boldsymbol{\Sigma}_{\boldsymbol{q}} + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_{\boldsymbol{q}}}^* \end{array}\right].$$

Defining $\boldsymbol{B} \in \mathbb{R}^{g \times u}$ as

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n_q} \end{bmatrix}, \tag{60}$$

where $\boldsymbol{I}_n$ denotes the identity matrix of dimension $n \times n$, we see that the expectation and covariance matrix of $\boldsymbol{\mathfrak{X}}$ can be equally expressed as

$$\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{B\mu} + \boldsymbol{\mu}^*_{\boldsymbol{\epsilon}} \end{bmatrix}, \tag{61}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma B}^T \\ \boldsymbol{B\Sigma} & \boldsymbol{B\Sigma B}^T + \boldsymbol{\Sigma}^*_{\boldsymbol{\epsilon}} \end{bmatrix}, \tag{62}$$

with

$$\boldsymbol{\mu}^*_{\boldsymbol{\epsilon}} = \begin{bmatrix} \boldsymbol{\mu}^*_{\boldsymbol{\epsilon_r}} \\ \boldsymbol{\mu}^*_{\boldsymbol{\epsilon_q}} \end{bmatrix},$$

and

$$\boldsymbol{\Sigma}^*_{\boldsymbol{\epsilon}} = \begin{bmatrix} \boldsymbol{\Sigma}^*_{\boldsymbol{\epsilon_r}} & \boldsymbol{0} \\ \boldsymbol{0}^T & \boldsymbol{\Sigma}^*_{\boldsymbol{\epsilon_q}} \end{bmatrix}.$$

Consider the model in Figure 9, where $\boldsymbol{x}_t \in \mathbb{R}^{u\times 1}$, $\boldsymbol{x}^*_t \in \mathbb{R}^{g\times 1}$, $\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}t} \in \mathbb{R}^{p\times 1}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}t} \in \mathbb{R}^{p\times p} \quad \forall\, t \in \{0, 1, \ldots, m\}$. Again,

$$p = u + g = (n_q + n_r) + (n_q + n^*_r).$$

As in the EnKF we assume that

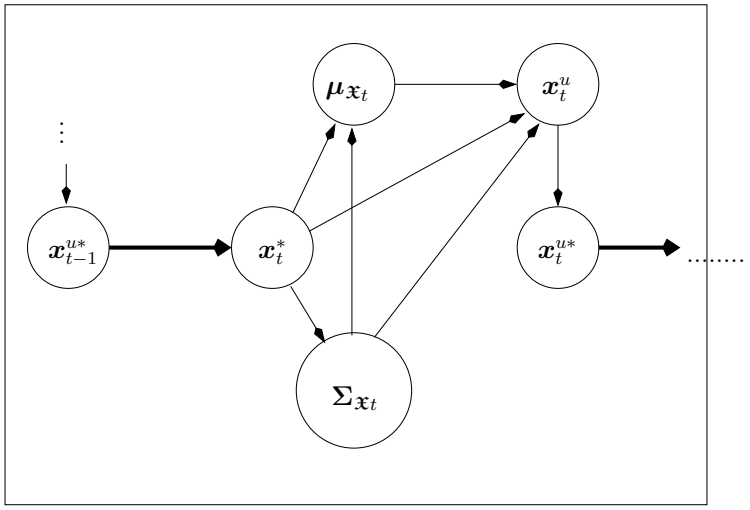


Figure 9: The figure shows the Directed Acyclic Graph (DAG) for one timestep in the Hierarchical Bayesian approximate fluid flow simulator. Thin arrows imply a stochastic relationship, while thick arrows imply a deterministic relationship.

$$\boldsymbol{\mathfrak{X}}_t | \boldsymbol{x}^o, \boldsymbol{x}^o_1, \ldots, \boldsymbol{x}^o_{t-1}, \boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}t}, \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}t} \sim \text{Gauss}_p(\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}t}, \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}t}), \quad \forall t \geq 1 \tag{63}$$

where, by omitting the subscript t, the parameters in Expression (63) are assigned the following prior distributions

$$\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}} | \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} \sim \text{Gauss}_p(\boldsymbol{\eta}_{\boldsymbol{\mathfrak{X}}}, \frac{1}{\zeta}\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}) \tag{64}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} \sim W^{-1}_p(\boldsymbol{\Psi}^{-1}_{\boldsymbol{\mathfrak{X}}}, \nu), \tag{65}$$

with

$$\boldsymbol{\eta_x} = \begin{bmatrix} \boldsymbol{\eta_r} \\ \boldsymbol{\eta_q} \\ \boldsymbol{A\eta_r} + \boldsymbol{\eta}^*_{\boldsymbol{\epsilon}_r} \\ \boldsymbol{\eta_q} + \boldsymbol{\eta}^*_{\boldsymbol{\epsilon}_q} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{B\eta} + \boldsymbol{\eta}^*_{\boldsymbol{\epsilon}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\eta}^* \end{bmatrix},$$

$$\begin{aligned} \boldsymbol{\Psi}^{-1}_{\boldsymbol{x}} &= \left[\begin{array}{cc|cc} \boldsymbol{\Psi_r} & \boldsymbol{\Psi_{rq}} & \boldsymbol{\Psi_r A}^T & \boldsymbol{\Psi_{rq}} \\ \boldsymbol{\Psi_{qr}} & \boldsymbol{\Psi_q} & \boldsymbol{\Psi_{qr} A}^T & \boldsymbol{\Psi_q} \\ \hline \boldsymbol{A\Psi_r} & \boldsymbol{A\Psi_{rq}} & \boldsymbol{A\Psi_r A}^T + \boldsymbol{\Psi}^*_{\boldsymbol{\epsilon}_r} & \boldsymbol{A\Psi_{rq}} \\ \boldsymbol{\Psi_{qr}} & \boldsymbol{\Psi_q} & \boldsymbol{\Psi_{qr} A}^T & \boldsymbol{\Psi_q} + \boldsymbol{\Psi}^*_{\boldsymbol{\epsilon}_q} \end{array}\right] \\ &= \begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{\Psi B}^T \\ \boldsymbol{B\Psi} & \boldsymbol{B\Psi B}^T + \boldsymbol{\Psi}^*_{\boldsymbol{\epsilon}} \end{bmatrix}, \end{aligned}$$

and ζ being known hyperparameters. Here the notation $\boldsymbol{U} \sim W_p^{-1}(\bullet)$ means that the matrix $\boldsymbol{U} \in \mathbb{R}^{p \times p}$ has the inverted Wishart distribution defined in Appendix A.1.

Since $\boldsymbol{\mu}^* = \boldsymbol{B\mu} + \boldsymbol{\mu}^*_{\boldsymbol{\epsilon}}$, by Result 1 and 2, Appendix A.2, it is sufficient to only assign a prior distributions on the fine scale reservoir $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^*_{\boldsymbol{\epsilon}}$. It should be noted, however, that a similar result does not exist for the prior distribution of $\boldsymbol{\Sigma_x}$. This is shown in Result 7, Appendix A.2.

Using the Bartlett decomposition (Bartlett 1933) we can, according to Lee & Zidek (1992), express the matrix above as

$$\boldsymbol{\Sigma_x} = \begin{bmatrix} \boldsymbol{\Sigma}_{1|*} + \boldsymbol{T\Sigma}_* \boldsymbol{T}^T & \boldsymbol{T\Sigma}_* \\ \boldsymbol{\Sigma}_* \boldsymbol{T}^T & \boldsymbol{\Sigma}_* \end{bmatrix},$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_{1|*} &= \boldsymbol{\Sigma} - \boldsymbol{\Sigma B}^T \left[\boldsymbol{B\Sigma B}^T + \boldsymbol{\Sigma}^*_{\boldsymbol{\epsilon}}\right]^{-1} \boldsymbol{B\Sigma}, \quad \in \mathbb{R}^{u \times u}, \\ \boldsymbol{T} &= \boldsymbol{\Sigma B}^T \left[\boldsymbol{B\Sigma B}^T + \boldsymbol{\Sigma}^*_{\boldsymbol{\epsilon}}\right]^{-1}, \quad \in \mathbb{R}^{u \times g}, \end{aligned}$$

and $\boldsymbol{\Sigma}_* = \left[\boldsymbol{B\Sigma B}^T + \boldsymbol{\Sigma}^*_{\boldsymbol{\epsilon}}\right]$. Similarly

$$\boldsymbol{\Psi}^{-1}_{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{\Psi}_{1|*} + \boldsymbol{\Gamma\Psi}_* \boldsymbol{\Gamma}^T & \boldsymbol{\Gamma\Psi}_* \\ \boldsymbol{\Psi}_* \boldsymbol{\Gamma}^T & \boldsymbol{\Psi}_* \end{bmatrix},$$

where

$$\begin{aligned} \boldsymbol{\Psi}_{1|*} &= \boldsymbol{\Psi} - \boldsymbol{\Psi B}^T \left[\boldsymbol{B\Psi B}^T + \boldsymbol{\Psi}^*_{\boldsymbol{\epsilon}}\right]^{-1} \boldsymbol{B\Psi}, \\ \boldsymbol{\Gamma} &= \boldsymbol{\Psi B}^T \left[\boldsymbol{B\Psi B}^T + \boldsymbol{\Psi}^*_{\boldsymbol{\epsilon}}\right]^{-1}, \end{aligned}$$

and $\boldsymbol{\Psi}_* = \left[\boldsymbol{B\Psi B}^T + \boldsymbol{\Psi}^*_{\boldsymbol{\epsilon}}\right]$.

As shown in Theorem 1, Appendix A.3, the new variables in the set $\boldsymbol{\Theta} = \{\boldsymbol{\Sigma}_*, \boldsymbol{T}, \boldsymbol{\Sigma}_{1|*}\}$, have the following distributions

$$\boldsymbol{\Sigma}_* \sim W_g^{-1}(\boldsymbol{\Psi}_*^{-1}, \nu - u) \tag{66}$$

$$\boldsymbol{\Sigma}_{1|*} \sim W_u^{-1}(\boldsymbol{\Psi}_{1|*}^{-1}, \nu) \tag{67}$$

$$\mathrm{Vec}(\boldsymbol{T})|\boldsymbol{\Sigma}_{1|*} \sim N_{ug}\left(\mathrm{Vec}(\boldsymbol{\Gamma}), (\boldsymbol{\Psi}_*^{-1} \otimes \boldsymbol{\Sigma}_{1|*})\right), \tag{68}$$

where the Kronecker product of two matrices, $\boldsymbol{A} \otimes \boldsymbol{B}$ and the Vec-operator is defined in Appendix A.3.

The theorem also shows that $\boldsymbol{\Sigma}_*$ will be independent of $\boldsymbol{\Sigma}_{1|*}$ and $\boldsymbol{T}|\boldsymbol{\Sigma}_{1|*}$ when $\nu > p$. This implies, by Definition 11, Appendix A.3, that the prior distribution in Expression (65) can be written as

$$f(\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}) = f(\boldsymbol{\Sigma}_*) \times f(\boldsymbol{T}|\boldsymbol{\Sigma}_{1|*}) \times f(\boldsymbol{\Sigma}_{1|*}).$$

The Hierarchical Bayesian approximate fluid flow simulator will, as mentioned above, only involve fluid flow simulation on a coarse scale. For notational convenience, let the ensemble on a coarse scale at timestep t be written as

$$\boldsymbol{L}_t = \left\{ \boldsymbol{x}_t^{*,1}, \ldots, \boldsymbol{x}_t^{*,n_s} \right\},$$

where $\boldsymbol{x}_t^{*,i} = \bar{\omega}_{\Delta_{t-1}}(\boldsymbol{x}_{t-1}^{*,i})$, is assumed to approximately follow a Gaussian distribution. Thus

$$\boldsymbol{x}_t^{*,i} \sim \text{Gauss}_g(\boldsymbol{B}\boldsymbol{\mu}_t + \boldsymbol{\mu}_{\boldsymbol{\epsilon}t}^*, \boldsymbol{B}\boldsymbol{\Sigma}_t\boldsymbol{B}^T + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^*),$$

for all ensemble members i. Using this ensemble, we can estimate the expectation, and covariance matrix by

$$\bar{\boldsymbol{x}}_t^* = \frac{1}{n_s} \sum_{i=1}^{n_s} \boldsymbol{x}_t^{*,i},$$

and

$$\boldsymbol{S}_t = \frac{1}{n_s} \sum_{i=1}^{n_s} (\boldsymbol{x}_t^{*,i} - \bar{\boldsymbol{x}}_t^*)(\boldsymbol{x}_t^{*,i} - \bar{\boldsymbol{x}}_t^*)^T.$$

Note that this corresponds to the maximum likelihood estimators of the mean value and covariance matrix of a multivariate Gaussian random variable (Johnson & Wichern 2002).

As shown in Appendix C, again omitting the subscript t, the posterior distribution of $\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}$ given $\boldsymbol{\Theta}$ and $\boldsymbol{L}$, $f(\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Theta}, \boldsymbol{L})$, will be Gaussian with parameters

$$\boldsymbol{\mu}_{\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\bar{\boldsymbol{x}}^*} = \begin{bmatrix} \boldsymbol{\eta} + \frac{n_s}{n_s+\zeta}\boldsymbol{T}(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*) \\ \boldsymbol{\eta}^* + \frac{n_s}{n_s+\zeta}(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{\Theta},\, \boldsymbol{L}} \\ \boldsymbol{\mu}_{\boldsymbol{x}^*|\boldsymbol{\Theta},\, \boldsymbol{L}} \end{bmatrix} \tag{69}$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\bar{\boldsymbol{x}}^*} = \frac{1}{\zeta}\left(\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}} - \frac{n_s}{n_s+\zeta} \begin{bmatrix} \boldsymbol{T}\boldsymbol{\Sigma}_*\boldsymbol{T}^T & \boldsymbol{T}\boldsymbol{\Sigma}_* \\ \boldsymbol{\Sigma}_*\boldsymbol{T}^T & \boldsymbol{\Sigma}_* \end{bmatrix} \right). \tag{70}$$

In addition, $\boldsymbol{\Sigma}_*$ has the inverted Wishart distribution with parameters

$$\hat{\boldsymbol{\Psi}}_*^{-1} = n_s\boldsymbol{S} + \boldsymbol{\Psi}_* + \frac{n_s\zeta}{n_s+\zeta}(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)^T, \tag{71}$$

and $\nu^* = \nu + n_s - u$ degrees of freedom. The posteriori distributions of $\boldsymbol{\Sigma}_{1|*}$ and $\boldsymbol{T}$ will be equal to the prior distributions.

Finally, the predictive distribution of $\boldsymbol{\mathfrak{X}}$, given the covariance matrices $\boldsymbol{\Theta}$ and the ensemble $\boldsymbol{L}$, $f(\boldsymbol{\mathfrak{X}}|\boldsymbol{\Theta}, \boldsymbol{L})$, will be Gaussian with expected value

$$\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}|\boldsymbol{\Theta}\ \boldsymbol{L}} = \boldsymbol{\mu}_{\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}|\bar{\boldsymbol{x}}^*}, \tag{72}$$

and covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}|\boldsymbol{\Theta}\ \boldsymbol{L}} = \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} + \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}|\bar{\boldsymbol{x}}^*}. \tag{73}$$

Then, by Result 2, Appendix A.2,

$$\begin{bmatrix} \boldsymbol{x}|\boldsymbol{\Theta}, \boldsymbol{L} \\ \boldsymbol{x}^*|\boldsymbol{\Theta}, \boldsymbol{L} \end{bmatrix} \sim \text{Gauss}_u \left(\begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{\Theta},\ \boldsymbol{L}} \\ \boldsymbol{\mu}_{\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{x}|\boldsymbol{\Theta},\ \boldsymbol{L}} & \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}} \\ \boldsymbol{\Sigma}^T_{\boldsymbol{x}\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}} & \boldsymbol{\Sigma}_{\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}} \end{bmatrix} \right), \tag{74}$$

where $\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{\Theta},\ \boldsymbol{L}}$ and $\boldsymbol{\mu}_{\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}}$ are given in Expression (69) and

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{x}|\boldsymbol{\Theta},\ \boldsymbol{L}} &= \left(1 + \frac{1}{\zeta}\right) \boldsymbol{\Sigma}_{1|*} + \left(1 - \frac{n_s}{\zeta(n_s + \zeta)}\right) \boldsymbol{T}\boldsymbol{\Sigma}_* \boldsymbol{T}^T, \\ \boldsymbol{\Sigma}_{\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}} &= \frac{n_s + 2\zeta}{n_s + \zeta} \boldsymbol{\Sigma}_*, \\ \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}} &= \frac{n_s + 2\zeta}{n_s + \zeta} \boldsymbol{T}\boldsymbol{\Sigma}_*. \end{aligned} \tag{75}$$

Moreover,

$$\boldsymbol{x}|\boldsymbol{x}^*, \boldsymbol{\Theta}, \boldsymbol{L} \sim \text{Gauss}_u \left(\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{x}^*}, \boldsymbol{\Sigma}_{\boldsymbol{x}|\boldsymbol{x}^*} \right),$$

where

$$\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{x}^*} = \boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{\Theta},\ \boldsymbol{L}} + \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}} (\boldsymbol{x}^* - \boldsymbol{\mu}_{\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}})$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{x}|\boldsymbol{x}^*} = \boldsymbol{\Sigma}_{\boldsymbol{x}|\boldsymbol{\Theta},\ \boldsymbol{L}} - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}} \boldsymbol{\Sigma}^T_{\boldsymbol{x}\boldsymbol{x}^*|\boldsymbol{\Theta},\ \boldsymbol{L}}.$$

By the same argument as in Section 6, this implies that downscaling of the reservoir properties can be done by

$$\boldsymbol{x} = \boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{x}^*} + \tilde{\boldsymbol{\epsilon}}^*, \tag{76}$$

where

$$\tilde{\boldsymbol{\epsilon}}^* \sim \text{Gauss}_u(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{x}|\boldsymbol{x}^*}).$$

We will now summarise the steps above, by outlining the algorithm for the approximate fluid flow simulator used in the HScCEnKF. As in Section 6, n_s denotes the size of the ensemble. The pseudo code for all of the subroutines in this algorithm can be found in Appendix D.

Algorithm 4: Approximate Fluid Flow Simulation In A Hierarchical Bayesian Setting

for $i = 1$ *to* n_s **do**
  Generate $\boldsymbol{x}_0^{c,i} \sim f(\boldsymbol{x}_0|\boldsymbol{x}_0^o)$
Generate $\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} \sim W_p^{-1}(\boldsymbol{\Psi}_{\boldsymbol{\mathfrak{X}}}{}^{-1}, \nu)$
for $t = 1$ *to* T **do**
  $\left[\boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}, \boldsymbol{\Sigma}_*\right] = \textbf{Bartlett}(\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}})$
  for $i = 1$ *to* n_s **do**
    $\boldsymbol{x}_{t-1}^{*,i} = \nu^*(\boldsymbol{x}_{t-1}^{c,i})$
    Generate $\boldsymbol{x}_t^{*,i} = \bar{\omega}_{\Delta_{t-1}}(\boldsymbol{x}_{t-1}^{*,i})$
  for $i = 1$ *to* n_s **do**
    $\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} = \textbf{getPosteriorCovMat}(\boldsymbol{x}_t^*, n, \nu, \zeta, \boldsymbol{T}, \boldsymbol{\Sigma}_{1|*}, \boldsymbol{\eta}_t^*, \boldsymbol{\Psi}_{*,t}, u)$
    $\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}t} = \textbf{getPosteriorMu}(\boldsymbol{x}_t^*, n_s, \zeta, \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}t}, \boldsymbol{\eta}_{\boldsymbol{\mathfrak{X}}t}, u, g)$
    $\boldsymbol{x}^t = \textbf{downScale}(\boldsymbol{x}_t^{*,i}, \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}t}, \boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}t}, u, g)$

Remark. In this algorithm we have assumed conditional independence between $\boldsymbol{x}_t^i$ and $\boldsymbol{x}_t^{*,j}, \quad \forall t,\ j \neq i$.

Remark. When the number of ensemble members is smaller than the dimension of the coarse scale reservoir, the posterior covariance matrix $\boldsymbol{\Sigma}_{*t} \sim f(\boldsymbol{\Sigma}_*|\boldsymbol{\Theta},\ \boldsymbol{L})$ might become ill-conditioned, meaning sensitive to roundoff error (Strang 1988). As a result of this unwanted feature, the posterior covariance matrices in Expression (75) might become negative definite, due to multiple matrix multiplications and inversion, of ill-conditioned matrices, which in theory should not happen when we select a positive definite prior covariance matrix.

A solution to this problem is to instead use the posterior mean of the covariance matrix given in Theorem 4, Appendix C, and use this in the further computations.

8 Hierarchical Scale-Corrected Ensemble Kalman Filter

In this section we present the methology and simplification used when the Hierarchical Scale-Corrected Ensemble Kalman Filter (HScCEnKF) is applied on a reservoir history matching problem containing both observed production and 4-D seismic.

Consider the hierarchical Bayesian model in Figure 10, where we as in Sections 6 and 7.3.2 assume that

$$
\begin{aligned}
\boldsymbol{\mathfrak{X}}_t &\sim f(\boldsymbol{\mathfrak{X}}_t|\boldsymbol{x}_0^o,\ldots,\boldsymbol{x}_{t-1}^o,\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}t},\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}t}) \sim \text{Gauss}_g(\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}t},\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}t}),\\
\boldsymbol{r}_t^{*q} &\sim f(\boldsymbol{r}_t^*|\boldsymbol{x}_0^o,\ldots,\boldsymbol{x}_{t-1}^o,\boldsymbol{q}_t^o,\boldsymbol{\mu}_{\boldsymbol{q}_t}^{\boldsymbol{x}^*},\boldsymbol{\Sigma}_{\boldsymbol{q}_t}^{\boldsymbol{x}^*}) \in \mathbb{R}^{n_{r^*}\times 1},\\
\boldsymbol{q}_t^{*q} &\sim f(\boldsymbol{q}_t^*|\boldsymbol{x}^o,\ldots,\boldsymbol{x}_{t-1}^o,\boldsymbol{q}_t^o,\boldsymbol{\mu}_{\boldsymbol{q}t}^{\boldsymbol{x}^*},\boldsymbol{\Sigma}_{\boldsymbol{q}_t}^{\boldsymbol{x}^*}) \in \mathbb{R}^{n_{q^*}\times 1},\\
\boldsymbol{r}_t^{q} &\sim f(\boldsymbol{r}_t|\boldsymbol{r}_t^{*q},\boldsymbol{\mu}_{\boldsymbol{r}_t}^{\boldsymbol{r}^*\boldsymbol{q}},\boldsymbol{\Sigma}_{\boldsymbol{r}_t}^{\boldsymbol{r}^*\boldsymbol{q}}) \in \mathbb{R}^{n_r\times 1},\\
\boldsymbol{r}_t^{qd} &\sim f(\boldsymbol{r}_t|\boldsymbol{r}_t^{q},\boldsymbol{d}_t^o,\boldsymbol{\Sigma}_{\boldsymbol{r}_t}^{\boldsymbol{r}^*\boldsymbol{q}}) \in \mathbb{R}^{n_r\times 1},\\
\boldsymbol{\mu}_{\boldsymbol{q}}^{\boldsymbol{q}^*} &= \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{q}t} \\ \boldsymbol{\mu}_{\boldsymbol{q}^*t} \end{bmatrix} \sim \text{Gauss}_{n_q+n_{q^*}}\left(\begin{bmatrix} \boldsymbol{\eta}_{\boldsymbol{q}t} \\ \boldsymbol{\eta}_{\boldsymbol{q}^*t} \end{bmatrix}, \frac{1}{\zeta_q}\begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{q}_t} & \boldsymbol{\Sigma}_{\boldsymbol{q}\boldsymbol{q}^*t} \\ \boldsymbol{\Sigma}_{\boldsymbol{q}^*\boldsymbol{q}_t} & \boldsymbol{\Sigma}_{\boldsymbol{q}^*t} \end{bmatrix}\right),\\
\boldsymbol{\Sigma}_{\boldsymbol{q}_t}^{\boldsymbol{x}^*} &= \left[\begin{array}{c|cc} \boldsymbol{\Sigma}_{\boldsymbol{q}_t} & \boldsymbol{\Sigma}_{\boldsymbol{q}\boldsymbol{r}^*t} & \boldsymbol{\Sigma}_{\boldsymbol{q}\boldsymbol{q}^*t} \\ \hline \boldsymbol{\Sigma}_{\boldsymbol{r}^*\boldsymbol{q}_t} & \boldsymbol{\Sigma}_{\boldsymbol{r}^*t} & \boldsymbol{\Sigma}_{\boldsymbol{r}^*\boldsymbol{q}^*t} \\ \boldsymbol{\Sigma}_{\boldsymbol{q}^*\boldsymbol{q}_t} & \boldsymbol{\Sigma}_{\boldsymbol{q}^*\boldsymbol{r}^*t} & \boldsymbol{\Sigma}_{\boldsymbol{q}^*t} \end{array}\right] \sim W_{n_q+n_{r^*}+n_{q^*}}^{-1}(\boldsymbol{\Psi}_{\boldsymbol{q}\ t}^{\boldsymbol{x}^*\ -1},\nu_q^{x^*}),\\
\boldsymbol{\mu}_{\boldsymbol{r}\ t}^{\boldsymbol{r}^*\boldsymbol{q}} &= \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{r}t} \\ \boldsymbol{\mu}_{\boldsymbol{r}^*|\boldsymbol{q}^o t} \end{bmatrix} \sim \text{Gauss}_{n_r+n_{r^*}}\left(\boldsymbol{\eta}_{\boldsymbol{r}}^{\boldsymbol{r}^*}, \frac{1}{\zeta_r}\boldsymbol{\Sigma}_{\boldsymbol{r}}^{\boldsymbol{r}^*\boldsymbol{q}}\right),\\
\boldsymbol{\Sigma}_{\boldsymbol{r}}^{\boldsymbol{r}^*\boldsymbol{q}} &= \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{r}_t} & \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{r}^*\boldsymbol{q}_t} \\ \boldsymbol{\Sigma}_{\boldsymbol{r}^*\boldsymbol{q}\boldsymbol{r}_t} & \boldsymbol{\Sigma}_{\boldsymbol{r}^*\boldsymbol{q}_t} \end{bmatrix} = W_{n_r+n_{r^*}}^{-1}\left(\boldsymbol{\Psi}_{\boldsymbol{r}_t}^{\boldsymbol{r}^*\boldsymbol{q}\ -1}, \nu_{\boldsymbol{r}_t}^{\boldsymbol{r}^*\boldsymbol{q}}\right),
\end{aligned}
$$

and $\boldsymbol{d}_t^o$ and $\boldsymbol{q}_t^o$ follow the same relationship as in Expressions (41) and (42). As in Section 7.3.2, $\boldsymbol{\eta}_{(\bullet)}$, $\boldsymbol{\Psi}_{(\bullet)}$, $\zeta_{(\bullet)}$ and $\nu_{(\bullet)}$ are known model parameters. Also note that $\boldsymbol{\mu}_{\boldsymbol{r}^*|\boldsymbol{q}^o t}$ and $\boldsymbol{\Sigma}_{\boldsymbol{r}^*|\boldsymbol{q}^o t}$ are similar to Expressions (47) and (48), replacing the fine scale expectation vectors and covariance matrices with their coarse scale equivalents. From now on we will skip the subscript t in our notation.

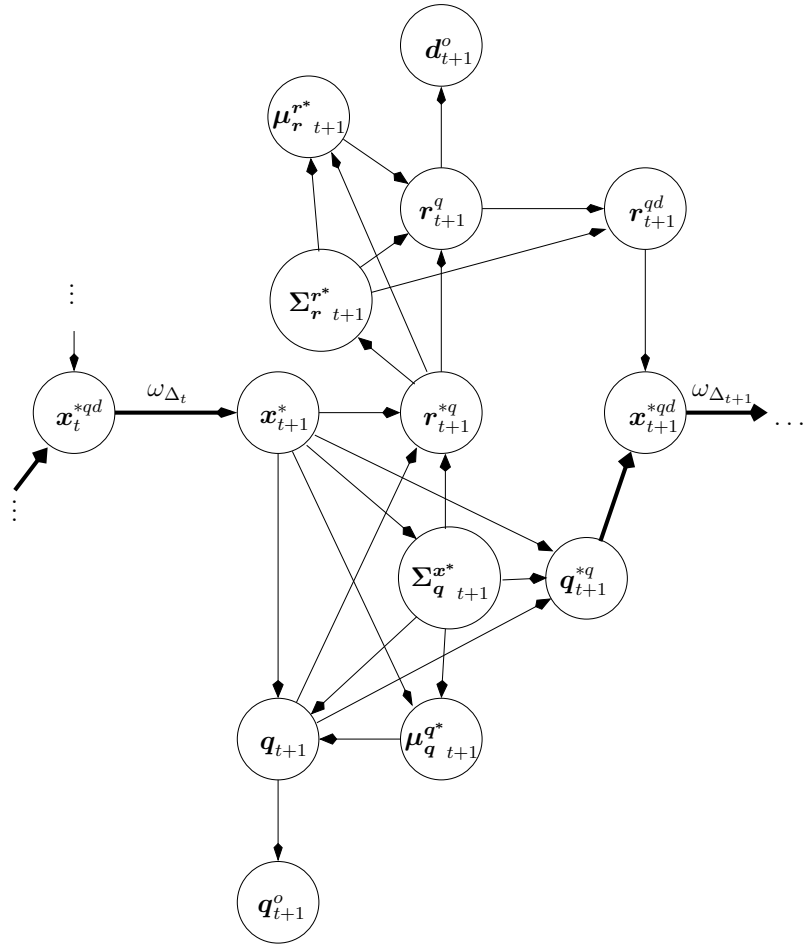


Figure 10: The figure shows the Directed Acyclic Graph (DAG) of the hierarchical Bayesian model considered. Single arrows mean that there is a stochastic relationship between the variables, while thick arrows imply that there is a deterministic relationship.

8.1 Assimilation With Observed Production

Rewriting $\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{x}^*}$ as

$$\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{x}^*} = \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{\boldsymbol{q}} & \boldsymbol{\Sigma}_{\boldsymbol{q}\boldsymbol{x}^*} \\ \hline \boldsymbol{\Sigma}_{\boldsymbol{x}^*\boldsymbol{q}} & \boldsymbol{\Sigma}_{\boldsymbol{x}^*} \end{array}\right],$$

we see that the production on a fine scale can be found through the downscaling procedure in Section 7.3.2, where we only consider $\boldsymbol{q}$ and $\boldsymbol{x}^*$. That is

$$\left[\boldsymbol{q}|\boldsymbol{q}^*, \boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*}, \boldsymbol{L}\right] = \boldsymbol{\mu}_{\boldsymbol{q}|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}} + \boldsymbol{\Sigma}_{\boldsymbol{q}^*|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{q}^*|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}} \left(\boldsymbol{q}^* - \boldsymbol{\mu}_{\boldsymbol{q}^*|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}}\right) + \boldsymbol{\epsilon}_{\boldsymbol{q}|\boldsymbol{q}^*}, \tag{77}$$

where

$$\boldsymbol{\epsilon}_{\boldsymbol{q}|\boldsymbol{q}^*} \sim \text{Gauss}_{n_q}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{q}|\boldsymbol{q}^*,\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}}\right),$$

$$\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*} = \left\{\boldsymbol{\Sigma}_{\boldsymbol{q}\ 1|*}^{\boldsymbol{x}^*}, \boldsymbol{T}_{\boldsymbol{q}}^{\boldsymbol{x}^*}, \boldsymbol{\Sigma}_{\boldsymbol{q}\ *}^{\boldsymbol{x}^*}\right\},$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{q}|\boldsymbol{q}^*,\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}} = \boldsymbol{\Sigma}_{\boldsymbol{q}|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}} - \boldsymbol{\Sigma}_{\boldsymbol{q}\boldsymbol{q}^*|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{q}^*|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}} \boldsymbol{\Sigma}^{T}_{\boldsymbol{q}\boldsymbol{q}^*|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}}.$$

Assimilating the coarse scale ensemble using the fine scale production can therefore be done by

$$\left[\begin{array}{c} \boldsymbol{r}^{q*} \\ \boldsymbol{q}^{q*} \end{array}\right] = \left[\begin{array}{c} \boldsymbol{r}^{q*} + \boldsymbol{\Sigma}_{\boldsymbol{r}^*\boldsymbol{q}|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{q}|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}} \left(\boldsymbol{q}^o + \boldsymbol{\epsilon}_{\boldsymbol{q}}{}^o - \boldsymbol{q}\right) \\ \boldsymbol{q}^{q*} + \boldsymbol{\Sigma}_{\boldsymbol{q}^*\boldsymbol{q}|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{q}|\boldsymbol{\Theta}_{\boldsymbol{q}}^{\boldsymbol{x}^*},\boldsymbol{L}} \left(\boldsymbol{q}^o + \boldsymbol{\epsilon}_{\boldsymbol{q}}{}^o - \boldsymbol{q}\right) \end{array}\right], \tag{78}$$

where $\boldsymbol{\epsilon}_{\boldsymbol{q}}{}^o \sim \text{Gauss}_{n_q}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{q}}{}^o)$. Note that this method corresponds to the method used in Lødøen & Omre (2005), which was proposed as an alternative to downscaling $\boldsymbol{r}^*$ to $\boldsymbol{r}$, assimilating $\boldsymbol{r}$, and finally upscaling $\boldsymbol{r}^q$ to $\boldsymbol{r}^{q*}$.

8.2 Assimilation With Seismic Observations

If we consider the mean of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}|\boldsymbol{L}$, given in Theorem 4, Appendix C, we see that the term $\text{tr}(\tilde{\boldsymbol{S}}\boldsymbol{\Psi}_*^{-1})$, where $\text{tr}(\boldsymbol{A})$ denotes the sum of the diagonal elements of the matrix $\boldsymbol{A}$, will be a multiplicative factor in the fine scale posterior covariance matrix. This implies that in the case where one of the reservoir properties has a very small prior variance, while another has a very large variance, the posterior covariance matrix of the reservoir property with the *a priori* small covariance will become unphysically large when we consider both properties simultaneously in the downscaling. Thus we choose to consider log-permeability/porosity, saturation and pressure as three variables assumed to be conditionally independent of each other in the downscaling of the reservoir. For notational convenience, let $\mathcal{D}_{\boldsymbol{r}}$ be the set containing these three new variables.

Consider the fine scale reservoir property $\boldsymbol{y} \in \mathcal{D}_{\boldsymbol{r}}$, and the coarse scale reservoir property $\boldsymbol{y}^* \in \mathcal{D}_{\boldsymbol{r}^*}$. Realisations from the posterior distribution

$$f(\boldsymbol{y}|\boldsymbol{y}^{*q}, \boldsymbol{\Theta}_{\boldsymbol{y}}^{\boldsymbol{y}^*\boldsymbol{q}}, \boldsymbol{L}_{\boldsymbol{y}}{}^{*q}),$$

can then be generated as follows:

$$\begin{aligned}\boldsymbol{y^q} &= \left[\boldsymbol{y}|\boldsymbol{y}^{*q}, \boldsymbol{\Sigma}_{\boldsymbol{y}}^{\boldsymbol{y}^*}, \boldsymbol{L_y}^{*q}\right] \\ &= \boldsymbol{\mu}_{\boldsymbol{y}|\boldsymbol{y}^{*q},\boldsymbol{\Theta}_{\boldsymbol{y}}^{\boldsymbol{y}^{*q}},\boldsymbol{L_y}^{*q}} + \boldsymbol{\Sigma}_{\boldsymbol{y}\boldsymbol{y}^{*q}|\boldsymbol{\Theta}_{\boldsymbol{y}}^{\boldsymbol{y}^{*q}},\boldsymbol{L_y}^{*q}} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{y}^{*q}|\boldsymbol{\Theta}_{\boldsymbol{y}}^{\boldsymbol{y}^{*q}},\boldsymbol{L_y}^{*q}} \\ &\quad \times \left(\boldsymbol{y}^{*q} - \boldsymbol{\mu}_{\boldsymbol{y}^{*q}|\boldsymbol{\Theta}_{\boldsymbol{y}}^{\boldsymbol{y}^{*q}},\boldsymbol{L_y}^{*q}}\right) + \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{y}|\boldsymbol{y}^{*q}}, \end{aligned} \tag{79}$$

where

$$\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{y}|\boldsymbol{y}^{*q}} \sim \text{Gauss}_{n_y}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{y}|\boldsymbol{y}^{*q},\boldsymbol{\Theta}_{\boldsymbol{y}}^{\boldsymbol{y}^{*q}},\boldsymbol{L_y}^{*q}}\right),$$

and

$$\boldsymbol{\Theta}_{\boldsymbol{y}}^{\boldsymbol{y}^{*q}} = \left\{\boldsymbol{\Sigma}_{\boldsymbol{y}}^{\boldsymbol{y}^{*q}}{}_{(1|*)}, \boldsymbol{T}_{\boldsymbol{y}}^{\boldsymbol{y}^{*q}}, \boldsymbol{\Sigma}_{\boldsymbol{y}}^{\boldsymbol{y}^{*q}}{}_{(*)}\right\}.$$

Here the posterior expectations and covariance matrices have the same structure as the posterior expectations covariance matrices given in Sections 6 and 7.3.2, when only considering the reservoir properties $\boldsymbol{y}$ in the corresponding coarse scale ensemble $\boldsymbol{L_y}^{*q}$.

Inversion of seismic data, can then be done on each trace, by the same procedure as in Algorithm 3 using posterior block covariance matrices instead of the empirical estimates.

We will now conclude this section by outlining the complete method used in the Hierarchical Ensemble Kalman Filter when applied on a reservoir containing both observed production and 4-D seismic in Algorithm 5. Note that the pseudo code for all subroutines in this algorithm can be found in Appendix D. Details concerning the seismic inversion can be found in Sections 4 and 6.

Algorithm 5: Hierarchical Scale-Corrected Ensemble Kalman Filter Applied on a Reservoir Containing both Observed Production and 4-D Seismic.

for $i = 1$ *to* n_s **do**
  Generate $\boldsymbol{r}_0^{qd,i} \sim f(\boldsymbol{r}_0|\boldsymbol{d}_0^o)$
$\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{x}^*} \sim W_{n_q+g}^{-1}(\boldsymbol{\Psi}_{\boldsymbol{q}}^{\boldsymbol{x}^*\,-1}, \nu_q^{x^*})$
$\left[\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{x}^*}{}_{1|*}, \boldsymbol{T}_{\boldsymbol{q}}^{\boldsymbol{x}^*}, \boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{x}^*}{}_{*}\right] = \textbf{Bartlett}(\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{x}^*}, n_q, g)$
$\boldsymbol{\Sigma}_{\boldsymbol{r}}^{\boldsymbol{r}^*\boldsymbol{q}}{}_{m} \sim W_{n_{r_m}+n_{r_m^*}}^{-1}(\boldsymbol{\Psi}_{\boldsymbol{r}}^{\boldsymbol{r}^*\boldsymbol{q}\,-1}, \nu_r^{r^*q})$
$\left[\boldsymbol{\Sigma}_{\boldsymbol{r}}^{\boldsymbol{r}^*\boldsymbol{q}}{}_{1|*,m}, \boldsymbol{T}_{\boldsymbol{r}}^{\boldsymbol{r}^*\boldsymbol{q}}{}_{m}, \boldsymbol{\Sigma}_{\boldsymbol{r}}^{\boldsymbol{r}^*\boldsymbol{q}}{}_{*,m}\right] = \textbf{Bartlett}(\boldsymbol{\Sigma}_{\boldsymbol{r}}^{\boldsymbol{r}^*\boldsymbol{q}}, n_{r_m}, n_{r_{m^*}})$
for $t = 1$ *to* T **do**
  for $i = 1$ *to* n_s **do**
    if $t-1 \in T_{d^o}$ **then**
      $\boldsymbol{r}_{t-1}^{*qd,i} = \nu^*(\boldsymbol{r}_{t-1}^{qd,i})$
    $\boldsymbol{x}_t^{*,i} = \bar{\omega}_{\Delta_{t-1}}(\boldsymbol{r}_{t-1}^{*qd,i}, \boldsymbol{q}_{t-1}^{q*,i})$
  for $i = 1$ *to* n_s **do**
    $\boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{x}^*}{}_{t} =$
    $\textbf{getPosteriorCovMat}(\boldsymbol{L}_t, n_s, \nu_q^{x^*}, \zeta_q, \boldsymbol{T}_{\boldsymbol{q}}^{\boldsymbol{x}^*}, \boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{x}^*}{}_{1|*}, \boldsymbol{\eta}_{\boldsymbol{q}}^{\boldsymbol{x}^*}{}_{t}, \boldsymbol{\Psi}_{\boldsymbol{q}}^{\boldsymbol{x}^*}{}_{*,t}, n_q)$
    $\boldsymbol{\mu}_{\boldsymbol{q}}^{\boldsymbol{x}^*}{}_{t} = \textbf{getPosteriorMu}(\boldsymbol{L}_t, n_s, \zeta_q, \boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{x}^*}{}_{t}, \boldsymbol{\eta}_{\boldsymbol{q}}^{\boldsymbol{x}^*}{}_{t}, n_q, g)$
    $\boldsymbol{q}_t^i = \textbf{downScale}(\boldsymbol{x}_t^{*,i}, \boldsymbol{\Sigma}_{\boldsymbol{q}}^{\boldsymbol{x}^*}{}_{t}, \boldsymbol{\mu}_{\boldsymbol{q}}^{\boldsymbol{x}^*}{}_{t}, n_q, g)$
    $\boldsymbol{r}_t^{*q,i} =$
    $\textbf{assimilateProduction}(\boldsymbol{q}_t^i, \boldsymbol{r}_t^{*,i}, \boldsymbol{q}_t^o, \boldsymbol{D}_{\boldsymbol{q}t}, \boldsymbol{\Sigma}_{\boldsymbol{r}^*t}^{\boldsymbol{q}}, \boldsymbol{\Sigma}_{\boldsymbol{q}t}^{o}, n_{r^*}, n_q)$
    $\boldsymbol{q}_t^{*q,i} =$
    $\textbf{assimilateProduction}(\boldsymbol{q}_t^i, \boldsymbol{q}_t^{*,i}, \boldsymbol{q}_t^o, \boldsymbol{D}_{\boldsymbol{q}t}, \boldsymbol{\Sigma}_{\boldsymbol{q}^*t}^{\boldsymbol{q}}, \boldsymbol{\Sigma}_{\boldsymbol{q}t}^{o}, n_{q^*}, n_q)$
  if $t \in T_{d^o}$ **then**
    for i *in* 1 *to* n_s **do**
      for *all* $\boldsymbol{y} \in \mathcal{D}_{\boldsymbol{r}}$ **do**
        $\boldsymbol{\Sigma}_{\boldsymbol{y}}^{\boldsymbol{y}^*\boldsymbol{q}}{}_{j,t} =$
        $\textbf{getPosteriorCovMat}(\boldsymbol{L}_{\boldsymbol{y}t}^{*q}, n_s, \nu_y^{y^*q}, \zeta_y, \boldsymbol{T}_{\boldsymbol{y}}^{\boldsymbol{y}^*\boldsymbol{q}}, \boldsymbol{\Sigma}_{\boldsymbol{y}}^{\boldsymbol{y}^*\boldsymbol{q}}{}_{1|*}, \boldsymbol{\eta}_{\boldsymbol{y}}^{\boldsymbol{y}^*\boldsymbol{q}}{}_{t}, \boldsymbol{\Psi}_{\boldsymbol{y}}^{\boldsymbol{y}^*\boldsymbol{q}}{}_{*,t}, n_y)$
        $\boldsymbol{\mu}_{\boldsymbol{y}}^{\boldsymbol{y}^*\boldsymbol{q}}{}_{t} =$
        $\textbf{getPosteriorMu}(\boldsymbol{L}_{\boldsymbol{y}}{}^{*q}, n_s, \zeta_y, \boldsymbol{\Sigma}_{\boldsymbol{y}}^{\boldsymbol{y}^*\boldsymbol{q}}{}_{t}, \boldsymbol{\eta}_{\boldsymbol{y}}^{\boldsymbol{y}^*\boldsymbol{q}}{}_{t}, n_y, n_{n_{y^*}})$
        $\boldsymbol{y}_t^{q,i} = \textbf{downScale}(\boldsymbol{y}_t^{*q,i}, \boldsymbol{\Sigma}_{\boldsymbol{y}}^{\boldsymbol{y}^*\boldsymbol{q}}{}_{t}, \boldsymbol{\mu}_{\boldsymbol{y}}^{\boldsymbol{y}^*\boldsymbol{q}}{}_{t}, n_y, n_{y^*})$
      $\boldsymbol{r}_t^{qd,i} = \textbf{seismicInversion}(\boldsymbol{r}_t^{q,i}, \boldsymbol{d}^o t, \boldsymbol{D}_{\boldsymbol{d}_t^o}, \boldsymbol{D}_{\boldsymbol{r}t},$
        $\boldsymbol{\mu}_{\boldsymbol{\epsilon}_{\boldsymbol{r}\boldsymbol{d}^o}}, \boldsymbol{\Sigma}_{\boldsymbol{r}}^{\boldsymbol{r}^*\boldsymbol{q}}{}_{t}, n_x \cdot n_y, n_\theta, S, \boldsymbol{\Sigma}_{\boldsymbol{\omega}t}, \boldsymbol{\Sigma}_{\boldsymbol{d}^o t})$
  else
    $\boldsymbol{r}_t^{*qd,i} = \boldsymbol{r}_t^{*q,i}, \ \forall i$

9 Synthetic Model

In the following we present a synthetic case study identical to the one used in Hegstad & Omre (2001) and Lødøen & Omre (2005). The reservoir which is inspired by the Troll field in the North Sea offshore Norway, has a size of $10^4 \times 10^4 \times 10^2$ feet^3. Further the reservoir has been discretised into $n = 50 \times 50 \times 15$ grid nodes, and we assume that the reservoir initially is fully saturated by oil. As can be seen in Figure 11, there are two horizontal production wells, and one vertical injection well in the reservoir.

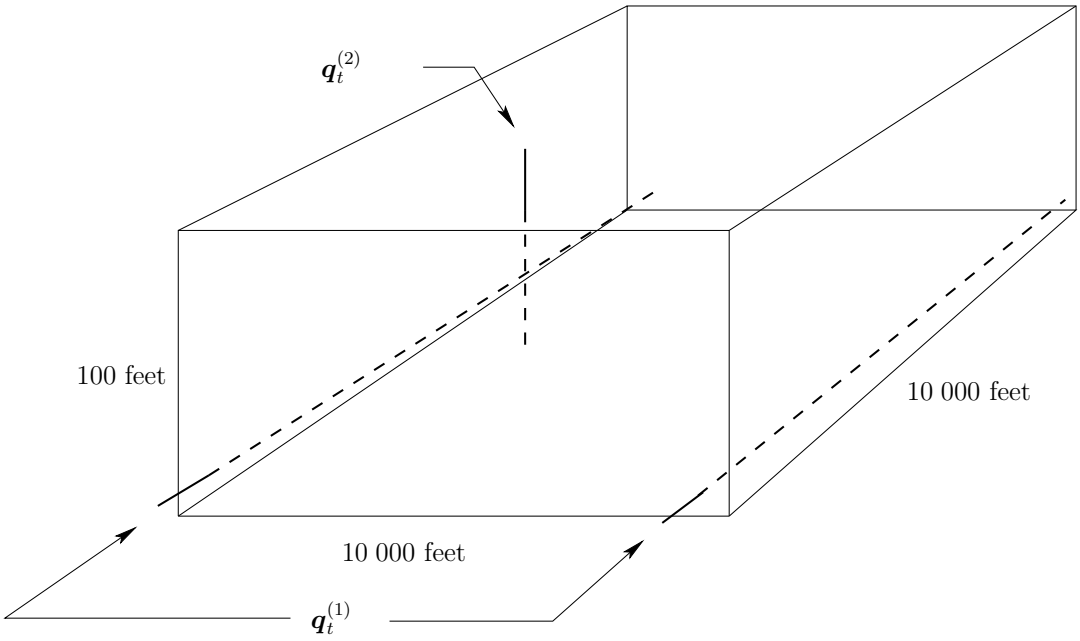


Figure 11: Outline of the well configuration of the reservoir.

In order to investigate some of the strengths and weaknesses of our method, we use a reference reservoir created by geoscientists familiar with the Troll field (Hegstad & Omre 2001, Omre & Lødøen 2004). Figure 12 displays vertical cross sections of the log permeability and porosity of the reference field. As we can see from this figure, the middle part of the reservoir contains higher permeability and porosity values than the lower and upper part.

The reference production $\boldsymbol{q}_t^o$, was found by running the fine scale simulator using the geology of the reference reservoir as input. Figure 13 displays the resulting reference production when the Eclipse 100 version 2005a simulator (GeoQuest 2004) is used as the fluid flow simulator ω. In this study we consider the oil production rates ($opr1$ and $opr2$), gas/oil ratio ($gor1$ and $gor2$) and bottom hole pressure ($bhp1$ and $bhp2$) in the production wells, and the bottom hole pressure in the injection well ($bhpi$). As we can see from this figure, the operating condition in each of the wells switches to a target bottom hole pressure of 4 100 psi, when the bhp reaches this value. Also note that the injected fluid is gas, and that the rate of the injection is 65 000 mscf/day. The reference

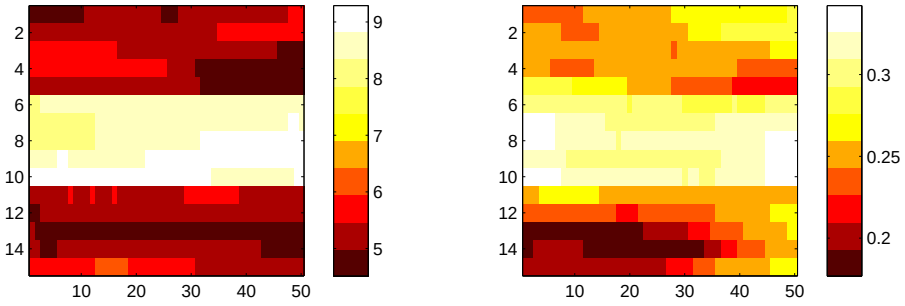


Figure 12: Vertical slice of log permeability and porosity of the reference reservoir.

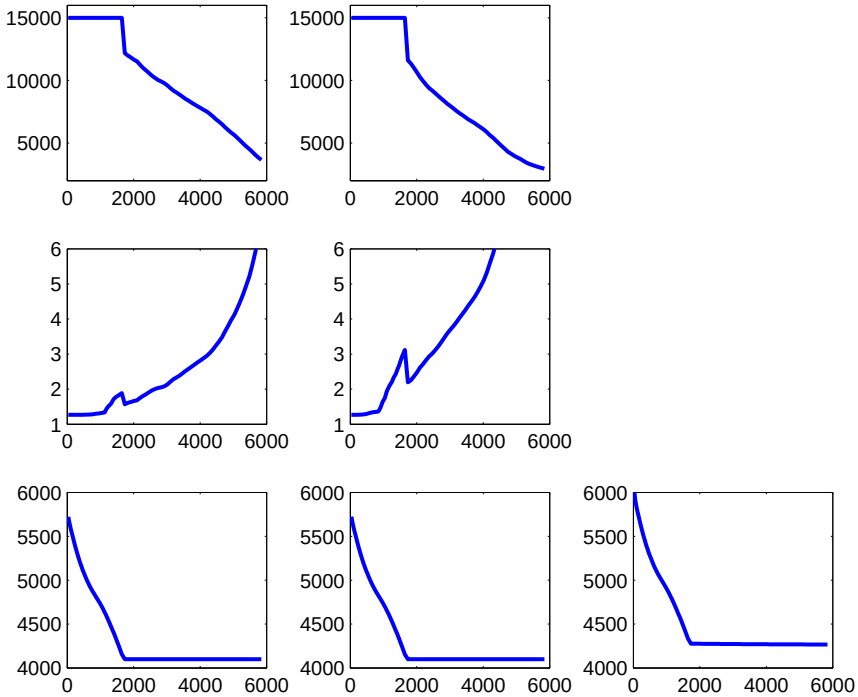


Figure 13: Reference production as a function of time [days]. The figure displays row-wise from top to bottom: $opr1$, $opr2$, $gor1$, $gor2$, $bhp1$, $bhp2$ and $bhpi$.

model was run for 5 844 days, collecting the reference saturation and pressure in each fine scale grid block every month. Synthetic seismic data was then found by using the reference reservoir as input in the nonlinear seismic forward model, described in Section 4.1, at 0, 300, 900 and 1 800 days. The geophysical parameters used in the forward model can be found in Table 1 and corresponds to the values found in Hoversten et al. (2006) and Hegstad & Omre (2001). In this Thesis we consider two AVO angles; $\theta_1 = 5°$ and $\theta_2 = 15°$. Other seismic parameters selected, were similar to the ones selected in Hegstad & Omre (2001).

Fixed Parameters		Regression Parameters	
Oil API	28.5	$K_{\max}$	17.0
Brine Salinity	0.07	$K_{\min}$	12.0
Gas gravity	0.59	$G_{\max}$	5.50
Temperature (°C)	65.0	$G_{\min}$	1.60
s_w	0.12	$\phi_{\max}$	0.10
		$\phi_{\min}$	0.40

Table 1: Geophysical model parameters and regression parameters used in seismic forward model.

Since this is a synthetic model, and no well-logs currently are available, we choose a linear model for the two functions in Expression (21),

$$K_{\text{sat2}} = K_{\max} + (\phi - \phi_{\min}) \cdot \frac{(K_{\min} - K_{\max})}{(\phi_{\max} - \phi_{\min})}$$
$$G_{\text{sat}} = G_{\max} + (\phi - \phi_{\min}) \cdot \frac{(G_{\min} - G_{\max})}{(\phi_{\max} - \phi_{\min})}.$$

The selected parameters above can be found in Table 1. We further selected the reference production and reservoir as the model parameters in the prior expectation $\boldsymbol{\eta}_{\boldsymbol{x}t}$, $\boldsymbol{\eta}_{\boldsymbol{q}t}$ and $\boldsymbol{\eta}_{\boldsymbol{r}t}$ respectively.

The correlation between each grid node in the fine scale reservoir was chosen using the same correlation function as in Hegstad & Omre (2001). Figure 14 displays the structure of the correlation matrix $\boldsymbol{\Psi}_{\boldsymbol{\rho}}$. As we can see from this figure, each lattice node has a correlation of about 0.1 to their two nearest vertical neighbours, while the correlation to their two nearest lateral neighbours is close to zero.

The prior covariance matrix for the reservoir $\boldsymbol{\Psi}_{\boldsymbol{r}}$, was then selected as

$$\boldsymbol{\Psi}_{\boldsymbol{r}t} = \boldsymbol{\Psi}_{\boldsymbol{\rho}} \otimes \begin{bmatrix} \sigma_\kappa^2 & \sigma_{\kappa\phi} & 0 & 0 \\ \sigma_{\kappa\phi} & \sigma_\phi^2 & 0 & 0 \\ 0 & 0 & \sigma_s^2 & 0 \\ 0 & 0 & 0 & \sigma_p^2 \end{bmatrix}, \quad \forall t$$

where the Kronecker product $\otimes$, is defined Appendix A.3. The prior covariance

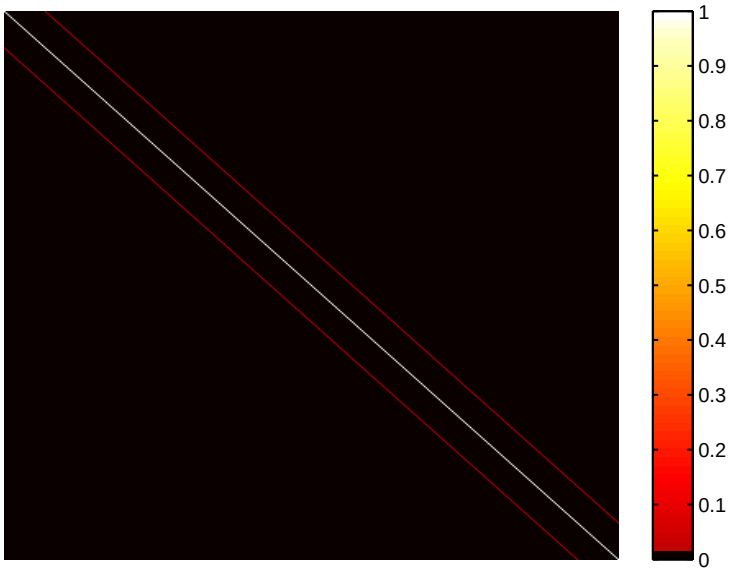


Figure 14: Structure of the correlation matrix used in the generation of the prior covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{r}}$.

matrix for the production was selected as

$$\boldsymbol{\Psi}_{\boldsymbol{q}_t} = \begin{bmatrix} \sigma^2_{opr1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{opr2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{gor1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{gor2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{bhp1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{bhp2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{bhpi} \end{bmatrix}, \quad \forall t.$$

The parameters $\sigma_{(\bullet)}$ can be found in Table 2.

Realisations from the prior distribution $f(\boldsymbol{r}_0|\boldsymbol{d}_0^o)$ can be obtained using Algorithm 1 in Section 4, selecting the same seismic model parameters as in Larsen et al. (2006), using the reference reservoir as the expected value $\boldsymbol{\mu}_{\boldsymbol{r}}$ and

$$\boldsymbol{\Sigma}_{\boldsymbol{r}} = \begin{bmatrix} 10 \cdot \boldsymbol{I}_n & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & 15 \cdot \boldsymbol{I}_n & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_n \end{bmatrix} \boldsymbol{\Psi}_{\boldsymbol{r}}.$$

Thus, assuming that there is a high uncertainty in the static reservoir properties log-permeability and porosity in the initial ensemble. Note that the covariance parameters are similar to the values found in Hegstad & Omre (2001), for log-permeability, porosity and the production data, while the covariance parameters

$\boldsymbol{\Psi_r}$		$\boldsymbol{\Psi_q}$	
σ^2_κ	0.55^2	σ^2_{opr1}	22 500
σ^2_ϕ	0.015^2	σ^2_{opr2}	22 500
σ^2_s	0.71^2	σ^2_{gor1}	0.5
σ^2_p	7.75^2	σ^2_{gor2}	0.5
$\sigma_{\kappa\phi}$	0.006	σ^2_{bhp1}	3 000
		σ^2_{bhp2}	3 000
		σ^2_{opri}	5 000

Table 2: Parameters used in the prior covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{r}t}$ and $\boldsymbol{\Psi}_{\boldsymbol{q}_t}$.

.

for saturation, pressure and the correlation between log-permeability and porosity are based on the reference model.

As in Lødøen & Omre (2005), the coarse scale fluid flow simulator $\bar{\omega}$, runs on a $n^* = 10 \times 10 \times 15$ lattice. In order to make the upscaling of the reservoir properties $\boldsymbol{r}$, a linear operation, we define the upscaling function $\nu^* : \mathbb{R}^{n_{\boldsymbol{r}}} \rightarrow \mathbb{R}^{n_{\boldsymbol{r}^*}}$ the following way:

- Permeability is mapped by geometric averaging,
- Porosity, saturation and pressure is mapped by arithmetic averaging.

This implies that

$$\boldsymbol{r}^* = \boldsymbol{A}\boldsymbol{r},$$

where $\boldsymbol{A} \in \mathbb{R}^{n_{\boldsymbol{r}^*} \times n_{\boldsymbol{r}}}$ is a sparse block diagonal matrix, with each element being either zero or 1/25. A reference coarse scale reservoir $\boldsymbol{r}^*$, and a reference coarse scale production, was then created by running the coarse scale fluid flow simulator with the upscaled static reservoir properties as input values.

As above, we selected the reference coarse scale production and reservoir as $\boldsymbol{\eta}_{\boldsymbol{q}t}^*$ and $\boldsymbol{\eta}_{\boldsymbol{r}t}^*$ respectively. Prior covariance matrices in the error terms of the upscaling, ${\boldsymbol{\Psi}_{\boldsymbol{\epsilon}_{\boldsymbol{r}}}}_t^*$ and ${\boldsymbol{\Psi}_{\boldsymbol{\epsilon}_{\boldsymbol{q}}}}_t^*$, were further selected as one tenth of their corresponding fine scale prior covariance matrices.

Empirical estimates of $Cov(\boldsymbol{q}^*)$ and $Cov(\boldsymbol{q}^*, \boldsymbol{r}^*)$ from the ensemble were also used as the prior covariance matrices $\boldsymbol{\Psi}_{\boldsymbol{q}\boldsymbol{q}^*t}$ and $\boldsymbol{\Psi}_{\boldsymbol{q}\boldsymbol{r}^*t}$, at each timestep t. The degrees of freedom in the prior distribution, were selected as $\nu_q^{x^*} = n_q + g + 2$ and $\nu_y^{y^*q} = n_{y_m^*} + n_{y_m} + 2$, $\forall \boldsymbol{y} \in \mathcal{D}_{\boldsymbol{r}}$, meaning that we made the prior distribution as flat as possible. Further we set the scaling parameters $\zeta_{(\bullet)}$ equal to 1.

Remark. If we assume that the number of grid nodes $n = o(10^5)$, Algorithm 4 would in the worst case require that we allocate $\mathcal{O}(10^{10})$ bytes of memory in the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{r}|\boldsymbol{r}^{*q}}$. Moreover a Cholesky decomposition of this matrix in order to generate $\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{r}|\boldsymbol{r}^{*q}} \sim \text{Gauss}_{n_r}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{r}|\boldsymbol{r}^{*q}})$ would require $o(10^{30})$ flops to complete (Golub & van Loan 1996). Here the notation $o(n)$ and $\mathcal{O}(n)$ means bounded above, and below n respectively (Cormen et al. 2001). In order to be able to run the Hierarchical Scale-Corrected Ensemble Kalman Filter on the fine scale reservoir, we therefore consider subsets of the reservoir in the downscaling

procedure described in Section 7.3.2. The subsets consists of $5 \times 5 \times 15$ fine scale grid blocks and $1 \times 1 \times 15$ corresponding coarse scale grid blocks. The motivation behind this approximation is that seismic inversion, as described in Algorithm 1, is performed trace by trace.

We also consider a model with $n = 10 \times 10 \times 15$ fine scale grid blocks and $n^* = 5 \times 5 \times 3$ coarse scale grid nodes, such that the HEnKF can be run on a 32−bit MATLAB system using Algorithm 5.

Note that the two fine scale models above from now on will be refered to as the large and small model respectively.

10 Results

In this section we consider the results obtained when we run a localised version of the Hierarchical Scale-Corrected Ensemble Kalman Filter (HScCEnKF) outlined in Section 8, conditioning on both production and 4-D seismic data, in the large model defined in Section 9, using an ensemble of $n_s = 100$. We also consider the results obtained on the small model running the HScCEnKF without any localisation scheme using an ensemble of $n_s = 20$, and the traditional Ensemble Kalman Filter (EnKF), conditioning on production data only, on the small model with an ensemble size of $n_s = 100$.

Note that the results have been obtained using the posterior mean of the covariance matrices instead of sampling from the posterior distributions. This was done in order to get stable and fast results without the concern of numerical errors. However, this issue will require a more thorough study.

Due to instabilities in the computation of the Jacobian used in the seismic inversion, very large values for the covariance matrix $\boldsymbol{\Sigma_m}$, was observed at certain seismic reflection times s. Since the expected values of P- and S-wave velocities involves taking the exponential of the these large values, we experienced unphysically large values for $\boldsymbol{\mu}_\alpha$ and $\boldsymbol{\mu}_\beta$. We therefore had to approximate these two expected values, by omitting $\boldsymbol{\Sigma_m}$ from the computation. Alternative solutions to this unwanted feature has not yet been investigated.

10.1 Hierarchical Scale-Corrected Ensemble Kalman Filter

Figure 15 shows the evolvement in time of a vertical slice of log-permeability and porosity for one ensemble member at the initial timestep, and after 300, 900 and 1 800 days of updating using the HScCEnKF. As expected, we can clearly see the added noise in the initial ensemble. Further, we see that the ensemble member is drawn towards the reference log-permeability and porosity as more production and seismic data becomes available for assimilation.

The estimated ensemble mean value of log-permeability and porosity at selected times, is shown in Figure 16. As we can see from this figure, the initial mean value of the porosity is close to the reference porosity. Further, we see that the assimilation of additional data in the model does not change the ensemble mean.

The same trend as in Figure 15 is seen when we look at the corresponding vertical slice of the estimated ensemble standard deviation for log-permeability and porosity, shown in Figure 17. Depending on the value of the prior model parameters, we experienced a layered structure in the estimated standard deviation of the porosity in the initial ensemble, generated using Algorithm 1. This feature, however, was removed as more data was assimilated into the ensemble. The position of these layers, corresponded to the locations where the P- and S-wave velocities were unphysically large. A primarily assumption is that this effect arise due to numerical instability in the computation of the Jacobian matrix. However, a more thorough stability analysis will have to be made in a future study.

To check if there might be inconsistencies between the updated static proper-

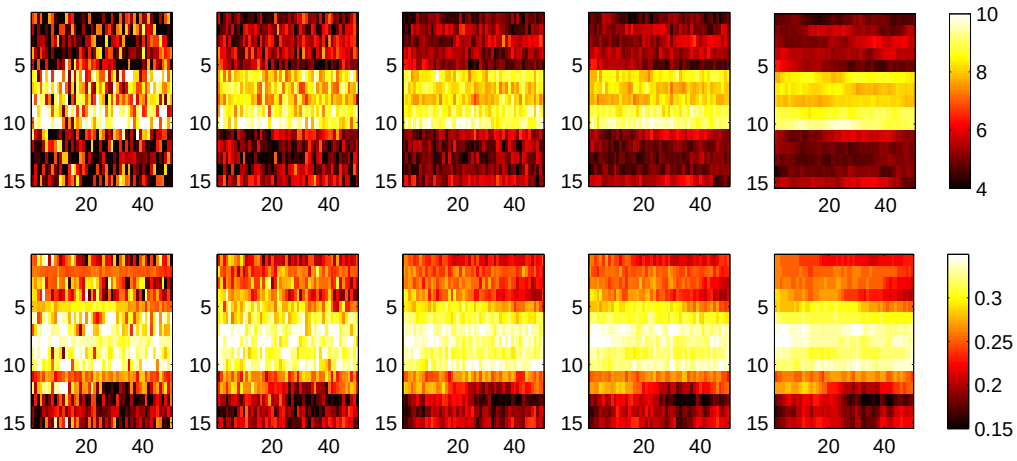


Figure 15: Localised Hierarchical Scale-Corrected Ensemble Kalman Filter, large model: Vertical slice of log-permeability and porosity for one ensemble member. The first row shows from left to right; log-permeability after 0, 300, 900 and 1 800 days of updating. The second row shows the porosity at the selected days. The rightmost plots show, from top to bottom, reference log-permeability and porosity.

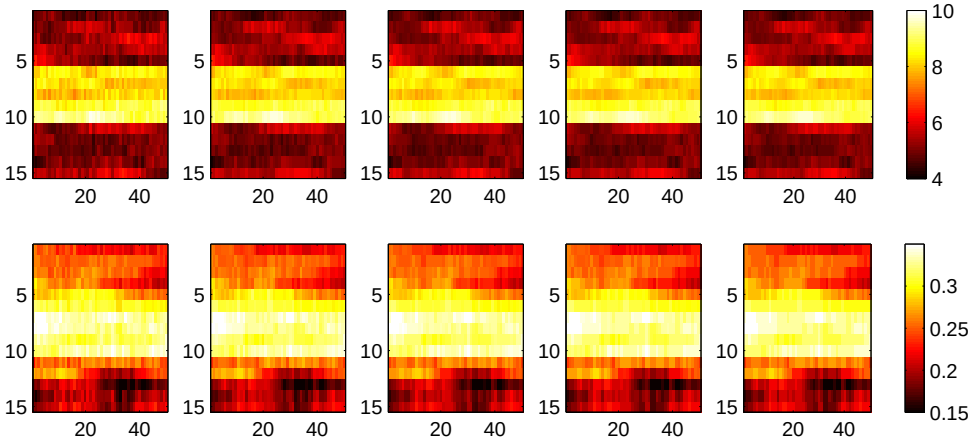


Figure 16: Localised Hierarchical Scale-Corrected Ensemble Kalman Filter, large model: Vertical slice of the ensemble mean for log-permeability and porosity together with the reference log-permeability and porosity. The first row shows from left to right; ensemble mean value of log-permeability after 0, 300, 900 and 1 800 days of updating. The second row shows the ensemble mean value of porosity at the selected days. The rightmost plots show, from top to bottom, reference log-permeability and porosity.

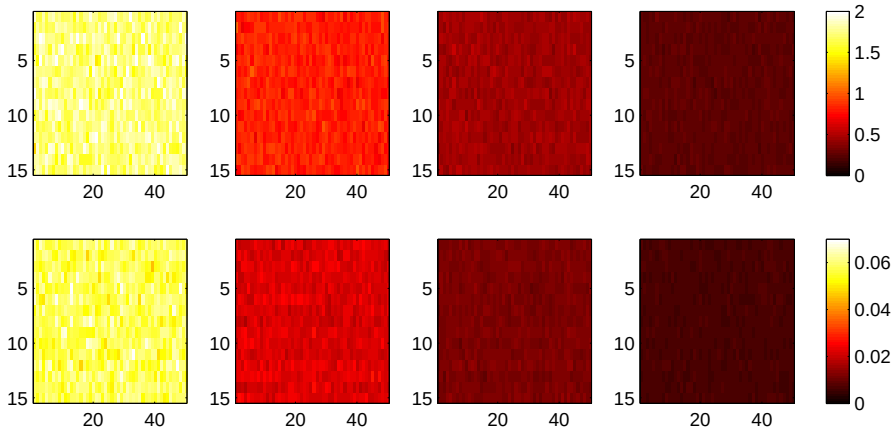


Figure 17: Localised Hierarchical Scale-Corrected Ensemble Kalman Filter, large model: Vertical slice of the ensemble standard deviation of log-permeability and porosity. The first row shows from left to right; ensemble standard deviation for log-permeability after 0, 300, 900 and 1 800 days of updating. The second row shows the ensemble standard deviation for porosity at the selected days.

ties, (permeability and porosity) and the spatio-temporal properties (saturation, pressure and production properties), we rerun the fine scale fluid flow simulator ω, from the initial timestep, using the updated static reservoir properties, without conditioning on the observations. The possible inconsistency might arise due to linear updates in the HScCEnKF, and the non-linear fluid flow partial differential equations.

Figure 18 displays the simulation results for all production properties, when rerunning after 0, 300, 900 and 1 800 days of updating, compared to the reference production. As we can see from this figure, the uncertainty is initially relatively large. Further, we observe that this uncertainty is reduced as more data is assimilated into the ensemble. Moreover, we see that all of the ensemble members reproduces the true reservoir production after 1 800 days of updating, which was expected when we compared the development in log-permeability and porosity of an ensemble member to the reference reservoir.

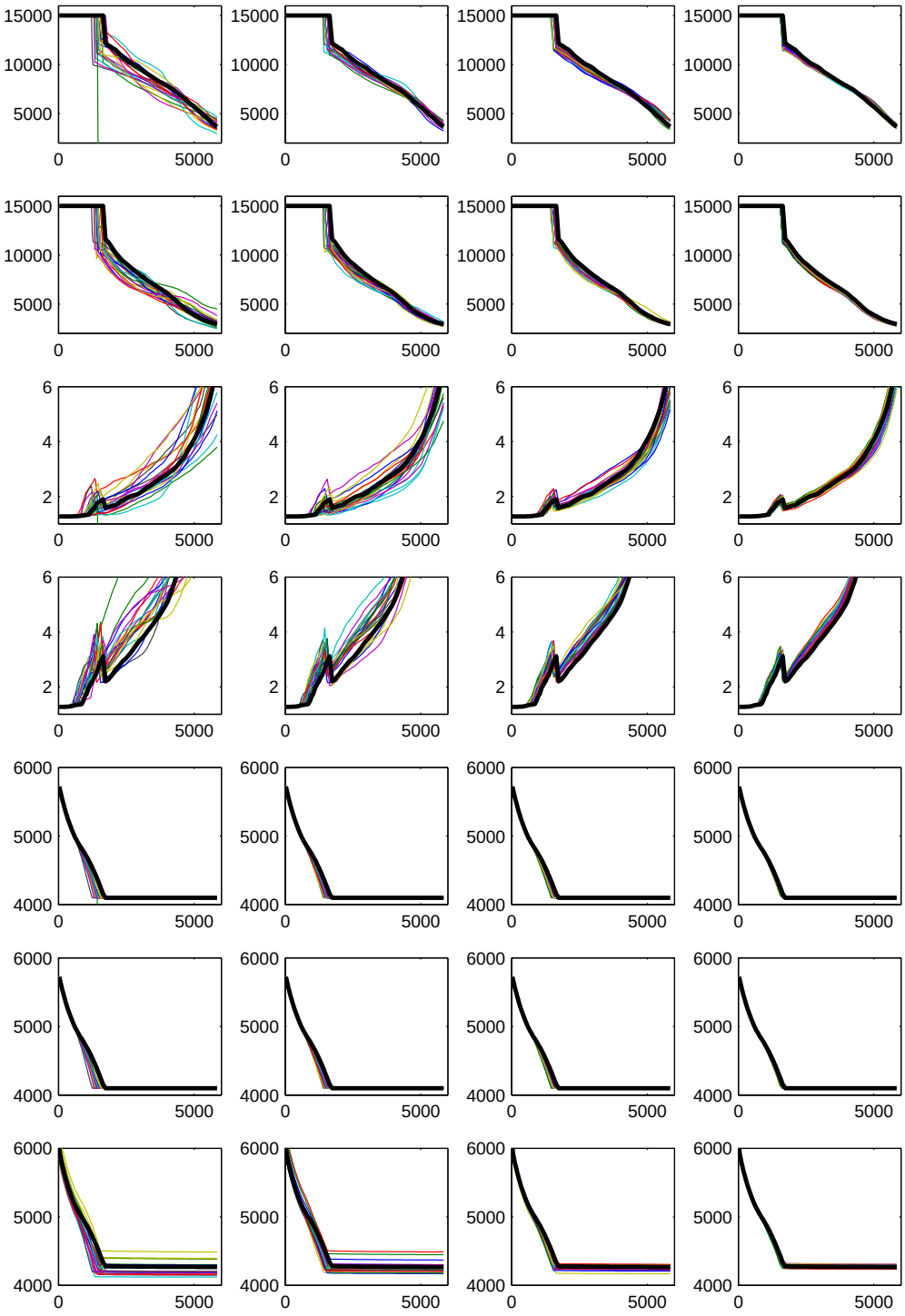


Figure 18: Localised Hierarchical Scale-Corrected Ensemble Kalman Filter, large model: Production properties from reruns of the 20 first updated fine scale ensemble members compared to the fine scale reference production (thick line). The figure shows from left to right; the initial ensemble, the ensemble after 300, 900 and 1 800 days of updating. The production properties considered are, from top to bottom; $opr1$, $opr2$, $gor1$, $gor2$, $bhp1$, $bhp2$, $bhpi$.

10.2 Comparison of Local and Global Downscaling Scheme

To investigate the possible loss of lateral connection in the downscaling when using the localised scheme, we run the HScCEnKF without performing any localisation on the smaller model outlined in Section 9, with $n_s = 20$ ensemble members.

Figure 19 shows the evolvement in time of a vertical slice of log-permeability and porosity for one ensemble member at the initial timestep, and after 300, 900 and 1 800 days of updating using the HScCEnKF. As above the ensemble member is drawn towards the reference log-permeability and porosity as more production and seismic data becomes available for assimilation.

The estimated mean value of log-permeability and porosity in the ensemble at selected times is shown in Figure 20. Similarly as above, using the larger model with the localised scheme, we can see from this figure that the initial mean value of the porosity is close to the reference porosity. Comparing this figure to Figure 16, we see that decreasing the number of ensemble members has led to a slight increase in deviation between the reference reservoir and the initial ensemble mean.

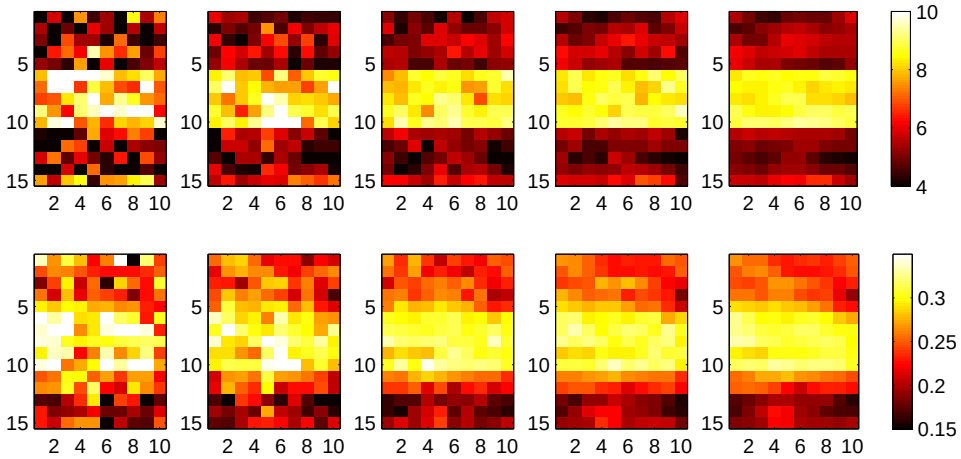


Figure 19: Hierarchical Scale-Corrected Ensemble Kalman Filter, small model: Vertical slice of log-permeability and porosity for one ensemble member. The first row shows from left to right; log-permeability after 0, 300, 900 and 1 800 days of updating. The second row shows the porosity at the selected days. The rightmost plots show, from top to bottom, reference log-permeability and porosity.

The same trend as in Figure 19 is seen when we look at the corresponding vertical slice of the estimated ensemble standard deviation for log-permeability and porosity, shown in Figure 21.

As above, we check if there might be inconsistencies between the updated static properties, and the dynamic properties, by rerunning the fine scale fluid flow simulator ω from the initial timestep, without conditioning on the observations,

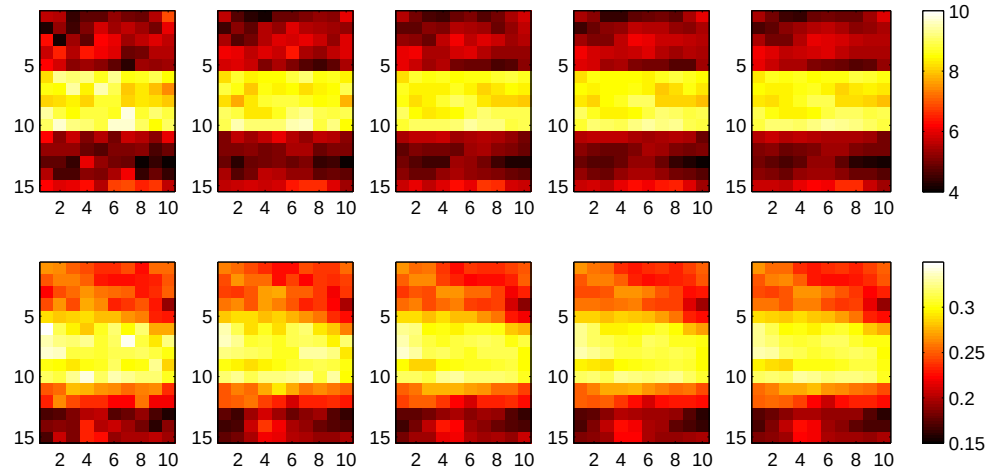


Figure 20: Hierarchical Scale-Corrected Ensemble Kalman Filter, small model: Vertical slice of the ensemble mean for log-permeability and porosity together with the reference log-permeability and porosity. The first row shows from left to right; ensemble mean value of log-permeability after 0, 300, 900 and 1 800 days of updating. The second row shows the ensemble mean value of porosity at the selected days. The rightmost plots show, from top to bottom, reference log-permeability and porosity.

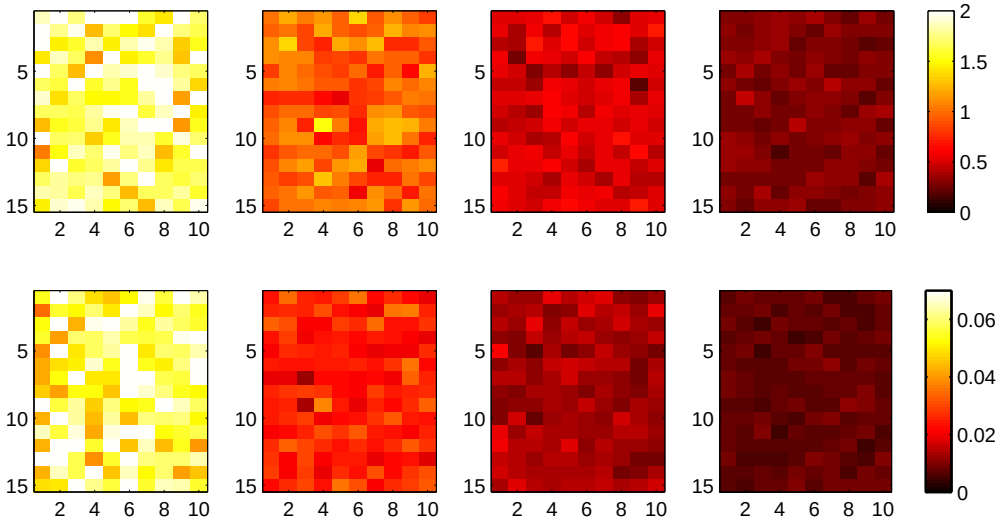


Figure 21: Hierarchical Scale-Corrected Ensemble Kalman Filter, small model: Vertical slice of the ensemble standard deviation of log-permeability and porosity. The first row shows from left to right; ensemble standard deviation for log-permeability after 0, 300, 900 and 1 800 days of updating. The second row shows the ensemble standard deviation for porosity at the selected days.

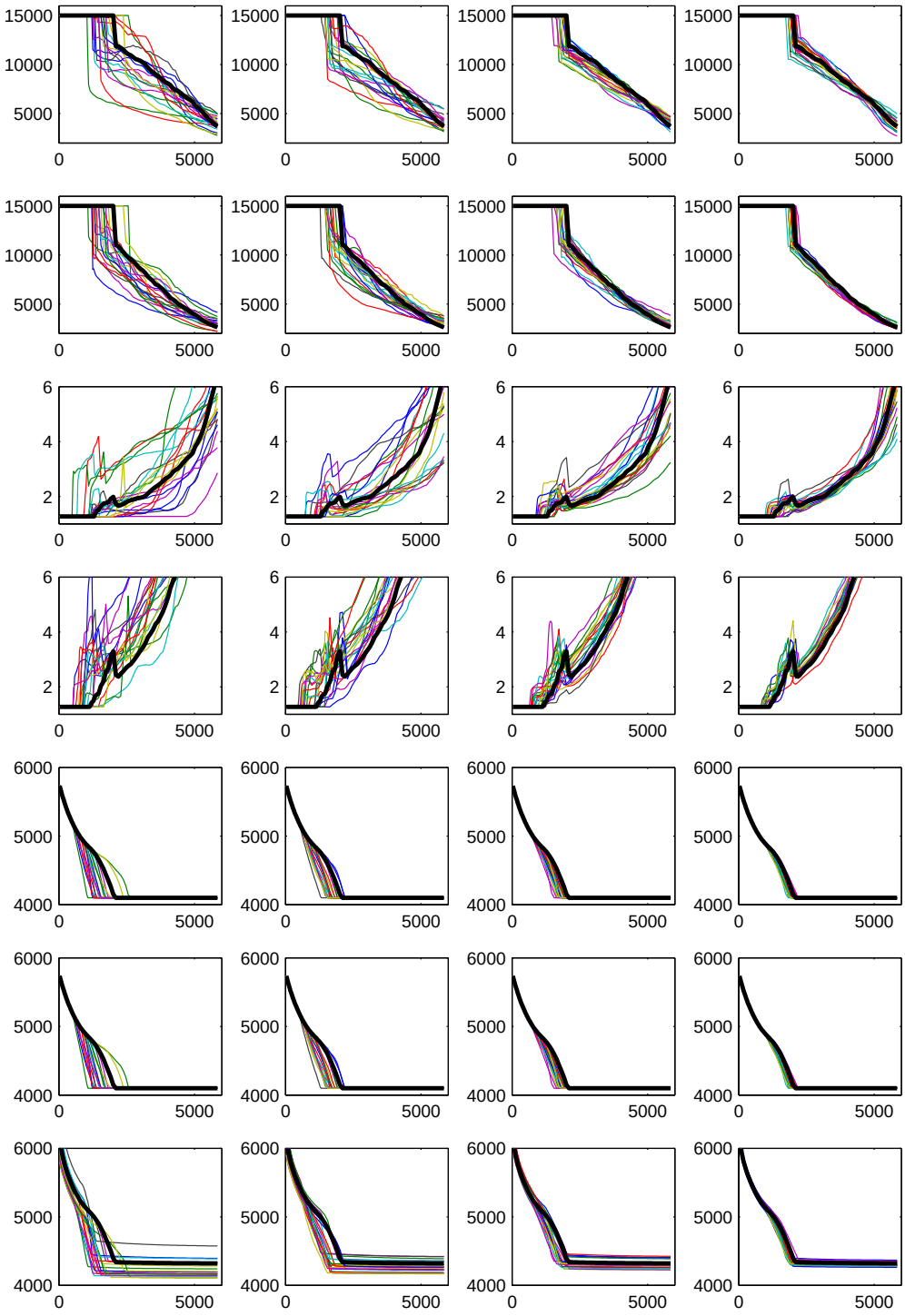


Figure 22: Hierarchical Scale-Corrected Ensemble Kalman Filter, small model: Production properties from reruns of the 20 first updated fine scale ensemble members compared to the fine scale reference production (thick line). The figure shows from left to right; the initial ensemble, the ensemble after 300, 900 and 1 800 days of updating. The production properties considered are, from top to bottom; $opr1$, $opr2$, $gor1$, $gor2$, $bhp1$, $bhp2$, $bhpi$.

using the updated static reservoir properties at the same report times as above. The outcome of the simulations is shown in Figure 22. As we can see from this figure, the uncertainty is initially large, and similarly to the results above, this uncertainty is reduced as more data becomes available. Since the gas-breakthrough times, and switching from target rate to bottom hole pressure in well controls, happen between 900 and 1 800 days, we see that the uncertainty in bottom hole pressure and the gas-oil ratio has the highest reduction between these two report times.

It is worth noticing that there seems to be a trend in the initial ensemble leading to an early gas-breakthrough. This trend is present until after 1 800 days when observations of gas-breakthrough have been made, and we get a seemingly unbiased production forecast using the updated ensemble. The reason for this trend is that our initial ensemble has too little uncertainty, and that the number of ensemble members is small. Rerunning using an ensemble that initially has a better coverage, removes this trend as shown in Figure 23.

If we compare Figures 22 and 23 with Figure 18, we can see that the initial uncertainty in production is larger in the small model than in the larger model, due to the method from which the initial ensemble was generated. Local deviations from the ensemble mean in each grid block for a fine scale model, will not influence the general fluid flow of the reservoir to the same degree as it would in a coarser model.

Since the localised downscaling scheme is performed on a different scale than the global downscaling it is difficult to draw any conclusions from the results above concerning the possible loss in lateral connection. However, we can see that the ensemble updated using the local scheme almost reproduces the reference production after only 300 days of updating, with a relatively small level of uncertainty compared to the ensemble from the smaller model, updated using the global scheme. This effect could be a result of the additional assumptions we have made in our model, and that the localised scheme is more sensitive to the prior model parameters. Therefore we rerun the model assuming that the initial expected values of the porosity is a linear function of the log-permeability, that is, $\boldsymbol{\mu}_{\boldsymbol{\phi}_0} = a_0\mathbf{1} + a_1\boldsymbol{\mu}_{\boldsymbol{\kappa}0}$. The results when we assimilate data, and rerun using the updated static properties, on both the large and small model, are shown in Figures 24 and 25. The model parameters a_0 and a_1 were chosen by regression using the reference reservoir.

As we can see when comparing these two figures, the localised scheme adjusts more quickly to the prior model parameters, than the global scheme, and there is a higher uncertainty in the predicted production properties for the ensemble running with the global scheme.

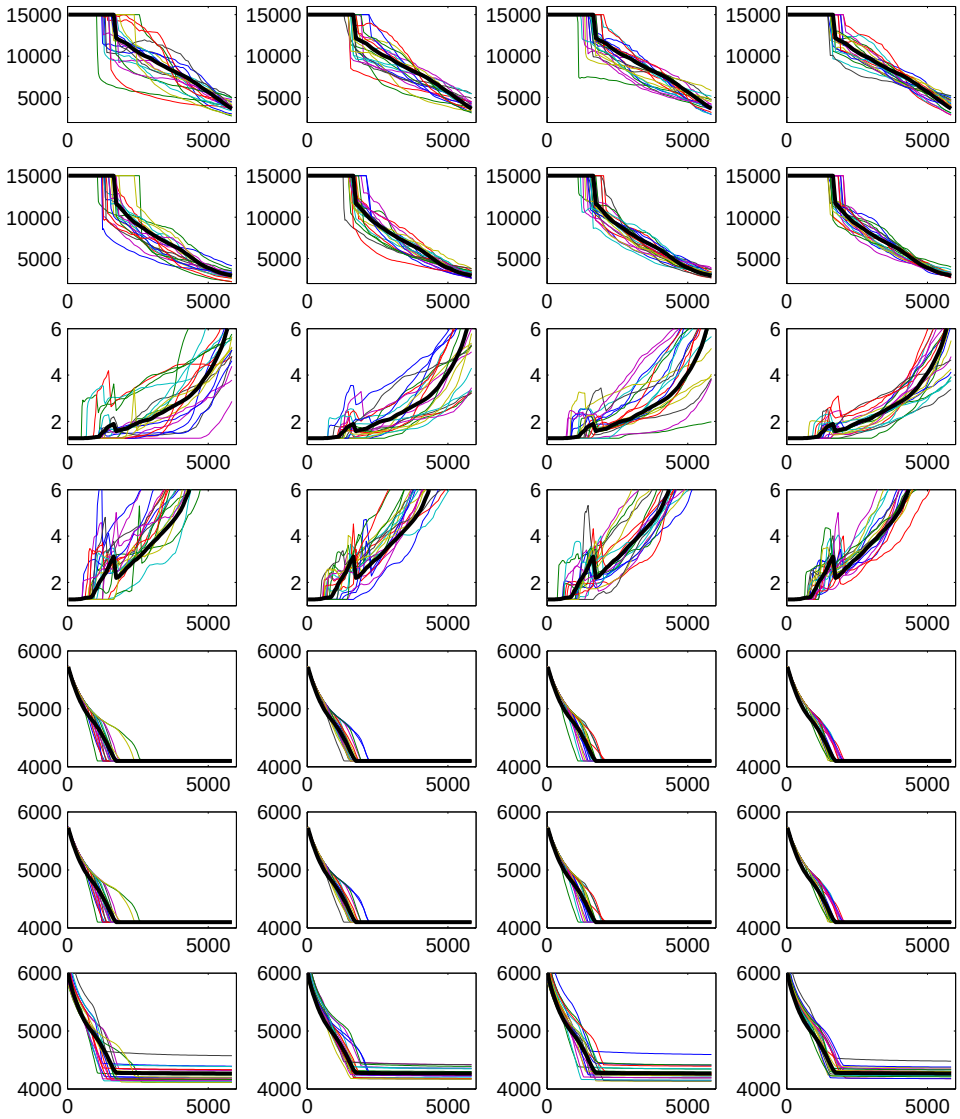


Figure 23: Hierarchical Scale-Corrected Ensemble Kalman Filter, small model: Production properties from reruns of the 20 first updated fine scale ensemble members compared to the fine scale reference production (thick line). The figure shows from left to right; the initial ensemble, the ensemble after 300, 900 and 1 800 days of updating. The production properties considered are, from top to bottom; $opr1$, $opr2$, $gor1$, $gor2$, $bhp1$, $bhp2$, $bhpi$.

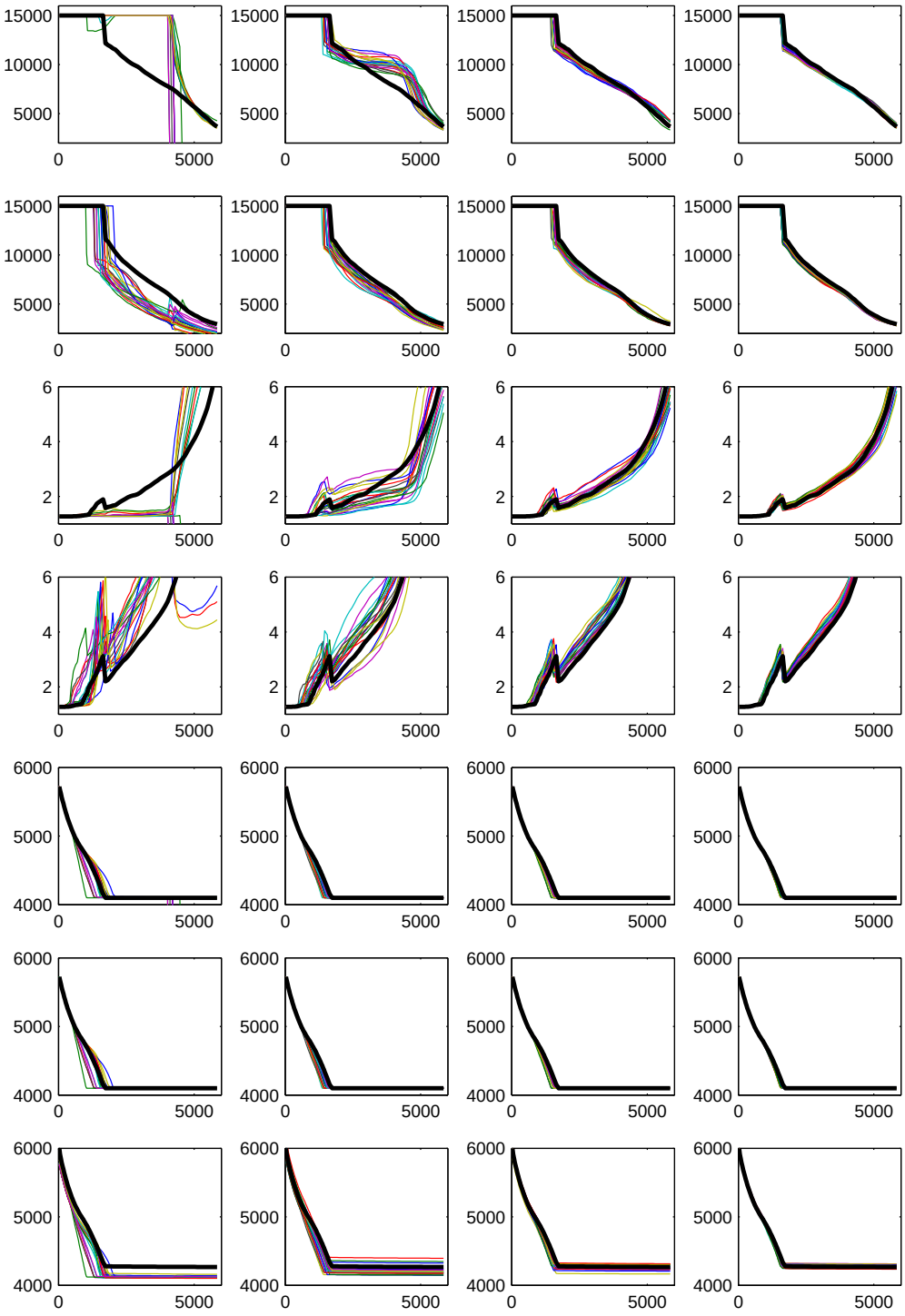


Figure 24: Localised Hierarchical Scale-Corrected Ensemble Kalman Filter, large model, assuming $\boldsymbol{\mu}_{\boldsymbol{\phi}_0} = a_0\mathbf{1} + a_1\boldsymbol{\mu}_{\boldsymbol{\kappa}_0}$: Production properties from reruns of the 20 first updated fine scale ensemble members compared to the fine scale reference production (thick line). The figure shows from left to right; the initial ensemble, the ensemble after 300, 900 and 1 800 days of updating. The production properties considered are, from top to bottom; $opr1$, $opr2$, $gor1$, $gor2$, $bhp1$, $bhp2$, $bhpi$.

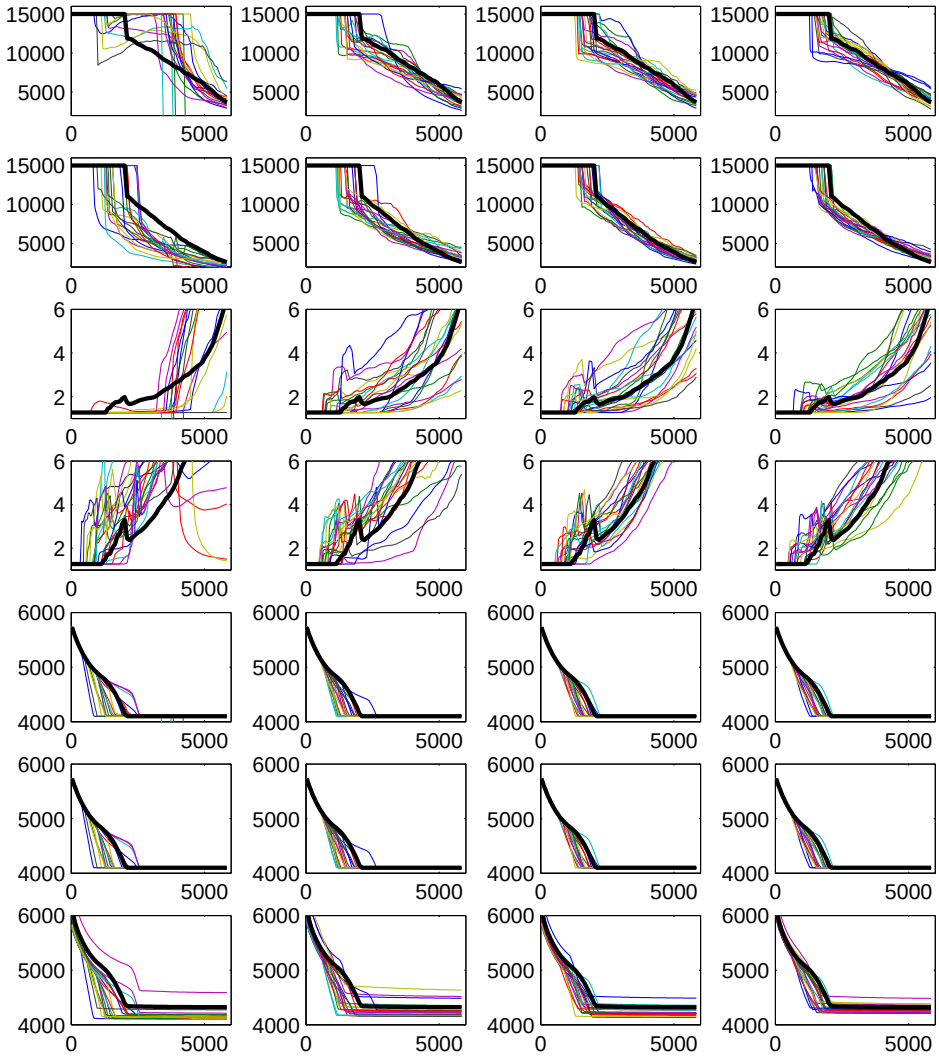


Figure 25: Hierarchical Scale-Corrected Ensemble Kalman Filter, small model, assuming $\boldsymbol{\mu}_{\phi_0} = a_0\mathbf{1} + a_1\boldsymbol{\mu}_{\boldsymbol{\kappa}0}$: Production properties from reruns of the 20 first updated fine scale ensemble members compared to the fine scale reference production (thick line). The figure shows from left to right; the initial ensemble, the ensemble after 300, 900 and 1 800 days of updating. The production properties considered are, from top to bottom; *opr*1, *opr*2, *gor*1, *gor*2, *bhp*1, *bhp*2, *bhpi*.

10.3 Traditional Ensemble Kalman Filter

The traditional EnKF conditioned on production data, with an ensemble size of $n_s = 100$, was run on the small model outlined in Section 9, in order to compare the performance of the HScCEnKF to the traditional approach. The evolvement in time of a vertical slice of log-permeability and porosity is shown in Figure 26. From this figure we can see that the ensemble member diverge from the reference reservoir as more data becomes available. A similar behaviour can be seen from the reruns of the updated static reservoir properties in the 20 first ensemble members, shown in Figure 27. As we can see from these two figures the small number of ensemble members leads to a poor estimate of the covariance matrix, $\boldsymbol{\Sigma_{rq}}$. Thus, the ensemble diverge from the reference reservoir.

These results illustrate one of the strengths with a hierarchical Bayesian approach, where the prior assumptions ensure a full rank representation of the posterior covariance matrices. However, the success of the HScCEnKF above, is closely related to the selected hyperparameters in the prior distributions. At the initial timesteps, we experienced a similar behaviour in the HScCEnKF ensemble posterior to data assimilation. This effect was due to the poor selection of $\boldsymbol{\Psi_{r^*q}}$, which was selected as the estimated covariance, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{r^*q^*}}$. As the ensemble was drawn towards the reference reservoir, this unwanted feature was corrected for, and data assimilation reduced the uncertainty in the ensemble instead of increasing it.

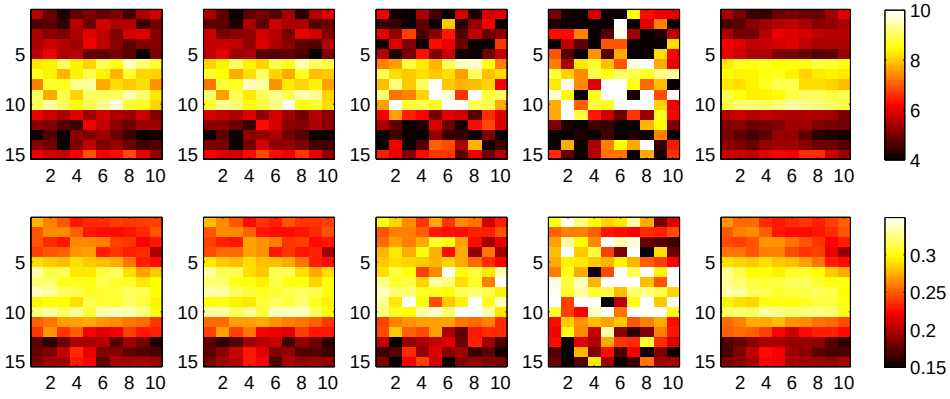


Figure 26: Traditional Ensemble Kalman Filter, small model: Vertical slice of log-permeability and porosity for one ensemble member. The first row shows from left to right; log-permeability after 0, 300, 900 and 1 800 days of updating. The second row shows the porosity at the selected days. The rightmost plots show, from top to bottom, reference log-permeability and porosity.

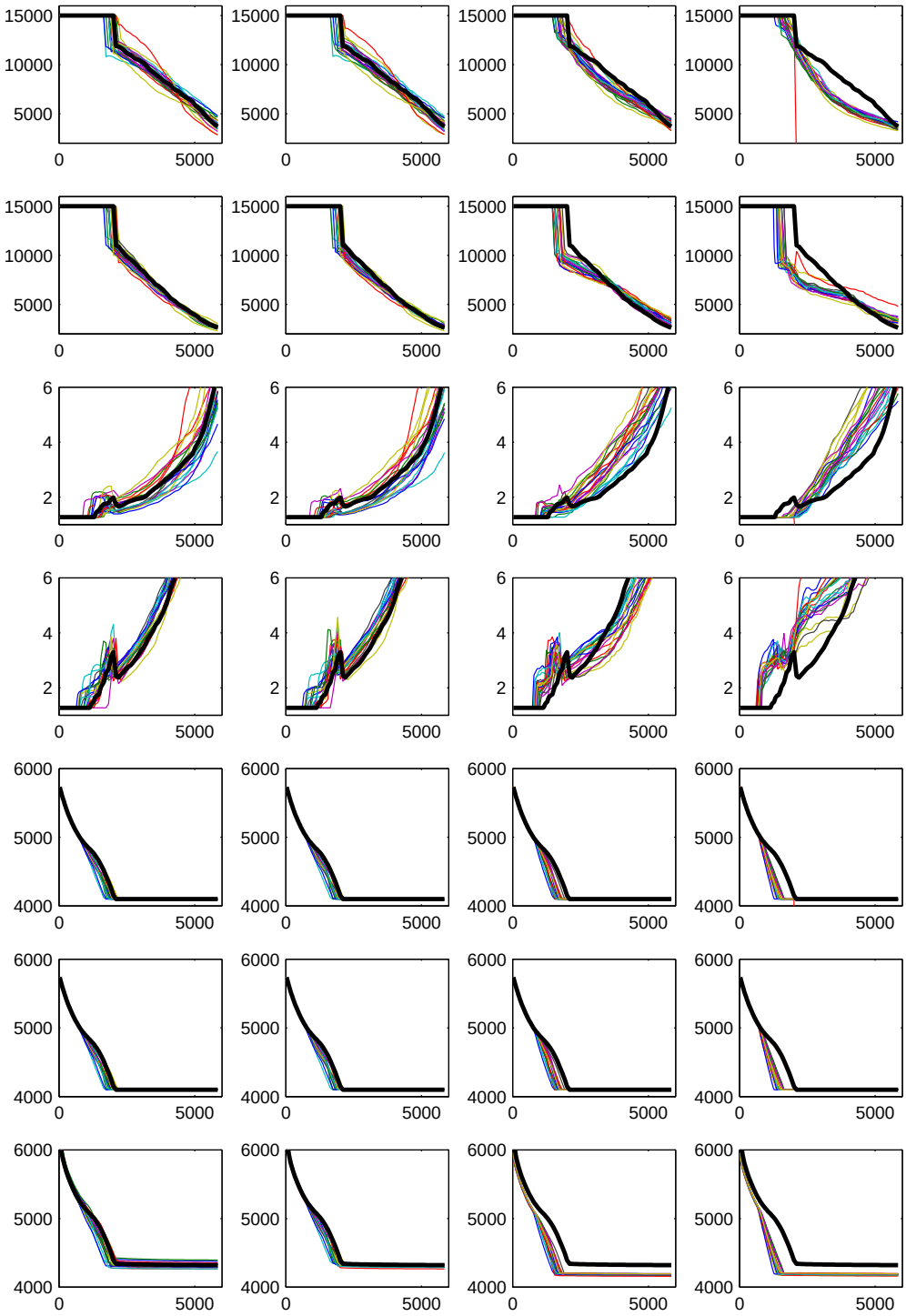


Figure 27: Traditional Ensemble Kalman Filter, small model: Production properties from reruns of the 20 first updated fine scale ensemble members compared to the fine scale reference production (thick line). The figure shows from left to right; the initial ensemble, the ensemble after 300, 900 and 1 800 days of updating. The production properties considered are, from top to bottom; $opr1$, $opr2$, $gor1$, $gor2$, $bhp1$, $bhp2$, $bhpi$.

10.4 Consistency of Approximate Fluid Flow Simulation

In order to check if the hierarchical Bayesian approximate fluid flow simulator defined in Section 7.3.2 is consistent, we perform fluid flow simulation on the small model, downscaling every fifth month using the global scheme, without conditioning on any data. If the simulator is consistent, we would expect the distribution of the downscaled ensemble to reflect the prior distribution. Moreover, we should not be able to spot any trends when we look at the difference between the coarse scale ensemble prior to downscaling, and the coarse scale ensemble posterior to downscaling and upscaling. That is $\boldsymbol{r}_t^* \sim \boldsymbol{r}_t^{u*}\ \forall t$ where, as in Section 7.3.2, $\boldsymbol{r}_t^*$ denotes the coarse scale ensemble prior to downscaling, and $\boldsymbol{r}_t^{u*}$ denotes the coarse scale ensemble posterior to downscaling/upscaling. Note that in this study, the parameters of the prior covariance matrix were three times larger than above.

Figure 28 displays line plots of $\boldsymbol{\kappa}_t^* - \boldsymbol{\kappa}_t^{u*}$ and $\boldsymbol{\phi}_t^* - \boldsymbol{\phi}_t^{u*}$ for the 20 ensemble members, after the first six updates. Initially we can see a higher uncertainty due to the seismic inversion performed in the generation of the initial ensemble. After four updates, the differences are stabilised, and we see a similar behaviour as we continue updating for 1 800 days, downscaling every 150 days.

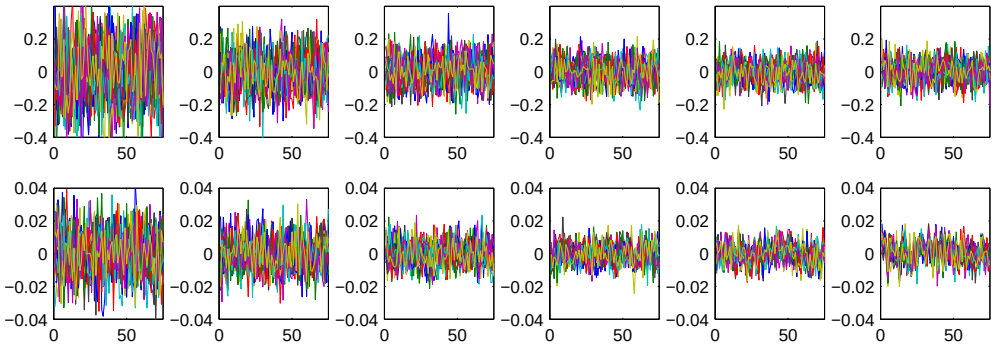


Figure 28: Hierarchical Bayesian approximate fluid flow simulation, small model: Difference between the coarse scale ensemble members prior to downscaling and posterior to downscaling and upscaling. Downscaling was performed every 150 days. The first row shows from left to right; difference in log-permeability after 150, 300, 450, 600, 750 and 900 days of updating. The second row shows the difference in porosity the selected times.

Figure 29 shows a vertical slice of the estimated standard deviation in the fine scale ensemble after 900, 1 200, 1 500 and 1 800 days of updating, downscaling every fifth month. As we can see from this figure, the standard deviation shows a similar, and stable behaviour for all four report times. The same trend can be seen in the vertical slice of log-permeability and porosity, for one ensemble member at the same four report times. This is shown in Figure 30.

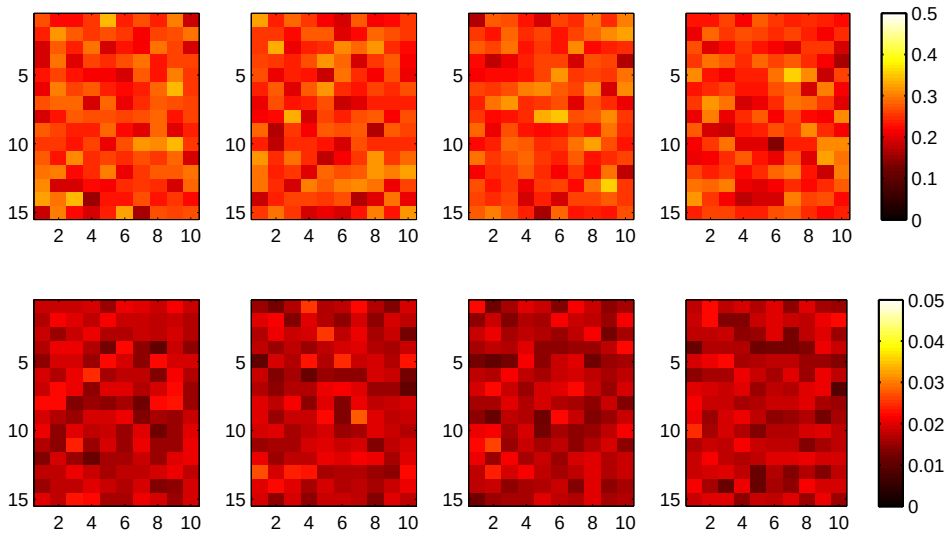


Figure 29: Hierarchical Bayesian approximate fluid flow simulation, small model: Vertical slice of the ensemble standard deviation of log-permeability and porosity. The first row shows from left to right; ensemble standard deviation for log-permeability after 900, 1 200, 1 500 and 1 800 days of updating. The second row shows the ensemble standard deviation for porosity the selected days.

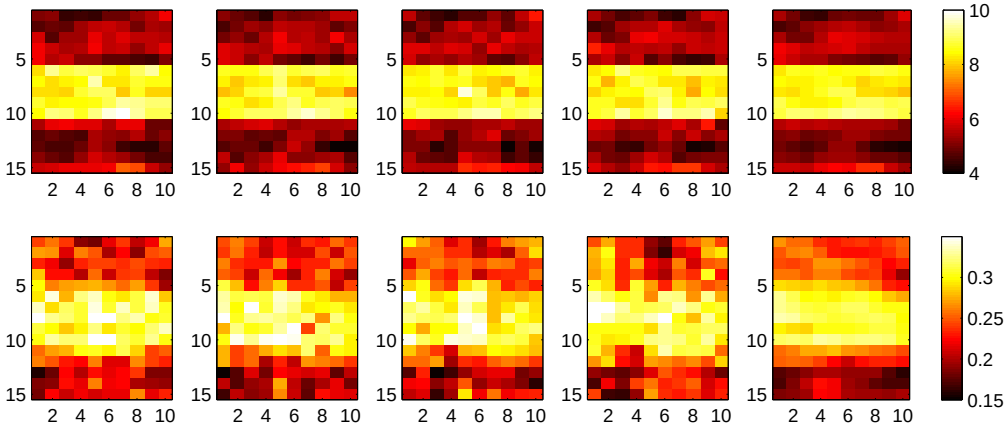


Figure 30: Hierarchical Bayesian approximate fluid flow simulation, small model: Vertical slice of log-permeability and porosity for one ensemble member. The first row shows from left to right; log-permeability after 900, 1 200, 1 500 and 1 800 days of updating. The second row shows the porosity the selected days. The rightmost plots show, from top to bottom, reference log-permeability and porosity.

10.5 Sensitivity to Prior Model Parameters

To investigate the sensitivity of the results with respect to the choice of prior model parameters, we redid the analysis on the small model, using the global downscaling scheme with two different selections of prior hyperparameters, refered to as the first and second case.

For the first case, the hyperparameter of the expected value was tuned such that:

$$\boldsymbol{\eta}_t = \boldsymbol{\eta}_t^r - \frac{1}{1.5}\left(\boldsymbol{\eta}_t^r - \bar{\boldsymbol{\eta}}_t^r\right) + \boldsymbol{\epsilon}_{\boldsymbol{\eta}^r},$$

where the superscript r indicates the corresponding values of the reference model, and $\boldsymbol{\epsilon}_{\boldsymbol{\eta}^r} \sim \text{Gauss}_{n_r}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}^r})$, with $\boldsymbol{\Sigma}_{\boldsymbol{\eta}^r}$ selected based on the estimated variance in the reference reservoir. This implies that all values were dampened towards the mean values before white noise was added. Figure 31 shows this effect for the log-permeability and porosity together with a realisation from the initial ensemble.

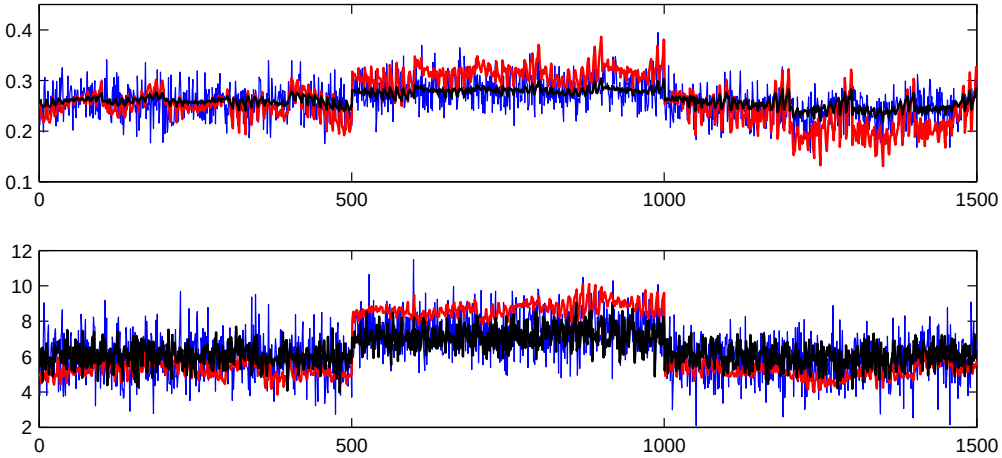


Figure 31: First case: Comparison of the model prior parameters $\boldsymbol{\eta}$ for porosity (top) and log-permeability (bottom) before and after modification, together with a realisation from the ensemble at the initial timestep shown in blue. The values of the reference reservoir are shown in red, while the dampened prior parameters with added noise are shown in black.

In the second case, we selected $\boldsymbol{\eta}_t$ as the mean value within the three separating layers of the reference reservoir, as can be seen from the line plot in Figure 32. The prior parameters of the covariance matrix, were in this study selected three times higher than the values found in Section 9.

If we consider the static properties of one ensemble member in the first case, shown in Figure 33, we see that high permeable middle layer is no longer as distinguishable as in the figures above. Similarly to the results above, the ensemble member moves towards the ensemble mean after 1 800 days of updating. This, however, is not shown here.

Figure 34 shows the evolvement of the static properties of one ensemble member for the second case. As above, the ensemble moves towards the ensemble

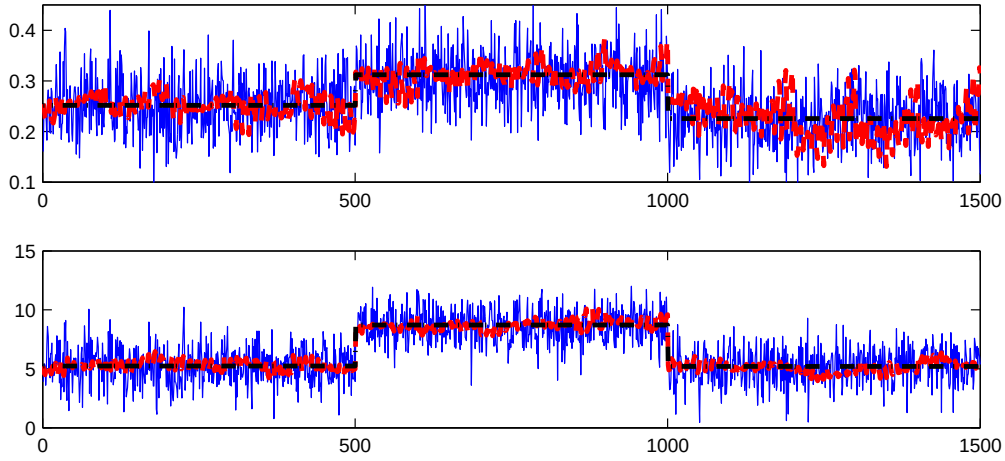


Figure 32: Second case: Comparison of the model prior parameters $\boldsymbol{\eta}$ for porosity (top) and log-permeability (bottom) before and after modification together with an ensemble at the initial timestep shown in blue. The values of the reference reservoir are shown in red, while the dampened prior parameters are shown in black.

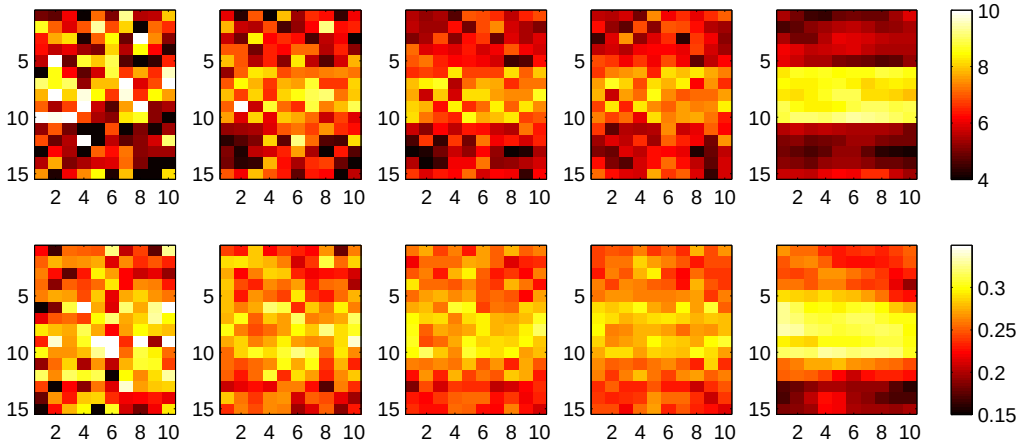


Figure 33: Hierarchical Scale-Corrected Ensemble Kalman Filter, small model, dampened prior parameter $\boldsymbol{\eta}$ with added white noise: Vertical slice of log-permeability and porosity for one ensemble member. The first row shows from left to right; log-permeability after 0, 300, 900 and 1 800 days of updating. The second row shows the porosity at the selected days. The rightmost plots show, from top to bottom, reference log-permeability and porosity.

mean, but due to the higher prior variance, the mean is not reproduced after 1 800 days of updating as above.

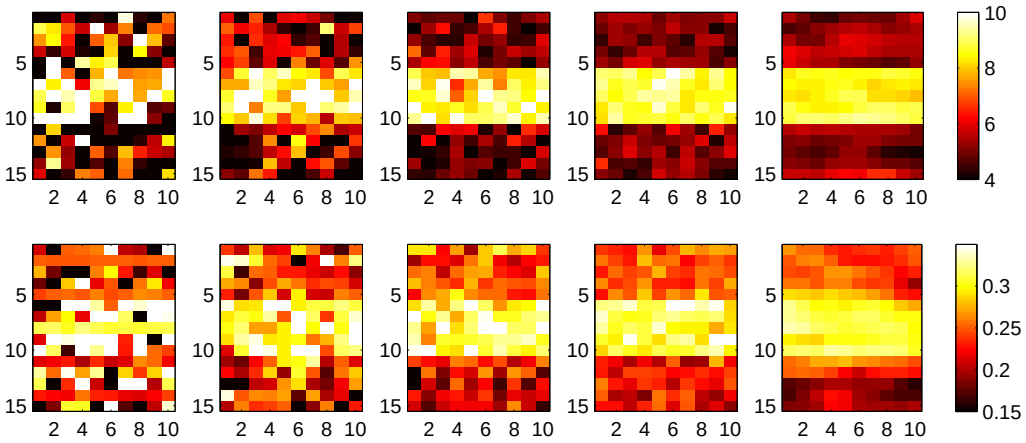


Figure 34: Hierarchical Scale-Corrected Ensemble Kalman Filter small model, dampened $\boldsymbol{\eta}$, increased prior variance: Vertical slice of log-permeability and porosity for one ensemble member. The first row shows from left to right; log-permeability after 0, 300, 900 and 1 800 days of updating. The second row shows the porosity at the selected days. The rightmost plots shows, from top to bottom, reference log-permeability and porosity.

The results obtained when rerunning the fine scale fluid flow simulator ω using the updated static variables for the first case as input, are shown in Figure 35. As we can see from this figure, shifting the prior values for the expectation leads to a bias in the estimated production for all of the reruns. Due to the smoothening of the permeability and porosity compared to the reference reservoir, we observe an early gas-breakthrough in the reruns.

Figure 36 shows results of the rerun using the updated static parameters for the second case. As we can see from this figure, the added prior variance leads to a higher uncertainty in the initial ensemble. As above, this uncertainty is reduced as more observations are available. Comparing the results in Figures 35 and 36, we notice that the second case provides a much better match with the reference production than the first case. This is both due to a better coverage of the sample space in the prior distribution through higher prior covariance, and a better selection of $\boldsymbol{\eta}$.

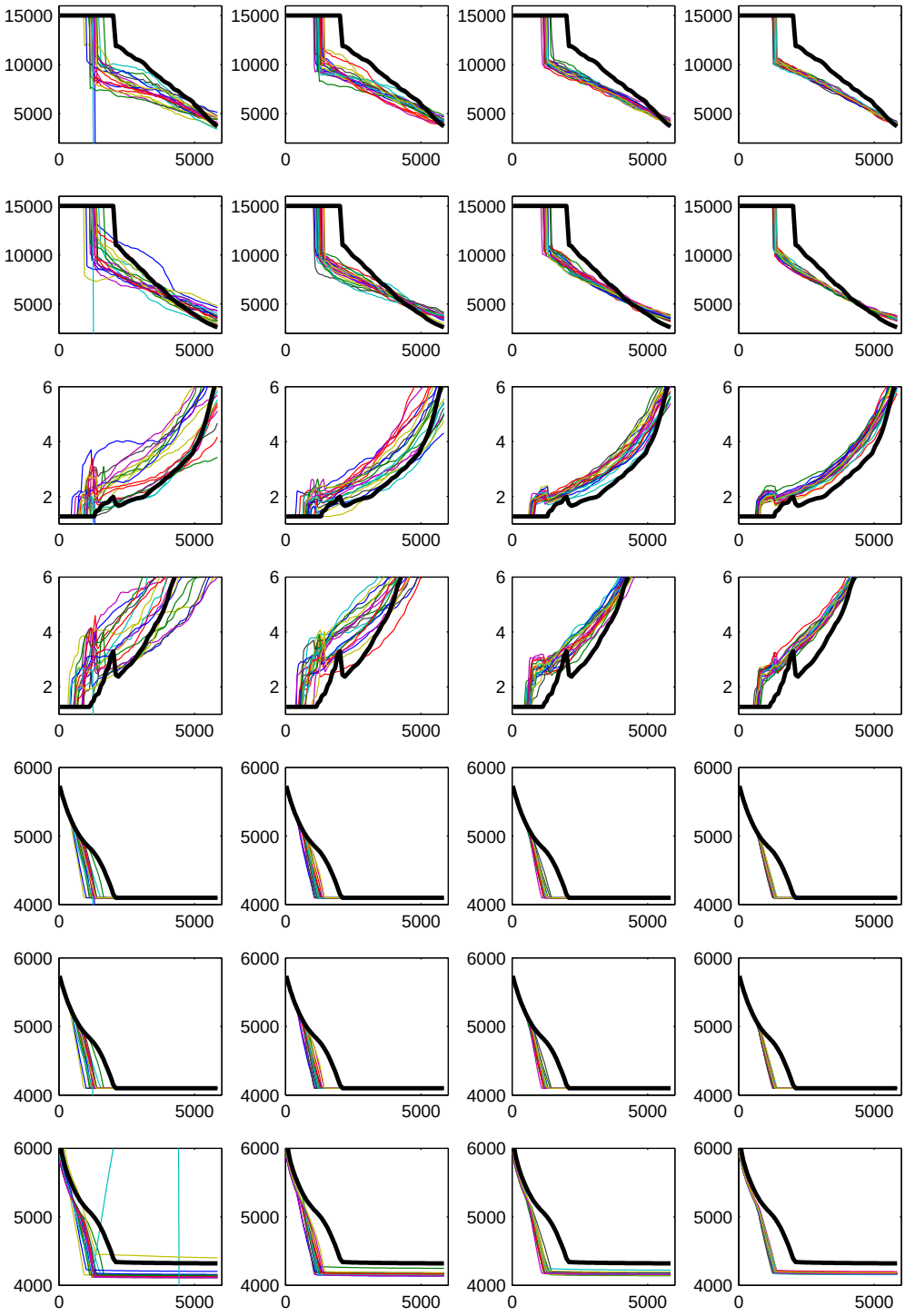


Figure 35: Hierarchical Scale-Corrected Ensemble Kalman Filter, small model, dampened prior parameter $\boldsymbol{\eta}$ with added white noise: Production properties from reruns of the 20 first updated fine scale ensemble members compared to the fine scale reference production (thick line). The figure shows from left to right; the initial ensemble, the ensemble after 300, 900 and 1 800 days of updating. The production properties considered are, from top to bottom; $opr1$, $opr2$, $gor1$, $gor2$, $bhp1$, $bhp2$, $bhpi$.

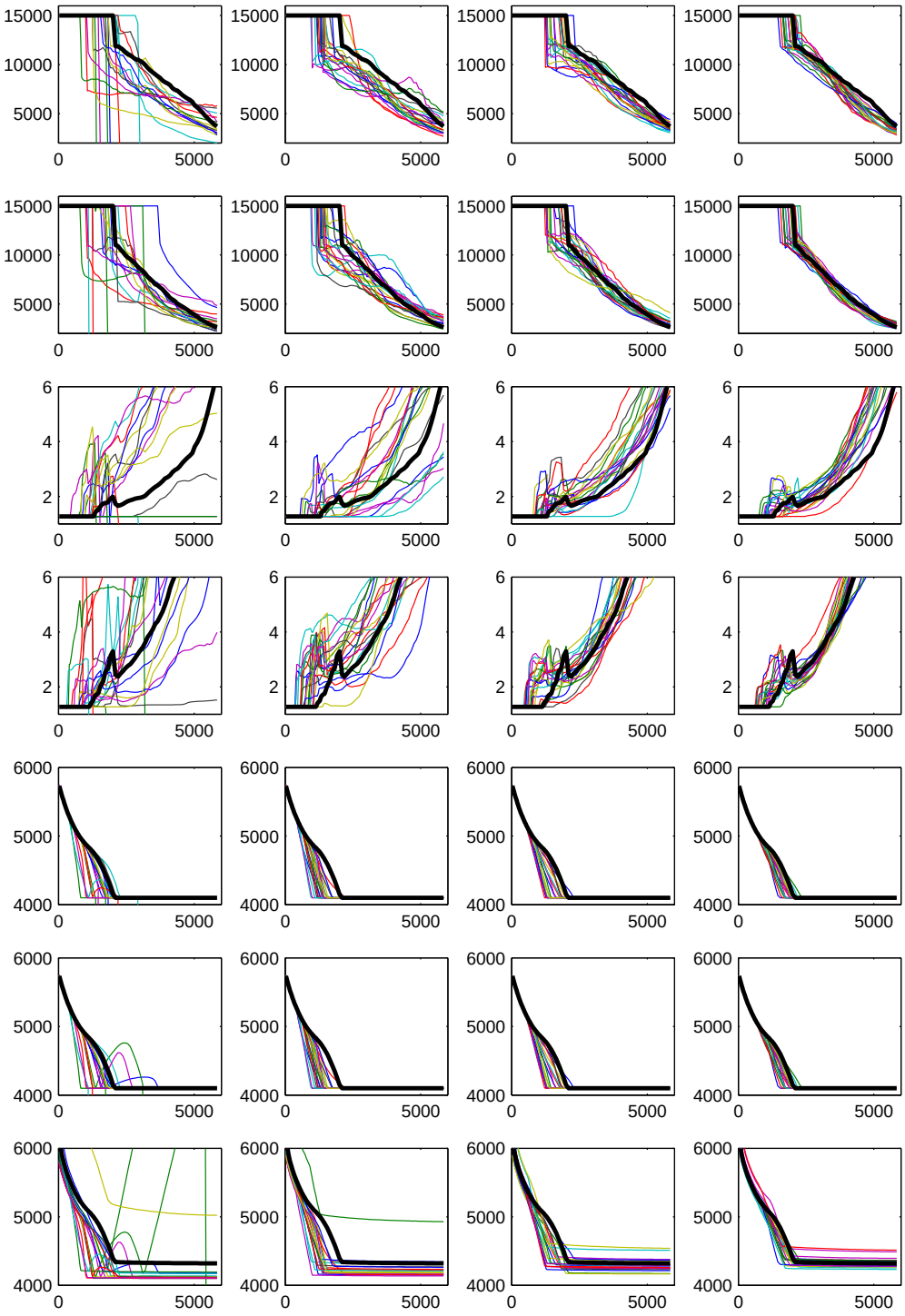


Figure 36: Hierarchical Scale-Corrected Ensemble Kalman Filter, small model, dampened $\boldsymbol{\eta}$, increased prior variance: Production properties from reruns of the 20 first updated fine scale ensemble members compared to the fine scale reference production (thick line). The figure shows from left to right; the initial ensemble, the ensemble after 300, 900 and 1 800 days of updating. The production properties considered are, from top to bottom; *opr*1, *opr*2, *gor*1, *gor*2, *bhp*1, *bhp*2, *bhpi*.

10.6 Summary

The results above have shown us some of the strengths and weaknesses of a Hierarchical approach to the EnKF. The major strength of the HScCEnKF is that all covariance matrices are ensured to have full rank, which is not the case in a traditional approach. As we saw in Section 10.3, a poor representation of the covariance matrix led to a divergence from the reference reservoir in the traditional EnKF. The major weakness is that additional prior assumptions have to be made through the selection of hyperparameters in the prior distributions. This will be a complicated task for a non-synthetic case study.

The use of a hierarchical Bayesian approximate fluid flow simulator in this study has shown that the success depends both on the selection of the $\boldsymbol{\eta}$ and $\boldsymbol{\Psi}$. During the downscaling phase of the reservoir properties, we compute the posterior mean, given in Expression (69), Section 7.3.2. If the difference between the coarse scale ensemble mean $\bar{\boldsymbol{x}}^*$, and the prior model parameter $\boldsymbol{\eta}^*$, is small compared to the values in the Bartlett decomposed posterior covariance matrix $\boldsymbol{T}_{\boldsymbol{r}}^{\boldsymbol{r}^*}$, the posterior expectation will only experience small changes. Thus, we experience results as shown in Figure 35, where data assimilation does not correct the error in the prior distribution. Better selection of the prior hyperparameters, however, led to a more correct representation of the reservoir production properties as shown in Figure 35.

When the prior assumptions concerning the reservoir deviate much from the true field, inclusion of seismic data and observed production, was not able to fully correct for this feature in the current study. However, a thorough sensitivity analysis with respect to selection of the prior covariance matrix has not been performed in the current study.

Tuning the prior model has shown that in a hierarchical EnKF, we sacrifice the unbiasedness of the EnKF, for consistency. In both the HScCEnKF and the traditional EnKF, the selection of prior model is therefore extremely important.

11 Conclusion

Reservoir history matching where the reservoir properties are presented on a fine scale, will, due to repeated fluid flow simulations, in most cases be computationally too demanding. Upscaling of the reservoir to a coarser scale must therefore be done, such that multiple fluid flow simulations can be afforded. The Ensemble Kalman Filter (EnKF) then gives a sequential Bayesian method for performing production history matching on a coarse scale.

The inclusion of 4-D seismic data, which typically appear on a very fine scale, requires a method of representing the reservoir on a fine scale given the reservoir on a coarser scale. Traditional downscaling procedures are, however, known to introduce bias in production forecasts. The Scale-Corrected EnKF was therefore introduced as a method for correcting this bias. This method, however, required fluid flow simulations both on a fine and coarse scale, and inversion of empirical covariance matrices which in most practical settings will be singular.

In this Thesis we have developed an hierarchical Bayesian extension to the EnKF and the Scale-Corrected EnKF, where only multiple fluid flow simulations on a coarse scale is required. The key assumption with this methology, is that the state of both fine and coarse scale reservoir, conditioned on previous observations of seismic and production, can be approximated by a multivariate Gaussian distribution. Moreover, by assigning natural conjugate prior distributions to the expectation and covariance matrix of this distribution, and defining linear Gaussian likelihood models, we ensured analytical tractability in all posterior distributions. Hence, a completely recursive method suitable for parallel computing was obtained.

The method was implemented and tested on a synthetic reservoir inspired by the Troll Field offshore Norway, including production history and synthetic 4-D seismic data from a reference model. High dimension of the reservoir, and limited computer resources, revealed some of the the weaknesses of this methology.

The primary concern is the need to generate high dimensional matrices having the inverted Wishart distribution, which is a computationally demanding task. Moreover, the downscaling of the reservoir involves inverting and multiple matrix multiplications of potentially ill-conditioned matrices, introducing numerical errors which could lead to negative definite covariance matrices when the number of ensemble members is small. An initial solution to these problems has been to use the posterior expectation, instead of sampling from the inverted Wishart distribution.

A localised technique performing downscaling sequentially on subsets of the reservoir was suggested in order to run the proposed algorithm within reasonable time on the full scale synthetic model. We also considered a model with reduced dimension such that the method could be tested without any localised computer scheme.

The results obtained has shown that the success of the HScCEnKF depends on the choice of initial parameters in the reservoir model. It is of utmost importance that the prior model parameters are selected such that they reflect the

true reservoir. This is a very complicated task which should be performed in collaboration with both geophysicists and reservoir engineers.

The estimation of hyperparameters, such that they reflect the true reservoir, will be a very important task for future development, as the the success of a hierarchical EnKF scheme, depends heavily on the choice of these parameters. The methology for both seismic inversion and approximate fluid flow simulation should further be applied to a real case study, in order to validate some of the assumptions made in the models. Well-logs needs to be included in the prior model, and approximations made should also be assessed through the use of MCMC sampling techniques.

The use of more sophisticated localisations schemes in the downscaling procedure should be considered, by e.g. considering larger neighbourhoods, which can be done without increasing the computational demands dramatically.

12 Bibliography

References

Aki, K. & Richards, P. G. (1980), *Quantitative Seismology: Theory and Methods*, W.H. Freeman.

Anderson, T. W. (2003), *An introduction to multivariate statistical analysis*, 3 edn, Wiley.

Avseth, P., Mukerji, T. & Mavko, G. (2005), *Quantitative Seismic Interpretation*, Cambridge.

Azis, K. & Settari, A. (1979), *Petroleum Reservoir Simulation*, Applied Science Publisher Ltd.

Bachrach, R. (2006), 'Joint estimation of porosity and saturation using stochastic rock-physics modelling', *Geophysics* **71**(5), 53–63.

Barndorff-Nielsen, O. E., Cox, D. R. & Kluppelberg, C. (2001), *Complex Stochastic Systems*, Chapman and Hall.

Bartlett, M. S. (1933), On the theory of statistical regression, *in* 'Proceedings of the Royal Society of Edinburgh', Vol. 53, pp. 260–283.

Batzel, M. & Wang, Z. (1992), 'Seismic properties of pore fluids', *Geophysics* **57**(11).

Biot, M. A. (1962), 'Mechanics of deformation and acoustic propagation in porous media', *Journal of Applied Physics* **33**.

Bissel, R., Kilough, L. & Sharma, Y. (1992), 'Reservoir History Matching Using the Methods of Gradients on a Workstation.'. Presented at the SPE European Petroleum Computer Conference, Stavanger, Norway, May 1992.

Brewer, J. W. (1978), 'Kronecker Products and Matrix Calculus in System Theory', *IEEE Transactions on circuits and systems* **25**.

Buland, A. & Omre, H. (2003), 'Bayesian linearized AVO inversion', *Geophysics* **68**(1), 185–198.

Casella, G. & Berger, R. L. (2002), *Statistical Inference*, Duxbury.

Caselton, W. F., Kan, L. & Zidek, J. V. (1992), Quality data networks that minimize entropy, *in* 'Statistics in the Environmental and Earth Sciences', Wiley.

Cerveny, V., Molotkov, I. & Psencik, I. (1977), 'Ray methods in seismology', University Karlova Press.

Chen, Y., Durlofsky, L. J., Gerritsen, M. & Wen, X. H. (2003), 'A Coupled Local-Global Upscaling Approach for Simulating Flow in Highly Hetrogeneous Formations', *Advances in Water Resources* .

Coles, S., Roberts, G. & Jarner, S. (2002), 'Computer Intensive Methods'.

Cormen, T. H., Clifford, S., Leiserson, C. E. & Rivest, R. L. (2001), *Introduction to Algorithms*, The MIT Press.

Darcy, H. (1856), 'Les Fontaines publiques de la Ville de Dijon', *Dalmount, Paris* .

Deemer, W. L. & Olkin, I. (1951), 'The Jacobians of certain matrix transformations useful in multivariate analysis', *Biometrika* **38**, 345–367.

Deutsch, C. V. (2002), *Geostatistical Reservoir Modeling*, Oxford University Press.

Dobrin, M. B. & Savit, C. H. (1988), *Introduction to geophysical prospecting*, New York: McGraw-Hill Book Company.

Durlofsky, L. J. (2003), Upscaling of Geocellular Models for Reservoir Flow Simulation: A Review of Recent Progess. Paper presented at 7$^{\text{th}}$ International Forum on Reservoir Simulation, Bühl/Baden-Baden, Germany, June 23-27, 2003.

Edwards, C. H. & Penney, D. E. (1998), *Calculus with Analytic Geometry*, Prentice Hall.

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

Evensen, G. (2003), 'The Ensemble Kalman Filter: Theoretical formulation and practical implementation', *Ocean Dynamics* pp. 343–367.

Evensen, G. (2007), *Data Assimilation. The Ensemble Kalman Filter.*, Springer.

Evensen, G., Burgers, G. & van Leeuwen, P. J. (1998), 'Analysis Scheme in the Ensemble Kalman Filter', *Monthly weather review* .

Ewing, R. E., ed. (1983), *The Mathematics of Reservoir Simulation*, SIAM: Philadelphia.

Ewing, R. & Lin, T. (1991), 'A class of parameter estimation techniques for fluid flow in porous media', *Adv. Water Resources* .

Farmer, C. L. (2002), 'Upscaling: a review', *International Journal for Numerical Methods in Fluids* .

Gassmann, F. (1951), 'Über die Elastizitat poroser Medien', *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich* **96**.

GeoQuest (2004), *ECLIPSE Reference Manual 2004A*, Schlumberger GeoQuest.

Golub, G. & van Loan, C. (1996), *Matrix Computations*, Johns Hopkins University Press.

Hegstad, B. K. & Omre, H. (2001), 'Uncertainty in production forecasts based on well observations, seismic data and production history', *Society of Petroleum Engineers Journal* pp. 409–425.

Hoversten, M. G., Cassassuce, F., Gasperikova, E., Newman, G. A., Chen, J., Rubin, Y., Hou, Z. & Vasco, D. (2006), 'Direct reservoir parameter estimation using joint inversion of marine seismic AVA and CSEM data', *Geophysics* **71**(3).

Høyland, A. (1988), *Sannsynlighetsregning*, Tapir.

Iserles, A., ed. (1996), *A first Course in the Numerical Analysis of Differential Equations*, Cambridge University Press.

Johnson, R. A. & Wichern, D. W. (2002), *Applied Multivariate Statistical Analysis*, Prentice Hall.

Kalman, R. E. (1960), 'A New Approach to Linear Filtering and Prediction Problems', *Transactions of the ASME - Journal of Basic Engineering* .

Keoderitz, L. F. (2005), *Lecture Notes on Applied Reservoir Simulation*, World Scientific.

Larsen, A. L., Ulvmoen, M., Omre, H. & Buland, A. (2006), 'Bayesian lithology/fluid prediction and simulation on the basis of a Markov-chain prior model', *Geophysics* **71**(5).

Lødøen, O. P. & Omre, H. (2005), 'Scale-corrected ensemble kalman filtering applied to production history conditioning in reservoir evaluation'. Paper submitted for publication.

Lødøen, O. P., Omre, H., Durlofsky, L. J. & Chen, Y. (2004), Assessment of Uncertainty in Reservoir Production Forcasts Using Upscaled Flow Models, *in* 'Proceedings from Geostat 2004; Banff, Canada'.

Lee, N. D. & Zidek, J. V. (1992), 'Interpolation with Uncertain Spatial Covariances: A Bayesian Alternative to Kriging', *Journal of Multivariate Analysis* **43**, 341–374.

Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979), *Multivariant Analysis*, Academic Press, London.

Mavko, G., Mukweji, T. & Dvorkin, J. (1998), *The Rock Physics Handbook*, Cambridge University Press.

Omre, H. (2000), Stochastic reservoir models conditioned to non-linear production history observations, *in* 'Geostatistics 2000, Cape Town', pp. 166–175. Proceedings of the Sixth International Geostatistical Congress held in Cape Town, South Africa, in April 2000.

Omre, H. & Lødøen, O. P. (2004), 'Improved production forecasts and history matching using approximate fluid-flow simulators', *Society of Petroleum Engineers Journal* **9**, 339–351.

Omre, H. & Tjelmeland, H. (1996), 'Petroleum geostatistics'.

Peaceman, D. W. (1977), *Fundamentals of numerical reservoir simulation (Developments in petroleum science)*, Elsevier Scientific Pub. Co.

Ripley, B. D. (1996), *Pattern Recognition and Neural Networks*, Cambridge University Press.

Ross, S. M. (2003), *Probability Models*, Academic Press.

Scales, J. A. & Snieder, R. (1997), 'To Bayes or not to Bayes?', *Geophysics* .

Sheriff, R. E. & Geldart, L. P. (1995), *Exploration Seismology*, Cambridge University Press.

Stolt, R. H. & Weglein, A. B. (1985), 'Migration and inversion of seismic data', *Geophysics* pp. 2458–2472.

Strang, G. (1988), *Linear Algebra And Its Applications*, Thomson Learning.

Tarantola, A. (2005), *Inverse Problem Theory*, Society for Industrial and Applied Mathematics.

Walpole, R. E., Myers, R. H., Myers, S. L. & Ye, K. (2002), *Probability and Statistics for Engineers and Scientists*, Prentice Hall.

White, R. E. (1986), *An Introduction to the Finite Element Method With Applications to Nonlinear Problems*, Wiley-Interscience.

Zoeppritz, K. (1919), 'Über Reflexion und Durchgang seismicher Wellen durch Unstetigkeitsflächen', Über Erdbebenwellen, VII B, Nachrichten Göttingen, Matematisch physikaliche Klasse.

13 Appendix

A Theoretical Results

First we present three definitions of the multivariate expectation and covariance.

Definition 1. The expectation of a multivariate random vector $\boldsymbol{x}$ given as

$$\boldsymbol{\mu_x} = E[\boldsymbol{x}] = \int \boldsymbol{x} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x},$$

Definition 2. The covariance of a multivariate random vector $\boldsymbol{x}$ is given as

$$\boldsymbol{\Sigma_x} = Cov(\boldsymbol{x}) = E\left[(\boldsymbol{x} - \boldsymbol{\mu_x})(\boldsymbol{x} - \boldsymbol{\mu_x})^T\right].$$

Definition 3. The covariance between two multivariate random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is given as

$$\boldsymbol{\Sigma_{xy}} = Cov(\boldsymbol{x}, \boldsymbol{y}) = E\left[(\boldsymbol{x} - \boldsymbol{\mu_x})(\boldsymbol{y} - \boldsymbol{\mu_y})^T\right].$$

A.1 Multivariate Probability Density Functions

We begin this Appendix by introducing some multivariate distributions, and their probability density functions (pdfs).

Definition 4 (Gaussian distribution)**.** Consider a k -dimensional random vector $\boldsymbol{y}$. Then

$$\boldsymbol{y} \sim \mathrm{Gauss}_k(\boldsymbol{\mu_y}, \boldsymbol{\Sigma_y}),$$

means that $\boldsymbol{y}$ has a multivariate Gaussian probability density function (pdf), with expectation $\boldsymbol{\mu_y}$ and covariance matrix $\boldsymbol{\Sigma_y}$. That is (see e.g Johnson & Wichern (2002))

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma_y}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu_y})^T {\boldsymbol{\Sigma_y}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu_y})\right\}. \tag{80}$$

Definition 5 (Wishart distribution)**.** Let

$$\boldsymbol{x}_1, \ldots \boldsymbol{x}_\nu \overset{\text{i.i.d}}{\sim} \mathrm{Gauss}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$$

Then

$$\boldsymbol{M} = \sum_{i=1}^{\nu} \boldsymbol{x}_i \boldsymbol{x}_i^T,$$

is said to have the Wishart distribution with scale matrix $\boldsymbol{\Sigma}$, and ν degrees of freedom. That is, $\boldsymbol{M} \sim W_p(\boldsymbol{\Sigma}, \nu)$. The pdf of $\boldsymbol{M}$ is then given by

$$f(\boldsymbol{M}) = \frac{|\boldsymbol{M}|^{-(\nu-p-1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M})\right\}}{2^{\nu p/2}\pi^{p(p-1)/4}|\boldsymbol{\Sigma_Z}|^{\nu/2}\prod_{i=1}^{p}\Gamma\left(\frac{1}{2}(\nu+1-i)\right)},$$

where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x. \tag{81}$$

and tr(•) denotes the trace operator, which in the case of a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is given by

$$\text{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}.$$

Definition 6 (Inverted Wishart distribution)**.** Let $\boldsymbol{M} \sim W_k(\boldsymbol{\Sigma}, \nu)$. Then $\boldsymbol{Z} = \boldsymbol{M}^{-1}$ is said to have the inverted Wishart distribution, $\boldsymbol{Z} \sim W_k^{-1}(\boldsymbol{\Sigma}, \nu)$, with corresponding pdf (Mardia et al. 1979),

$$f(\boldsymbol{Z}) = \frac{|\boldsymbol{Z}|^{-(\nu+k+1)/2} \exp\left\{-\frac{1}{2}\text{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}^{-1})\right\}}{2^{\nu k/2}\pi^{k(k-1)/4}|\boldsymbol{\Sigma}|^{\nu/2}\prod_{i=1}^{k}\Gamma\left(\frac{1}{2}(\nu+1-i)\right)}. \tag{82}$$

Definition 7 (Multivariate student t)**.** Let $\boldsymbol{w}$ be a k -dimensional random vector. Then

$$\boldsymbol{w} \sim t_k(\boldsymbol{\mu_w}, \boldsymbol{\Sigma_w}, \nu)$$

means that $\boldsymbol{w}$ has the multivariate student t pdf. That is (see e.g Mardia et al. (1979))

$$f(\boldsymbol{w}) = \frac{\Gamma((\nu+k)/2)|\boldsymbol{\Sigma_w}|^{-1/2}}{(\pi\nu)^{k/2}\Gamma(\nu/2)[1+\nu^{-1}(\boldsymbol{w}-\boldsymbol{\mu_w})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{w}-\boldsymbol{\mu_w})]^{(\nu+k)/2}}. \tag{83}$$

Definition 8 (Multivariate log-Gaussian)**.** Let $\boldsymbol{x}$ be a p-dimensional vector having the Gaussian distribution with parameters $\boldsymbol{\mu_x}$ and $\boldsymbol{\Sigma_x}$. Then $\boldsymbol{u} = \exp\{\boldsymbol{x}\}$ is said to have the multivariate log-Gaussian distribution with parameters $\boldsymbol{\mu_u}$ and $\boldsymbol{\Sigma_u}$, denoted by $\boldsymbol{u} \sim \Lambda_p(\boldsymbol{\mu_u}, \boldsymbol{\Sigma_u})$.

A.2 Properties of Some Multivariate Distributions

We now outline some important results for the multivariate distribution above.

Result 1 (Linear Combinations)**.** *Let* $\boldsymbol{x} \sim \text{Gauss}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Further, let* $\boldsymbol{A} \in \mathbb{R}^{g\times p}$*, then*

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \sim \text{Gauss}_g(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T).$$

Proof. See (Johnson & Wichern 2002, Chapter 4) □

Result 2. *Let*

$$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \sim \text{Gauss}_p\left(\begin{bmatrix} \boldsymbol{\mu_x} \\ \boldsymbol{\mu_y} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma_x} & \boldsymbol{\Sigma_{xy}} \\ \boldsymbol{\Sigma_{yx}} & \boldsymbol{\Sigma_y} \end{bmatrix}\right).$$

Then

$$\begin{aligned} \boldsymbol{x} &\sim \text{Gauss}_g(\boldsymbol{\mu_x}, \boldsymbol{\Sigma_x}) \\ \boldsymbol{y} &\sim \text{Gauss}_m(\boldsymbol{\mu_y}, \boldsymbol{\Sigma_y}) \\ \boldsymbol{x}|\boldsymbol{y} &\sim \text{Gauss}_g(\boldsymbol{\mu_{x|y}}, \boldsymbol{\Sigma_{x|y}}), \end{aligned}$$

where

$$\boldsymbol{\mu_{x|y}} = \boldsymbol{\mu_x} + \boldsymbol{\Sigma_{xy}}\boldsymbol{\Sigma_y}^{-1}(\boldsymbol{y}-\boldsymbol{\mu_y})$$

and

$$\boldsymbol{\Sigma_{x|y}} = \boldsymbol{\Sigma_x} - \boldsymbol{\Sigma_{xy}}\boldsymbol{\Sigma_y}^{-1}\boldsymbol{\Sigma_{yx}}$$

Proof. See (Mardia et al. 1979, Chapter 3). □

Result 3 (Properties of the Wishart distribution)**.** *Let $\boldsymbol{U} \in \mathbb{R}^{k\times k}$ be a random matrix having the Wishart distribution with parameters $\boldsymbol{\Sigma_U}$ and ν. Then*

$$E[\boldsymbol{U}] = \nu\boldsymbol{\Sigma_U} \tag{84}$$

Proof. From Definition 5, we see that

$$E[\boldsymbol{U}] = \sum_{i=1}^{\nu} E[\boldsymbol{x}_i\boldsymbol{x}_i^T] = \nu\boldsymbol{\Sigma_U}$$

□

Result 4 (Properties of the inverted Wishart distribution)**.** *Let $\boldsymbol{U} \in \mathbb{R}^{k\times k}$ be a random matrix having the inverted Wishart distribution with parameters $\boldsymbol{\Sigma_U}$ and ν. Then*

$$E[\boldsymbol{U}] = \frac{{\boldsymbol{\Sigma_U}}^{-1}}{\nu - k - 1} \tag{85}$$

Proof. See Mardia et al. (1979, Chapter 3) or Anderson (2003, Chapter 6). □

Result 5. *Given a random variable $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$ with corresponding pdf $f(\boldsymbol{x})$. Assume there exists functions $g : \boldsymbol{x} \to \boldsymbol{y}$ and $h : \boldsymbol{y} \to \boldsymbol{x}$ such that the transformation $\boldsymbol{y} = g(\boldsymbol{x})$ and $\boldsymbol{x} = h(\boldsymbol{y}) = g^{-1}(\boldsymbol{y})$ is one-to-one. Then the pdf of $\boldsymbol{y} \in \mathbb{R}^{n\times 1}$ is given by*

$$f(\boldsymbol{y}) = f\left(g^{-1}(\boldsymbol{y})\right)|J| \tag{86}$$

where J is the determinant of the Jacobian given as

$$J = \left|\frac{\partial g^{-1}(\boldsymbol{y})}{\partial \boldsymbol{y}}\right|,$$

and $|J|$ denotes the absolute value of J.

Proof. See Edwards & Penney (1998) □

Result 6 (Properties of the multivariate log-Gaussian distribution)**.** *Let $\boldsymbol{u} \sim \Lambda_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then the pdf of $\boldsymbol{u}$ is given by*

$$f(\boldsymbol{u}) = \frac{|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{p/2}} \prod_{i=1}^{p} \frac{1}{u_i} \exp\left\{-\frac{1}{2}(\log(\boldsymbol{u}) - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\log(\boldsymbol{u}) - \boldsymbol{\mu})\right\}$$

Moreover, the expectation is given by

$$\boldsymbol{\mu_u} = \exp\left\{\boldsymbol{\mu} + \frac{\mathrm{diag}_p(\boldsymbol{\Sigma})}{2}\right\},$$

where $diag_p(\boldsymbol{\Sigma})$ is a p -dimensional vector containing the diagonal elements of $\boldsymbol{\Sigma}$, and element (i, j) of the covariance matrix is given by

$$\boldsymbol{\Sigma}_{\boldsymbol{u}(i,j)} = \exp\left\{\mu_i + \mu_j + \frac{1}{2}(\sigma_{ii} + \sigma_{jj})\right\}[\exp\{\sigma_{ij}\} - 1]$$

Proof. By Definition 8 and Result 5

$$f(\boldsymbol{u}) = f(\boldsymbol{x}(\boldsymbol{u}))|J_{\boldsymbol{xu}}|$$

where $\boldsymbol{x} \sim \text{Gauss}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and

$$J_{\boldsymbol{xu}} = \begin{vmatrix} \frac{1}{u_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{u_2} & 0 & \dots & 0 \\ & & \ddots & & \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & \frac{1}{u_p} \end{vmatrix} = \prod_{i=1}^{p} \frac{1}{u_i}$$

Thus,

$$\frac{|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{p/2}} \prod_{i=1}^{p} \frac{1}{u_i} \exp\left\{-\frac{1}{2}(\log(\boldsymbol{u}) - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\log(\boldsymbol{u}) - \boldsymbol{\mu})\right\}.$$

Element i of the $\boldsymbol{\mu}$ is given by

$$E[u_i] = E[\exp\{x_i\}] = E\left[\exp\{\mathbf{1}_i^T \boldsymbol{x}\}\right]$$

where $\mathbf{1}_i$ is a vector where all elements are zero, except element i which is one. By the definition of the multivariate expectation (Anderson 2003), this gives

$$\begin{aligned} E[u_i] &= \int \frac{|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{p/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) + \mathbf{1}_i^T \boldsymbol{x}\right\} \mathrm{d}\boldsymbol{x} \\ &= \int \frac{|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{p/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \tilde{\boldsymbol{\mu}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \tilde{\boldsymbol{\mu}})\right\} \mathrm{d}\boldsymbol{x} \\ &\quad \exp\left\{-\frac{1}{2}\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \frac{1}{2}\tilde{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\mu}}\right\}, \end{aligned}$$

where $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{1}_i$. Since the integral above simply is the integral of a multivariate Gaussian pdf, we get

$$\begin{aligned} E[u_i] &= \exp\left\{-\frac{1}{2}\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \frac{1}{2}\tilde{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\mu}}\right\} \\ &= \exp\left\{\mathbf{1}_i^T \boldsymbol{\mu} + \frac{\mathbf{1}_i^T \boldsymbol{\Sigma} \mathbf{1}_i}{2}\right\} \\ &= \exp\left\{\mu_i + \frac{\sigma_{ii}}{2}\right\}. \end{aligned}$$

Thus,

$$E[\boldsymbol{u}] = \exp\left\{\boldsymbol{\mu} + \frac{\text{diag}_p(\boldsymbol{\Sigma})}{2}\right\}.$$

By a similar calculation as above,

$$\begin{aligned} E[u_i u_j] &= E\left[\exp\{(\mathbf{1}_i + \mathbf{1}_j)^T \boldsymbol{x}\}\right] \\ &= \int \frac{|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{p/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \check{\boldsymbol{\mu}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \check{\boldsymbol{\mu}})\right\} \mathrm{d}\boldsymbol{x} \\ &\quad \exp\left\{-\frac{1}{2}\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \frac{1}{2}\check{\boldsymbol{\mu}}^T \boldsymbol{\Sigma}^{-1} \check{\boldsymbol{\mu}}\right\}, \end{aligned}$$

where $\check{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\Sigma}(\mathbf{1}_i + \mathbf{1}_j)$. Since $\text{Cov}(u_i, u_j) = E[u_i u_j] - E[u_i]E[u_j]$, we get

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{u}(i,j)} &= \exp\left\{(\mathbf{1}_i + \mathbf{1}_j)^T\boldsymbol{\mu} + \frac{\mathbf{1}_i^T\boldsymbol{\Sigma}\mathbf{1}_i + \mathbf{1}_j^T\boldsymbol{\Sigma}\mathbf{1}_j}{2} + \mathbf{1}_i^T\boldsymbol{\Sigma}\mathbf{1}_j\right\} \\ &\quad - \exp\left\{(\mathbf{1}_i + \mathbf{1}_j)^T\boldsymbol{\mu} + \frac{\mathbf{1}_i^T + \mathbf{1}_i^T\boldsymbol{\Sigma} + \mathbf{1}_j\boldsymbol{\Sigma}\mathbf{1}_j}{2}\right\} \\ &= \exp\left\{\mu_i + \mu_j + \frac{1}{2}(\sigma_{ii} + \sigma_{jj})\right\}[\exp\{\sigma_{ij}\} - 1]\end{aligned}$$

□

Result 7 (Linear combinations of inverted Wishart distributions)**.** *Assume*

$$\boldsymbol{\Sigma} \sim W_u^{-1}(\boldsymbol{\Psi}, \nu)$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \sim W_g^{-1}(\boldsymbol{\Psi}_{\boldsymbol{\epsilon}}, \nu)$$

Further, let

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\boldsymbol{B}^T \\ \boldsymbol{B}\boldsymbol{\Sigma} & \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \end{bmatrix},$$

where $\boldsymbol{B} \in \mathbb{R}^{g\times u}$*, and*

$$\boldsymbol{R} = \begin{bmatrix} -\boldsymbol{B}^T\boldsymbol{M}_{12}^T - \boldsymbol{M}_{12}\boldsymbol{B} & \boldsymbol{M}_{12} \\ \boldsymbol{M}_{12}^T & \mathbf{0} \end{bmatrix},$$

with

$$\boldsymbol{M}_{12}^T = \sum_{i=1}^{\nu} \boldsymbol{y}_i \boldsymbol{z}_i^T,$$

$\boldsymbol{y}_i \sim \text{Gauss}_u(\mathbf{0}, \boldsymbol{\Psi})$*, and* $\boldsymbol{z}_j \sim \text{Gauss}_g(\mathbf{0}, \boldsymbol{\Psi}_{\boldsymbol{\epsilon}})$*, independent* $\forall i, j$*.*

Then

$$[\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}{}^{-1} + \boldsymbol{R}]^{-1} \sim W_{u+g}^{-1}(\boldsymbol{\Psi}_{\boldsymbol{\mathfrak{X}}}, \nu),$$

where

$$\boldsymbol{\Psi}_{\boldsymbol{\mathfrak{X}}} = \begin{bmatrix} \boldsymbol{\Psi} + \boldsymbol{B}^T\boldsymbol{\Psi}_{\boldsymbol{\epsilon}}\boldsymbol{B} & -\boldsymbol{B}^T\boldsymbol{\Psi}_{\boldsymbol{\epsilon}} \\ -\boldsymbol{\Psi}_{\boldsymbol{\epsilon}}\boldsymbol{B} & \boldsymbol{\Psi}_{\boldsymbol{\epsilon}} \end{bmatrix},$$

Proof. By Definition 6, $\boldsymbol{M} = \boldsymbol{\Sigma}^{-1} \sim W_u(\boldsymbol{\Psi}, \nu)$ and $\boldsymbol{M}_{\boldsymbol{\epsilon}} = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}{}^{-1} \sim W_u(\boldsymbol{\Psi}_{\boldsymbol{\epsilon}}, \nu)$. In addition, by Definition 5,

$$\boldsymbol{M} = \sum_{i=1}^{\nu} \boldsymbol{y}_i \boldsymbol{y}_i^T$$

and

$$\boldsymbol{M}_{\boldsymbol{\epsilon}} = \sum_{i=1}^{\nu} \boldsymbol{z}_i \boldsymbol{z}_i^T,$$

where

$$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_\nu \overset{\text{i.i.d}}{\sim} \text{Gauss}_u(\boldsymbol{0}, \boldsymbol{\Psi})$$
$$\boldsymbol{z}_1, \ldots, \boldsymbol{z}_\nu \overset{\text{i.i.d}}{\sim} \text{Gauss}_g(\boldsymbol{0}, \boldsymbol{\Psi}_{\boldsymbol{\epsilon}}).$$

By (Mardia et al. 1979, Appendix A),

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}{}^{-1} = \begin{bmatrix} \boldsymbol{M} + \boldsymbol{B}^T \boldsymbol{M}_{\boldsymbol{\epsilon}} \boldsymbol{B} & -\boldsymbol{B}^T \boldsymbol{M}_{\boldsymbol{\epsilon}} \\ -\boldsymbol{M}_{\boldsymbol{\epsilon}} \boldsymbol{B} & \boldsymbol{M}_{\boldsymbol{\epsilon}} \end{bmatrix},$$

Now let $\boldsymbol{x}_i = \boldsymbol{y}_i - \boldsymbol{B}^T \boldsymbol{z}_i$, then by Result 2,

$$\boldsymbol{w}_i = \begin{bmatrix} \boldsymbol{x}_i \\ \boldsymbol{z}_i \end{bmatrix} \sim \text{Gauss}_p \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Psi} + \boldsymbol{B}^T \boldsymbol{\Psi}_{\boldsymbol{\epsilon}} \boldsymbol{B} & -\boldsymbol{B}^T \boldsymbol{\Psi}_{\boldsymbol{\epsilon}} \\ -\boldsymbol{\Psi}_{\boldsymbol{\epsilon}} \boldsymbol{B} & \boldsymbol{\Psi}_{\boldsymbol{\epsilon}} \end{bmatrix} \right).$$

Thus,

$$\boldsymbol{M}_{\boldsymbol{\mathcal{X}}} = \sum_{i=1}^{\nu} \boldsymbol{w}_i \boldsymbol{w}_i^T = \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}{}^{-1} + \boldsymbol{R} \sim W_p(\boldsymbol{\Psi}_{\boldsymbol{\mathcal{X}}}, \nu),$$

and finally

$$[\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}{}^{-1} + \boldsymbol{R}]^{-1} \sim W_p^{-1}(\boldsymbol{\Psi}_{\boldsymbol{\mathcal{X}}}, \nu)$$

□

A.3 General Results

We now outline some Definitions, and general results needed in the proofs of the main theorems.

Definition 9. The Vec-operator applied on a matrix $\boldsymbol{x} \in \mathbb{R}^{n \times m}$ gives

$$\text{Vec}(\boldsymbol{X})^T = [X_{11}, \ldots, X_{n1}, X_{12}, \ldots, X_{n2}, \ldots, X_{1m}, \ldots, X_{nm}]$$

Also let $\text{Vec}^{-1}\text{Vec}(\boldsymbol{X}) = \boldsymbol{X}$.

Definition 10. Given two matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{B} \in \mathbb{R}^{m \times m}$. Then the Kronecker product $\boldsymbol{A} \otimes \boldsymbol{B} \in \mathbb{R}^{nm \times nm}$, and is given by

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A} b_{11} & \boldsymbol{A} b_{12} & \ldots & \boldsymbol{A} b_{1m} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{A} b_{m1} & \boldsymbol{A} b_{m2} & \ldots & \boldsymbol{A} b_{mm} \end{bmatrix}.$$

Definition 11. We say that multivariate random variable $\boldsymbol{x}$ is statistically independent of $\boldsymbol{y}$ if

$$f(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}) f(\boldsymbol{y}) \quad \forall \boldsymbol{x}, \boldsymbol{y}$$

Result 8. *Given three matrices* $\boldsymbol{A}$, $\boldsymbol{B}$ $\boldsymbol{C}$ *and* $\boldsymbol{D}$ *such that* $\boldsymbol{A}^T \boldsymbol{B} \boldsymbol{C} \boldsymbol{D}^T$ *exists. Then*

$$\text{tr}(\boldsymbol{A}^T \boldsymbol{B} \boldsymbol{C} \boldsymbol{D}^T) = \text{Vec}(\boldsymbol{A})^T (\boldsymbol{D} \otimes \boldsymbol{B}) \text{Vec}(\boldsymbol{C}) \tag{87}$$

Proof. Using properties of the Kronecker product and the Vec operator (Brewer 1978), namely

$$\begin{aligned}\text{tr}(\boldsymbol{AXB}) &= \text{Vec}(\boldsymbol{A}^T)^T(\boldsymbol{I}\otimes\boldsymbol{X})\text{Vec}(\boldsymbol{B}),\\ (\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D}) &= (\boldsymbol{AC}\otimes\boldsymbol{BD})\end{aligned}$$

and

$$\begin{aligned}\text{Vec}(\boldsymbol{AB}) &= (\boldsymbol{I}\otimes\boldsymbol{A})\text{Vec}(\boldsymbol{B})\\ &= (\boldsymbol{B}^T\otimes\boldsymbol{I})\text{Vec}(\boldsymbol{A}),\end{aligned}$$

we get

$$\begin{aligned}\text{tr}(\boldsymbol{A}^T\boldsymbol{BCD}^T) &= \text{Vec}(\boldsymbol{A})^T\text{Vec}(\boldsymbol{BCD}^T)\\ &= \text{Vec}(\boldsymbol{A})^T(\boldsymbol{D}\otimes\boldsymbol{I})\text{Vec}(\boldsymbol{BC})\\ &= \text{Vec}(\boldsymbol{A})^T(\boldsymbol{D}\otimes\boldsymbol{I})(\boldsymbol{I}\otimes\boldsymbol{B})\text{Vec}(\boldsymbol{C})\\ &= \text{Vec}(\boldsymbol{A})^T(\boldsymbol{D}\otimes\boldsymbol{B})\text{Vec}(\boldsymbol{C})\end{aligned}$$

□

Result 9.

$$E[\boldsymbol{X}] = E[E[\boldsymbol{X}|\boldsymbol{Y}]] \tag{88}$$

Proof. For the univariate case, this is shown in Ross (2003, Chapter 3). From the definition of the expectation of a vector, or matrix (Johnson & Wichern 2002, Chapter 2), the result follows. □

Result 10.

$$Cov(\boldsymbol{X}) = E[Cov(\boldsymbol{X}|\boldsymbol{Y})] + Cov(E[\boldsymbol{X}|\boldsymbol{Y}]) \tag{89}$$

Proof.

$$\begin{aligned}Cov(\boldsymbol{X}) &= E[(\boldsymbol{X}-E[\boldsymbol{X}])(\boldsymbol{X}-E[\boldsymbol{X}])^T]\\ &= E[\boldsymbol{XX}^T] - E[\boldsymbol{X}]E[\boldsymbol{X}]^T.\end{aligned}$$

Moreover,

$$\begin{aligned}E[Cov(\boldsymbol{X}|\boldsymbol{Y})] &= E\left[E[(\boldsymbol{X}|\boldsymbol{Y}-E[\boldsymbol{X}|\boldsymbol{Y}])(\boldsymbol{X}|\boldsymbol{Y}-E[\boldsymbol{X}|\boldsymbol{Y}])^T]\right]\\ &= E\left[E[(\boldsymbol{X}|\boldsymbol{Y})(\boldsymbol{X}|\boldsymbol{Y})^T]\right] - E[\boldsymbol{X}|\boldsymbol{Y}]E[\boldsymbol{X}|\boldsymbol{Y}]^T\\ &= E[\boldsymbol{XX}^T] - E[\boldsymbol{X}|\boldsymbol{Y}]E[\boldsymbol{X}|\boldsymbol{Y}]^T,\end{aligned}$$

where the last expression follows since for a matrix $\boldsymbol{X}\in\mathbb{R}^{n\times p}$

$$\begin{aligned}E\left[E[(\boldsymbol{X}|\boldsymbol{Y})(\boldsymbol{X}|\boldsymbol{Y})^T]\right]_{i,j} &= E\left[E\left[\sum_{k=1}^{p}X_{ik}X_{jk}|\boldsymbol{Y}\right]\right]\\ &= E\left[\sum_{k=1}^{p}E[X_{ik}X_{jk}|\boldsymbol{Y}]\right]\\ &= \sum_{k=1}^{p}E[X_{ik}X_{jk}]\\ &= E[\boldsymbol{XX}^T]_{i,j},\quad \forall(i,j)\in\{(1,\ldots,n)\times(1,\ldots,p)\}.\end{aligned}$$

Again using Expression (88) we get

$$\begin{aligned} Cov(E[\boldsymbol{X}|\boldsymbol{Y}]) &= E\left[(E[\boldsymbol{X}|\boldsymbol{Y}] - E[E[\boldsymbol{X}|\boldsymbol{Y}]])(E[\boldsymbol{X}|\boldsymbol{Y}] - E[E[\boldsymbol{X}|\boldsymbol{Y}]])^T\right] \\ &= E[\boldsymbol{X}|\boldsymbol{Y}]E[\boldsymbol{X}|\boldsymbol{Y}]^T - E[\boldsymbol{X}]E[\boldsymbol{X}]^T. \end{aligned}$$

The result follows by inserting these last two expressions into Expression (89). □

Theorem 1. *Let*

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

have the inverted Wishart distribution with parameters

$$\boldsymbol{\Psi}^{-1} = \begin{bmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{12}^T & \boldsymbol{\Psi}_{22} \end{bmatrix}$$

and ν. Further, let

$$\begin{aligned} \boldsymbol{\Sigma}_{1|*} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T \\ \boldsymbol{T} &= \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \boldsymbol{\Sigma}_{*} &= \boldsymbol{\Sigma}_{22}. \end{aligned}$$

Then

$$\begin{aligned} \boldsymbol{\Sigma}_{1|*} &\sim W_u^{-1}(\boldsymbol{\Psi}_{1|*}, \nu) \\ \mathrm{Vec}(\boldsymbol{T})|\boldsymbol{\Sigma}_{1|*} &\sim \mathrm{Gauss}_{ug}(\mathrm{Vec}(\boldsymbol{\Gamma}), (\boldsymbol{\Psi}_*^{-1} \otimes \boldsymbol{\Sigma}_{1|*})) \\ \boldsymbol{\Sigma}_{*} &\sim W_g^{-1}(\boldsymbol{\Psi}_*, \nu - u), \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\Psi}_{1|*} &= \boldsymbol{\Psi}_{11} - \boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{12}^T \\ \boldsymbol{\Gamma} &= \boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_{22}^{-1} \\ \boldsymbol{\Psi}_{*} &= \boldsymbol{\Psi}_{22}. \end{aligned}$$

Moreover, $\boldsymbol{\Sigma}_*$ is statistically independent of $\boldsymbol{\Sigma}_{1|*}$ and $\mathrm{Vec}(\boldsymbol{T})|\boldsymbol{\Sigma}_{1|*}$ when $\nu > u + g + 1$.

Remark. This result is different from the one obtained by Caselton et al. (1992), and used in Lee & Zidek (1992). They claimed that

$$\mathrm{Vec}(\boldsymbol{T})|\boldsymbol{\Sigma}_{1|*} \sim \mathrm{Gauss}_{ug}(\mathrm{Vec}(\boldsymbol{\Gamma}), (\boldsymbol{\Sigma}_{1|*} \otimes \boldsymbol{\Psi}_*^{-1}))$$

Proof. Since the block matrix $\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}$ only depends on the three matrices $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{22}$, we can write

$$f(\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}) = f(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{22})$$

As shown in Caselton et al. (1992), we can use the Bartlett decomposition (Bartlett 1933), to express $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ as

$$\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{T}\boldsymbol{\Delta}\boldsymbol{T}^T \\
&\equiv \begin{bmatrix} \boldsymbol{I}_u & \boldsymbol{T} \\ \boldsymbol{0} & \boldsymbol{I}_g \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{1|*} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_* \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_u & \boldsymbol{T} \\ \boldsymbol{0} & \boldsymbol{I}_g \end{bmatrix}^T \\
&= \begin{bmatrix} \boldsymbol{\Sigma}_{1|*} + \boldsymbol{T}\boldsymbol{\Sigma}_*\boldsymbol{T}^T & \boldsymbol{T}\boldsymbol{\Sigma}_* \\ \boldsymbol{\Sigma}_*\boldsymbol{T}^T & \boldsymbol{\Sigma}_* \end{bmatrix} \\
\boldsymbol{\Psi} &= \boldsymbol{N}\boldsymbol{\Lambda}\boldsymbol{N}^T \\
&\equiv \begin{bmatrix} \boldsymbol{I}_u & \boldsymbol{\Gamma} \\ \boldsymbol{0} & \boldsymbol{I}_g \end{bmatrix} \begin{bmatrix} \boldsymbol{\Psi}_{1|*} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Psi}_* \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_u & \boldsymbol{\Gamma} \\ \boldsymbol{0} & \boldsymbol{I}_g \end{bmatrix}^T,
\end{aligned}$$

where the different matrices are defined in the theorem. Now, by general properties of the determinant and the trace operator (Mardia et al. 1979, Appendix A),

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{1|2}^{-1} & -\boldsymbol{\Sigma}_{1|2}^{-1}\boldsymbol{T} \\ -\boldsymbol{T}^T\boldsymbol{\Sigma}_{1|2}^{-1} & \boldsymbol{\Sigma}_*^{-1} + \boldsymbol{T}^T\boldsymbol{\Sigma}_{1|2}^{-1}\boldsymbol{T} \end{bmatrix},$$

and

$$|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_*| \cdot |\boldsymbol{\Sigma}_{1|2}|.$$

Thus,

$$\begin{aligned}
f\left(\boldsymbol{\Sigma}_{11}(\bullet), \boldsymbol{\Sigma}_{12}(\bullet), \boldsymbol{\Sigma}_{22}(\bullet)\right) &\propto |\boldsymbol{\Sigma}|^{-(\nu+p+1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}^{-1}\boldsymbol{\Sigma}^{-1}\right)\right\} \\
&\propto |\boldsymbol{\Sigma}_*|^{-(\nu+p+1)/2} \times |\boldsymbol{\Sigma}_{1|*}|^{-(\nu+p+1)/2} \times \\
&\quad \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}_{1|*}\boldsymbol{\Sigma}_{1|*}^{-1} + \boldsymbol{\Psi}_*\boldsymbol{\Sigma}_*^{-1} + \boldsymbol{\Psi}_*(\boldsymbol{T}-\boldsymbol{\Gamma})^T\boldsymbol{\Sigma}_{1|*}^{-1}(\boldsymbol{T}-\boldsymbol{\Gamma})\right)\right\}
\end{aligned}$$

Then, by Result 5,

$$f(\boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}, \boldsymbol{\Sigma}_*) = f\left(\boldsymbol{\Sigma}_{11}(\bullet), \boldsymbol{\Sigma}_{12}(\bullet), \boldsymbol{\Sigma}_{22}(\bullet)\right)|J|,$$

where by Deemer & Olkin (1951) and Mardia et al. (1979, Appendix B)

$$\begin{aligned}
J &= \begin{vmatrix} \frac{\partial \boldsymbol{\Sigma}_{11}(\bullet)}{\partial \boldsymbol{\Sigma}_{1|*}} & \frac{\partial \boldsymbol{\Sigma}_{11}(\bullet)}{\partial \boldsymbol{T}} & \frac{\partial \boldsymbol{\Sigma}_{11}(\bullet)}{\partial \boldsymbol{\Sigma}_*} \\ \frac{\partial \boldsymbol{\Sigma}_{12}(\bullet)}{\partial \boldsymbol{\Sigma}_{1|*}} & \frac{\partial \boldsymbol{\Sigma}_{12}(\bullet)}{\partial \boldsymbol{T}} & \frac{\partial \boldsymbol{\Sigma}_{12}(\bullet)}{\partial \boldsymbol{\Sigma}_*} \\ \frac{\partial \boldsymbol{\Sigma}_{22}(\bullet)}{\partial \boldsymbol{\Sigma}_{1|*}} & \frac{\partial \boldsymbol{\Sigma}_{22}(\bullet)}{\partial \boldsymbol{T}} & \frac{\partial \boldsymbol{\Sigma}_{22}(\bullet)}{\partial \boldsymbol{\Sigma}_*} \end{vmatrix} \\
&= \begin{vmatrix} \boldsymbol{I}_{u^2} & (\star) & (\star\star) \\ \boldsymbol{0} & \mathrm{diag}_u(\boldsymbol{\Sigma}_*) & (\star\star\star) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_g^2 \end{vmatrix} \\
&= |\boldsymbol{I}_{u^2}||\mathrm{diag}_u(\boldsymbol{\Sigma}_*)||\boldsymbol{I}_{g^2}| \\
&= |\boldsymbol{\Sigma}_*|^u.
\end{aligned}$$

Here $\text{diag}_u(\boldsymbol{W})$ is a diagonal block matrix, where the u diagonal elements are the matrices $\boldsymbol{W}$. Hence,

$$f(\boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}, \boldsymbol{\Sigma}_*) \propto |\boldsymbol{\Sigma}|^{-(\nu-2u+p+1)/2} \times |\boldsymbol{\Sigma}_{1|*}|^{-(\nu+p+1)/2} \exp\left\{-\frac{1}{2}\text{tr}\left(\boldsymbol{\Psi}_{1|*}\boldsymbol{\Sigma}_{1|*}^{-1} + \boldsymbol{\Psi}_*\boldsymbol{\Sigma}_*^{-1} + \boldsymbol{\Psi}_*(\boldsymbol{T}-\boldsymbol{\Gamma})^T\boldsymbol{\Sigma}_{1|*}^{-1}(\boldsymbol{T}-\boldsymbol{\Gamma})\right)\right\} \tag{90}$$

From this last expression we see that

$$f(\boldsymbol{\Sigma}_*) \propto |\boldsymbol{\Sigma}|^{-(\nu-u+g+1)/2} \times \exp\left\{-\frac{1}{2}\text{tr}\left(\boldsymbol{\Psi}_*\boldsymbol{\Sigma}_*^{-1}\right)\right\},$$

which we recognise as the kernel of the inverted Wishart pdf. Therefore

$$\boldsymbol{\Sigma}_* \sim W_g^{-1}(\boldsymbol{\Psi}_*^{-1}, \nu-u)$$

Similarly, using Result 8 and other general properties of the Kronecker product (Brewer 1978, Mardia et al. 1979),

$$\begin{aligned} f(\boldsymbol{T}|\boldsymbol{\Sigma}_{1|*}) &\propto \exp\left\{-\frac{1}{2}\text{tr}\left(\boldsymbol{\Psi}_*(\boldsymbol{T}-\boldsymbol{\Gamma})^T\boldsymbol{\Sigma}_{1|*}^{-1}(\boldsymbol{T}-\boldsymbol{\Gamma})\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\text{tr}\left((\boldsymbol{T}-\boldsymbol{\Gamma})^T\boldsymbol{\Sigma}_{1|*}^{-1}(\boldsymbol{T}-\boldsymbol{\Gamma})\boldsymbol{\Psi}_*\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\text{tr}\left(\text{Vec}(\boldsymbol{T}-\boldsymbol{\Gamma})^T(\boldsymbol{\Psi}_*\otimes\boldsymbol{\Sigma}_{1|*}^{-1})\text{Vec}(\boldsymbol{T}-\boldsymbol{\Gamma})\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\text{tr}\left(\text{Vec}\left(\boldsymbol{T}-\boldsymbol{\Gamma}\right)^T\left(\boldsymbol{\Psi}_*^{-1}\otimes\boldsymbol{\Sigma}_{1|*}\right)^{-1}\text{Vec}\left(\boldsymbol{T}-\boldsymbol{\Gamma}\right)\right)\right\} \end{aligned}$$

Since the expression above is the kernel of a multivariate Gaussian pdf,

$$\text{Vec}(\boldsymbol{T})|\boldsymbol{\Sigma}_{1|*} \sim \text{Gauss}_{ug}\left(\text{Vec}(\boldsymbol{\Gamma}), \left(\boldsymbol{\Psi}_*^{-1}\otimes\boldsymbol{\Sigma}_{1|*}\right)\right).$$

Again using Expression (90) and properties of the Kronecker product (Anderson 2003)

$$\begin{aligned} f(\boldsymbol{\Sigma}_{1|*}) &\propto |\boldsymbol{\Sigma}_{1|*}|^{-(\nu+p+1)/2}\exp\left\{-\frac{1}{2}\text{tr}\left(\boldsymbol{\Psi}_{1|*}\boldsymbol{\Sigma}_{1|*}^{-1}\right)\right\} \\ &\quad\times\int\exp\left\{-\frac{1}{2}\text{tr}\left(\text{Vec}\left(\boldsymbol{T}-\boldsymbol{\Gamma}\right)^T\left(\boldsymbol{\Psi}_*^{-1}\otimes\boldsymbol{\Sigma}_{1|*}\right)^{-1}\text{Vec}\left(\boldsymbol{T}-\boldsymbol{\Gamma}\right)\right)\right\}\text{d}\boldsymbol{T} \\ &\propto |\boldsymbol{\Sigma}_{1|*}|^{-(\nu+p+1)/2}\exp\left\{-\frac{1}{2}\text{tr}\left(\boldsymbol{\Psi}_{1|*}\boldsymbol{\Sigma}_{1|*}^{-1}\right)\right\}\left|\left(\boldsymbol{\Psi}_*^{-1}\otimes\boldsymbol{\Sigma}_{1|*}\right)\right|^{-1/2} \\ &\propto |\boldsymbol{\Sigma}_{1|*}|^{-(\nu+u+1)/2}\exp\left\{-\frac{1}{2}\text{tr}\left(\boldsymbol{\Psi}_{1|*}\boldsymbol{\Sigma}_{1|*}^{-1}\right)\right\} \end{aligned}$$

Thus,

$$\boldsymbol{\Sigma}_{1|*} \sim W_u^{-1}(\boldsymbol{\Psi}_{1|*}^{-1}, \nu).$$

Finally, since

$$f(\boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}, \boldsymbol{\Sigma}_*) = f(\boldsymbol{T}|\boldsymbol{\Sigma}_{1|*})f(\boldsymbol{\Sigma}_{1|*})f(\boldsymbol{\Sigma}_*),$$

by Definition 11 the theorem follows. □

Result 11. *Assume*

$$\text{Vec}(\boldsymbol{T}) \sim \text{Gauss}_{ug}\left(\text{Vec}(\boldsymbol{\Gamma}), (\boldsymbol{A} \otimes \boldsymbol{\Sigma})\right)$$

Where $\boldsymbol{T} \in \mathbb{R}^{u\times g}$, $\boldsymbol{\Gamma} \in \mathbb{R}^{u\times g}$, $\boldsymbol{A} \in \mathbb{R}^{g\times g}$ *and* $\boldsymbol{\Sigma} \in \mathbb{R}^{u\times u}$. *Note that this implies that* $\boldsymbol{A} = \boldsymbol{A}^T$ *and* $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^T$. *Further, let* $\boldsymbol{B} \in \mathbb{R}^{g\times g}$ *be a symmetric matrix. Then*

$$E[\boldsymbol{T}\boldsymbol{B}\boldsymbol{T}^T] = \text{tr}(\boldsymbol{A}\boldsymbol{B})\boldsymbol{\Sigma} + \boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{\Gamma}^T$$

Proof. From Definitions 9 and 10,

$$\text{Vec}(\boldsymbol{T}) = \begin{bmatrix} \boldsymbol{T}_{(1)} \\ \boldsymbol{T}_{(2)} \\ \vdots \\ \boldsymbol{T}_{(g)} \end{bmatrix} \sim \text{Gauss}_{ug}\left(\begin{bmatrix} \boldsymbol{\Gamma}_{(1)} \\ \boldsymbol{\Gamma}_{(2)} \\ \vdots \\ \boldsymbol{\Gamma}_{(g)} \end{bmatrix}, \begin{bmatrix} a_{11}\boldsymbol{\Sigma} & a_{12}\boldsymbol{\Sigma} & \dots & a_{1g}\boldsymbol{\Sigma} \\ a_{21}\boldsymbol{\Sigma} & a_{22}\boldsymbol{\Sigma} & \dots & a_{2g}\boldsymbol{\Sigma} \\ \vdots & \vdots & \vdots & \vdots \\ a_{g1}\boldsymbol{\Sigma} & a_{g2}\boldsymbol{\Sigma} & \dots & a_{gg}\boldsymbol{\Sigma} \end{bmatrix}\right),$$

where $\boldsymbol{x}_{(i)}$ is the i'th column vector of the matrix $\boldsymbol{X}$, we have by Result 2

$$\begin{aligned} Cov(\boldsymbol{T}_{(i)}, \boldsymbol{T}_{(j)}) &= a_{ij}\boldsymbol{\Sigma} \\ E[\boldsymbol{T}_{(i)}] &= \boldsymbol{\Gamma}_{(i)} \end{aligned}$$

Thus,

$$\begin{aligned} E[\boldsymbol{T}\boldsymbol{B}\boldsymbol{T}^T] &= E\left[\sum_{i=1}^{g}\sum_{j=1}^{g} \boldsymbol{T}_{(j)} b_{ji} \boldsymbol{T}_{(i)}^T\right] \\ &= \sum_{i=1}^{g}\sum_{j=1}^{g} b_{ji} E\left[\boldsymbol{T}_{(j)}\boldsymbol{T}_{(i)}^T\right] \\ &= \sum_{i=1}^{g}\sum_{j=1}^{g} b_{ji}\left(Cov(\boldsymbol{T}_{(j)}, \boldsymbol{T}_{(i)}) + E[\boldsymbol{T}_{(j)}]E[\boldsymbol{T}_{(i)}]^T\right) \\ &= \sum_{i=1}^{g}\sum_{j=1}^{g} b_{ji}\left(a_{ji}\boldsymbol{\Sigma} + \boldsymbol{\Gamma}_{(j)}\boldsymbol{\Gamma}_{(i)}^T\right) \\ &= \sum_{i=1}^{g}\sum_{j=1}^{g} b_{ij}a_{ji}\boldsymbol{\Sigma} + \sum_{i=1}^{g}\sum_{j=1}^{g} \boldsymbol{\Gamma}_{(j)} b_{ji} \boldsymbol{\Gamma}_{(i)}^T \\ &= \text{tr}(\boldsymbol{A}\boldsymbol{B})\boldsymbol{\Sigma} + \boldsymbol{\Gamma}\boldsymbol{B}\boldsymbol{\Gamma}^T. \end{aligned}$$

□

Result 12. *The likelihood function in a multivariate Gaussian model can be written as*

$$f(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto f(\bar{\boldsymbol{x}}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \times g(\boldsymbol{S}; \boldsymbol{\Sigma}) \tag{91}$$

Proof. By adding and subtracting $\bar{\boldsymbol{x}}$ we get (Mardia et al. 1979)

$$\begin{aligned} f(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{1}{2}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})^T \left(\frac{\boldsymbol{\Sigma}}{n}\right)^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu})\right\} \\ \times \exp\left\{-\frac{1}{2}\text{tr}\left(n\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right)\right\} \\ \propto f(\bar{\boldsymbol{x}}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \times g(\boldsymbol{S}; \boldsymbol{\Sigma}), \end{aligned}$$

where $f(\bar{\boldsymbol{x}}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a pdf similar to that defined in Expression (106), and $g(\boldsymbol{S}; \boldsymbol{\Sigma})$ is a function dependent only on $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$. □

Remark. In the case when the number of samples is larger then the number of variables, $g(\ ;\ )$ is replaced by the pdf of $\boldsymbol{S}$, namely the Wishart (Mardia et al. 1979).

Result 13. *Consider a multivariate stochastic vector*

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}^* \end{bmatrix}$$

where $\boldsymbol{x} \in \mathbb{R}^{u \times 1}$ *and* $\boldsymbol{x}^* \in \mathbb{R}^{g \times 1}$ *Further, let* $\hat{\boldsymbol{x}} = g(\boldsymbol{x}^*)$ *be a predictor of* $\boldsymbol{x}$ *given* $\boldsymbol{x}^*$*. Then* $\hat{\boldsymbol{x}} = E[\boldsymbol{x}|\boldsymbol{x}^*]$ *minimises the Mean Squared Prediction Error (MSPE) given by*

$$E\left[(\boldsymbol{x} - g(\boldsymbol{x}^*))^T(\boldsymbol{x} - g(\boldsymbol{x}^*))\right]$$

Proof. By Bayes' rule,

$$f(\boldsymbol{x}, \boldsymbol{x}^*) = f(\boldsymbol{x}|\boldsymbol{x}^*)f(\boldsymbol{x}^*).$$

Hence,

$$\begin{aligned}
\text{MSPE} &= \int\int (\boldsymbol{x} - g(\boldsymbol{x}^*))^T(\boldsymbol{x} - g(\boldsymbol{x}^*))f(\boldsymbol{x}|\boldsymbol{x}^*)f(\boldsymbol{x}^*)\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{x}^* \\
&= \int f(\boldsymbol{x}^*)\left[\int (\boldsymbol{x} - g(\boldsymbol{x}^*))^T(\boldsymbol{x} - g(\boldsymbol{x}^*))f(\boldsymbol{x}|\boldsymbol{x}^*)\mathrm{d}\boldsymbol{x}\right]\mathrm{d}\boldsymbol{x}^* \\
&= \int \left[E[\boldsymbol{x}^T\boldsymbol{x}|\boldsymbol{x}^*] - 2g(\boldsymbol{x}^*)^T E[\boldsymbol{x}|\boldsymbol{x}^*] + g(\boldsymbol{x}^*)^T g(\boldsymbol{x}^*)\right] f(\boldsymbol{x}^*)\mathrm{d}\boldsymbol{x}^*.
\end{aligned}$$

Since $f(\boldsymbol{x}^*) > 0, \quad \forall \boldsymbol{x}^*$, the expression above is minimised by

$$\hat{\boldsymbol{x}} = \arg\ \min_{\boldsymbol{a}} \left\{E[\boldsymbol{x}^T\boldsymbol{x}|\boldsymbol{x}^*] - 2\boldsymbol{a}^T E[\boldsymbol{x}|\boldsymbol{x}^*] + \boldsymbol{a}^T\boldsymbol{a}\right\},$$

where $\boldsymbol{a} = g(\boldsymbol{x}^*)$.

Using the results of (Mardia et al. 1979, Appendix B), we get

$$\frac{\partial}{\partial \boldsymbol{a}}\left(E[\boldsymbol{x}^T\boldsymbol{x}|\boldsymbol{x}^*] - 2\boldsymbol{a}^T E[\boldsymbol{x}|\boldsymbol{x}^*] + \boldsymbol{a}^T\boldsymbol{a}\right)\Big|_{\boldsymbol{a}=\hat{\boldsymbol{x}}} = -2E[\boldsymbol{x}|\boldsymbol{x}^*] + 2\hat{\boldsymbol{x}} = 0.$$

Thus,

$$\hat{\boldsymbol{x}} = E[\boldsymbol{x}|\boldsymbol{x}^*].$$

□

B Ensemble Kalman Filter

In this section, we show that with the assumption of a Gaussian distribution in the forward model, the EnKF algorithm generates realisations from the correct posterior distribution. Note that here we only have considered the model where observations of the reservoir production properties $\boldsymbol{q}_t^o$ has been obtained, but a similar argument holds when we include both seismic observations $\boldsymbol{r}_t^o$, and reservoir production properties.

Let

$$\boldsymbol{x}_t = \begin{bmatrix} \boldsymbol{r}_t \\ \boldsymbol{q}_r \end{bmatrix} \in \mathbb{R}^{n_r+n_q},$$

be a vector containing the reservoir state, and production properties at time t. Assume that observations production properties of reservoir have been made prior to the current timestep.

Let

$$\boldsymbol{x}_t|\boldsymbol{r}_0^o, \boldsymbol{q}_1^o, \ldots, \boldsymbol{q}_{t-1}^o \sim \text{Gauss}_{n_r+n_q}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t).$$

Further, we assume

$$\begin{aligned} \boldsymbol{q}_t^o &= \boldsymbol{D}_t\boldsymbol{q}_t + \boldsymbol{\epsilon}_t \qquad (92) \\ \boldsymbol{r}_0^o &= \boldsymbol{D}_0\boldsymbol{r}_0 + \boldsymbol{\epsilon}_0, \end{aligned}$$

where

$$\boldsymbol{\epsilon}_t \sim \text{Gauss}_{n_q}(0, \boldsymbol{\Sigma}_t^o), \quad \text{independent of } \boldsymbol{x}_t, \boldsymbol{\epsilon}_s \quad \forall t, s \neq t, \qquad (93)$$

$$\boldsymbol{\epsilon}_0 \sim \text{Gauss}_{n_r}(0, \boldsymbol{\Sigma}_0^o), \quad \text{independent of } \boldsymbol{x}_t, \boldsymbol{\epsilon}_s \quad \forall t, s \neq t, \qquad (94)$$

$\boldsymbol{D}_t \in \mathbb{R}^{n_q \times n_q}$, $\quad i \in \{1, 2, \ldots\}$, and $\boldsymbol{D}_0 \in \mathbb{R}^{n_r \times n_r}$. Result 1 then gives

$$\begin{aligned} \boldsymbol{q}_t^o|\boldsymbol{q}_t &\sim \text{Gauss}_{n_q}(\boldsymbol{D}_t\boldsymbol{q}_t, \boldsymbol{\Sigma}_t^o), \quad \text{independent of } \boldsymbol{r}_t, \qquad (95) \\ \boldsymbol{r}_0^o|\boldsymbol{r}_0 &\sim \text{Gauss}_{n_r}(\boldsymbol{D}_0\boldsymbol{r}_0, \boldsymbol{\Sigma}_0^o), \quad \text{independent of } \boldsymbol{q}_0. \qquad (96) \end{aligned}$$

Using the definition of the multivariate expectation (Johnson & Wichern 2002) we can write

$$\boldsymbol{\mu}_t = \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{r}t} \\ \boldsymbol{\mu}_{\boldsymbol{q}t} \end{bmatrix}.$$

Similarly, using Definition 2, the covariance matrix of $\boldsymbol{x}_t$ can then be partitioned into

$$\boldsymbol{\Sigma}_t = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{r}t} & \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}t} \\ \boldsymbol{\Sigma}_{\boldsymbol{q}\boldsymbol{r}t} & \boldsymbol{\Sigma}_{\boldsymbol{q}t} \end{bmatrix}.$$

Result 14. *Let*

$$\boldsymbol{x}_t|\boldsymbol{r}_0^o, \boldsymbol{q}_1^o, \ldots, \boldsymbol{q}_{t-1}^o \sim \text{Gauss}_{n_r+n_q}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t),$$

and

$$\boldsymbol{q}_t^o|\boldsymbol{q}_t \sim \text{Gauss}_{n_q}(\boldsymbol{D}_t\boldsymbol{q}_t, \boldsymbol{\Sigma}_t^o), \quad \textit{independent of } \boldsymbol{r}_t.$$

Then

$$\begin{aligned}\boldsymbol{r}_t^c &= \boldsymbol{r}_t|\boldsymbol{r}_0^o, \boldsymbol{q}_1^o, \ldots, \boldsymbol{q}_{t-1}^o, \boldsymbol{q}_t^o \sim \text{Gauss}_{n_r}(\boldsymbol{\mu}_{\boldsymbol{r}_t^c}, \boldsymbol{\Sigma}_{\boldsymbol{r}_t^c}) \\ \boldsymbol{q}_t^c &= \boldsymbol{q}_t|\boldsymbol{r}_0^o, \boldsymbol{q}_1^o, \ldots, \boldsymbol{q}_{t-1}^o, \boldsymbol{q}_t^o \sim \text{Gauss}_{n_q}(\boldsymbol{\mu}_{\boldsymbol{q}_t^c}, \boldsymbol{\Sigma}_{\boldsymbol{q}_t^c})\end{aligned} \tag{97}$$

where

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{r}_t^c} &= \boldsymbol{\mu}_{\boldsymbol{r}t} + \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o - \boldsymbol{D}_t\boldsymbol{\mu}_{\boldsymbol{q}_t}) \\ \boldsymbol{\mu}_{\boldsymbol{q}_t^c} &= \boldsymbol{\mu}_{\boldsymbol{q}_t} + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o - \boldsymbol{D}_t\boldsymbol{\mu}_{\boldsymbol{q}_t}),\end{aligned} \tag{98}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{r}_t^c} &= \boldsymbol{\Sigma}_{\boldsymbol{r}t} - \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t \\ \boldsymbol{\Sigma}_{\boldsymbol{q}_t^c} &= \boldsymbol{\Sigma}_{\boldsymbol{q}_t} - \boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t\end{aligned} \tag{99}$$

Proof. By applying Result 1, to Expression (92), we know that

$$\boldsymbol{q}_t^o \sim \text{Gauss}_{n_q}(\boldsymbol{D}_t\boldsymbol{\mu}_{\boldsymbol{q}}, \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T + \boldsymbol{\Sigma}_t^o).$$

Then, by Result 2

$$\begin{aligned}\boldsymbol{r}_t^c &\sim \text{Gauss}_{n_r}(\boldsymbol{\mu}_{\boldsymbol{r}_t^c}, \boldsymbol{\Sigma}_{\boldsymbol{r}_t^c}) \\ \boldsymbol{q}_t^c &\sim \text{Gauss}_{n_q}(\boldsymbol{\mu}_{\boldsymbol{q}_t^c}, \boldsymbol{\Sigma}_{\boldsymbol{q}_t^c})\end{aligned}$$

with

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{r}_t^c} &= E[\boldsymbol{r}_t] + Cov(\boldsymbol{r}_t, \boldsymbol{q}_t^o)Cov(\boldsymbol{q}_t^o)^{-1}(\boldsymbol{q}_t^o - E[\boldsymbol{q}_t^o]) \\ \boldsymbol{\mu}_{\boldsymbol{q}_t^c} &= E[\boldsymbol{q}_t] + Cov(\boldsymbol{q}_t, \boldsymbol{q}_t^o)Cov(\boldsymbol{q}_t^o)^{-1}(\boldsymbol{q}_t^o - E[\boldsymbol{q}_t^o])\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{r}_t^c} &= Cov(\boldsymbol{r}_t) - Cov(\boldsymbol{r}_t, \boldsymbol{q}_t^o)Cov(\boldsymbol{q}_t^o)^{-1}Cov(\boldsymbol{r}_t, \boldsymbol{q}_t^o)^T \\ \boldsymbol{\Sigma}_{\boldsymbol{q}_t^c} &= Cov(\boldsymbol{q}_t) - Cov(\boldsymbol{q}_t, \boldsymbol{q}_t^o)Cov(\boldsymbol{q}_t^o)^{-1}Cov(\boldsymbol{q}_t, \boldsymbol{q}_t^o)^T.\end{aligned}$$

Since

$$\begin{aligned}Cov(\boldsymbol{q}_t, \boldsymbol{q}_t^o) &= Cov(\boldsymbol{q}_t, \boldsymbol{D}_t\boldsymbol{q}_t + \boldsymbol{\epsilon}) \\ &= \boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T,\end{aligned}$$

and

$$\begin{aligned}Cov(\boldsymbol{r}_t, \boldsymbol{q}_t^o) &= Cov(\boldsymbol{r}_t, \boldsymbol{D}_t\boldsymbol{q}_t + \boldsymbol{\epsilon}) \\ &= \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T,\end{aligned}$$

the result follows.

□

Result 15. *Given realisations*

$$\boldsymbol{x}_t = \begin{bmatrix} \boldsymbol{r}_t \\ \boldsymbol{q}_t \end{bmatrix} \sim f(\boldsymbol{x}_t|\boldsymbol{r}_0^o, \boldsymbol{q}_1^o, \ldots, \boldsymbol{q}_{t-1}^o) = \text{Gauss}_{n_r+n_q}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t).$$

Realisations

$$\boldsymbol{r}_t^c \sim f(\boldsymbol{r}_t|\boldsymbol{r}_0^o, \boldsymbol{q}_1^o, \ldots, \boldsymbol{q}_{t-1}^o, \boldsymbol{q}_t^o),$$

and

$$\boldsymbol{q}_t^c \sim f(\boldsymbol{q}_t|\boldsymbol{r}_0^o, \boldsymbol{q}_1^o, \ldots, \boldsymbol{q}_{t-1}^o, \boldsymbol{q}_t^o),$$

can be generated by

$$\begin{aligned} \boldsymbol{r}_t^c &= \boldsymbol{r}_t + \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t) \\ \boldsymbol{q}_t^c &= \boldsymbol{q}_t + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t), \end{aligned} \tag{100}$$

where

$$\boldsymbol{\epsilon}_t \sim \text{Gauss}_{n_q}(\boldsymbol{0}, \boldsymbol{\Sigma}_t^o).$$

Proof. Result 2 states that $\boldsymbol{r}_t^c$ and $\boldsymbol{q}_t^c$ both will follow the multivariate Gaussian distribution with expectation and Covariance given by

$$\begin{aligned} E[\boldsymbol{r}_t^c] &= E[\boldsymbol{r}_t + \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t)] \\ &= \boldsymbol{\mu}_{\boldsymbol{r}_t^c} \\ E[\boldsymbol{q}_t^c] &= E[\boldsymbol{q}_t + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t)] \\ &= \boldsymbol{\mu}_{\boldsymbol{q}_t^c} \end{aligned}$$

and

$$\begin{aligned} Cov(\boldsymbol{r}_t^c) &= Cov\left(\boldsymbol{r}_t + \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_t\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t)\right) \\ &= Cov(\boldsymbol{r}_t) + \boldsymbol{K}_{\boldsymbol{r}}Cov(\boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t)\boldsymbol{K}_{\boldsymbol{r}}{}^T + 2Cov\left(\boldsymbol{r}_t, \boldsymbol{K}_{\boldsymbol{r}}\left[\boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t\right]\right) \\ &= \boldsymbol{\Sigma}_{\boldsymbol{r}t} + \boldsymbol{K}_{\boldsymbol{r}}(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)\boldsymbol{K}_{\boldsymbol{r}}{}^T - 2\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T\boldsymbol{K}_{\boldsymbol{r}}{}^T \\ &= \boldsymbol{\Sigma}_{\boldsymbol{r}_t^c} \\ Cov(\boldsymbol{q}_t^c) &= Cov\left(\boldsymbol{q}_t + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_t\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t)\right) \\ &= Cov(\boldsymbol{q}_t) + \boldsymbol{K}_{\boldsymbol{q}}Cov(\boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t)\boldsymbol{K}_{\boldsymbol{q}}{}^T + 2Cov\left(\boldsymbol{q}_t, \boldsymbol{K}_{\boldsymbol{q}}\left[\boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t\right]\right) \\ &= \boldsymbol{\Sigma}_{\boldsymbol{q}_t} + \boldsymbol{K}_{\boldsymbol{q}}(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)\boldsymbol{K}_{\boldsymbol{q}}{}^T - 2\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T\boldsymbol{K}_{\boldsymbol{q}}{}^T \\ &= \boldsymbol{\Sigma}_{\boldsymbol{q}_t^c} \end{aligned}$$

Here

$$\boldsymbol{K}_{\boldsymbol{r}t} = \left[\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}\right],$$

and

$$\boldsymbol{K}_{\boldsymbol{q}_t} = \left[\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}\right]. \tag{101}$$

□

Corollary 1.

$$\boldsymbol{x}_t|\bullet, \boldsymbol{q}_t^o \sim \text{Gauss}_{n_r+n_q}\left(\begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{r}_t^c} \\ \boldsymbol{\mu}_{\boldsymbol{q}_t^c} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{r}_t^c} & \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}(\boldsymbol{I} - \boldsymbol{D}_t^T\boldsymbol{K}_{\boldsymbol{q}_t}{}^T) \\ (\boldsymbol{I} - \boldsymbol{K}_{\boldsymbol{q}_t}\boldsymbol{D}_t)\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}{}^T & \boldsymbol{\Sigma}_{\boldsymbol{q}_t^c} \end{bmatrix}\right), \tag{102}$$

with $\boldsymbol{\mu}_{\boldsymbol{r}_t^c}$, $\boldsymbol{\mu}_{\boldsymbol{q}_t^c}$, $\boldsymbol{\Sigma}_{\boldsymbol{r}_t^c}$, $\boldsymbol{\Sigma}_{\boldsymbol{q}_t^c}$ *and* $\boldsymbol{K}_{\boldsymbol{q}}$ *given in Expressions* (98), (99) *and* (101).

Proof. Using Result 15 we see that

$$\begin{aligned} Cov(\boldsymbol{r}_t^c, \boldsymbol{q}_t^c) &= Cov\left(\boldsymbol{r}_t + \boldsymbol{K_r}(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t), \boldsymbol{q}_t + \boldsymbol{K_q}(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_t - \boldsymbol{D}_t\boldsymbol{q}_t)\right) \\ &= \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t} - \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T\boldsymbol{K}_{\boldsymbol{q}_t}^T - \boldsymbol{K}_{\boldsymbol{r}t}\boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t} \\ &\quad + \boldsymbol{K}_{\boldsymbol{r}t}(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)\boldsymbol{K}_{\boldsymbol{q}_t}^T \\ &= \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}(\boldsymbol{I} - \boldsymbol{D}_t^T\boldsymbol{K}_{\boldsymbol{q}_t}^T). \end{aligned}$$

Then by Result 2, the Corollary follows. □

Result 16. *Let*

$$\boldsymbol{x}_t^1, \ldots, \boldsymbol{x}_t^n \overset{\text{i.i.d}}{\sim} \text{Gauss}_{n_r+n_q}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t),$$

and

$$\boldsymbol{\epsilon}_t^1, \ldots, \boldsymbol{\epsilon}_t^n \overset{\text{i.i.d}}{\sim} \text{Gauss}_{n_r+n_q}(\boldsymbol{0}, \boldsymbol{\Sigma}_t^o).$$

Further, let

$$\hat{\boldsymbol{\Sigma}}_t = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}t} & \hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}\boldsymbol{q}_t} \\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}\boldsymbol{r}_t} & \hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}_t} \end{bmatrix} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_t^i - \bar{\boldsymbol{x}}_t)(\boldsymbol{x}_t^i - \bar{\boldsymbol{x}}_t)^T,$$

where $\bar{\boldsymbol{x}}_t$ is the arithmetic mean. Then

$$\begin{aligned} \boldsymbol{x}_t^{c,i} &= \begin{bmatrix} \boldsymbol{r}_t^i + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T(\hat{\boldsymbol{\Sigma}}_t^o + \boldsymbol{D}_t\hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_t^i - \boldsymbol{D}_t\boldsymbol{q}_t^i) \\ \boldsymbol{q}_t^i + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T(\hat{\boldsymbol{\Sigma}}_t^o + \boldsymbol{D}_t\hat{\boldsymbol{\Sigma}}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o + \boldsymbol{\epsilon}_t^i - \boldsymbol{D}_t\boldsymbol{q}_t^i) \end{bmatrix} \\ &\xrightarrow{\text{d}} \text{Gauss}_{n_r+n_q}\left(\begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{r}_t^c} \\ \boldsymbol{\mu}_{\boldsymbol{q}_t^c} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{r}_t^c} & \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}(\boldsymbol{I} - \boldsymbol{D}_t^T\boldsymbol{K}_{\boldsymbol{q}_t}^T) \\ (\boldsymbol{I} - \boldsymbol{K}_{\boldsymbol{q}_t}\boldsymbol{D}_t)\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}^T & \boldsymbol{\Sigma}_{\boldsymbol{q}_t^c} \end{bmatrix}\right), \end{aligned}$$

where $\xrightarrow{\text{d}}$ denotes "convergence in distribution ",

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{r}_t^c} &= \boldsymbol{\mu}_{\boldsymbol{r}t} + \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o - \boldsymbol{D}_t\boldsymbol{\mu}_{\boldsymbol{q}_t}) \\ \boldsymbol{\mu}_{\boldsymbol{q}_t^c} &= \boldsymbol{\mu}_{\boldsymbol{q}_t} + \boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}(\boldsymbol{q}_t^o - \boldsymbol{D}_t\boldsymbol{\mu}_{\boldsymbol{q}_t}) \end{aligned}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{r}_t^c} &= \boldsymbol{\Sigma}_{\boldsymbol{r}t} - \boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{r}\boldsymbol{q}_t}\boldsymbol{D}_t \\ \boldsymbol{\Sigma}_{\boldsymbol{q}_t^c} &= \boldsymbol{\Sigma}_{\boldsymbol{q}_t} - \boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t \end{aligned}$$

and

$$\boldsymbol{K}_{\boldsymbol{q}_t} = \left[\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T(\boldsymbol{\Sigma}_t^o + \boldsymbol{D}_t\boldsymbol{\Sigma}_{\boldsymbol{q}_t}\boldsymbol{D}_t^T)^{-1}\right].$$

Remark. The result above simply restates the procedure used in the Ensemble Kalman filter (Evensen et al. 1998), which was an extension of the original EnKF proposed by Evensen (1994).

Proof. Since $\hat{\boldsymbol{\Sigma}}_t$ is a consistent estimator for $\boldsymbol{\Sigma}_t$, then by the Central Limit Theorem (Mardia et al. 1979, Theorem 2.9.1), Strong Law of Large Numbers (Casella & Berger 2002) and Corollary 1, the result follows. □

C Hierarchical Scale-Corrected Ensemble Kalman Filter

We now consider the Hierarchical Bayesian model described in Figure C.1.

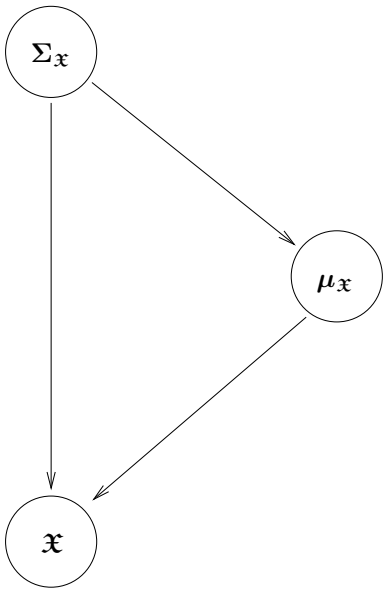


Figure C.1: The figure shows the Directed Acyclic Graph (DAG) of the Hierarchical Bayesian model considered. Here $\boldsymbol{\mathfrak{X}} \in \mathbb{R}^{p\times 1}$, $\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}} \in \mathbb{R}^{p\times 1}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} \in \mathbb{R}^{p\times p}$ are distributed as in Expressions (103) to (105).

$$\boldsymbol{\mathfrak{X}} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}^* \end{bmatrix} \sim \text{Gauss}_p\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}^* \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*} \\ \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}^T & \boldsymbol{\Sigma}_* \end{bmatrix} \right), \tag{103}$$

where $\boldsymbol{x} \in \mathbb{R}^{u\times 1}$ and $\boldsymbol{x}^* \in \mathbb{R}^{g\times 1}$. Further, let

$$\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}^* \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*} \\ \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}^T & \boldsymbol{\Sigma}_* \end{bmatrix},$$

such that

$$\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}} \sim \text{Gauss}_p(\boldsymbol{\eta}_{\boldsymbol{\mathfrak{X}}}, \frac{1}{\zeta}\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}) \tag{104}$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} \sim W_p^{-1}(\boldsymbol{\Psi}_{\boldsymbol{\mathfrak{X}}}{}^{-1}, \nu), \tag{105}$$

where

$$\boldsymbol{\eta}_{\boldsymbol{\mathfrak{X}}} = \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\eta}^* \end{bmatrix},$$

and

$$\boldsymbol{\Psi}_{\mathfrak{X}}{}^{-1} = \begin{bmatrix} \boldsymbol{\Psi} & \boldsymbol{\Psi}_{\boldsymbol{x}\boldsymbol{x}^*} \\ \boldsymbol{\Psi}_{\boldsymbol{x}\boldsymbol{x}^*}^T & \boldsymbol{\Psi}_* \end{bmatrix}.$$

As shown in Theorem 1, we can write the prior distribution in Expression (105) in terms of new variables in the set $\boldsymbol{\Theta} = \{\boldsymbol{\Sigma}_{1|*},\ \boldsymbol{T},\ \boldsymbol{\Sigma}_*\}$ as

$$\begin{aligned} \boldsymbol{\Sigma}_{1|*} &= \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}\boldsymbol{\Sigma}_*^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}^T \\ \boldsymbol{T} &= \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}\boldsymbol{\Sigma}_*^{-1} \\ \boldsymbol{\Sigma}_* &= \boldsymbol{\Sigma}_*. \end{aligned}$$

Where these variables follow the distribution given in Theorem 1. Assume we have generated realisations

$$\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_n^* \overset{\text{i.i.d}}{\sim} \text{Gauss}_g(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_*),$$

where i.i.d means independent identically distributed.

The sample mean is given as

$$\bar{\boldsymbol{x}}^* = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i^*.$$

Then

$$\begin{aligned} E[\bar{\boldsymbol{x}}^*] &= E\left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i^*\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E[\boldsymbol{x}_i^*] \\ &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\mu}^* = \boldsymbol{\mu}^*, \end{aligned}$$

and

$$\begin{aligned} Cov(\bar{\boldsymbol{x}}^*) &= Cov\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i^*\right) \\ &\overset{\text{i.i.d}}{=} \frac{1}{n^2}\sum_{i=1}^{n} Cov(\boldsymbol{x}_i^*) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} \boldsymbol{\Sigma}_* = \frac{1}{n}\boldsymbol{\Sigma}_*. \end{aligned}$$

By using the general multivariate Gaussian theory (Mardia et al. 1979, Chapter 3), we see that

$$\bar{\boldsymbol{x}}^* | \boldsymbol{\mu}^*, \boldsymbol{\Sigma}_* \sim \text{Gauss}_g\left(\boldsymbol{\mu}^*, \frac{1}{n}\boldsymbol{\Sigma}_*\right). \tag{106}$$

For notational convenience we now define the matrix of samples $\boldsymbol{L} \in \mathbb{R}^{g \times n}$ as

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{x}_1^*, & \ldots & , \boldsymbol{x}_n^* \end{bmatrix}.$$

The sample covariance matrix is given by

$$\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i^* - \bar{\boldsymbol{x}}^*)(\boldsymbol{x}_i^* - \boldsymbol{\mu}^*)^T.$$

This corresponds to the maximum likelihood estimator of the covariance matrix when the samples are i.i.d. Gaussian. Note that when the number of elements in the random vector $\boldsymbol{x}^*$ is large, which is typically the case in reservoir simulation, the the number of samples you to generate tend to be restricted. This implies that the matrix $\boldsymbol{S}$ will not in most practical applications have full rank (Johnson & Wichern 2002, Result 3.3). Which again means (Strang 1988) that $\boldsymbol{S}$ is singular, consequently its inverse $\boldsymbol{S}^{-1}$, does not exist, and we can not use the general results concerning the relationship between the sample mean vector and the sample covariance matrix in the multivariate Gaussian case (Johnson & Wichern 2002, Chapter 4).

In the following theorems, we consider a hierarchical Bayesian setting as the one described in Figure C.1. We also assume to have generated realisations $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_n^* \overset{\text{i.i.d}}{\sim} \text{Gauss}_g(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_*)$, then we have the following properties of the various posterior distributions.

Theorem 2.

$$\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Theta}, \boldsymbol{L} \sim \text{Gauss}_p(\boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}) \tag{107}$$

where

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\bar{\boldsymbol{x}}^*} &= \boldsymbol{\eta}_{\boldsymbol{\mathcal{X}}} + \left(\begin{bmatrix}\boldsymbol{T}\\ \boldsymbol{I}_g\end{bmatrix} \otimes \left(\frac{n}{n+\zeta}\right)\right)(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*) \\ &= \begin{bmatrix}\boldsymbol{\eta} + \frac{n}{n+\zeta}\boldsymbol{T}(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*) \\ \boldsymbol{\eta}^* + \frac{n}{n+\zeta}(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)\end{bmatrix} = \begin{bmatrix}\boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*} \\ \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}\end{bmatrix}\end{aligned} \tag{108}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\bar{\boldsymbol{x}}^*} = \frac{1}{\zeta}\left(\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}} - \frac{n}{n+\zeta}\begin{bmatrix}\boldsymbol{T}\boldsymbol{\Sigma}_*\boldsymbol{T}^T & \boldsymbol{T}\boldsymbol{\Sigma}_* \\ \boldsymbol{\Sigma}_*\boldsymbol{T}^T & \boldsymbol{\Sigma}_*\end{bmatrix}\right). \tag{109}$$

Theorem 3.

$$\boldsymbol{\Sigma}_*|\boldsymbol{L} \sim W_g^{-1}(\hat{\boldsymbol{\Psi}}_*^{-1}, \nu + n - u). \tag{110}$$

where

$$\hat{\boldsymbol{\Psi}}_*^{-1} = n\boldsymbol{S} + \boldsymbol{\Psi}_* + \frac{n\zeta}{n+\zeta}(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)^T \tag{111}$$

$$\boldsymbol{\Sigma}_{1|*}|\boldsymbol{L} \sim W_u^{-1}(\boldsymbol{\Psi}_{1|*}^{-1}, \nu) \tag{112}$$

$$\text{Vec}(\boldsymbol{T})|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{L} \sim \text{Gauss}_{ug}\left(\text{Vec}(\boldsymbol{\Gamma}), (\boldsymbol{\Psi}_*^{-1} \otimes \boldsymbol{\Sigma}_{1|*})\right) \tag{113}$$

Theorem 4. *The posterior mean of* $\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}$ *is given by*

$$\hat{\boldsymbol{\Sigma}} = \alpha\frac{\boldsymbol{\Psi}_{\boldsymbol{\mathcal{X}}}}{\nu - p - 1} + (1-\alpha)\tilde{\boldsymbol{\Sigma}} \tag{114}$$

where

$$\alpha = \frac{\nu - p - 1}{\nu + n - p - 1}, \tag{115}$$

$$\tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \left[\frac{1+\mathrm{tr}(\tilde{\boldsymbol{S}}\boldsymbol{\Psi}_*^{-1})}{\nu-u-1}\right]\boldsymbol{\Psi}_{1|*} + \boldsymbol{\Gamma}\tilde{\boldsymbol{S}}\boldsymbol{\Gamma}^T & \boldsymbol{\Gamma}\tilde{\boldsymbol{S}} \\ \tilde{\boldsymbol{S}}\boldsymbol{\Gamma}^T & \tilde{\boldsymbol{S}} \end{bmatrix}, \tag{116}$$

and

$$\tilde{\boldsymbol{S}} = \frac{1}{n}\left(n\boldsymbol{S} + \frac{n\zeta}{n+\zeta}(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)^T\right).$$

Theorem 5. *Let $\boldsymbol{\mathfrak{X}}_f$ denote the future observations of the reservoir then*

$$\boldsymbol{\mathfrak{X}}_f|\boldsymbol{\Theta}, \boldsymbol{L} \sim \mathrm{Gauss}_p(\boldsymbol{\mu}_{\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}|\bar{\boldsymbol{x}}^*}, \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} + \boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}|\bar{\boldsymbol{x}}^*}) \tag{117}$$

with $\boldsymbol{\mu}_{\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}}|\bar{\boldsymbol{x}}^*$ and $\boldsymbol{\Sigma}_{\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}|\bar{\boldsymbol{x}}^*}$ given in Expressions (108) *and* (109) *respectively.*

Theorem 6. *The predictive distribution of $\boldsymbol{\mathfrak{X}}_f^T = [\boldsymbol{x}_f^T, \boldsymbol{x}_f^{*T}]$, given realisations from the course scale reservoir $\boldsymbol{L}$ is*

$$\boldsymbol{x}_f^*|\boldsymbol{L} \sim t_g\left(\boldsymbol{\eta}^* + \frac{n}{n+\zeta}(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*), \frac{c-d}{q}\hat{\boldsymbol{\Psi}}_*, q\right) \tag{118}$$

$$\boldsymbol{x}_f|\boldsymbol{x}_f^*, \boldsymbol{L} \sim t_u\left(\boldsymbol{\eta} + \boldsymbol{\Gamma}(\boldsymbol{x}_f^* - \boldsymbol{\eta}^*), \frac{c + (\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)^T\boldsymbol{\Psi}_*^{-1}(\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)}{\nu + 1 - u}\boldsymbol{\Psi}_{1|*}, \nu + 1 - u\right) \tag{119}$$

where

$$\begin{aligned} q &= \nu + n + 1 - p \\ c &= 1 + \frac{1}{\zeta} \\ d &= \frac{n}{\zeta(n+\zeta)} \end{aligned}$$

Remark (Theorem 2). If we set $\boldsymbol{F} = 1$, $\boldsymbol{A} = n$, $\hat{\boldsymbol{\beta}}_2 = \bar{\boldsymbol{x}}^*$, $\boldsymbol{\beta} = \boldsymbol{\mu}$ and $\boldsymbol{\beta}^0 = \boldsymbol{\eta}$, this result is equivalent to Theorem 2.1 in Lee & Zidek (1992). Also note that this implies that they actually did not need to restrict their Kriging predictor to cases when the the number of samples n is larger than the number of covariates k. This because the Gaussian multivariate likelihood function always can be written as shown in Result 12 which need not be a pdf.

Remark (Theorem 6). This theorem is stated for completeness, although it is not used in this Thesis. As shown in Theorem 1 the proof of the equivalent theorem given in Lee & Zidek (1992) was based on a non-valid assumption concerning the distribution of $\boldsymbol{T}|\boldsymbol{\Sigma}_{1|*}$. We therefore give a detailed proof were this is corrected for.

Proof Theorem 2. Consider the pdf of $\boldsymbol{\mu_{\mathfrak{X}}}|\boldsymbol{\Theta}, \boldsymbol{L}$ namely $f(\boldsymbol{\mu_{\mathfrak{X}}}|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}, \boldsymbol{\Sigma}_*, \boldsymbol{L})$. Using Expressions (91) we get

$$\begin{aligned}
f(\boldsymbol{\mu_{\mathfrak{X}}}|\boldsymbol{\Theta}, \boldsymbol{L}) &\propto f(\boldsymbol{L}|\boldsymbol{\mu_{\mathfrak{X}}}, \boldsymbol{\Sigma}_*, \boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}) \times f(\boldsymbol{\mu_{\mathfrak{X}}}|\boldsymbol{\Sigma}_*, \boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}) \\
&\propto |\boldsymbol{\Sigma}_*|^{-n/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{x}_i^* - \boldsymbol{\mu}^*)^T \boldsymbol{\Sigma}_*^{-1} (\boldsymbol{x}_i^* - \boldsymbol{\mu}^*) \right\} \\
&\quad \times f(\boldsymbol{\mu_{\mathfrak{X}}}|\boldsymbol{\Sigma}_*, \boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}) \\
&\propto f(\bar{\boldsymbol{x}}^*|\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_*) \times g(\boldsymbol{S}; \boldsymbol{\Sigma}_*) \times f(\boldsymbol{\mu_{\mathfrak{X}}}|\boldsymbol{\Sigma}_*, \boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}) \\
&\propto f(\bar{\boldsymbol{x}}^*|\boldsymbol{\mu_{\mathfrak{X}}}, \boldsymbol{\Sigma_{\mathfrak{X}}}) \times f(\boldsymbol{\mu_{\mathfrak{X}}}|\boldsymbol{\Sigma_{\mathfrak{X}}}) \\
&\propto f(\bar{\boldsymbol{x}}^*, \boldsymbol{\mu_{\mathfrak{X}}}, \boldsymbol{\Sigma_{\mathfrak{X}}}) \\
&\propto f(\boldsymbol{\mu_{\mathfrak{X}}}|\bar{\boldsymbol{x}}^*, \boldsymbol{\Sigma_{\mathfrak{X}}})
\end{aligned}$$

Let the matrix $\boldsymbol{E} \in \mathbb{R}^{g \times p}$ be given by

$$\boldsymbol{E} = \left[\begin{array}{cc} \boldsymbol{0} & \boldsymbol{I}_g \end{array} \right],$$

where $\boldsymbol{0} \in \mathbb{R}^{g \times u}$ is a matrix containing only zeros, and $\boldsymbol{I}_g$ denotes the identity matrix $\in \mathbb{R}^{g \times g}$. Using this matrix we see that

$$\bar{\boldsymbol{x}}^*|\boldsymbol{\mu_{\mathfrak{X}}}, \boldsymbol{\Sigma_{\mathfrak{X}}} \sim \text{Gauss}_g \left(\boldsymbol{E}\boldsymbol{\mu_{\mathfrak{X}}}, \frac{1}{n} \boldsymbol{E}\boldsymbol{\Sigma_{\mathfrak{X}}}\boldsymbol{E}^{\boldsymbol{T}} \right). \tag{120}$$

From the expression above we see that

$$\begin{aligned}
f(\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma_{\mathfrak{X}}}) &= \int f(\bar{\boldsymbol{x}}^*|\boldsymbol{\mu}^*, \boldsymbol{\Sigma_{\mathfrak{X}}}) f(\boldsymbol{\mu}^*|\boldsymbol{\Sigma_{\mathfrak{X}}}) \mathrm{d}\boldsymbol{\mu}^* \\
&\propto \int \exp\left\{ -\frac{1}{2} (\bar{\boldsymbol{x}}^* - \boldsymbol{\mu}^*)^T \left(\frac{\boldsymbol{\Sigma_*}}{n} \right)^{-1} (\bar{\boldsymbol{x}}^* - \boldsymbol{\mu}^*) \right\} \\
&\quad \times \exp\left\{ -\frac{1}{2} (\boldsymbol{\mu}^* - \boldsymbol{\eta}^*)^T \left(\frac{\boldsymbol{\Sigma}_*}{\zeta} \right)^{-1} (\boldsymbol{\mu}^* - \boldsymbol{\eta}^*) \right\} \mathrm{d}\boldsymbol{\mu}^* \\
&\propto \exp\left\{ -\frac{1}{2} \bar{\boldsymbol{x}}^{*T} \left(\frac{\boldsymbol{\Sigma_*}}{n} \right)^{-1} \bar{\boldsymbol{x}}^* \right\} \\
&\quad \times \int \exp\left\{ -\frac{1}{2} \boldsymbol{\mu}^{*T} \left(\frac{\boldsymbol{\Sigma}_*}{n+\zeta} \right)^{-1} \boldsymbol{\mu}^* - \frac{1}{n+\zeta} (n\bar{\boldsymbol{x}}^* + \zeta\eta^*)^T \left(\frac{\boldsymbol{\Sigma}_*}{n+\zeta} \right)^{-1} \right\} \mathrm{d}\boldsymbol{\mu}^*.
\end{aligned}$$

Let

$$\tilde{\boldsymbol{\Sigma}} = \left(\frac{\boldsymbol{\Sigma}_*}{n+\zeta} \right)^{-1},$$

and

$$\tilde{\boldsymbol{\mu}} = \frac{1}{n+\zeta} (n\bar{\boldsymbol{x}}^* + \zeta\eta^*).$$

Since the integral of the kernel of a multivariate Gaussian distributed variable is proportional to one in distribution, we get

$$\begin{aligned} f(\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}) &\propto \exp\left\{-\frac{1}{2}\bar{\boldsymbol{x}}^{*T}\left(\frac{\boldsymbol{\Sigma}_*}{n}\right)^{-1}\bar{\boldsymbol{x}}^*\right\} \\ &\quad\times\int\exp\left\{-\frac{1}{2}(\boldsymbol{\mu}^*-\tilde{\boldsymbol{\mu}})^T\tilde{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mu}^*-\tilde{\boldsymbol{\mu}})\right\}\mathrm{d}\boldsymbol{\mu}^*\times\exp\left\{+\frac{1}{2}\tilde{\boldsymbol{\mu}}^T\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}}\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\bar{\boldsymbol{x}}^{*T}\left[n-\frac{n^2}{n+\zeta}\right]\boldsymbol{\Sigma}_*^{-1}\bar{\boldsymbol{x}}^*-2\boldsymbol{\eta}^{*T}\frac{n\zeta}{n+\zeta}\boldsymbol{\Sigma}_*^{-1}\bar{\boldsymbol{x}}^*\right)\right\}. \end{aligned} \tag{121}$$

The multivariate Gaussian pdf as defined in Expression (80) with parameters $\boldsymbol{\mu}_{\boldsymbol{Y}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{Y}}$ can always be written in the following manner:

$$f(\boldsymbol{y})\propto\exp\left\{-\frac{1}{2}(\boldsymbol{y}^T{\boldsymbol{\Sigma}_{\boldsymbol{Y}}}^{-1}\boldsymbol{y}-2{\boldsymbol{\mu}_{\boldsymbol{Y}}}^T{\boldsymbol{\Sigma}_{\boldsymbol{Y}}}^{-1}\boldsymbol{y})\right\}.$$

Comparing this last expression with the one in Expression (121), we observer that

$$\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}\sim\mathrm{Gauss}_g(\boldsymbol{E}\boldsymbol{\eta}_{\boldsymbol{\mathcal{X}}},\frac{n+\zeta}{n\zeta}\boldsymbol{E}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}\boldsymbol{E}^{\boldsymbol{T}}). \tag{122}$$

Now since both $\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}$ and $\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}$ are multivariate Gaussian with parameters given in Expressions (104) and (122), then by Result 2,

$$\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\bar{\boldsymbol{x}}^*,\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}\sim\mathrm{Gauss}_p(\boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*},\boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}). \tag{123}$$

Where

$$\boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}=E[\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}]+Cov(\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}},\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}})Cov(\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}})^{-1}(\bar{\boldsymbol{x}}^*-E[\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}])$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}=Cov(\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}})-Cov(\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}},\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}})Cov(\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}})^{-1}Cov(\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}},\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}})^T.$$

The only unknown element in the two expressions above is $Cov(\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}},\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}})$, but since

$$\begin{aligned} E[\bar{\boldsymbol{x}}^*|\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}},\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}] &= E[\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}]+Cov(\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}},\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}})^TCov(\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}})^{-1}(\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}-E[\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}]) \\ &= \boldsymbol{E}\boldsymbol{\eta}_{\boldsymbol{\mathcal{X}}}+\zeta\boldsymbol{\Sigma}^T_{\bar{\boldsymbol{x}}^*\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}}{\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}}^{-1}(\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}-\boldsymbol{\eta}_{\boldsymbol{\mathcal{X}}})=\boldsymbol{E}\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}, \end{aligned}$$

we see that

$$Cov(\boldsymbol{\mu}_{\boldsymbol{\mathcal{X}}}|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}},\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}})=\frac{1}{\zeta}(\boldsymbol{E}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}})^T=\frac{1}{\zeta}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}\boldsymbol{E}^T.$$

Inserting this into the expressions above, we get

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*} &= \boldsymbol{\eta}_{\boldsymbol{\mathcal{X}}}+\frac{1}{\zeta}\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{X}}}\boldsymbol{E}^T\frac{n\zeta}{n+\zeta}\boldsymbol{\Sigma}_*^{-1}(\bar{\boldsymbol{x}}^*-\boldsymbol{E}\boldsymbol{\eta}_{\boldsymbol{\mathcal{X}}}) \\ &= \boldsymbol{\eta}_{\boldsymbol{\mathcal{X}}}+\begin{bmatrix}\boldsymbol{T}\\ \boldsymbol{I}_g\end{bmatrix}\otimes\left(\frac{n}{n+\zeta}\right)(\bar{\boldsymbol{x}}^*-\boldsymbol{E}\boldsymbol{\eta}_{\boldsymbol{\mathcal{X}}}) \end{aligned}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*} &= \frac{1}{\zeta}\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} - \frac{1}{\zeta^2}\begin{bmatrix} \boldsymbol{T}\boldsymbol{\Sigma}_* \\ \boldsymbol{\Sigma}_* \end{bmatrix}\left(\frac{n\zeta}{n+\zeta}\right)\boldsymbol{\Sigma}_*^{-1}\begin{bmatrix} \boldsymbol{\Sigma}_*\boldsymbol{T}^T & \boldsymbol{\Sigma}_* \end{bmatrix} \\ &= \frac{1}{\zeta}\left(\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} - \frac{n}{n+\zeta}\begin{bmatrix} \boldsymbol{T}\boldsymbol{\Sigma}_*\boldsymbol{T}^T & \boldsymbol{T}\boldsymbol{\Sigma}_* \\ \boldsymbol{\Sigma}_*\boldsymbol{T}^T & \boldsymbol{\Sigma}_* \end{bmatrix}.\right)\end{aligned}$$

□

Proof Theorem 3. Consider the posterior pdf $f(\boldsymbol{\Theta}|\boldsymbol{L})$. Using the results in Chapter 3 of Ross (2003)

$$\begin{aligned} f(\boldsymbol{\Theta}|\boldsymbol{L}) &\propto f(\boldsymbol{\Theta},\boldsymbol{L}) \propto f(\boldsymbol{L}|\boldsymbol{\Theta})f(\boldsymbol{\Theta}) \\ &\propto \int f(\boldsymbol{L}|\boldsymbol{\Theta},\boldsymbol{\mu})f(\boldsymbol{\mu}|\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\mu} f(\boldsymbol{\Theta}) \\ &\propto f(\boldsymbol{\Theta})g(\boldsymbol{S};\boldsymbol{\Sigma}_*)\int f(\bar{\boldsymbol{x}}^*|\boldsymbol{\mu},\boldsymbol{\Theta})f(\boldsymbol{\mu}|\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\mu} \\ &\propto f(\boldsymbol{\Theta})g(\boldsymbol{S};\boldsymbol{\Sigma}_*)f(\bar{\boldsymbol{x}}^*|\boldsymbol{\Theta}) \\ &\propto f(\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_*)g(\boldsymbol{S};\boldsymbol{\Sigma}_*)f(\boldsymbol{T}|\boldsymbol{\Sigma}_{1|*})f(\boldsymbol{\Sigma}_{1|*})f(\boldsymbol{\Sigma}_*)\end{aligned}$$

Inserting the prior distribution in Expression (105), and the likelihood function given in Expression (91), we get

$$\begin{aligned} f(\boldsymbol{\Sigma}_*|\boldsymbol{L}) &\propto f(\bar{\boldsymbol{x}}^*|\boldsymbol{\Sigma}_*)g(\boldsymbol{S};\boldsymbol{\Sigma}_*)f(\boldsymbol{\Sigma}_*) \\ &\propto |\boldsymbol{\Sigma}_*|^{-1/2}\exp\left\{-\frac{1}{2}(\bar{\boldsymbol{x}}^*-\boldsymbol{\eta}^*)^T\frac{n\zeta}{n+\zeta}\boldsymbol{\Sigma}_*^{-1}(\bar{\boldsymbol{x}}^*-\boldsymbol{\eta}^*)\right\} \\ &\quad\times|\boldsymbol{\Sigma}_*|^{-(n-1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(n\boldsymbol{S}\boldsymbol{\Sigma}_*^{-1})\right\}\times|\boldsymbol{\Sigma}_*|^{-(\nu-u+g+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}_*\boldsymbol{\Sigma}_*^{-1})\right\} \\ &\propto |\boldsymbol{\Sigma}_*|^{-(\nu+n-u+g+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\left[n\boldsymbol{S}+\boldsymbol{\Psi}_*+\frac{n\zeta}{n+\zeta}(\bar{\boldsymbol{x}}^*-\boldsymbol{\eta}^*)(\bar{\boldsymbol{x}}^*-\boldsymbol{\eta}^*)^T\right]\boldsymbol{\Sigma}_*^{-1}\right)\right\}.\end{aligned}$$

Comparing this last expression with the pdf in Expression (82), we can see that $\boldsymbol{\Sigma}_*|\boldsymbol{L}$ has the distribution given in Expression (110). Further, we find that

$$f(\boldsymbol{T}|\boldsymbol{L},\boldsymbol{\Sigma}_{1|*}) \propto f(\boldsymbol{T}|\boldsymbol{\Sigma}_{1|*}) \tag{124}$$

$$\begin{aligned} f(\boldsymbol{\Sigma}_{1|*}|\boldsymbol{L}) &\propto \int f(\boldsymbol{T}|\boldsymbol{\Sigma}_{1|*})f(\boldsymbol{\Sigma}_{1|*})\mathrm{d}\boldsymbol{T} \\ &\propto f(\boldsymbol{\Sigma}_{1|*})\times\int f(\boldsymbol{T}|\boldsymbol{\Sigma}_{1|*})\mathrm{d}\boldsymbol{T} \\ &\propto f(\boldsymbol{\Sigma}_{1|*}). \end{aligned} \tag{125}$$

Where the probability density functions in Expressions (124), and (125) are given in Expressions (113) and (112) respectively. □

Proof Theorem 4. Using Expressions (110) to (113), together with Expression (85) and the *a priori* assumption of independence between $\boldsymbol{T}$ and $\boldsymbol{\Sigma}_*$, we get

$$E[\boldsymbol{\Sigma}_*|\boldsymbol{L}] = \frac{\hat{\boldsymbol{\Psi}}}{\nu+n-p-1} = \frac{\hat{\boldsymbol{\Psi}}_*}{a}, \tag{126}$$

$$E[\boldsymbol{\Sigma}_{1|*}|\boldsymbol{L}] = \frac{\boldsymbol{\Psi}_{1|*}}{\nu - u - 1}, \tag{127}$$

$$\begin{aligned} E[\boldsymbol{T}\boldsymbol{\Sigma}_*|\boldsymbol{L}] &= E[\boldsymbol{T}|\boldsymbol{L}]E[\boldsymbol{\Sigma}_*|\boldsymbol{L}] \\ &= E\left[E[\boldsymbol{T}|\boldsymbol{L},\boldsymbol{\Sigma}_{1|*}]|\boldsymbol{L}\right]\frac{\hat{\boldsymbol{\Psi}}_*}{a} \\ &= \frac{\boldsymbol{\Gamma}\hat{\boldsymbol{\Psi}}_*}{a}. \end{aligned} \tag{128}$$

Expression (87), together with Expression (88) and Result 11 further gives

$$\begin{aligned} E[\boldsymbol{T}\boldsymbol{\Sigma}_*\boldsymbol{T}^T|\boldsymbol{L}] &= E\left[E[\boldsymbol{T}\boldsymbol{\Sigma}_*\boldsymbol{T}^T|\boldsymbol{L},\boldsymbol{T}]|\boldsymbol{L}\right] \\ &= E\left[\boldsymbol{T}E[\boldsymbol{\Sigma}_*|\boldsymbol{L}]\boldsymbol{T}^T|\boldsymbol{L}\right] \\ &= \frac{1}{a}E\left[E[\boldsymbol{T}\hat{\boldsymbol{\Psi}}_*\boldsymbol{T}^T|\boldsymbol{L},\boldsymbol{\Sigma}_{1|*}]|\boldsymbol{L}\right] \\ &= \frac{1}{a}E\left[\mathrm{tr}\left(\hat{\boldsymbol{\Psi}}_*\boldsymbol{\Psi}_*^{-1}\right)\boldsymbol{\Sigma}_{1|*} + \boldsymbol{\Gamma}\hat{\boldsymbol{\Psi}}_*\boldsymbol{\Gamma}^T|\boldsymbol{L}\right]. \\ &= \frac{1}{a}\mathrm{tr}(\boldsymbol{\Psi}_*^{-1}\hat{\boldsymbol{\Psi}}_*)E[\boldsymbol{\Sigma}_{1|*}|\boldsymbol{L}] + \frac{1}{a}\boldsymbol{\Gamma}\hat{\boldsymbol{\Psi}}_*\boldsymbol{\Gamma}^T \\ &= \frac{1}{a}\mathrm{tr}(\boldsymbol{\Psi}_*^{-1}\hat{\boldsymbol{\Psi}}_*)\frac{\boldsymbol{\Psi}_{1|*}}{\nu - u - 1} + \frac{1}{a}\boldsymbol{\Gamma}\hat{\boldsymbol{\Psi}}_*\boldsymbol{\Gamma}^T. \end{aligned} \tag{129}$$

The posterior mean of $\boldsymbol{\Sigma}$ is thus given by

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} E[\boldsymbol{\Sigma}_{1|*}|\boldsymbol{L}] + E[\boldsymbol{T}\boldsymbol{\Sigma}_*\boldsymbol{T}^T|\boldsymbol{L}] & E[\boldsymbol{T}\boldsymbol{\Sigma}_*|\boldsymbol{L}] \\ E[\boldsymbol{\Sigma}_*\boldsymbol{T}^T|\boldsymbol{L}] & E[\boldsymbol{\Sigma}_*|\boldsymbol{L}] \end{bmatrix}.$$

Inserting Expressions (126) to (129) into the expression above, we get

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} &= \frac{1}{a}\begin{bmatrix} \frac{a+\mathrm{tr}(\boldsymbol{\Psi}_*^{-1}\hat{\boldsymbol{\Psi}}_*)}{\nu-u-1}\boldsymbol{\Psi}_{1|*} + \boldsymbol{\Gamma}\hat{\boldsymbol{\Psi}}_*\boldsymbol{\Gamma}^T & \boldsymbol{\Gamma}\hat{\boldsymbol{\Psi}}_* \\ \hat{\boldsymbol{\Psi}}_*\boldsymbol{\Gamma}^T & \hat{\boldsymbol{\Psi}}_* \end{bmatrix} \\ &= \alpha\frac{\boldsymbol{\Psi}_{\boldsymbol{\mathfrak{X}}}}{\nu - p - 1} + (1-\alpha)\tilde{\boldsymbol{\Sigma}}, \end{aligned}$$

with α and $\tilde{\boldsymbol{\Sigma}}$ defined in Expressions (115) and (116). □

Proof Theorem 5. Using Result 2, together with Expressions (103) and (107), we see that

$$\begin{aligned} E[\boldsymbol{\mathfrak{X}}_f|\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}},\boldsymbol{L},\boldsymbol{\Theta}] = \boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}} &= E[\boldsymbol{\mathfrak{X}}_f|\boldsymbol{L},\boldsymbol{\Theta}] + \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}_f\boldsymbol{\mu}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}^{-1}(\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}} - \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}) \\ &= E\left[E[\boldsymbol{\mathfrak{X}}_f|\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}},\boldsymbol{L},\boldsymbol{\Theta}]|\boldsymbol{L},\boldsymbol{\Theta}\right] + \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}_f\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}^{-1}(\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}} - \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}) \\ &= \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*} + \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}_f\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}^{-1}(\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}} - \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}) \\ Cov(\boldsymbol{\mathfrak{X}}_f|\boldsymbol{\mu},\boldsymbol{L},\boldsymbol{\Theta}) = \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} &= Cov(\boldsymbol{\mathfrak{X}}_f|\boldsymbol{L},\boldsymbol{\Theta}) - \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}_f\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}}\boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}_f\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}}^T \end{aligned}$$

Solving this equation with respect to $\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}_f\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}}$ yields $\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}_f\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}} = \boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}$. Inserting this into the second equation we get.

$$Cov(\boldsymbol{\mathfrak{X}}_f|\boldsymbol{L},\boldsymbol{\Theta}) = \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} + \boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*}.$$

Again using Result 2, we see that

$$\begin{bmatrix} \boldsymbol{\mathfrak{X}}_f|\boldsymbol{L},\boldsymbol{\Theta} \\ \boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}|\boldsymbol{L},\boldsymbol{\Theta} \end{bmatrix} \sim \text{Gauss}_{p+p}\left(\begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*} \\ \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} + \boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*} & \boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*} \\ \boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*} & \boldsymbol{\Sigma}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}^*} \end{bmatrix}\right).$$

Thus, $\boldsymbol{\mathfrak{X}}_f|\boldsymbol{L},\boldsymbol{\Theta}$ follows the distribution in Expression (117). □

Proof Theorem 6.

$$\begin{aligned} f(\boldsymbol{\mathfrak{X}}_f|\boldsymbol{L}) = \int\int\int & f(\boldsymbol{\mathfrak{X}}_f|\boldsymbol{Z},\boldsymbol{L}) \times f(\boldsymbol{\Sigma}_*|\boldsymbol{L}) \times f(\boldsymbol{T}|\boldsymbol{\Sigma}_{1|*},\boldsymbol{L}) \\ & \times f(\boldsymbol{\Sigma}_{1|*}|\boldsymbol{L})\mathrm{d}\boldsymbol{\Sigma}_*\mathrm{d}\boldsymbol{\Sigma}_{1|*}\mathrm{d}\boldsymbol{T} \end{aligned} \tag{130}$$

As shown in Lee & Zidek (1992), we can rewrite Expression (117) as

$$f(\boldsymbol{\mathfrak{X}}_f|\boldsymbol{\Theta},\boldsymbol{L}) \propto |\boldsymbol{\Sigma}_*|^{-1/2}|\boldsymbol{\Sigma}_{1|*}|^{-1/2}\exp\left\{-\frac{1}{2}(A+B)\right\}, \tag{131}$$

where

$$\begin{aligned} A &= \frac{1}{c-d}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})^T\boldsymbol{\Sigma}_*^{-1}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}}), \\ B &= \frac{1}{c}[\boldsymbol{T}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}}) - (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})]^T\boldsymbol{\Sigma}_{1|*}^{-1} \\ &\quad \times [\boldsymbol{T}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}}) - (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})], \\ d &= \frac{n}{\zeta(n+\zeta)} \end{aligned}$$

and $c = 1 + \frac{1}{\zeta}$. Inserting the results of Expressions (110) to (113) and (131) into Expression (130) further gives

$$\begin{aligned} f(\boldsymbol{\mathfrak{X}}_f|\boldsymbol{L}) \propto & \int |\boldsymbol{\Sigma}_*|^{-(\nu+n+1-u+g+1)/2} \\ & \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\left[\frac{1}{c-d}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})^T + \hat{\boldsymbol{\Psi}}_*\right]\boldsymbol{\Sigma}_*^{-1}\right)\right\}\mathrm{d}\boldsymbol{\Sigma}_* \end{aligned} \tag{132}$$

$$\begin{aligned} & \times \int\int |\boldsymbol{\Sigma}_{1|*}|^{-1/2}\exp\left\{-\frac{1}{2c}[\boldsymbol{T}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}}) - (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})]^T\right. \\ & \quad \left.\times\ \boldsymbol{\Sigma}_{1|*}^{-1}[\boldsymbol{T}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}}) - (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})]\right\} \\ & \quad \times |\boldsymbol{\Psi}_*^{-1} \otimes \boldsymbol{\Sigma}_{1|*}|^{-1/2} \\ & \quad \times \exp\left\{-\frac{1}{2}(\boldsymbol{T} - \boldsymbol{\Gamma})^T(\boldsymbol{\Psi}_*^{-1} \otimes \boldsymbol{\Sigma}_{1|*})^{-1}(\boldsymbol{T} - \boldsymbol{\Gamma})\right\}\mathrm{d}\boldsymbol{T} \end{aligned} \tag{133}$$

$$\times |\boldsymbol{\Sigma}_{1|*}|^{-(\nu+u+1)/2} \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}_{1|*}\boldsymbol{\Sigma}_{1|*}^{-1}\right)\right\}\mathrm{d}\boldsymbol{\Sigma}_{1|*} \tag{134}$$

The first integral we recognise as the integral of the kernel of an inverse Wishart pdf with parameters

$$\boldsymbol{C} = \left[\frac{1}{c-d}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})^T + \hat{\boldsymbol{\Psi}}_*\right]^{-1} \tag{135}$$

and degrees of freedom

$$\nu_{\boldsymbol{C}} = \nu + n + 1 - u$$

This implies that the the integral in Expression (132) is proportional to $|\boldsymbol{C}^{-1}|^{\nu_{\boldsymbol{C}}/2} = |\boldsymbol{C}|^{-\nu_{\boldsymbol{C}}/2}$.

As noted in Lee & Zidek (1992), the second integral is equal to the pdf of $(\boldsymbol{x}_f - \boldsymbol{\eta})|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*$ when

$$(\boldsymbol{x}_f - \boldsymbol{\eta})|\boldsymbol{T}, \boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^* \sim \text{Gauss}_u(\boldsymbol{T}(\boldsymbol{x}_f^* - \boldsymbol{\eta}^*), c\boldsymbol{\Sigma}_{1|*})$$

and $\boldsymbol{T}|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*$ has the distribution given in Expression (113). From Johnson & Wichern (2002, Result 4.6), we then know that

$$(\boldsymbol{x}_f - \boldsymbol{\eta})|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^* \sim \text{Gauss}_u(\boldsymbol{\mu}_D, \boldsymbol{\Sigma}_D), \tag{136}$$

where, by using Expressions (88), (89) and some general properties of the Kronecker product, (Brewer 1978), and the Vec operator, namely

$$\text{Vec}(\boldsymbol{x}) = \boldsymbol{x},$$

when $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$,

$$\begin{aligned}
\boldsymbol{\mu}_D &= E[(\boldsymbol{x}_f - \boldsymbol{\eta})|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*] \\
&= E[E[(\boldsymbol{x}_f - \boldsymbol{\eta})|\boldsymbol{T}, \boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*]|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*] \\
&= E[\boldsymbol{T}(\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*] \\
&= \boldsymbol{\Gamma}(\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Sigma}_D &= Cov\left((\boldsymbol{x}_f - \boldsymbol{\eta})|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*\right) \\
&= E\left[Cov\left((\boldsymbol{x}_f - \boldsymbol{\eta})|\boldsymbol{T}, \boldsymbol{\Sigma}_{1|*}\right)|\boldsymbol{\Sigma}_{1|*}\right] \\
&\quad + Cov\left(E[(\boldsymbol{x}_f - \boldsymbol{\eta})|\boldsymbol{T}, \boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*]|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*\right) \\
&= E[c\boldsymbol{\Sigma}_{1|*}|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*] + Cov\left(\boldsymbol{T}(\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*\right) \\
&= c\boldsymbol{\Sigma}_{1|*} + Cov\left(\text{Vec}\left(\boldsymbol{T}(\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)\right)|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*\right)^T \\
&= c\boldsymbol{\Sigma}_{1|*} + Cov\left(\left((\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)^T \otimes \boldsymbol{I}_u\right)\text{Vec}(\boldsymbol{T})|\boldsymbol{\Sigma}_{1|*}, \boldsymbol{x}_f^*\right) \\
&= c\boldsymbol{\Sigma}_{1|*} + \left((\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)^T \otimes \boldsymbol{I}_u\right) Cov(\text{Vec}(\boldsymbol{T})|\boldsymbol{\Sigma}_{1|*}) \left((\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)^T \otimes \boldsymbol{I}_u\right)^T \\
&= (c \otimes \boldsymbol{\Sigma}_{1|*}) + \left((\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)^T \otimes \boldsymbol{I}_u\right)(\boldsymbol{\Psi}_*^{-1} \otimes \boldsymbol{\Sigma}_{1|*}) \left((\boldsymbol{x}_f^* - \boldsymbol{\eta}^*) \otimes \boldsymbol{I}_u\right) \\
&= \left(\left[c + (\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)^T \boldsymbol{\Psi}_*^{-1}(\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)\right] \otimes \boldsymbol{\Sigma}_{1|*}\right) \\
&\equiv (w \otimes \boldsymbol{\Sigma}_{1|*}) = w\boldsymbol{\Sigma}_{1|*}.
\end{aligned}$$

Writing $\boldsymbol{v} = (\boldsymbol{x}_f - \boldsymbol{\eta}) - \boldsymbol{\Gamma}(\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)$ we get

$$\begin{aligned}
f(\boldsymbol{\mathfrak{X}}_f|\boldsymbol{L}) &\propto |\boldsymbol{C}|^{-\nu_{\boldsymbol{C}}/2} \\
&\quad \times \int |w\boldsymbol{\Sigma}_{1|*}|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{v}^T(w\boldsymbol{\Sigma}_{1|*})^{-1}\boldsymbol{v}\right\} \\
&\qquad \times |\boldsymbol{\Sigma}_{1|*}|^{-(\nu+u+1)/2} \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}_{1|*}\boldsymbol{\Sigma}_{1|*}^{-1}\right)\right\} \mathrm{d}\boldsymbol{\Sigma}_{1|*} \\
&\propto |\boldsymbol{C}|^{-\nu_{\boldsymbol{C}}/2} w^{-u/2} \\
&\quad \times \int |\boldsymbol{\Sigma}_{1|*}|^{-(\nu+u+1+1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\left[\frac{1}{w}\boldsymbol{v}\boldsymbol{v}^T + \boldsymbol{\Psi}_{1|*}\right]\boldsymbol{\Sigma}_{1|*}^{-1}\right)\right\} \mathrm{d}\boldsymbol{\Sigma}_{1|*}
\end{aligned}$$

This last integral we recognise as the kernel of an inverse Wishart distributed variable with parameters

$$\boldsymbol{G} = \left[\frac{1}{w}\boldsymbol{v}\boldsymbol{v}^T + \boldsymbol{\Psi}_{1|*}\right]^{-1}$$

and $\nu_{\boldsymbol{G}} = \nu + 1$ degrees of freedom. Therefore the integral is proportional to $|\boldsymbol{G}|^{-\nu_{\boldsymbol{G}}/2}$, and by using Result A.2.3 in Mardia et al. (1979)

$$\begin{aligned}
f(\boldsymbol{\mathfrak{X}}_f|\boldsymbol{L}) &\propto |\boldsymbol{C}|^{-\nu_{\boldsymbol{C}}/2} w^{-u/2} |\boldsymbol{G}|^{-\nu_{\boldsymbol{G}}/2} \\
&\propto \left|\hat{\boldsymbol{\Psi}}_* + \frac{1}{c-d}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})(\boldsymbol{x}^* - \boldsymbol{\mu}_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})^T\right|^{-\nu_{\boldsymbol{C}}/2} \\
&\quad \times w^{-u/2}\left|\frac{1}{w}\boldsymbol{v}\boldsymbol{v}^T + \boldsymbol{\Psi}_{1|*}\right|^{-\nu_{\boldsymbol{G}}/2} \\
&\propto \left|\hat{\boldsymbol{\Psi}}_*\right|^{-1/2}\left(1 + (\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})^T((c-d)\hat{\boldsymbol{\Psi}}_*)^{-1}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})\right)^{-\nu_{\boldsymbol{C}}/2} \\
&\quad \times w^{-u/2}\left|\frac{1}{w}\right|^{u(\nu+1)/2} |w\boldsymbol{\Sigma}_{1|*}|^{-\nu/2}|w\boldsymbol{\Sigma}_{1|*}|^{-1/2} \\
&\quad \times \left(1 + \boldsymbol{v}^T(w\boldsymbol{\Psi}_{1|*})^{-1}\boldsymbol{v}\right)^{-(\nu+1-p+p)/2} \\
&\propto \left|\hat{\boldsymbol{\Psi}}_*\right|^{-1/2}\left(1 + (\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})^T((c-d)\hat{\boldsymbol{\Psi}}_*)^{-1}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})\right)^{-\nu_{\boldsymbol{C}}/2} \\
&\quad \times \left(1 + [(\boldsymbol{x}_f - \boldsymbol{\eta}) - \boldsymbol{\Gamma}(\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)]^T(w\boldsymbol{\Psi}_{1|*})^{-1}\right. \\
&\qquad \left.[(\boldsymbol{x}_f - \boldsymbol{\eta}) - \boldsymbol{\Gamma}(\boldsymbol{x}_f^* - \boldsymbol{\eta}^*)]\right)^{-(\nu+1)/2}.
\end{aligned}$$

Comparing this last expression to the pdf of the multivariate student t distribution in Expression (83), we see that $\boldsymbol{x}_f|\boldsymbol{x}_f^*, \boldsymbol{L}$ has the distribution in Expression (119). Similarly by Bayes' rule

$$f(\boldsymbol{x}_f^*|\boldsymbol{L}) \propto \left|\hat{\boldsymbol{\Psi}}_*\right|^{-1/2}\left(1 + (\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})^T((c-d)\hat{\boldsymbol{\Psi}}_*)^{-1}(\boldsymbol{x}^* - \boldsymbol{\mu}^*_{\boldsymbol{\mu}|\bar{\boldsymbol{x}}})\right)^{-(\nu+n-u+1)/2}$$

By a similar argument as above, we see that $f(\boldsymbol{x}_f^*|\boldsymbol{L})$ has the multivariate student t distribution given in Expression (118). □

D Algorithms

In this section we present the pseudo code for some of the subroutines used in the Hierarchical Scale-Corrected Ensemble Kalman Filter. Note that these algorithms have not been optimised.

Algorithm 6: Bartlett Decomposition

function $\left[\boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}, \boldsymbol{\Sigma}_{*}\right] = \textbf{Bartlett}(\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}, u, g)$
$[\boldsymbol{\Sigma}, \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}, \boldsymbol{\Sigma}_{*}] = \textbf{partitionMatrix}(\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}, u, g)$
$\boldsymbol{\Sigma}_{1|*} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}\boldsymbol{\Sigma}_{*}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}^{T}$
$\boldsymbol{T} = \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}\boldsymbol{\Sigma}_{*}^{-1}$
$\boldsymbol{\Sigma}_{*} = \boldsymbol{\Sigma}_{*}$

Algorithm 7: Inverse Bartlett Decomposition

function $[\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}] = \textbf{invBartlett}(\boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}, \boldsymbol{\Sigma}_{*})$
$\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} = \begin{bmatrix} \boldsymbol{\Sigma}_{1|*} + \boldsymbol{T}\boldsymbol{\Sigma}_{*}\boldsymbol{T}^{T} & \boldsymbol{T}\boldsymbol{\Sigma}_{*} \\ \boldsymbol{\Sigma}_{*}\boldsymbol{T}^{T} & \boldsymbol{\Sigma}_{*} \end{bmatrix}$

Algorithm 8: Partition Matrix

function$[\boldsymbol{\Sigma}, \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}, \boldsymbol{\Sigma}_{*}] = \textbf{partitionMatrix}(\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}, u, g)$
$p = u + g$
$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}(1 : u, 1 : u)$
$\boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*} = \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}(1 : u, u + 1 : p)$
$\boldsymbol{\Sigma}_{*} = \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}(u + 1 : p, u + 1 : p)$

Algorithm 9: Sample Posterior Covariance Matrix

function$[\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}] =$ **getPosteriorCovMat**$(\boldsymbol{L}, n, \nu, \zeta, \boldsymbol{T}, \boldsymbol{\Sigma}_{1|*}, \boldsymbol{\eta}*, \boldsymbol{\Psi}_*, u)$

$\boldsymbol{L} = [\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*]$

$\bar{\boldsymbol{x}}^* = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}^{*,i}$

$\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}^{*,i} - \bar{\boldsymbol{x}}^*)(\boldsymbol{x}^{*,i} - \bar{\boldsymbol{x}}^*)^T$

$\hat{\boldsymbol{\Psi}}_{\boldsymbol{*}}^{\ -1} = n\boldsymbol{S} + \boldsymbol{\Psi}_{\boldsymbol{*}} + \frac{n\zeta}{n+\zeta}(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)^T$

Generate $\boldsymbol{\Sigma}_{\boldsymbol{*}} \sim W_g^{-1}(\hat{\boldsymbol{\Psi}}_*^{\ -1}, \nu + n - u)$

$\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}} =$ **invBartlett** $\left(\boldsymbol{\Sigma}_{1|*}, \boldsymbol{T}, \boldsymbol{\Sigma}_*\right)$

Algorithm 10: Posterior Expectation

function$[\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}] =$ **getPosteriorMu**$(\boldsymbol{L}, n, \zeta, \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}, \boldsymbol{\eta}_{\boldsymbol{\mathfrak{X}}}, u, g)$

$\boldsymbol{L} = [\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*]$

$[\boldsymbol{\Sigma}_{\boldsymbol{x}}, \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}, \boldsymbol{\Sigma}_{\boldsymbol{x}^*}] =$ **partitionMatrix**$(\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}, u, g)$

$\bar{\boldsymbol{x}}^* = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}^{*,i}$

$\boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}} = \boldsymbol{\eta}_{\boldsymbol{\mathfrak{X}}} + \left(\frac{n}{n+\zeta}\right)\begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}\boldsymbol{\Sigma}_{x^*}^{-1} \\ \boldsymbol{I}_g \end{bmatrix}(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)$

Algorithm 11: Posterior Mean Covariance Matrix

function$[\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}] =$ **getPosteriorMeanCovMat**$(\boldsymbol{L}, n, \nu, \zeta, \boldsymbol{\eta}^*, \boldsymbol{\Psi}_{\boldsymbol{\mathfrak{X}}}, u, g)$

$p = u + g$

$\boldsymbol{L} = [\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*]$

$\left[\boldsymbol{\Psi}_{1|*}, \boldsymbol{T}, \boldsymbol{\Psi}_*\right] =$ **Bartlett**$(\boldsymbol{\Psi}_{\boldsymbol{\mathfrak{X}}}, u, g)$

$\bar{\boldsymbol{x}}^* = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}^{*,i}$

$\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}^{*,i} - \bar{\boldsymbol{x}}^*)(\boldsymbol{x}^{*,i} - \bar{\boldsymbol{x}}^*)^T$

$\tilde{\boldsymbol{S}} = \frac{1}{n}\left(n\boldsymbol{S} + \boldsymbol{\Psi}_{\boldsymbol{*}} + \frac{n\zeta}{n+\zeta}(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)(\bar{\boldsymbol{x}}^* - \boldsymbol{\eta}^*)^T\right)$

$\alpha = \frac{\nu - p - 1}{\nu + n - p - 1}$

$\tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \left[\frac{1+\mathrm{tr}(\tilde{\boldsymbol{S}}\boldsymbol{\Psi}_*^{-1})}{\nu-u-1}\right]\boldsymbol{\Psi}_{1|*} + \boldsymbol{\Gamma}\tilde{\boldsymbol{S}}\boldsymbol{\Gamma}^T & \boldsymbol{\Gamma}\tilde{\boldsymbol{S}} \\ \tilde{\boldsymbol{S}}\boldsymbol{\Gamma}^T & \tilde{\boldsymbol{S}} \end{bmatrix}$

$\hat{\boldsymbol{\Sigma}} = \alpha\frac{\boldsymbol{\Psi}_{\boldsymbol{\mathfrak{X}}}}{\nu-p-1} + (1-\alpha)\tilde{\boldsymbol{\Sigma}}$

Algorithm 12: Downscale

function$[\boldsymbol{x}] =$ **downScale**$(\boldsymbol{x}^*, \boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}, \boldsymbol{\mu}_{\boldsymbol{\mathfrak{X}}}, u, g)$

$[\boldsymbol{\Sigma}_{\boldsymbol{x}}, \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}, \boldsymbol{\Sigma}_{\boldsymbol{x}^*}] =$ **partitionMatrix**$(\boldsymbol{\Sigma}_{\boldsymbol{\mathfrak{X}}}, u, g)$

$\boldsymbol{\Sigma}_{\boldsymbol{x}|\boldsymbol{x}^*} = \boldsymbol{\Sigma}_{\boldsymbol{x}} - \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}\boldsymbol{\Sigma}_{\boldsymbol{x}^*}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}^T$

Generate $\boldsymbol{\epsilon}_{\boldsymbol{x}|\boldsymbol{x}^*} \sim \mathrm{Gauss}_u\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{x}|\boldsymbol{x}^*}\right)$

$\boldsymbol{x} = \boldsymbol{\mu}_{\boldsymbol{x}} + \boldsymbol{\Sigma}_{\boldsymbol{x}\boldsymbol{x}^*}\boldsymbol{\Sigma}_{\boldsymbol{x}^*}^{-1}\left(\boldsymbol{x}^* - \boldsymbol{\mu}_{\boldsymbol{x}^*}\right) + \boldsymbol{\epsilon}_{\boldsymbol{x}|\boldsymbol{x}^*}$

Algorithm 13: Assimilate With Production Data

function$[\boldsymbol{x}]$ = **assimilateProduction**$(\boldsymbol{q}, \boldsymbol{x}, \boldsymbol{q}^o, \boldsymbol{D_q}, \boldsymbol{\Sigma_q^x}, \boldsymbol{\Sigma_q}^o, n_q, n_x)$
Generate $\boldsymbol{\epsilon_q}^o \sim \text{Gauss}_{n_q}(\boldsymbol{0}, \boldsymbol{\Sigma_q}^o)$
$[\boldsymbol{\Sigma_x}, \boldsymbol{\Sigma_{xq}}, \boldsymbol{\Sigma_q}] =$ **partitionMatrix**$(\boldsymbol{\Sigma_q^x}, n_x, n_q)$
$\boldsymbol{x}^q = \boldsymbol{x} + \boldsymbol{\Sigma_{xq}}\boldsymbol{\Sigma_q}^{-1}\left(\boldsymbol{q}^o + \boldsymbol{\epsilon_q}^o - \boldsymbol{D_q q}\right)$

Algorithm 14: Assimilate With Seismic Data

function$[\boldsymbol{r}^{qd}] =$
seismicInversion$(\boldsymbol{r}^q, \boldsymbol{d}^o, \boldsymbol{D_{d^o}}, \boldsymbol{D_r}, \boldsymbol{\mu_{\epsilon_{rd^o}}}, \boldsymbol{\Sigma_r}, n_{xy}, n_\theta, S, \boldsymbol{\Sigma_\omega}, \boldsymbol{\Sigma_{d^o}})$
for $j = 1$ *to* n_{xy} **do**
Generate: $\boldsymbol{\epsilon_\omega} \sim \text{Gauss}_{n_\theta \cdot S}\left(\boldsymbol{0}, \boldsymbol{\Sigma_\omega}\right)$
Generate: $\boldsymbol{\epsilon_{d^o}} \sim \text{Gauss}_{n_\theta \cdot S}\left(\boldsymbol{0}, \boldsymbol{\Sigma_{d^o}}\right)$
$\boldsymbol{\Sigma_{d^o}} = \boldsymbol{D_{r}}_j \boldsymbol{\Sigma_{r,j}} \boldsymbol{D_{r}}_j^T + \boldsymbol{D_{d^o,j}} \boldsymbol{\Sigma_\omega} \boldsymbol{D}_{\boldsymbol{d^o},j}^T + \boldsymbol{\Sigma_{\epsilon_{d^o}}}$
$\boldsymbol{\epsilon_r} = \boldsymbol{D_{d^o}}\left(\boldsymbol{\mu_{\epsilon_{rd^o}}} + \boldsymbol{\epsilon_\omega}\right) + \boldsymbol{\epsilon_{d^o}}$
$\boldsymbol{r}_j^{qd} = \boldsymbol{r}_j^q + \boldsymbol{\Sigma_{r,j}} \boldsymbol{D_{r}}_j^T \boldsymbol{\Sigma}_{\boldsymbol{d^o}j}^{-1}\left(\boldsymbol{d}_j^o + \boldsymbol{\epsilon_r}_j - \boldsymbol{D_{r}}_j \boldsymbol{r}_j^q\right)$