THE UNIVERSITY OF TULSA
THE GRADUATE SCHOOL

HISTORY MATCHING AND UNCERTAINTY CHARACTERIZATION USING ENSEMBLE-BASED METHODS

by
Alexandre Anozé Emerick

A dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
in the Discipline of Petroleum Engineering

The Graduate School
The University of Tulsa

2012

THE UNIVERSITY OF TULSA

THE GRADUATE SCHOOL

HISTORY MATCHING AND UNCERTAINTY CHARACTERIZATION USING ENSEMBLE-BASED METHODS

by
Alexandre Anozé Emerick

A DISSERTATION

APPROVED FOR THE DISCIPLINE OF

PETROLEUM ENGINEERING

By Dissertation Committee

_____, Chairperson
Albert C. Reynolds

_____
Leslie G. Thompson

_____
Mengjiao Yu

_____
Richard A. Redner

ABSTRACT

Alexandre Anozé Emerick (Doctor of Philosophy in Petroleum Engineering)

History Matching and Uncertainty Characterization Using Ensemble-based Methods

Directed by Albert C. Reynolds

339 pp., Chapter 10: Conclusions

(554 words)

In the last decade, ensemble-based methods have been widely investigated and applied for data assimilation of flow problems associated with atmospheric physics and petroleum reservoir history matching. Among these methods, the ensemble Kalman filter (EnKF) is the most popular one for history-matching applications. The main advantages of EnKF are computational efficiency and easy implementation. Moreover, because EnKF generates multiple history-matched models, EnKF can provide a measure of the uncertainty in reservoir performance predictions. However, because of the inherent assumptions of linearity and Gaussianity and the use of limited ensemble sizes, EnKF does not always provide an acceptable history-match and does not provide an accurate characterization of uncertainty. In this work, we investigate the use of ensemble-based methods, with emphasis on the EnKF, and propose modifications that allow us to obtain a better history match and a more accurate characterization of the uncertainty in reservoir description and reservoir performance predictions.

When EnKF is applied to reservoir history matching, it is necessary to keep the size of the ensemble small in order to obtain computational efficiency. However, a small ensemble size introduces sampling errors and limits the degrees of freedom to assimilate data, which deteriorate the results of the data assimilation. In this

work, we introduce a distance-based covariance localization procedure to reduce these problems. The proposed method is applied to several synthetic cases and one real field case, and the results show significant improvements compared to the standard EnKF and other choices of covariance localization. In addition, we investigate other methods proposed in the literature to ameliorate the negative effects of a small ensemble. We conclude that distance-based covariance localization is, to date, the most effective method.

Another problem that often occurs when EnKF is applied to reservoir history-matching problems is that the values of the objective function obtained by the final ensemble are relatively high, especially when compared to gradient-based history matching. High values of the objective function are associated with poor data matches. More importantly, a model that results in high value of the objective function gives a small value of the posterior probability density function, which suggests that this model is a sample from a low probability region. In this work, we introduce a procedure that combines EnKF and Markov chain Monte Carlo (MCMC) for the purpose of improving the final data matches and obtaining a more accurate characterization of uncertainty. We also introduce a procedure based on multiple assimilations of the same data with an inflated covariance of the measurement errors. This procedure forms the basis of a new iterative form of ensemble smoother (ES-MDA). We applied ES-MDA to history match production and/or seismic data in synthetic reservoir problems and in a real field case. The results show that ES-MDA outperforms EnKF in terms of the quality of data matches with a computational cost comparable with EnKF.

In this work, we also investigate the use of an adjoint-based implementation of the randomized maximum likelihood (RML) method and propose a new parameterization based on an ensemble of prior realizations to reduce the computational cost of RML.

Finally, we present a comparative study among eight ensemble-based methods in terms of the quality of the data matches, characterization of uncertainty and computational cost. Among the ensemble-based methods, ES-MDA obtained the best performance and resulted in a quantification of uncertainty comparable to an adjoint-based RML.

ACKNOWLEDGEMENTS

I would like to thank to my advisor and friend Prof. Albert Reynolds for his guidance, assistance and patience during the my graduate studies. I wish to extend my thanks to Prof. Leslie Thompson, Prof. Mengjiao Yu and Prof. Richard Redner for serving as members of my Dissertation Committee.

I am very grateful to Petrobras for sponsoring my PhD studies. Special thanks to my bosses and friends at Petrobras, Alvaro Peres, Mauro Becker and Flavia Pacheco for their incentive and support.

Thanks also to my friends at TU, Diego Oliveira, Emilio Coutinho, Fahim Forouzanfar, Mehrdad Shirangi, Mohammad Sharifi, Reza Tavakoli and Sy Do. It was really fun working with you guys!

This dissertation is dedicated to my wife, Tatiana, and my sons, Rafael and Gabriel, for their unconditional love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Reservoir simulation is a valuable tool for the decision-making process involved in the development and management of petroleum reservoir exploitation projects. A reservoir simulation model combines rock and fluid properties with a mathematical formulation to describe the fluid flow in the porous media. This model is then used to predict the performance of the reservoir under various operating conditions. However, in order to improve the predictive capability of a reservoir model, it is necessary to incorporate in this model all relevant information available about the field. The process of incorporating dynamic data in reservoir models is known in the petroleum literature as history matching.

History matching is an ill-posed problem because the amount of independent data available is much less than the number of variables. Hence, there exists an infinite number of combinations of the unknown reservoir properties that results in reservoir models which are able to match the observations. Besides that, the information available about the reservoir is always inaccurate and sometimes inconsistent. As a result, reservoir models are constructed with uncertain parameters; consequently, their predictions are also uncertain.

Over the last decade, increased importance has been attributed to the quantification of uncertainty in reservoir performance predictions and reservoir description in order to manage risk. Because of this interest in the characterization of uncertainty, it is now more commonplace to generate multiple history-matched models. However, generating multiple history-matched models in of itself does not necessarily lead to a correct assessment of uncertainty. In reality, uncertainty has no scientific

meaning outside the realm of statistics and probability. Bayesian statistics provides a straightforward theory for dealing with uncertainty. In the application of interest in this dissertation, Bayes' theorem allows one to write down the posterior probability density function (pdf) for reservoir model parameters conditional to field measurements such as production and seismic data. Then, the problem of characterizing the uncertainty in reservoir model parameters is reduced to the problem of sampling this posterior pdf. If a set of realizations of the vector of model parameters represents a set of samples from the posterior pdf, then a correct assessment of the uncertainty in specific outcomes of reservoir performance predictions can be generated by making a prediction with each model and then constructing statistics for the set of outcomes. For example, one can estimate the pdf for a well's predicted oil rate at each time from a histogram of the set of predictions.

Markov chain Monte Carlo (MCMC) provides a theoretically attractive method for sampling the posterior pdf for reservoir model parameters. It is well known that a properly designed MCMC method will sample this pdf correctly in the limit as the number of states in the chain goes to infinity [73, 169]. However, for high-dimensional problems, MCMC typically requires a large number of iterations to provide a reasonable sampling of the target pdf. When using MCMC to estimate the posterior pdf of reservoir model parameters conditional to observed production data, calculation of the probability of accepting the transition from the current state to the proposed state requires a run of the reservoir simulator to evaluate the likelihood part of the posterior pdf. This requirement makes the direct application of MCMC to realistic reservoir problems prohibitively expensive in terms of the computational cost.

The randomized maximum likelihood (RML) [96, 134, 142] is a method designed for sampling a posterior pdf of a vector of model parameters conditional to a set of measurements. Although RML can be proved to sample correctly only if the predicted data are linearly related to the vector of model parameters [142], RML

generates samples from different modes of the distribution so that it can approximate a non-Gaussian posterior pdf. The application of RML to generate N_e realizations requires the minimization of N_e objective functions, which, in general, can only be done efficiently with a gradient-based optimization algorithm, e.g., a quasi-Newton method [192, 59]. However, the efficient calculation of the gradients needed in a quasi-Newton method requires the implementation of the adjoint method [21, 19, 188, 105], and, unfortunately, adjoint code is not commonly available in commercial reservoir simulators.

Recently, the ensemble Kalman filter (EnKF) [43, 18, 80] has emerged as an attractive option for reservoir history-matching problems because it is easy to implement and computationally efficient. Moreover, because EnKF generates multiple history-matched models, EnKF can provide a measure of the uncertainty in reservoir performance predictions. Perhaps more importantly, when applied to field cases, EnKF appears to perform reasonably well and has typically given better results than a model based on a manual history matching [158, 14, 48, 74, 37]. In addition, EnKF does not require adjoint implementation for computing gradients, which makes the method easy to adapt to different types of model parameters and commercial reservoir simulators. Another attractive feature of the EnKF is the sequential data assimilation, which makes the method well suited for closed-loop reservoir management problems [87, 88, 180, 23, 20].

Even though there has been an intense interest and investigation of EnKF for reservoir history-matching problems in the last decade, because of the inherent assumptions of linearity and Gaussianity and the use of limited ensemble sizes, some problems remain with the method, including the following:

- Excessive and incorrect reduction in the ensemble variance after data assimilation.

- Limited number of degrees of freedom to history-match data.

- Nonphysical updates in model parameters and state variables.
- Inconsistency between updated model parameters and primary reservoir simulator variables.
- Relatively poor data matches when compared to gradient-based history matching methods.
- Difficulty to preserve more complex geological features, e.g., channels and facies.
- Inability to correctly characterize uncertainty in reservoir model parameters and production predictions.

1.1 Literature Review

1.1.1 Ensemble Kalman filter

Since its introduction by Evensen [43], the number of publications about EnKF became quite extensive. EnKF has been applied in diverse research fields, including oceanography [13, 94], atmospheric modeling [187], numerical weather prediction [82, 163], hydrology [140, 25, 107, 161] and petroleum reservoir history matching [1, 132]. Evensen [46] presents a chronological list of applications of EnKF. The first application for petroleum problems was presented by Lorentzen et al. [111], where EnKF was applied to a two-phase flow in a wellbore to improve predictions of pressure behavior. The first reservoir application of EnKF was presented by Nævdal et al. [125], where EnKF was used to update permeability fields for near-well reservoir models. After these pioneering applications, the interest and frequency of use of EnKF as a history-matching technique increased significantly. Some recent field applications of EnKF for history matching can be found in [158, 14, 48, 74, 37]. Two recent review papers [1, 132] summarize the main developments and applications of the EnKF in reservoir problems from 2001 to early 2010.

The sequential data assimilation characteristic of EnKF requires the modification of the traditional history-matching problem from a parameter-estimation problem to a parameter-state-estimation problem. Specifically, when applying EnKF for history matching, it is necessary to update a combined parameter-state vector, which includes the reservoir model parameters (uncertain reservoir rock properties) and the primary variables of the reservoir simulator (typically gridblock pressure, fluid saturations and bubble-point pressure in a standard black-oil reservoir simulator). The reason for updating primary variables, which represent the state of the dynamical system, is to avoid running the reservoir simulations from time zero (initial reservoir condition) after every data assimilation time-step. The underlying assumption is that the updated primary variables are statistically consistent with the ones that would be obtained by running the reservoir simulator with the updated set of model parameters from time zero. However, this consistency can be proved only for problems with Gaussian statistics, linear relation between model and predicted data and negligible model error [167]. However, the reservoir simulator equations are highly nonlinear. Hence, the assumption of consistency is invalid, which may deteriorate the performance of the data assimilation [167, 182]. In practice, after data assimilation at time t_n, the reservoir simulations may be restarted at time t_n with wrong pressure and saturation values, thereby violating the historical material balance of the field. In the extreme case, the analysis step in the EnKF may result in non-physical values for these variables, e.g., negative pressures or saturations larger than one. Wen and Chen [184] introduced the "confirming step," which consists of rerunning the simulator starting from the previous data assimilation time-step with the updated set of model parameters obtained at the current data assimilation time-step to obtain physically plausible state variables. However, Zafari and Reynolds [191] showed that this procedure is inconsistent for the linear case and should not be used. Some iterative procedures have been proposed to overcome inconsistency

because of the nonlinearity [143, 70, 103]. Most of these methods can be viewed as ensemble approximations of the RML method [1]. Reynolds et al. [143] and Li and Reynolds [103] presented two iterative forms of EnKF based on adjoint-gradients. The ensembles are used to approximate the Hessian in a Gauss-Newton type update equation. Because each ensemble member is updated based on a different gradient, these two methods are expected to be able to sample multimodal distributions, as the standard RML does. The main difficulty encountered in these two iterative methods is the dependency on the adjoint implementation. Gu and Oliver [70] proposed an iterative method called ensemble randomized maximum likelihood (EnRML), which uses the ensemble to estimate an average sensitivity matrix. With this procedure, Gu and Oliver [70] avoided the need of adjoint-based gradients. However, because all models are updated using the same average sensitivity, EnRML is not expected to sample multiple modes of a posterior distribution. In fact, EnKF is also not expected to sample multimodal distributions correctly because all models are updated based on the same Kalman gain. Moreover, because EnRML uses a rough approximation of the sensitivity matrix, the search direction is not guaranteed to be "downhill." If EnRML encounters an "uphill" search direction, the iterative procedure fails to improve the data matches. The computational cost of these iterative procedures is typically much higher than that of the standard EnKF. Perhaps the simplest procedure to overcome problems with inconsistency between updated model parameters and primary variables is to rerun the reservoir simulator with the latest ensemble of model parameters from time zero after each data assimilation. This procedure ensures consistency but requires a considerable increase in the computational time. Wang et al. [182] refers to this procedure as half-iteration EnKF (HI-EnKF). To improve computational efficiency, Wang et al. [182] proposed rerunning the ensemble from time zero after a data assimilation time-step only if the average relative change in the ensemble mean model is larger than a threshold value. Using the PUNQ-S3

case [55], Wang et al. [182] concluded that HI-EnKF and EnRML resulted in similar results. However, HI-EnKF requires much less computational time than EnRML.

One important advantage of EnKF is the fact that it does not require adjoint implementation for computing gradients. This makes EnKF easy to adapt to different types of reservoir model parameters, different types of data and different commercial reservoir simulators. Oliver and Chen [132] pointed out that there is a tendency to increase the types of model parameters estimated with EnKF. Gridblock porosities and permeabilities are the typical model parameters considered. However, other types of parameters including net-to-gross ratio [137], fluid contacts [167, 156, 182], fault transmissibilities [48, 156], end points of relative-permeability curves [23, 182], facies [3, 195], prior means [102, 27] and even structural parameters [155], where the simulation grid is deformed to update the top and base of the model, have been considered as model parameters using EnKF.

Although EnKF has been successfully applied to history match field cases, EnKF often fails to provide a reasonable characterization of uncertainty. In fact, to theoretically establish that the ensemble generated by assimilation of data with EnKF represents a correct sampling of the posterior pdf for the state vector requires one to assume a Gaussian prior model for the state vector, a linear relation between predicted data and the state vector, Gaussian measurement errors, which are uncorrelated in time, and that the dynamical system (forward model) represents a first-order Markov process. Even when the preceding assumptions hold, it is also necessary to let the size of the ensemble approach infinity to show that EnKF samples correctly because covariances are estimated from the ensemble of state vectors. Mandel et al. [119] present a rigorous proof that, under the above conditions, EnKF converges to the standard Kalman filter (KF) when the ensemble size goes to infinity. Mandel et al. [119] also point out that EnKF introduces dependence in the ensemble because the covariances are estimated based on all ensemble members. As

a result, the members of the final ensemble do not represent independent samples of the posterior pdf. In fact, earlier in 1998, Houtekamer and Mitchell [80] pointed out that because EnKF uses the same Kalman gain to update all ensemble members, EnKF introduces an inbreeding in the analysis. Houtekamer and Mitchell [80] proposed a variant of EnKF using two ensembles, where the statistics computed in one ensemble is used to update the other. In a comment on this paper, van Leeuwen [172] gave a theoretical justification for the inbreeding effect. However, van Leeuwen [172] pointed out that the main concern in the EnKF is related to the use of small ensemble sizes. Small ensembles are necessary for computational efficiency, but they introduce sampling errors and limit the degrees of freedom to assimilate data [1]. As a result, the ensemble variance obtained after data assimilation with EnKF tends to be greatly underestimated [172, 57, 6]. The underestimation of posterior variances is an important limitation of EnKF. Firstly, the low variance in the ensemble can make it difficult to assimilate new independent data. Secondly, but equally or more important, underestimation of posterior variances effectively indicates underestimation of uncertainty in the reservoir model parameters after data assimilation.

Besides the possible underestimation of uncertainty in the reservoir model parameters, EnKF also seems to fail to provide reasonable uncertainty quantification in the reservoir performance predictions. Using the PUNQ-S3 case, Lorentzen et al. [113] found that running EnKF with different initial ensembles resulted in cumulative distributions of forecasted field oil production which are not mutually consistent, i.e., the estimated distributions for each ensemble are too different to represent approximations of the same cumulative density function (cdf). Thulin et al. [168] also concluded that with only a single EnKF run, there is no control over the Monte Carlo error in the estimate of the conditional pdf. Thulin et al. [168] proposed repeating the data assimilation with EnKF multiple times using different prior ensembles and using the average cdf computed from all EnKF runs as the final distribution of pre-

dictions. Comparing EnKF with RML, Tavakoli and Reynolds [166] concluded that EnKF results in unreasonably high values of the objective function after data assimilation. A high value of the objective function is typically associated to poor data matches. However, and perhaps more importantly, high values of the objective function mean that the realizations of vectors of model parameters are of extremely low "probability," i.e., give a low value of the posterior pdf compared to those obtained with gradient-based optimization algorithms.

1.1.2 Sampling errors and rank deficiency in the EnKF

The relatively small ensemble size used in practice to represent covariances in the EnKF introduces sampling errors that produce spurious long-distance correlations between elements of the state vector and predicted data, whereas in reality, variables and data at gridblocks far apart should be uncorrelated. Spurious correlations can result in a non-negligible change in a component of the EnKF state vector due to assimilation of data at a location far away from the spatial location of this component, whereas if covariances were accurately represented, no change in the component would occur. When this incorrect change occurs during the EnKF analysis step, the variance of the state component is also incorrectly reduced. This loss of variance can make it difficult to modify properly the state component by assimilating later time data [80, 81, 46, 1].

Representing covariances with finite samples also limits the degrees of freedom available to update the state vector. In particular, the subvector corresponding to model parameters in any realization of the EnKF analyzed state vector at any data assimilation time-step is a linear combination of the corresponding initial N_e vectors of model parameters [1]. Thus, there are at most N_e coefficients that can be adjusted to make any particular realization consistent with all observed data. More generally, any updated vector of model parameters lies in the subspace spanned by the members of the prior ensemble. In addition, Lorenc [110] showed that if we

assimilate a perfect observation, i.e., noise free with the variances of measurement errors set to zero in the update equations, we lose at least one degree of freedom. Although we never assimilate perfect data, it is reasonable to expect that, after several data assimilation time-steps, we will experience a reduction in the available degrees of freedom left to assimilate subsequent data. This problem is critical in situations in which it is necessary to assimilate a large number of independent data, e.g., reservoirs with several wells [137] or time-lapse seismic data [158]. Although increasing the ensemble size reduces problems caused by sampling errors and limited degrees of freedom, computational efficiency requires the use of a small ensemble.

A standard procedure for reducing the spurious correlations due to sampling errors is to replace the prior (forecast) covariance at each data assimilation time-step by the Schur (element-wise) product of a correlation matrix, having compact support, and the forecast covariance matrix. This procedure is referred to as covariance localization. The first application of localization was presented in Houtekamer and Mitchell [80], where a distance cutoff was applied to the Kalman gain such that only model parameters within a critical distance of the observation were updated. According to the results in [80], the optimal radius of the cutoff increased as the ensemble size increased. The use of the Schur product for covariance localization was introduced by Houtekamer and Mitchell [81]. In this work, the authors pointed out that localization using the Schur product resulted in relatively smoother models compared to those obtained using the distance cutoff or using no localization at all. Another standard procedure to attenuate problems related to sampling errors and limited degrees of freedom is called local analysis [46]. In the local analysis procedure, we update a component of the state vector associated with a particular gridblock location using only the data in some local neighborhood of this gridblock. However, because components are updated independently, local analysis may introduce some discontinuities in the updated states. Nevertheless, Sakov and Bertino

[150] showed that if we also apply a correlation function for tapering the updates in the local analysis procedure, the results obtained are similar to those obtained from covariance localization. Both procedures, local analysis and covariance localization, require choosing a "localization region," and the performance of these methods is highly dependent on this choice.

Arroyo-Negrete et al. [10] proposed a localization procedure for petroleum reservoir applications based on streamlines. In this paper, the authors used streamlines to associate gridblocks with regions of influence of well data for each model of the ensemble. Then, a common region of influence that encompasses all individual regions of influence is defined and then used for covariance localization. In this procedure, an entry of the correlation matrix is defined as one if the gridblock is within the region of influence of the well and, otherwise, the entry is set equal to zero. In a synthetic problem, this procedure reduced the effect of spurious correlations on the updated permeability field. Devegowda et al. [31] modified the procedure proposed by Arroyo-Negrete et al. [10] by using the values of normalized streamline sensitivities so that the values of the correlation matrix for localization vary between zero and one. Streamline-based localization is intuitive and attractive because it incorporates some of the physics of the fluid transport into the localization procedure. However, if the localization is based only on those regions of influence, there is a possibility of keeping the model changes too localized around wells, especially when the prior model covariance has a long correlation range [1].

There exists some attempts to define localization schemes for general covariance structures without making the assumption that covariance depends on the distance between variables. One potential advantage of these non-distance dependent localization schemes is that, unlike distance-based localization methods, they can be applied to "localize" non-spatial variables, e.g., aquifer strength and end points of relative permeability curves. Furrer and Bengtsson [57] showed that it is possible to

minimize term-by-term the norm of the difference between the true forecast covariance matrix and the localized estimate ignoring the positive-definiteness constraint. Based on this procedure, Furrer and Bengtsson [57] obtained a simple expression that can be used as a non-distance localization procedure. Anderson [6] also presented a non-distance based localization procedure, which is called hierarchical ensemble filter. In the hierarchical ensemble filter, a small number of different ensembles is used to compute confidence factors, which are used to correct the regression coefficients (entries of the Kalman gain matrices). Zhang and Oliver [194] modified the method proposed by Anderson [6], such that instead of using multiple ensembles, we resample a single ensemble with replacement to generate bootstrapped ensembles.

Another procedure that has appeared in the literature to compensate for the underestimation of posterior variances in the EnKF is covariance inflation [7]. Covariance inflation is often used in oceanography and numerical weather prediction applications [7, 71, 138, 181, 5, 104], but has rarely been applied in reservoir history-matching problems. Covariance inflation uses an inflation factor to increase the covariance in the forecast ensemble without changing the mean. The inflation factor is a tuning parameter, and the "optimal" inflation factor is problem dependent. Evensen [47] proposed an adaptive covariance inflation procedure which computes the inflation factor based on the size of the ensemble and the measurements configuration.

1.1.3 Ensemble square root filters

Ensemble square root filter (EnSRF) is a generic terminology for a set of ensemble-based implementations of the Kalman filter which do not require to perturb the observations during the data assimilation [151]. There is an extensive literature about square root filters in the oceanography and numerical weather prediction research areas; see, e.g., [170, 151] and reference therein. Similar to covariance inflation, these methods have rarely been used in reservoir history-matching applications. The main motivation for these methods is to avoid additional sampling errors caused

by the perturbed observation scheme used in the EnKF [186, 170].

1.1.4 Ensemble smoother

The ensemble smoother (ES) was proposed by van Leeuwen and Evensen [174]. Unlike EnKF, ES does not assimilate data sequentially in time. Instead, ES computes a global update by simultaneously assimilating all data available. Other than that, the ES formulation is similar to EnKF. van Leeuwen and Evensen [174] found that EnKF performed better than ES when applied to an ocean circulation model. Evensen and van Leeuwen [49] compared ES and EnKF with Lorenz equations and concluded that EnKF outperforms ES because the recursive updates in the EnKF keep the ensemble of states on track and closer to the true solution. Recently, Skjervheim et al. [159] also compared ES and EnKF and concluded that both methods gave similar results for the reservoir history-matching problems considered in that paper. The main advantage of ES is that it avoids restarts of the reservoir simulator, which are necessary with the EnKF sequential data assimilation procedure. This makes ES much faster and easier to implement than EnKF when applied to reservoir history-matching problems. The elimination of simulation restarts also makes ES an attractive option for data assimilation workflows which integrate different parts of the reservoir modeling process, including seismic, structural and geological modeling with flow simulation; see, e.g., Zachariassen et al. [190]. These workflows typically require one to integrate different geomodeling softwares and may include upscaling of the rock properties, which makes the simulation restarts required by EnKF very inconvenient if not impossible.

Unlike oceanic and atmospheric models, which present chaotic and unstable dynamics [46, Chap. 6], reservoir simulation models are typically stable functions of the rock property fields. If we also neglect model uncertainty, which is a common assumption in reservoir history-matching problems, we only need to consider the parameter-estimation problem when applying ES. In this case, ES removes the

parameter-state consistency issue observed in the sequential data assimilation with EnKF. However, Reynolds et al. [143] showed that EnKF is similar to applying, at each data assimilation time-step, one Gauss-Newton iteration with a full step with the sensitivity matrix replaced by an average sensitivity matrix estimated from the ensemble. Sequential data assimilation seems to be one of the reasons why, in most cases, EnKF performs better than ES. Specifically with EnKF, at each data assimilation time-step, one Gauss-Newton correction is done to each realization in the ensemble of states, but because there are production data for several consecutive data assimilation time-steps, which typically have overlapping information content, EnKF accumulates several Gauss-Newton corrections, keeping the ensemble conditioned to the production history. With ES, on the other hand, all data are assimilated simultaneously, which means that a single Gauss-Newton correction is applied to condition the ensemble to all data available. Hence, ES may not be able to provide reasonable data matches when applied to reservoir history-matching problems.

1.1.5 Markov chain Monte Carlo

In the MCMC, a sequence of realizations (states) is obtained by simulating a Markov chain, i.e., the probability of introducing a new state in the chain depends only on the current state. Feller [53], Tierney [169] and Nummelin [128] present theoretical results that guarantee that asymptotically the states of the Markov chain represent samples of the target distribution. Ripley [144] provides a summary of theoretical results. However, for high-dimensional problems, MCMC typically requires a large number of proposals to provide a reasonable sampling of the target pdf.

Inspired by the early work of Tjelmeland et al. [171] and Hegstad and Omre [77, 78], Oliver et al. [133] explored the application of MCMC methods to sample the posterior pdf for the log-permeability field conditional to pressure data for a small two-dimensional single-phase flow problem. The forward model was a finite-difference simulator on a 15×15 grid. The proposal mechanisms considered were based on local

and global perturbations drawn from the prior Gaussian pdf for the log-permeability field, as well as from a Gaussian pdf centered at the maximum a posteriori (MAP) estimate with covariance matrix given by the inverse Hessian matrix evaluated at the MAP estimate. If the Gaussian pdf linearized about the MAP estimate is a good approximation of the target pdf (the posterior pdf for the log-permeability field), then proposing new states from this Gaussian pdf should result in a higher acceptance rate than would be achieved by proposing from the prior Gaussian pdf for the log-permeability field. Although for the example considered by Oliver et al. [133], this higher acceptance rate was achieved, the number of independent states in the Markov chains generated was quite small. In "the high variance case," (prior variance of each gridblock log-permeability equal to 1.0), for each proposal probability distribution considered, fewer than 15 independent states (log-permeability fields) were obtained in the Markov chain generated with 50,000 proposed transitions.

Ma et al. [117] and Ma et al. [118] introduced a two-stage MCMC method for sampling a posteriori distribution for the permeability field conditional to water-cut and gas-oil ratio data. In these two papers, the objective was to sample the conditional distribution for the permeability field defined on a fine-grid model. For each new proposed permeability field, they compute the data mismatch term by running a very coarse grid simulation model using an upscaled permeability field and then use either a linear relation or a nonparametric regression-based statistical function to compute what the data match on the fine grid would be. If the estimated fine-grid data match is better than the fine-grid data match obtained for the current state in the Markov chain, they run the fine grid reservoir simulation model to generate the true fine-grid data matches, which are needed to apply the Metropolis-Hastings condition for deciding whether to accept the proposed permeability field as the new state in the chain. By using the coarse grid to "pre-screen" proposed realizations, they improved computational efficiency. For a synthetic two-dimensional

problem, they achieved an acceptance rate of approximately 30% and use only about 1,000 proposals to generate the chain. It is not clear, however, that the chain of permeability fields generated by this procedure is well-mixed. For example, the four samples of the permeability field (four states in the chain) shown in Fig. 13 of [118] are very similar. If the chain is, in fact, poorly mixed, then it means that MCMC has not explored the full posterior distribution, i.e., MCMC generated samples from only a local region of the posterior pdf.

1.1.6 Randomized maximum likelihood

RML was independently proposed by Kitanidis [96] and Oliver et al. [134] as an approximate sampling method. Oliver et al. [134] originally introduced the method as a proposal mechanism for MCMC. However, because the acceptance criterion was difficult to evaluate and the acceptance rate was very high, the authors suggested accepting all proposals, which results in the procedure now known as RML method. Reynolds et al. [142] showed that when the prior pdf is Gaussian and the relationship between the vector of model parameters and predicted data is linear, RML provides a correct sampling of the posterior pdf. Reynolds et al. [142] also investigated the case in which the prior mean is uncertain and showed that if the prior mean is properly incorporated as a model variable, realistic realizations of the rock properties fields are obtained with RML. For the nonlinear case, RML is not guaranteed to sample the posterior pdf correctly. However, it has been shown that RML is able to sample multiple modes in nonlinear univariate "toy problems" [134, 191, 165], which suggests that RML can approximate non-Gaussian distributions.

Liu and Oliver [109] considered a one-dimensional single-phase flow problem on a simulation grid consisting of twenty gridblocks. Pressure data were available at a single producing well and two observation wells. The objective was to compare methods for sampling the posterior pdf for the porosity and log-permeability fields. Markov chains were generated from a Metropolis-Hastings MCMC implementation

with proposals for the rock property fields generated from the prior Gaussian pdf for these fields using a local perturbation procedure. The chains generated displayed extremely slow mixing. For example, Fig. 6 of [109] indicates that the values of the objection function may be correlated over as many as one million successive states of the chain. Because of this, they generated a very long Markov chain consisting of 320 million states and assumed that states in this chain obtained subsequent to the burn-in period represented the correct target distribution. They were able to generate this long chain because running the reservoir simulator for their problem required less than 1/4 of a second of CPU time. They used the results of this very long chain as a reference to compare the following approximate sampling methods: linearization about the MAP, RML and pilot point method [29, 99, 100, 189, 185]. Liu [108] included the gradual deformation method [148, 83] in this comparative study. The results presented in [109] and [108] show that RML was the only approximate sampling method which gave an acceptable characterization of uncertainty.

RML requires one minimization of an objective function to generate each sample of the posterior pdf. Conceptually, any optimization method could be applied to minimize this objective function. However, it seems that only gradient-based optimization methods are computationally feasible for large-scale reservoir models. Newton-like methods, including Gauss-Newton (GN) and Levenberg-Marquardt (LM) have been applied for generating conditional realizations with RML [188, 105, 192]. These methods require the computation of the whole sensitivity matrix, i.e., they require the derivative of predicted data with respect to each model parameter in order to build the Hessian matrix. Two methods commonly used for computing sensitivity coefficients are the direct method, also known as gradient simulator method [9, 147], and the adjoint method [21, 19, 188, 105]. In the direct method, the computation of the sensitivity matrix requires the solution of one matrix problem at the end of each reservoir simulation time-step for each reservoir model pa-

rameter [147]. In the adjoint method, the computation of the sensitivity matrix requires solving one matrix problem for each datum [188, 105, 147]. Thus, unless the number of data points or the number of reservoir model parameters is small, methods based on the whole sensitivity matrix are impractical for large field cases. Zhang and Reynolds [192] compared the computational efficiency of GN and LM with a nonlinear conjugate gradient method [54] and the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) method [126]. The last two methods require only the gradient vector of the objective function during the minimization process. The gradient vector can be computed efficiently with the adjoint method with the solution of a single matrix problem for each time-step (backward in time). Zhang and Reynolds [192] concluded that LBFGS is the most promising optimization method for large-scale reservoir problems. Perhaps the main difficulty associated with the application of gradient-based RML to large reservoir problems is the dependency on adjoint-code implementation, which, unfortunately, is not commonly available in commercial reservoir simulators.

In spite of the fact that the number of publications on reservoir applications of the EnKF is quite large, very few [60, 191, 165] compare EnKF with RML. Zafari and Reynolds [191] concluded that EnKF failed to provide a reasonable sampling of a bi-modal posterior pdf of a nonlinear "toy problem." For the same problem, RML resulted in a nearly perfect sampling. Gao et al. [60] concluded that EnKF and RML resulted in similar uncertainty quantifications for reservoir production predictions in the PUNQ-S3 test case [55]. However, Tavakoli and Reynolds [165] concluded that RML resulted in much better data matches and uncertainty characterization than EnKF when applied to estimate the permeability field in a simple reservoir problem.

1.1.7 Parameterization methods for history matching

In a history-matching problem, the number of uncertain reservoir model parameters can be on the order of tens to hundreds of thousands. For this reason,

parameterizing the history matching to search for solutions in a lower dimensional space is very common in the petroleum literature. Perhaps the first parameterization method applied for history matching is zonation [85, 86, 63, 157]. In zonation, the reservoir engineer chooses regions (zones) of the model in which the rock properties are allowed to be adjusted during history matching. Zonation is also the base procedure for manual history matching [2, 69, 129]. The main criticisms about zonation include the sometimes arbitrary choices of the zones, which may not reproduce the geological understanding of the reservoir; discontinuities introduced in the rock properties between zones; and the inability to achieve good data matches because of the limited number of parameters. The gradzone [16] and multiscale [68, 67] parameterization methods can be considered as "automatic" implementations of the zonation parameterization. Another parameterization commonly used in the petroleum and ground water literature is the pilot point method [29, 99, 100, 189, 185]. In the pilot point method, the properties of a few reservoir gridblocks are selected as model parameters (pilot points). The values at the remaining gridblocks are obtained by kriging. However, this parameterization tends to result in extreme values of parameters in the pilot points and to create nonphysical artifacts in the rock properties [109]. Oliver [130] and Reynolds et al. [141] applied the Karhunen-Lòeve expansion (eigen-decomposition) of the prior covariance matrix for parameterization of history-matching problems. However, the eigen-decomposition of a covariance matrix can be very expensive, or even not computationally feasible, for large reservoir problems. Sarma et al. [152] avoided this difficulty in the Karhunen-Lòeve expansion using a kernel eigenvalue decomposition. In this approach, the covariance matrix is replaced by an ensemble approximation based on a large number of unconditional realizations. Although parameterizations based on eigen-decomposition of the covariance matrix can be very efficient in reducing the dimensionality of the problem, there is a question of whether keeping only the main eingenpairs of the prior covariance results in

models which are "too smooth" because the high frequency components, which are associated with local reservoir heterogeneities, are discarded; see, e.g., Oliver and Chen [131], where it is shown that realizations resulting from the main eigenvectors of the prior covariance matrix are too smooth to be considered samples of the prior pdf. Gradual deformation [148, 83] also parameterizes the history matching in terms of the prior covariance. This method tends to result in reasonable geological models, however, it seems to encounter problems in obtaining good data matches [58]. Sarma et al. [153] introduced the use of kernel principal component analysis as a parameterization for preserving higher order statistical moments. In this paper, the authors claim that using polynomial kernels of order n corresponds to preserving statistical moments of order $2n$ in a transformed space. The method requires inverting the model back from the transformed to the original space, which is called the pre-image problem. However, there is no guarantee that this inversion is unique or that an inverse exists [154]. Besides that, there is no assurance that the higher-order statistical moments will be preserved after the "solution" of the pre-image problem. Sarma et al. [153] applied this procedure to channelized models generated with multipoint geostatistics [162]. Although the method was able to obtain a model with channelized features after history matching, the model does not give a good match of water cut data. Rodrigues [147] presented a parameterization based on the main singular values of the dimensionless sensitivity matrix G_{D} [193]. He also described how to compute the singular vectors corresponding to the largest singular values using the Lanczos algorithm [65, 177] without explicitly forming G_{D}. Tavakoli and Reynolds [166] extended the ideas presented in Rodrigues [147] and provided a theoretical argument showing that the parameterization in terms of the truncated singular value decomposition (TSVD) of G_{D} is optimal in terms of reduction of uncertainty. In [165], the same authors implemented the TSVD-based parameterization with the RML method, which seems to be the most efficient RML implementation presented

to date. In the same paper, they also suggest that EnKF can be used as a "first guess" for RML. EnKF can also be interpreted as a parameterization method in which each realization of the final ensemble is represented by a linear combination of the N_e realizations in the initial ensemble.

1.1.8 History matching of time-lapse seismic data

Time-lapse (4D) seismic consists of a sequence of 3D seismic acquisitions obtained at different times of the production period of a reservoir. By computing the difference between time-lapsed surveys, it is possible to identify changes in the reservoir caused by production. For example, time-lapse seismic allows us to identify waterflooded regions and areas with bypassed oil in the reservoir, which helps to define target locations for infill wells. The number of publications on field applications of time-lapse seismic in the petroleum literature is large; see, e.g., [160, 116, 178, 91, 90, 129, 56].

Using time-lapse seismic data for reservoir history matching is not a new subject in the petroleum literature. The number of publications on this subject is quite extensive. One of the first applications of time-lapse seismic data for history matching was presented in Huang et al. [84]. This paper uses a perturbation method similar to simulated annealing [95, 175] to match production and time-lapse amplitude data. A similar method was also applied by Waggoner et al. [179] to history match production and acoustic impedance data in a gas-condensate reservoir in the Gulf of Mexico. Mezghani et al. [123] presented a procedure to history match time-lapse seismic data by perturbing the geostatistical model using the gradual deformation method [148, 83]. The derivatives required by the optimization method were computed using a finite-difference approximation. Gosselin et al. [66] implemented the Levenberg-Marquardt method for the simultaneous history matching of production and time-lapse seismic data. The derivatives were calculated using the gradient simulator method [9, 147]. Dong and Oliver [32] presented a gradient-based imple-

mentation for history matching time-lapse impedance data using the adjoint method for computing gradients and the LBFGS method for optimization. Skjervheim et al. [158] presented the first application of EnKF for history matching time-lapse seismic data. In this paper, the authors applied EnKF to a small synthetic case and concluded that the incorporation of seismic data resulted in a better estimation of the permeability field. They also presented an application to a real field case, where they were able to improve the seismic data match without deterioration of the production data match.

History matching time-lapse seismic data requires the capability to compute seismic data from a given reservoir model. This task can be accomplished by introducing a rock-fluid model to convert the reservoir properties and the simulator primary variables (pressure and saturations) into modeled elastic properties. The most widely used model to predict the seismic response of a reservoir due to production is the Gassmann model [62]. Among the elastic properties typically used for seismic data history matching, the most common choices are pressure-wave impedance (P-impedance or acoustic impedance) and Poisson's ratio; see, e.g, [179, 66, 75, 158, 50]. However, other seismic attributes, such as amplitudes [76] and time-shifts [97, 159], have also been used. Fahimuddin et al. [51] investigated different kinds of seismic data for history matching with EnKF. They concluded that time-difference impedance data performed better than time-difference amplitude data.

1.2 Research Objectives

The main objectives of this research are as follows: (i) investigate the use of ensemble-based methods, with emphasis on the EnKF, for history matching and uncertainty characterization in petroleum reservoir models; (ii) propose improvements for the problems commonly observed with the EnKF, with a focus on preserving the main advantages of the method, namely, easy implementation and computational efficiency.

1.3 Dissertation Organization

This dissertation is organized into ten chapters, including the Introduction and Conclusions. Chapter 2 presents the theoretical fundamentals of history matching as a Bayesian estimation problem. In this chapter, we present the formulation of the EnKF and its variants. Chapter 3 is dedicated to covariance localization. In this chapter, we introduce a distance-based covariance localization procedure for reservoir history-matching problems. Chapter 4 presents a comparison between covariance localization and other methods for dealing with the underestimation of posterior variances in the EnKF. Chapter 5 presents a method which combines EnKF with MCMC for improving the sampling results obtained by EnKF. Chapter 6 introduces a procedure in which the same data is assimilated multiple times with an inflated covariance of the measurement errors. In this chapter, we investigate the use of multiple data assimilation of time-lapse seismic and production data using the ensemble smoother. Chapter 7 presents a real field history-matching case. For this field, we assimilate 3D and 4D seismic data together with production data using EnKF and ES with multiple data assimilation. Chapter 8 introduces an ensemble-based parameterization to reduce the computational cost of an adjoint-based RML. Chapter 9 presents the results of a comparative study among eight ensemble-based methods and an adjoint-based RML for a small but highly nonlinear reservoir history-matching problem. The objective of this study is to compare the performance of these methods in terms of the quality of the data matches, uncertainty characterization and computational cost. The last chapter of this dissertation presents the conclusions.

CHAPTER 2

BAYESIAN FORMULATION OF THE HISTORY MATCHING PROBLEM

Bayesian statistics provides an adequate framework for dealing with poorly known model parameters and conditioning these parameters to inaccurate data. The starting point is Bayes' theorem which allows us to write the conditional pdf, $f(m|d_{\text{obs}})$, of a N_m-dimensional vector of model parameters, m, given a N_d-dimensional vector of observations, d_{obs}, as

$$\pi(m) \equiv f(m|d_{\text{obs}}) = \frac{f(d_{\text{obs}}|m)f(m)}{f(d_{\text{obs}})} = \frac{f(d_{\text{obs}}|m)f(m)}{\int_{\text{D}} f(d_{\text{obs}}|m)f(m)dm} = aL(m|d_{\text{obs}})f(m). \tag{2.1}$$

In the above equation, $f(m)$ is the prior pdf of the vector of model parameters and $f(d_{\text{obs}})$ is the pdf of the vector of observations. $f(d_{\text{obs}}|m)$ is the conditional pdf of d_{obs} given m. This pdf corresponds to the likelihood function, which is denoted by $L(m|d_{\text{obs}})$; a is a normalizing constant. Assuming that the prior pdf is Gaussian and that measurement errors have also a Gaussian distribution, we can write $f(m|d_{\text{obs}})$ as

$$
\begin{aligned}
f(m|d_{\text{obs}}) &= a \exp\left\{-\frac{1}{2}\left(m - m_{\text{pr}}\right)^{\text{T}} C_{\text{M}}^{-1}\left(m - m_{\text{pr}}\right)\right\} \\
&\times \exp\left\{-\frac{1}{2}\left(g(m) - d_{\text{obs}}\right)^{\text{T}} C_{\text{D}}^{-1}\left(g(m) - d_{\text{obs}}\right)\right\} \\
&= a \exp\left\{-\frac{1}{2}\left(m - m_{\text{pr}}\right)^{\text{T}} C_{\text{M}}^{-1}\left(m - m_{\text{pr}}\right)\right. \\
&- \left.\frac{1}{2}\left(g(m) - d_{\text{obs}}\right)^{\text{T}} C_{\text{D}}^{-1}\left(g(m) - d_{\text{obs}}\right)\right\} \\
&= a \exp\left\{-O(m)\right\}, \qquad (2.2)
\end{aligned}
$$

where

$$
O(m) = O_m(m) + O_d(m), \qquad (2.3)
$$

with

$$
O_m(m) = \frac{1}{2}\left(m - m_{\text{pr}}\right)^{\text{T}} C_{\text{M}}^{-1}\left(m - m_{\text{pr}}\right) \qquad (2.4)
$$

and

$$
O_d(m) = \frac{1}{2}\left(g(m) - d_{\text{obs}}\right)^{\text{T}} C_{\text{D}}^{-1}\left(g(m) - d_{\text{obs}}\right). \qquad (2.5)
$$

In the above equations, $O(m)$ is the so-called objective function. In a history-matching problem, we typically want to find a minimum of this objective function because minimizing $O(m)$ is equivalent to maximizing the posterior pdf $f(m|d_{\text{obs}})$. In the above equations, m_{pr} is the N_m-dimensional vector containing the prior mean; C_{M} is the $N_m \times N_m$ prior covariance matrix of model parameters; C_{D} is the $N_d \times N_d$ covariance matrix of measurements errors and $g(m)$ is the vector of predicted data for a given vector m. For the applications of interest in this dissertation, $g(m)$ is obtained from a reservoir simulation run. Note that the objective function is composed of two parts: $O_m(m)$, which is the model mismatch part and $O_d(m)$, which is the data mismatch part. Typically, m_{pr} corresponds to a very smooth model, often flat (when

the prior mean is constant). Hence, if an estimate m gives a high value of $O_m(m)$, this estimate is a rough model, i.e., the model possibly presents unrealistically high (overshooting) or low (undershooting) values of gridblock rock properties. If m gives a high value of $O_d(m)$, this estimate corresponds to a poor data match. Perhaps more importantly, if m corresponds to a high value of $O(m) = O_m(m) + O_d(m)$, then m has low "probability," i.e., m gives a low value of the posterior pdf.

In the posterior pdf derived in Eq. 2.2, we neglected the uncertainty associated with model errors. However, it can be shown that Eq. 2.2 is still valid for the case with model errors if we assume that the model errors also follow a Gaussian distribution [164, Chap. 1]. In this case, the covariance matrix C_{D} in the objective function of Eq. 2.3 corresponds to the sum of the covariance matrix of the measurement errors and the covariance matrix of the model error.

Typically in a reservoir history-matching problem, the number of unknown model parameters greatly exceeds the number of independent data. This makes the problem of minimizing the mismatch between predicted and observed data ill-posed, and some regularization is often needed. However, the objective function (2.3), which was obtained from a direct application of Bayes' theorem, already contains a built-in regularization term $O_m(m)$. Besides regularizing the minimization problem, $O_m(m)$ has another very important significance, namely, it represents the prior geostatistical knowledge of the reservoir.

2.1 The Maximum a Posteriori Estimate

The maximum a posteriori (MAP) estimate is the model that maximizes $f(m|d_{\mathrm{obs}})$, or equivalently minimizes the objective function (Eq. 2.3) [135, p. 143], i.e.,

$$m_{\mathrm{MAP}} = \arg\min_{m} O(m). \tag{2.6}$$

Because the functional $g(m)$ is nonlinear, the posterior pdf $f(m|d_{\text{obs}})$ is not Gaussian even when the prior $f(m)$ is Gaussian. In this case, $f(m|d_{\text{obs}})$ may have multiple modes, where each mode corresponds to a local minimum of $O(m)$. In fact, $O(m)$ may even have multiple global minima, and, in this case, the MAP estimate is not unique.

If the theoretical relation between data and model parameters is linear, the posterior pdf given by Eq. 2.2 is also Gaussian. In the linear case, we write the relation between model and predicted data as

$$g(m) = Gm, \tag{2.7}$$

where G is the $N_d \times N_m$ sensitivity matrix. For the linear case, the posterior mean is equivalent to the MAP estimate, which can be obtained by using Eq. 2.7 in Eq. 2.3 and requiring the gradient of $O(m)$ to vanish, i.e.,

$$\begin{aligned} 0 &= \nabla_m O(m) \\ &= C_{\text{M}}^{-1}\left(m - m_{\text{pr}}\right) + G^{\text{T}} C_{\text{D}}^{-1}\left[Gm - d_{\text{obs}} - Gm_{\text{pr}} + Gm_{\text{pr}}\right] \\ &= \left(C_{\text{M}}^{-1} + G^{\text{T}} C_{\text{D}}^{-1} G\right)\left(m - m_{\text{pr}}\right) + G^{\text{T}} C_{\text{D}}^{-1}\left(Gm_{\text{pr}} - d_{\text{obs}}\right). \end{aligned} \tag{2.8}$$

Solving Eq. 2.8 for m and denoting the result as m_{MAP}, we obtain

$$m_{\text{MAP}} = m_{\text{pr}} - \left(C_{\text{M}}^{-1} + G^{\text{T}} C_{\text{D}}^{-1} G\right)^{-1} G^{\text{T}} C_{\text{D}}^{-1}\left(Gm_{\text{pr}} - d_{\text{obs}}\right). \tag{2.9}$$

Following Oliver et al. [135, Chap. 7], Eq. 2.9 can be written as

$$m_{\text{MAP}} = m_{\text{pr}} + C_{\text{M}} G^{\text{T}}\left(C_{\text{D}} + G C_{\text{M}} G^{\text{T}}\right)^{-1}\left(d_{\text{obs}} - Gm_{\text{pr}}\right), \tag{2.10}$$

which is particularly useful when the number of data is considerably less than the

number of model parameters, i.e., $N_d \ll N_m$. Note that in Eq. 2.10, we have the inverse of a $N_d \times N_d$ matrix while in Eq. 2.9, we have the inverses of $N_m \times N_m$ matrices.

Expanding $O(m)$ about m_{MAP}, we can write

$$\begin{aligned} O(m) &= O(m_{\text{MAP}}) + \left(\nabla_m O(m_{\text{MAP}})\right)^{\text{T}} (m - m_{\text{MAP}}) \\ &+ \frac{1}{2} (m - m_{\text{MAP}})^{\text{T}} B_{\text{MAP}} (m - m_{\text{MAP}}), \end{aligned} \tag{2.11}$$

with B_{MAP} denoting the Hessian matrix. Using the linear relationship (2.7) in the objective function (2.3) and taking the gradient of $O(m)$ with respect to m twice, we obtain

$$B_{\text{MAP}} = \nabla_m \left[\left(\nabla_m O(m_{\text{MAP}})\right)^{\text{T}} \right] = C_{\text{M}}^{-1} + G^{\text{T}} C_{\text{D}}^{-1} G. \tag{2.12}$$

Eq. 2.12 shows that for the linear case, B_{MAP} is a constant matrix. Hence, the expansion (2.11) in exact. Using $\nabla_m O(m_{\text{MAP}}) = 0$ and using the expansion (2.11), we can write the posterior pdf as

$$\begin{aligned} f(m|d_{\text{obs}}) &= a \exp\left\{-O(m_{\text{MAP}})\right\} \exp\left\{ -\frac{1}{2} (m - m_{\text{MAP}})^{\text{T}} B_{\text{MAP}} (m - m_{\text{MAP}}) \right\} \\ &= \widehat{a} \exp\left\{ -\frac{1}{2} (m - m_{\text{MAP}})^{\text{T}} B_{\text{MAP}} (m - m_{\text{MAP}}) \right\}, \end{aligned} \tag{2.13}$$

which is a Gaussian distribution for m with mean m_{MAP} and covariance given by

$$
\begin{aligned}
C_{\mathrm{MAP}} &= B_{\mathrm{MAP}}^{-1} \\
&= \left(C_{\mathrm{M}}^{-1} + G^{\mathrm{T}} C_{\mathrm{D}}^{-1} G\right)^{-1} \\
&= C_{\mathrm{M}} - C_{\mathrm{M}} G^{\mathrm{T}} \left(G C_{\mathrm{M}} G^{\mathrm{T}} + C_{\mathrm{D}}\right)^{-1} G C_{\mathrm{M}}, \qquad (2.14)
\end{aligned}
$$

where the last equality can be obtained by using the matrix inversion lemmas as described in [135, Chap. 7].

2.2 Markov Chain Monte Carlo

For the nonlinear case, the posterior pdf given by Eq. 2.2 is not Gaussian and cannot be fully described by the mean and covariance. Nevertheless, for the purpose of evaluating uncertainty, it suffices to generate a sampling of this pdf. MCMC is a procedure for generating samples of a target pdf, $\pi(m)$, using a Markov chain mechanism, i.e., the probability of introducing a new state in the chain depends only on the current state. The proposal/acceptance criteria are constructed so that states of the chain are more likely to remain for a longer time, i.e., for more iterations, in high probability regions of the target pdf. Relevant theoretical results that guarantee that, asymptotically, the states of the chain are samples of $\pi(m)$ can be found in Feller [53], Tierney [169] and Nummelin [128]. Ripley [144] provides a summary of theoretical results. Most MCMC implementations define transition probabilities based on the idea of Metropolis et al. [122], who proposed defining the probability of a transition from state m to $\widehat{m}$ as the product of proposing this transition, $q(m, \widehat{m})$, times the probability of accepting the proposed transition, $\alpha(m, \widehat{m})$. In a recent survey, the Metropolis algorithm was selected as one of the ten algorithms that have had the greatest influence on the development and practice of science and engineering in the 20th Century [26]. Hastings [73] extended the Metropolis algorithm to what is perhaps the most popular MCMC method, the Metropolis-Hastings algorithm. The

Metropolis-Hastings MCMC method is presented in the Algorithm 2.1.

Algorithm 2.1 Metropolis-Hastings MCMC

1. Set $\ell = 0$, where ℓ is the iteration index of the Markov chain. Select the initial state of the chain, $m_\ell = m_0$.

2. Generate a sample $\widehat{m}_{\ell+1}$ from the pdf $q(m_\ell, \widehat{m}_{\ell+1})$. $\widehat{m}_{\ell+1}$ represents the new proposed state in the chain and $q(m_\ell, \widehat{m}_{\ell+1})$ represents the "probability" of proposing a transition from state m_ℓ to $\widehat{m}_{\ell+1}$.

3. Compute the probability of accepting the proposed state as

$$\alpha(m_\ell, \widehat{m}_{\ell+1}) = \min\left\{1, \frac{q(\widehat{m}_{\ell+1}, m_\ell)\pi(\widehat{m}_{\ell+1})}{q(m_\ell, \widehat{m}_{\ell+1})\pi(m_\ell)}\right\}. \tag{2.15}$$

4. Sample u from a uniform distribution on $[0, 1]$. If $u \leq \alpha(m_\ell, \widehat{m}_{\ell+1})$, accept the new state, i.e., set $m_{\ell+1} = \widehat{m}_{\ell+1}$. Otherwise, repeat the old state in the chain, i.e., set $m_{\ell+1} = m_\ell$.

5. Set $\ell = \ell + 1$ and return to step 2.

As long as the proposal mechanism is such that it is possible to reach any state m that has $\pi(m) \neq 0$ in a finite number of transitions, the states in the chain constructed from the Metropolis-Hastings algorithm will asymptotically represent samples from $\pi(m)$ independent of the initial state, m_0, as long as $\pi(m_0) \neq 0$. Note that the computation of the acceptance probability requires that we know the target pdf only up to the normalization constant. This is a distinct advantage when $\pi(m)$ is a conditional pdf generated from Bayes' theorem with an unknown normalizing constant. The main difficultly encountered when using MCMC is the slow convergence typically observed for high-dimensional problems. However, some choices of the proposal distribution, $q(m, \widehat{m})$, may result in higher rates of acceptance of new states in the chain and reduce the number of iterations (proposals) to converge to the target distribution; see, e.g., Oliver et al. [133].

If the proposal mechanism is based on a prior Gaussian pdf, the transition probability in the Markov chain, $q(m, \widehat{m})$, can be written as

$$q(m, \widehat{m}) = b \exp\left(-\frac{1}{2}(\widehat{m} - m_{\mathrm{pr}})^{\mathrm{T}} C_{\mathrm{M}}^{-1}(\widehat{m} - m_{\mathrm{pr}})\right) = b \exp\left(-\frac{1}{2} z_m^{\mathrm{T}} z_m\right), \quad (2.16)$$

where b is a normalizing constant, which we do not need to compute, and $z \sim \mathcal{N}(0, I_{N_m})$. Note that the probability of proposing a transition from the current state, m, to $\widehat{m}$ is independent of m. In this case, we can write the proposal as

$$\widehat{m} = m_{\mathrm{pr}} + C_{\mathrm{M}}^{1/2} z_m. \quad (2.17)$$

From Eq. 2.17, the proposal of a new state in the chain can be done by sampling the random vector z_m. We refer to this procedure as global perturbation. However, for high-dimensional problems, this procedure typically leads to very low acceptance rates in the chain. In this case, most of the proposals occur in low-probability regions of the target pdf, causing the Markov chain to repeat states for several iterations. To avoid the large number of rejected proposals, Oliver et al. [133] applied a local perturbation procedure where, at each iteration, they randomly select one component of the random vector z_m corresponding to the current realization of the chain, and then replace this component by a sample from $\mathcal{N}(0, 1)$. Gelman et al. [64] presents an alternative local perturbation procedure where instead of changing one single component of the random vector z_m, we write the transition probability, $q(m, \widehat{m})$, as a Gaussian centered at the current state m with a scaled covariance $\sigma^2 C_{\mathrm{M}}$, where $\sigma < 1$ denotes the scaling factor. For this case, the proposed state can be written as

$$\widehat{m} = m + C_{\mathrm{M}}^{1/2} \delta z. \quad (2.18)$$

The random vector δz is a sample from $\mathcal{N}(0, \sigma^2 I_{N_m})$. This procedure leads to a symmetric transition probability, i.e., $q(m, \widehat{m}) = q(\widehat{m}, m)$, and thus, the probability

of accepting the proposed state simplifies to

$$\alpha(m, \widehat{m}) = \min\left\{1, \frac{\pi(\widehat{m})}{\pi(m)}\right\}. \tag{2.19}$$

The scaling factor, σ, controls the performance of the chain; small values of σ lead to high acceptance rates, but the chain will move very slowly through the target pdf and thus gives poor mixing. High values for σ, on the other hand, will lead to low acceptance rates. Although the optimal choice of σ depends on the target pdf, several authors (e.g., Gelman et al. [64], Roberts et al. [145], Roberts and Rosenthal [146]) have provided theoretical and empirical results which indicate that in high dimensions, it is optimal to choose σ such that the acceptance rate of the chain is approximately 0.234.

2.3 The Randomized Maximum Likelihood Method

The RML method [96, 134, 142] provides a correct sampling of the posterior pdf for the linear-Gaussian case [134]. For the nonlinear case, computational evidence indicates that RML provides an approximate sampling [134, 109, 135]. In fact, RML was designed to generate samples of different modes of the pdf so that it can approximate non-Gaussian distributions.

To obtain the jth sample, $m_{\mathrm{c},j}$, of the posterior pdf using RML, we solve the minimization problem

$$m_{\mathrm{c},j} = \arg\min_{m} O_j(m), \tag{2.20}$$

where

$$O_j(m) = \frac{1}{2}\left(m - m_{\mathrm{uc},j}\right)^{\mathrm{T}} C_{\mathrm{M}}^{-1}\left(m - m_{\mathrm{uc},j}\right) + \frac{1}{2}\left(g(m) - d_{\mathrm{uc},j}\right)^{\mathrm{T}} C_{\mathrm{D}}^{-1}\left(g(m) - d_{\mathrm{uc},j}\right). \tag{2.21}$$

In the above equation, $m_{\mathrm{uc},j}$ is a sample from $\mathcal{N}(m_{\mathrm{pr}}, C_{\mathrm{M}})$ and $d_{\mathrm{uc},j}$ is a sample from $\mathcal{N}(d_{\mathrm{obs}}, C_{\mathrm{D}})$. A conditional sample, m_j, obtained with RML will be close to the prior sample, $m_{\mathrm{uc},j}$, unless the minimization of $O_j(m)$ requires large corrections in m_j. Conceptually, the RML method can be applied with any optimization method. However, efficient implementations often require gradient-based optimization algorithms [192, 60] with adjoint-based gradient calculation [21, 19, 188, 105].

2.4 Sequential Data Assimilation

Sequential data assimilation (SDA) is the process of updating dynamical systems by incorporating data sequentially in time. Because the dynamic data in petroleum reservoirs, e.g., production, pressure and time-lapse seismic data, are available sequentially in time, the reservoir history-matching problem fits in the SDA context. Moreover, SDA can also be adequately described in a Bayesian framework.

So far, the history-matching problem was presented in terms of a parameter estimation problem, i.e., find the posterior distribution of a vector of model parameters m, given a set of observations d_{obs}. However, in a SDA process, it is convenient to rewrite the problem as a parameter-state estimation problem [46], i.e., we define a state vector y as

$$y^n = \begin{bmatrix} m^n \\ p^n \end{bmatrix}, \tag{2.22}$$

where, as before, m is a N_m-dimensional vector of model parameters and p is a N_p-dimensional vector representing the dynamical state of the system. The superscript n was introduced to indicate the time dependence on the model and states, which we assumed to be discretized in time. As before, the conditional pdf can be written using the Bayes' theorem

$$f(y|d_{\mathrm{obs}}) \propto f(y) f(d_{\mathrm{obs}}|y). \tag{2.23}$$

Let us assume that at the time t_n, we have updated the distribution of y by incorporating all data available up to the time t_n simultaneously. In this case, we write the posterior pdf as

$$f(y^n|d^n_{\text{obs}},\ldots,d^1_{\text{obs}}) \propto f(y^n)f(d^n_{\text{obs}},\ldots,d^1_{\text{obs}}|y^n), \tag{2.24}$$

where $f(y^n)$ represents our prior knowledge about the distribution of the state y at time-step t_n. Now if we want to incorporate a new observation at time t_{n+1}, we could start afresh and write

$$f(y^{n+1}|d^{n+1}_{\text{obs}},d^n_{\text{obs}},\ldots,d^1_{\text{obs}}) \propto f(y^{n+1})f(d^{n+1}_{\text{obs}},d^n_{\text{obs}},\ldots,d^1_{\text{obs}}|y^{n+1}), \tag{2.25}$$

or we could claim that just before the time t_{n+1}, the knowledge about y is summarized in the distribution $f(y^n|d^n_{\text{obs}},\ldots,d^1_{\text{obs}})$. Hence, we can just use $f(y^{n+1}|d^n_{\text{obs}},\ldots,d^1_{\text{obs}})$ as the distribution prior the assimilation of d^{n+1}_{obs} and update the pdf of y as

$$f(y^{n+1}|d^{n+1}_{\text{obs}},d^n_{\text{obs}},\ldots,d^1_{\text{obs}}) \propto f(y^{n+1}|d^n_{\text{obs}},\ldots,d^1_{\text{obs}})f(d^{n+1}_{\text{obs}}|d^n_{\text{obs}},\ldots,d^1_{\text{obs}},y^{n+1}). \tag{2.26}$$

It turns out that these two procedures, Eqs. 2.25 and 2.26, are equivalent under the assumption that the measurement errors of the individual observation vectors $\{d^{n+1}_{\text{obs}},\ldots,d^1_{\text{obs}}\}$ are uncorrelated in time and that the model evolution is a first order Markov process, i.e., y^{n+1} depends only on y^n and not on any of the preceding states $y^{n-1},\ldots,y^0$ [46, Chap. 7]. This equivalence follows from

$$
\begin{aligned}
f(y^{n+1}|d^{n+1}_{\text{obs}}, d^{n}_{\text{obs}}, \ldots, d^{1}_{\text{obs}}) &\propto f(y^{n+1}|d^{n}_{\text{obs}}, \ldots, d^{1}_{\text{obs}}) f(d^{n+1}_{\text{obs}}|d^{n}_{\text{obs}}, \ldots, d^{1}_{\text{obs}}, y^{n+1}) \\
&\propto f(y^{n+1}) f(d^{n}_{\text{obs}}, \ldots, d^{1}_{\text{obs}}|y^{n+1}) f(d^{n+1}_{\text{obs}}|d^{n}_{\text{obs}}, \ldots, d^{1}_{\text{obs}}, y^{n+1}) \\
&\propto f(y^{n+1}) f(d^{n+1}_{\text{obs}}, d^{n}_{\text{obs}}, \ldots, d^{1}_{\text{obs}}|y^{n+1}), \qquad (2.27)
\end{aligned}
$$

where the last part was obtained by the assumption of uncorrelated measurement errors in time, i.e.,

$$
f(d^{n+1}_{\text{obs}}|d^{n}_{\text{obs}}, \ldots, d^{1}_{\text{obs}}, y^{n+1}) = f(d^{n+1}_{\text{obs}}|y^{n+1}) \qquad (2.28)
$$

and

$$
f(d^{n}_{\text{obs}} \ldots, d^{1}_{\text{obs}}|y^{n+1}) f(d^{n+1}_{\text{obs}}|y^{n+1}) = f(d^{n+1}_{\text{obs}} \ldots, d^{1}_{\text{obs}}|y^{n+1}). \qquad (2.29)
$$

In summary, under assumptions of a first-order Markovian process with uncorrelated measurement errors, assimilating data sequentially in time is equivalent to assimilating all data simultaneously.

2.4.1 Sequential data assimilation for linear problems with Gaussian prior – The Kalman filter

Under the restrictions of a Gaussian prior, a linear relation between state and predicted data, Gaussian noise in the measurements and Gaussian model error, the Kalman filter (KF) [92] is the optimal sequential data-assimilation scheme. In the KF, the mean, μ^n_y, and the covariance, C^n_{Y}, of the state vector, y^n, are updated sequentially in time using

$$
\mu^{n,a}_y = \mu^{n,f}_y + K_n \left(d^n_{\text{obs}} - H_n \mu^{n,f}_y \right) \qquad (2.30)
$$

and

$$C_{\mathrm{Y}}^{n,a} = \left(I_{N_y} - K_n H_n\right) C_{\mathrm{Y}}^{n,f}, \tag{2.31}$$

where

$$K_n \equiv C_{\mathrm{Y}}^{n,f} H_n^{\mathrm{T}} \left(H_n C_{\mathrm{Y}}^{n,f} H_n^{\mathrm{T}} + C_{\mathrm{D}}^{n}\right)^{-1}. \tag{2.32}$$

Eqs. 2.30–2.32 are known as the KF analysis equations. In these equations, K_n is known as the Kalman gain matrix. H_n is the $N_n \times N_y$ sensitivity matrix, which defines the linear relation between the state vector and predicted data, i.e.,

$$d^{n,f} = H_n y^{n,f}. \tag{2.33}$$

C_{D}^{n} is the covariance matrix of the measurement errors at the time t_n. I_{N_y} is the $N_y \times N_y$ identity matrix, where N_y denotes the dimension of the state vector y^n. The superscripts a and f denote analysis and forecast, respectively.

Kalman [92] derived the KF as a minimum variance estimator. However, the KF equations can also be obtained from a Bayesian formulation [46]. In fact, if we write the KF analysis equations only for the model parameters part, m, and use the Kalman gain definition (2.32) in Eqs. 2.30 and 2.31, we obtain

$$\mu_m^{n,a} = \mu_m^{n,f} + C_{\mathrm{M}}^{n,f} G_n^{\mathrm{T}} \left(G_n C_{\mathrm{M}}^{n,f} G_n^{\mathrm{T}} + C_{\mathrm{D}}^{n}\right)^{-1} \left(d_{\mathrm{obs}}^{n} - G_n \mu_m^{n,f}\right) \tag{2.34}$$

and

$$C_{\mathrm{M}}^{n,a} = C_{\mathrm{M}}^{n,f} - C_{\mathrm{M}}^{n,f} G_n^{\mathrm{T}} \left(G_n C_{\mathrm{M}}^{n,f} G_n^{\mathrm{T}} + C_{\mathrm{D}}^{n}\right)^{-1} G_n C_{\mathrm{M}}^{n,f}. \tag{2.35}$$

In Eqs. 2.34 and 2.35, we use G_n to denote the linear relationship between the vector of model parameters m and the predicted data. Comparing Eqs. 2.34 and 2.35 with Eqs. 2.10 and 2.14 and noting the equivalence between simultaneous and sequential data assimilation previously established, we conclude that the KF updated mean, $\mu_m^{n,a}$, corresponds to the MAP estimate, i.e., $\mu_m^{n,a} = m_{\mathrm{MAP}}$, and the analyzed model

covariance, $C_{\mathrm{M}}^{n,a}$, corresponds to C_{MAP}.

2.5 The Ensemble Kalman Filter

In the KF, both the mean and covariance are updated whenever new data are available. However, even for linear problems, updating the states covariance matrix using Eq. 2.31 may not be computationally feasible if the dimension of the state is large. Moreover, if the problem is nonlinear, the posterior pdf is not Gaussian, and the KF equations are no longer applicable. For nonlinear problems, the extended Kalman filter (EKF) is an alternative to the KF. The EKF uses linearizations of the model and observation equations around the estimated mean of the state. However, for highly nonlinear problems, these linearizations may lead to unbounded instabilities in the covariance updates [46, Chap. 4]. Moreover, the EKF still requires updating the whole states covariance matrix, which becomes computationally infeasible for high-dimensional problems.

Evensen [43] introduced the ensemble Kalman filter (EnKF) as an alternative to overcome the problems of the EKF in high-dimensional nonlinear dynamical systems. The EnKF is a Monte Carlo method in which an ensemble of states is employed to represent the mean and covariance, which are updated sequentially in time. Although the EnKF was originally introduced in [43], the method was later clarified in [18, 80], where the concept of updating each ensemble member with independently perturbed observations was introduced, resulting in the current implementation of the EnKF. Typically in the EnKF, the number of ensemble members is much smaller than the number of unknowns so that the problem of updating a large state covariance matrix is managed by using the low-rank ensemble approximation. Hence, in the EnKF, the uncertainty is represented and propagated using an ensemble of states. This is different from KF and EKF, where uncertainty is propagated updating the states covariance matrix. The initial ensemble in the EnKF represents a sampling from the prior distribution, which, to a certain degree, tends to reduce the depen-

dence on Gaussian assumptions. However, the analysis step is still based on the first and second order moments of the distributions, so the performance of EnKF may deteriorate if the distributions are too far from Gaussian [1]. Unfortunately, the ensemble representation is also the main approximation introduced by EnKF. In order to achieve computational efficiency, it is necessary to limit the size of the ensemble. However, small ensembles introduce sampling errors, which lead to spurious correlations. Moreover, the size of the ensemble limits the space in which the solutions can be represented. As a result, we observe excessive reduction in posterior covariances after the assimilation of data [80, 81, 46, 1]. This problem will be further discussed in Chapters 3 and 4 of this dissertation.

The EnKF equations are typically introduced by defining an augmented state vector, y^n, in which the predicted data vector, d^n, is also included, i.e.,

$$y_j^n = \begin{bmatrix} m_j^n \\ p_j^n \\ d_j^n \end{bmatrix}, \tag{2.36}$$

where the subscript j denotes the jth ensemble member. As before, m_j^n and p_j^n are, respectively, the vectors of model parameters and states of the dynamical system. For reservoir applications, m includes all model parameters required in the history matching, typically gridblock porosities, permeabilities, end points of relative permeability curves, etc.; p includes the primary variables of the reservoir simulator, typically gridblock pressure, fluid saturations and bubble-point pressure in a standard black-oil reservoir simulator. For simplicity in notation, we assume that data assimilation steps correspond to reservoir simulation time-steps. Then, we define the matrix H_n as

$$H_n \equiv \begin{bmatrix} O & I_{N_n} \end{bmatrix}, \tag{2.37}$$

where O is the $N_n \times (N_m + N_p)$ null matrix and I_{N_n} is the $N_n \times N_n$ identity matrix. Note that the dimension of H_n depends on the number of data to be assimilated at the nth assimilation time-step, denoted by N_n. Using the definition (2.37), we can write the predicted data vector, $d_j^{n,f}$, as we did for KF, i.e.,

$$d_j^{n,f} = H_n y_j^{n,f}. \tag{2.38}$$

This "trick" of augmenting the state vector with the predicted data allows us to derive the EnKF equations in a similar fashion to the standard KF equations. The trick turns a nonlinear relation between data and the original state vector into a linear relationship between the vector of predicted data and the augmented state vector, making it possible to write down the formula for the analysis step analytically [103]. However, this trick does not remove the effect of the nonlinearity. As shown in Li and Reynolds [103], augmenting the state vector with data results in a correct procedure for sampling the pdf if, and only if, at every data assimilation time-step, the predicted data vector is a linear function of the combined (unaugmented) state vector.

Using the augmented state vector, the EnKF analysis equations can be written in a similar form as the KF equations. Unlike the KF, where we update directly the mean and covariance matrix, in the EnKF, we update each ensemble member using

$$y_j^{n,a} = y_j^{n,f} + \widetilde{K}_n \left(d_{\mathrm{uc},j}^n - d_j^{n,f} \right), \text{ for } j = 1, 2, \cdots N_e, \tag{2.39}$$

and

$$\widetilde{K}_n = \widetilde{C}_{\mathrm{Y}}^{n,f} H_n^{\mathrm{T}} \left(H_n \widetilde{C}_{\mathrm{Y}}^{n,f} H_n^{\mathrm{T}} + C_{\mathrm{D}}^n \right)^{-1}. \tag{2.40}$$

In the above equations, N_e is the number of state vectors in the ensemble, i.e., the ensemble size; $d_{\mathrm{uc},j}^n$ is a sample from the Gaussian distribution $\mathcal{N}(d_{\mathrm{obs}}^n, C_{\mathrm{D}}^n)$. We can

interpret $d^n_{\mathrm{uc},j}$ as a perturbed (or randomized) observed data vector as we do in the RML method. We introduced the tilde in the matrices $\widetilde{K}_n$ and $\widetilde{C}^{n,f}_{\mathrm{Y}}$ to emphasize that these matrices are estimated from an ensemble. Throughout this dissertation, we use the tilde whenever a matrix is estimated from an ensemble. The ensemble approximation of the states covariance matrix is given by

$$\widetilde{C}^{n,f}_{\mathrm{Y}} = \frac{1}{N_e - 1}\sum_{j=1}^{N_e}(y^{n,f}_j - \overline{y}^{n,f})(y^{n,f}_j - \overline{y}^{n,f})^{\mathrm{T}} = \frac{\Delta Y^{n,f}\left(\Delta Y^{n,f}\right)^{\mathrm{T}}}{N_e - 1}, \tag{2.41}$$

where the columns of $\Delta Y^{n,f}$ are the ensemble members minus the mean, i.e.,

$$\Delta Y^{n,f} = Y^{n,f} - \overline{Y}^{n,f} = \left[y^{n,f}_1 - \overline{y}^{n,f}, \ldots, y^{n,f}_{N_e} - \overline{y}^{n,f}\right], \tag{2.42}$$

where

$$\overline{y}^{n,f} = \frac{1}{N_e}\sum_{j=1}^{N_e} y^{n,f}_j. \tag{2.43}$$

One advantage of the EnKF is that the states covariance matrix $C^{n,f}_{\mathrm{Y}}$ does not need to be explicitly formed. Instead, we partition $\widetilde{C}^{n,f}_{\mathrm{Y}}$ as

$$\widetilde{C}^{n,f}_{\mathrm{Y}} = \begin{bmatrix} \widetilde{C}^{n,f}_{\mathrm{MM}} & \widetilde{C}^{n,f}_{\mathrm{MP}} & \widetilde{C}^{n,f}_{\mathrm{MD}} \\ \widetilde{C}^{n,f}_{\mathrm{PM}} & \widetilde{C}^{n,f}_{\mathrm{PP}} & \widetilde{C}^{n,f}_{\mathrm{PD}} \\ \widetilde{C}^{n,f}_{\mathrm{DM}} & \widetilde{C}^{n,f}_{\mathrm{DP}} & \widetilde{C}^{n,f}_{\mathrm{DD}} \end{bmatrix}, \tag{2.44}$$

and note that in Eq. 2.40, we only need the products

$$\widetilde{C}^{n,f}_{\mathrm{Y}} H^{\mathrm{T}}_n = \begin{bmatrix} \widetilde{C}^{n,f}_{\mathrm{MM}} & \widetilde{C}^{n,f}_{\mathrm{MP}} & \widetilde{C}^{n,f}_{\mathrm{MD}} \\ \widetilde{C}^{n,f}_{\mathrm{PM}} & \widetilde{C}^{n,f}_{\mathrm{PP}} & \widetilde{C}^{n,f}_{\mathrm{PD}} \\ \widetilde{C}^{n,f}_{\mathrm{DM}} & \widetilde{C}^{n,f}_{\mathrm{DP}} & \widetilde{C}^{n,f}_{\mathrm{DD}} \end{bmatrix} \begin{bmatrix} O \\ I_{N_n} \end{bmatrix} = \begin{bmatrix} \widetilde{C}^{n,f}_{\mathrm{MD}} \\ \widetilde{C}^{n,f}_{\mathrm{PD}} \\ \widetilde{C}^{n,f}_{\mathrm{DD}} \end{bmatrix} = \widetilde{C}^{n,f}_{\mathrm{YD}} \tag{2.45}$$

and

$$H_n \widetilde{C}_{\mathrm{Y}}^{n,f} H_n^{\mathrm{T}} = \widetilde{C}_{\mathrm{DD}}^{n,f}. \tag{2.46}$$

Thus, the EnKF analysis equation directly involves only the two cross-covariance matrices, $\widetilde{C}_{\mathrm{YD}}^{n,f}$ and $\widetilde{C}_{\mathrm{DD}}^{n,f}$, which are represented by the following approximations:

$$\widetilde{C}_{\mathrm{YD}}^{n,f} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (y_j^{n,f} - \overline{y}^{n,f})(d_j^{n,f} - \overline{d}^{n,f})^{\mathrm{T}} \tag{2.47}$$

and

$$\widetilde{C}_{\mathrm{DD}}^{n,f} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (d_j^{n,f} - \overline{d}^{n,f})(d_j^{n,f} - \overline{d}^{n,f})^{\mathrm{T}}. \tag{2.48}$$

Using Eqs. 2.45 and 2.46, we can rewrite the EnKF analysis equation (2.39) as

$$y_j^{n,a} = y_j^{n,f} + \widetilde{C}_{\mathrm{YD}}^{n,f} \left(\widetilde{C}_{\mathrm{DD}}^{n,f} + C_{\mathrm{D}}^{n} \right)^{-1} \left(d_{\mathrm{uc},j}^{n} - d_j^{n,f} \right), \text{ for } j = 1, 2, \cdots N_e. \tag{2.49}$$

In terms of practical implementation, even the cross-covariance matrix $\widetilde{C}_{\mathrm{YD}}^{n,f}$ does not need to be completely formed or stored. In fact, after computing the $N_n \times N_n$ matrix $\left(\widetilde{C}_{\mathrm{DD}}^{n,f} + C_{\mathrm{D}}^{n} \right)$, we can compute the vector x_j by solving the linear problem

$$x_j \equiv \left(\widetilde{C}_{\mathrm{DD}}^{n,f} + C_{\mathrm{D}}^{n} \right)^{-1} \left(d_{\mathrm{uc},j}^{n} - d_j^{n,f} \right). \tag{2.50}$$

Then, the analysis equation (2.49) can be written as

$$
\begin{aligned}
y_j^{n,a} &= y_j^{n,f} + \widetilde{C}_{\mathrm{YD}}^{n,f} x_j \\
&= y_j^{n,f} + \frac{1}{N_e - 1} \sum_{k=1}^{N_e} (y_k^{n,f} - \overline{y}^{n,f})(d_k^{n,f} - \overline{d}^{n,f})^{\mathrm{T}} x_j \\
&= y_j^{n,f} + \frac{1}{N_e - 1} \sum_{k=1}^{N_e} \alpha_k (y_k^{n,f} - \overline{y}^{n,f}),
\end{aligned} \tag{2.51}
$$

where the α_k's are real numbers given by

$$
\alpha_k = (d_k^{n,f} - \overline{d}^{n,f})^{\mathrm{T}} x_j; \tag{2.52}
$$

thus the calculation of $y_j^{n,a}$ requires only vector sums and inner products.

2.5.1 Inversion and rescaling

As noted before, the EnKF analysis equation (Eq. 2.49) requires the inversion of the $N_n \times N_n$ matrix C given by

$$
C = \widetilde{C}_{\mathrm{DD}}^{f} + C_{\mathrm{D}}. \tag{2.53}
$$

In order to simplify the notation, we dropped the superscript n in the equations of this and the next two sections (Sections 2.6 and 2.7.4), noting that all equations refer to the same data assimilation time-step. Because $\widetilde{C}_{\mathrm{DD}}^{f}$ is a real-symmetric positive semi-definite matrix, C given by Eq. 2.53 is real-symmetric positive-definite as long as C_{D} is positive-definite, which means that C is invertible. However, C may be poorly conditioned [46, Chap. 14]. Hence, EnKF implementations typically use a pseudo-inverse of C computed using a truncated singular value decomposition (TSVD), i.e,

$$
C^{+} = U_r \Lambda_r^{-1} U_r^{\mathrm{T}}, \tag{2.54}
$$

where C^+ denotes the pseudo-inverse of C. U_r is the $N_n \times N_r$ matrix with its jth column equal to the left singular vector of C corresponding to the jth singular value. Λ_r is a diagonal matrix containing the N_r largest nonzero singular values of C. N_r is typically defined by sorting the singular values, λ_i, in a decreasing order and finding the maximum N_r such that

$$\frac{\sum_{i=1}^{N_r} \lambda_i}{\sum_{i=1}^{N_n} \lambda_i} \leq \xi, \tag{2.55}$$

where ξ is a number typically between 0.9 and 1.0. However, C may also be poorly scaled as it may be constructed based on data with different magnitudes, e.g., pressure and water-cut data. In this case, one may lose the information necessary to match data when truncating small singular values; see, e.g., Wang et al. [182] where an example is shown in which water-cut data could not be matched because of truncation. For this reason, it is important to rescale the components of the matrix C before applying the TSVD. This rescaling can be done by using the Cholesky decomposition of C_{D}, i.e.,

$$C_{\mathrm{D}} = C_{\mathrm{D}}^{1/2} C_{\mathrm{D}}^{\mathrm{T}/2} \tag{2.56}$$

and writing Eq. 2.53 as

$$C = C_{\mathrm{D}}^{1/2} \widehat{C} C_{\mathrm{D}}^{\mathrm{T}/2}, \tag{2.57}$$

where

$$\widehat{C} = C_{\mathrm{D}}^{-1/2} \widetilde{C}_{\mathrm{DD}}^{f} C_{\mathrm{D}}^{-\mathrm{T}/2} + I_{N_n}. \tag{2.58}$$

The TSVD is now applied to the matrix $\widehat{C}$ to obtain

$$\widehat{C} = \widehat{U}_r \widehat{\Lambda}_r \widehat{U}_r^{\mathrm{T}}, \tag{2.59}$$

and the pseudo-inverse of C becomes

$$C^{+} = C_{\mathrm{D}}^{-\mathrm{T}/2}\widehat{U}_r\widehat{\Lambda}_r^{-1}\widehat{U}_r^{\mathrm{T}}C_{\mathrm{D}}^{-1/2}. \tag{2.60}$$

This rescaling procedure can be justified by defining the dimensionless sensitivity matrix, G_{D}, [193] as

$$G_{\mathrm{D}} = C_{\mathrm{D}}^{-1/2}GC_{\mathrm{Y}}^{1/2}, \tag{2.61}$$

where G is the $N_n \times N_y$ sensitivity matrix. For the linear case, we can write the vector of predicted data d^f in terms of the state vector y^f as

$$d^f = Gy^f, \tag{2.62}$$

so that the cross-covariance matrix C_{DD}^f can be written as

$$C_{\mathrm{DD}}^f = GC_{\mathrm{Y}}^f G^{\mathrm{T}}. \tag{2.63}$$

Using Eqs. 2.61 and 2.63 in Eq. 2.58, we can write $\widehat{C}$ as

$$\widehat{C} = G_{\mathrm{D}}G_{\mathrm{D}}^{\mathrm{T}} + I_{N_n}. \tag{2.64}$$

Let ω_i be the ith singular value of G_{D}. It is straightforward to show that the ith singular value of $\widehat{C}$, denoted by $\widehat{\lambda}_i$, is given by

$$\widehat{\lambda}_i = \omega_i^2 + 1. \tag{2.65}$$

The results of Tavakoli and Reynolds [166] imply that the singular values of G_{D} govern the reduction in uncertainty in the state vector due to the assimilation of data with the Kalman filter. In addition, Tavakoli and Reynolds [166] showed that small singular values of G_{D} have a negligible influence on the reduction of uncertainty.

From Eq. 2.65, we note that the singular values of $\widehat{C}$ are defined by the singular values of G_{D} so that truncating small values of $\widehat{\lambda}_i$ corresponds to eliminating small singular values of G_{D}, which have the smallest influence on the reduction of uncertainty. In this sense, the rescaling procedure presented in this section is optimal.

2.5.2 Subspace inversion

The inversion procedure presented in the previous section requires the TSVD of a $N_n \times N_n$ matrix. However, when the number of data points is large, as in the case when assimilating seismic data, the computational cost of this SVD procedure is too high. Evensen [45] introduced a subspace inversion procedure, which is computationally more efficient in the case where $N_e \ll N_n$. Here, we briefly review the subspace inversion proposed by Evensen [45] and rewrite this procedure for the case in which rescaling is applied.

In the subspace inversion, we define $C^\star = (N_e - 1)C$, where C is given by Eq. 2.53 and write $C^\star$ in the following form:

$$C^\star = \Delta D^f (\Delta D^f)^{\mathrm{T}} + (N_e - 1) C_{\mathrm{D}}. \tag{2.66}$$

Here, $\Delta D^f = D^f - \overline{D}^f$, where D^f denotes the $N_n \times N_e$ matrix of predicted data, i.e., the columns of D^f correspond to the vectors of predicted data obtained by the ensemble members. $\overline{D}^f$ is the $N_n \times N_e$ matrix with all its columns equal to the mean ensemble prediction. Instead of computing the pseudo-inverse of $C^\star$ directly by TSVD, we apply SVD to ΔD^f and truncate with the N_r largest singular values, i.e.,

$$\Delta D^f \approx U_r W_r V_r^{\mathrm{T}}, \tag{2.67}$$

where U_r is the $N_n \times N_r$ matrix containing as its columns the left singular vectors of ΔD^f corresponding to its first N_r largest singular values; V_r is the $N_e \times N_r$

corresponding matrix of right singular vectors, and W_r is a diagonal matrix with the N_r largest singular values of ΔD^f as its diagonal entries. Because we can choose any $N_r \leq \min\{N_n, N_e - 1\}$, we write Eq. 2.67 as an approximation; equality holds when we keep all nonzero singular values. Using Eq. 2.67 in Eq. 2.66, we obtain

$$C^\star \approx U_r W_r \left[I_{N_r} + (N_e - 1) W_r^{-1} U_r^{\mathrm{T}} C_{\mathrm{D}} U_r W_r^{-1} \right] W_r U_r^{\mathrm{T}}. \tag{2.68}$$

In Eq. 2.68, we introduced the additional approximation that $U_r U_r^{\mathrm{T}} \approx I_{N_n}$. Defining the $N_r \times N_r$ symmetric matrix X as

$$X \equiv (N_e - 1) W_r^{-1} U_r^{\mathrm{T}} C_{\mathrm{D}} U_r W_r^{-1}, \tag{2.69}$$

we can write Eq. 2.68 as

$$C^\star \approx U_r W_r \left[I_{N_r} + X \right] W_r U_r^{\mathrm{T}}. \tag{2.70}$$

Because X is real-symmetric positive semidefinite, the SVD of X is equivalent to a Schur decomposition,

$$X = Z_r \Gamma_r Z_r^{\mathrm{T}}, \tag{2.71}$$

where Γ_r is a $N_r \times N_r$ diagonal matrix with the ith diagonal entry equal to the ith eigenvalue of X, which is real and non-negative, and Z_r is a $N_r \times N_r$ orthogonal matrix with its ith column equal to the ith eigenvector of X. Using Eq. 2.71 in Eq. 2.70 and the fact that Z_r is an orthogonal matrix, we obtain

$$C^\star \approx \left(U_r W_r Z_r \right) \left[I_{N_r} + \Gamma_r \right] \left(U_r W_r Z_r \right)^{\mathrm{T}}. \tag{2.72}$$

Because W_r and Γ_r are diagonal matrices, it is trivial to compute their inverses and the pseudo-inverse of $C^\star$ becomes

$$(C^\star)^+ = \left(U_r W_r^{-1} Z_r\right) \left[I_{N_r} + \Gamma_r\right]^{-1} \left(U_r W_r^{-1} Z_r\right)^{\mathrm{T}}, \tag{2.73}$$

which is the pseudo-inverse using the subspace inversion scheme of Evensen [45].

As before, we may still have scaling problems when truncating the small singular values of ΔD^f. Hence, we present a subspace inversion scheme for the case in which we rescale ΔD^f before applying TSVD. This rescaling can be done by rewriting Eq. 2.66 as

$$C^\star = C_{\mathrm{D}}^{1/2} \left[C_{\mathrm{D}}^{-1/2} \Delta D^f (\Delta D^f)^{\mathrm{T}} C_{\mathrm{D}}^{-\mathrm{T}/2} + (N_e - 1) I_{N_n}\right] C_{\mathrm{D}}^{\mathrm{T}/2}. \tag{2.74}$$

Instead of using the TSVD of ΔD^f, we calculate the following TSVD:

$$C_{\mathrm{D}}^{-1/2} \Delta D^f \approx \widehat{U}_r \widehat{W}_r \widehat{V}_r^{\mathrm{T}}. \tag{2.75}$$

Using the same procedure as before, we obtain

$$C^\star \approx C_{\mathrm{D}}^{1/2} \widehat{U}_r \widehat{W}_r \left[I_{N_r} + (N_e - 1) \widehat{W}_r^{-1} \widehat{W}_r^{-1}\right] \widehat{W}_r \widehat{U}_r^{\mathrm{T}} C_{\mathrm{D}}^{\mathrm{T}/2}. \tag{2.76}$$

Finally, the pseudo-inverse of $C^\star$ can be written as

$$(C^\star)^+ = \left(C_{\mathrm{D}}^{-\mathrm{T}/2} \widehat{U}_r \widehat{W}_r^{-1}\right) \left[I_{N_r} + \widehat{\Gamma}_r\right]^{-1} \left(C_{\mathrm{D}}^{-\mathrm{T}/2} \widehat{U}_r \widehat{W}_r^{-1}\right)^{\mathrm{T}}, \tag{2.77}$$

where $\widehat{\Gamma}_r$ is the $N_r \times N_r$ diagonal matrix given by

$$\widehat{\Gamma}_r = (N_e - 1) \widehat{W}_r^{-1} \widehat{W}_r^{-1}. \tag{2.78}$$

This rescaling procedure requires that we factor the covariance matrix of the measurement errors as $C_{\mathrm{D}} = C_{\mathrm{D}}^{1/2} C_{\mathrm{D}}^{\mathrm{T}/2}$. However, this factorization can be computationally expensive if the number of measurement errors is large, which is the case when we have seismic data. An alternative procedure is to perform rescaling

based only on the diagonal elements of C_{D}. In this case, we write

$$C_{\mathrm{D}} = S\widehat{C}_{\mathrm{D}}S, \tag{2.79}$$

where S is the $N_n \times N_n$ matrix containing the square root of the diagonal components of C_{D}, i.e., $S = \mathrm{diag}\left[C_{\mathrm{D}}^{1/2}\right]$; $\widehat{C}_{\mathrm{D}}$ is the correlation matrix of the measurement errors. Using Eq. 2.79, we rewrite Eq. 2.74 as

$$C^{\star} = S\left[S^{-1}\Delta D^{f}(\Delta D^{f})^{\mathrm{T}}S^{-1} + (N_e - 1)\widehat{C}_{\mathrm{D}}\right]S, \tag{2.80}$$

and calculate the following TSVD:

$$S^{-1}\Delta D^{f} \approx \widehat{U}_r\widehat{W}_r\widehat{V}_r^{\mathrm{T}}. \tag{2.81}$$

Similarly to Eq. 2.70, we can write Eq. 2.80 as

$$C^{\star} \approx S\widehat{U}_r\widehat{W}_r\left[I_{N_r} + \widehat{X}\right]\widehat{W}_r\widehat{U}_r^{\mathrm{T}}S, \tag{2.82}$$

where $\widehat{X}$ is the $N_r \times N_r$ symmetric matrix given by

$$\widehat{X} \equiv (N_e - 1)\widehat{W}_r^{-1}\widehat{U}_r^{\mathrm{T}}\widehat{C}_{\mathrm{D}}\widehat{U}_r\widehat{W}_r^{-1}. \tag{2.83}$$

Writing the Schur decomposition

$$\widehat{X} = \widehat{Z}_r\widehat{\Gamma}_r\widehat{Z}_r^{\mathrm{T}}, \tag{2.84}$$

and using Eq. 2.84 in Eq. 2.82, we obtain

$$C^{\star} \approx \left(S\widehat{U}_r\widehat{W}_r\widehat{Z}_r\right)\left[I_{N_r} + \widehat{\Gamma}_r\right]\left(S\widehat{U}_r\widehat{W}_r\widehat{Z}_r\right)^{\mathrm{T}}. \tag{2.85}$$

Thus, the pseudo-inverse of $C^{\star}$ becomes

$$(C^{\star})^{+} = \left(S^{-1}\widehat{U}_r\widehat{W}_r^{-1}\widehat{Z}_r\right)\left[I_{N_r} + \widehat{\Gamma}_r\right]^{-1}\left(S^{-1}\widehat{U}_r\widehat{W}_r^{-1}\widehat{Z}_r\right)^{\mathrm{T}}. \tag{2.86}$$

2.5.3 Ensemble representation of C_D

Evensen [44] suggested that the use of a low-rank ensemble representation of the covariance matrix of the measurement errors, C_{D}, presents numerical advantages in terms of reduction of the computational cost of the analysis in the EnKF. The ensemble representation of C_{D} can be written as

$$\widetilde{C}_{\mathrm{D}} = \frac{1}{N_e - 1}\Delta D_{\mathrm{uc}}\Delta D_{\mathrm{uc}}^{\mathrm{T}}, \tag{2.87}$$

where $\Delta D_{\mathrm{uc}} = D_{\mathrm{uc}} - \overline{D}_{\mathrm{uc}}$, with D_{uc} denoting the $N_n \times N_e$ matrix containing the ensemble of perturbed observations, i.e., the jth column of D_{uc} is equal to $d_{\mathrm{uc},j}$ which corresponds to the vector of perturbed observations used to update jth ensemble member; and $\overline{D}_{\mathrm{uc}}$ denotes the $N_n \times N_e$ with all columns equal to the average of the ensemble perturbed observations, $\overline{d}_{\mathrm{uc}}$.

However, Kepert [93] showed that the use of $\widetilde{C}_{\mathrm{D}}$ leads to loss of rank in the ensemble if $N_n \geq N_e$ and ensemble collapse if $N_n \geq 2N_e - 2$. In fact, if $N_n \geq N_e$, $\widetilde{C}_{\mathrm{D}}$ has at most $N_e - 1$ nonzero singular values. In this case, it can be shown that the use of $\widetilde{C}_{\mathrm{D}}$ in the analysis is equivalent to assimilating $N_n - N_e + 1$ perfect observations, i.e., observation where the corresponding measurement errors are zero. Lorenc [110] showed that the assimilation of a perfect observation removes one degree of freedom of the ensemble. Therefore, for $N_n \geq N_e$, we lose $N_n - N_e + 1$ degrees of freedom and if $N_n = 2N_e - 2$, we lose all degrees of freedom so the ensemble collapses. Evensen [45] showed that if we project the matrix ΔD_{uc} onto the subspace spanned by the left singular vectors of ΔD^f, we do not observe the loss of rank. In this case, the EnKF analysis with subspace inversion corresponds to updating the ensemble using the vectors of data mismatches, $\delta d_j = d_{\mathrm{uc},j} - d_j^f$, also projected onto the subspace

spanned by the left singular vectors of ΔD^f. According to Skjervheim et al. [158], this procedure tends to reduce loss of ensemble rank when assimilating a large number of independent data, which can be useful, for example, when we assimilate seismic data. However, Aanonsen et al. [1] pointed out that there is no assurance that this procedure generates realizations of model parameters that honor all observed data, and, in this case, the estimate of model parameters may be considerably less accurate than the estimate that could be obtained by using a larger ensemble size.

2.6 Ensemble Square Root Filter

The EnKF analysis equation is based on the KF equations (Eqs. 2.30–2.32). In the KF, the forecast covariance matrix of the state vector is updated using

$$C_{\mathrm{Y}}^a = \left(I_{N_y} - KH\right) C_{\mathrm{Y}}^f. \tag{2.88}$$

According to Whitaker and Hamill [186], because EnKF uses an updating scheme based on perturbed observations, the KF covariance analysis equation (2.88) is satisfied in a statistical sense only. This may result in a suboptimal filter behavior, which is particularly evident for small ensembles [186]. In an attempt to avoid this suboptimal behavior, several implementations of ensemble square root filters (EnSRF) have been proposed in the literature [15, 4, 186, 45]. These methods allow deterministic updates of the ensemble so that the analyzed covariance exactly satisfies Eq. 2.88 [170, 151]. For EnSRF, we first update the ensemble mean using

$$\overline{y}^a = \overline{y}^f + \widetilde{K}(d_{\mathrm{obs}} - \overline{d}^f); \tag{2.89}$$

then, we update the ensemble changes, ΔY^a, using

$$\Delta Y^a = \Delta Y^f T_R. \tag{2.90}$$

In Eq. 2.90, T_R is the $N_e \times N_e$ ensemble transform matrix (ETM). This EnSRF scheme is referred to as a right-multiplied scheme [170, 151]. This scheme requires the factorization (square root) of a $N_e \times N_e$ matrix, which is computationally feasible for the typical ensemble sizes employed in reservoir history-matching applications. The ETM is defined such that Eq. 2.88 is exactly satisfied. The computation of the ETM depends on the square root implementation [170]. One difficulty with right-multiplied schemes is that it is not clear how to apply the Schur product for covariance localization, as discussed in Chapters 3 and 4 of this dissertation. However, there also exists left-multiplied schemes of EnSRF [151]. These schemes require an ETM, T_L, such that

$$\Delta Y^a = T_L \Delta Y^f, \tag{2.91}$$

where T_L is a $N_y \times N_y$ matrix. The following sections present two EnSRF formulations, the right-multiplied scheme proposed by Evensen [45] and the left-multiplied scheme proposed by Whitaker and Hamill [186].

2.6.1 Right-multiplied ensemble square root filter

In order to derive the square root scheme proposed by Evensen [45], we first use the definition of the Kalman gain (Eq. 2.32) in the covariance analysis equation (Eq. 2.88) and rewrite Eq. 2.88 in the following form:

$$C_{\mathrm{Y}}^a = C_{\mathrm{Y}}^f - C_{\mathrm{Y}}^a H^{\mathrm{T}} \left[H C_{\mathrm{Y}}^f H^{\mathrm{T}} + C_{\mathrm{D}} \right]^{-1} H C_{\mathrm{Y}}^f. \tag{2.92}$$

Then, we use the ensemble representations of the states covariance matrices, $\widetilde{C}_{\mathrm{Y}}^a = \frac{1}{N_e-1}\Delta Y^a (\Delta Y^a)^{\mathrm{T}}$, $\widetilde{C}_{\mathrm{Y}}^f = \frac{1}{N_e-1}\Delta Y^f (\Delta Y^f)^{\mathrm{T}}$, $\widetilde{C}_{\mathrm{Y}}^f H^{\mathrm{T}} = \frac{1}{N_e-1}\Delta Y^f (\Delta D^f)^{\mathrm{T}}$ and $H\widetilde{C}_{\mathrm{Y}}^f H^{\mathrm{T}} = \frac{1}{N_e-1}\Delta D^f (\Delta D^f)^{\mathrm{T}}$ in Eq. 2.92, which results in

$$\begin{aligned}\Delta Y^a(\Delta Y^a)^{\mathrm{T}} &= \Delta Y^f(\Delta Y^f)^{\mathrm{T}} - \Delta Y^f(\Delta D^f)^{\mathrm{T}} \\ &\times \left[\Delta D^f(\Delta D^f)^{\mathrm{T}} + (N_e-1)C_{\mathrm{D}}\right]^{-1}\Delta D^f(\Delta Y^f)^{\mathrm{T}}. \end{aligned} \tag{2.93}$$

As before, defining the $N_n \times N_n$ matrix

$$C^\star = \Delta D^f(\Delta D^f)^{\mathrm{T}} + (N_e-1)C_{\mathrm{D}}, \tag{2.94}$$

we write Eq. 2.93 as

$$\Delta Y^a(\Delta Y^a)^{\mathrm{T}} = \Delta Y^f\left[I_{N_e} - (\Delta D^f)^{\mathrm{T}}(C^\star)^{-1}\Delta D^f\right](\Delta Y^f)^{\mathrm{T}}. \tag{2.95}$$

Because $C^\star$ is real-symmetric positive-definite, we can write its Schur decomposition as

$$C^\star = X\Lambda X^{\mathrm{T}}, \tag{2.96}$$

and its inverse as

$$(C^\star)^{-1} = X\Lambda^{-1}X^{\mathrm{T}}, \tag{2.97}$$

where X is the $N_n \times N_n$ matrix containing the orthonormal eigenvectors of $C^\star$ and Λ is a $N_n \times N_n$ diagonal matrix with the eigenvalues of $C^\star$ as its diagonal entries. Using Eq. 2.97 in Eq. 2.95 and writing $\Lambda^{-1} = \Lambda^{-1/2}\Lambda^{-1/2}$, we obtain

$$\Delta Y^a(\Delta Y^a)^{\mathrm{T}} = \Delta Y^f\left[I_{N_e} - (\Delta D^f)^{\mathrm{T}}X\Lambda^{-1/2}\Lambda^{-1/2}X^{\mathrm{T}}\Delta D^f\right](\Delta Y^f)^{\mathrm{T}}, \tag{2.98}$$

or equivalently

$$\Delta Y^a(\Delta Y^a)^{\mathrm{T}} = \Delta Y^f \left[I_{N_e} - (\Lambda^{-1/2}X^{\mathrm{T}}\Delta D^f)^{\mathrm{T}}(\Lambda^{-1/2}X^{\mathrm{T}}\Delta D^f)\right](\Delta Y^f)^{\mathrm{T}}. \quad (2.99)$$

Now, we write the SVD of $\Lambda^{-1/2}X^{\mathrm{T}}\Delta D^f$ as

$$\Lambda^{-1/2}X^{\mathrm{T}}\Delta D^f = UWV^{\mathrm{T}}, \quad (2.100)$$

where U is the $N_n \times N_n$ orthogonal matrix of left singular vectors, W is the $N_n \times N_e$ matrix with non-zero values corresponding to the singular values and V is the $N_e \times N_e$ matrix of right singular vectors. Using Eq. 2.100 in Eq. 2.99,

$$\begin{aligned}
\Delta Y^a(\Delta Y^a)^{\mathrm{T}} &= \Delta Y^f \left[I_{N_e} - (UWV^{\mathrm{T}})^{\mathrm{T}}(UWV^{\mathrm{T}})\right](\Delta Y^f)^{\mathrm{T}} \\
&= \Delta Y^f \left[I_{N_e} - VW^{\mathrm{T}}U^{\mathrm{T}}UWV^{\mathrm{T}}\right](\Delta Y^f)^{\mathrm{T}} \\
&= \Delta Y^f V \left[I_{N_e} - W^{\mathrm{T}}W\right] V^{\mathrm{T}}(\Delta Y^f)^{\mathrm{T}} \\
&= \left[\Delta Y^f V \left(I_{N_e} - W^{\mathrm{T}}W\right)^{1/2}\right]\left[\Delta Y^f V \left(I_{N_e} - W^{\mathrm{T}}W\right)^{1/2}\right]^{\mathrm{T}} \quad (2.101)
\end{aligned}$$

Thus, we can define

$$\Delta Y^a \equiv \Delta Y^f V \left(I_{N_e} - W^{\mathrm{T}}W\right)^{1/2}, \quad (2.102)$$

and the ETM as

$$T_R \equiv V \left(I_{N_e} - W^{\mathrm{T}}W\right)^{1/2}. \quad (2.103)$$

Sakov and Oke [151] showed that the above square root scheme is biased, i.e., it does not preserve the mean. Leeuwenburgh et al. [101] concluded that this scheme also leads to outliers that contain most of the ensemble variance. Redefining the

ETM as

$$T_R = V\left(I_{N_e} - W^{\mathrm{T}}W\right)^{1/2} V^{\mathrm{T}}, \tag{2.104}$$

makes the analysis unbiased and resolves the issue with outliers.

2.6.2 Left-multiplied ensemble square root filter

The left-multiplied EnSRF requires an ETM, T_L, such that

$$\Delta Y^a = T_L \Delta Y^f, \tag{2.105}$$

where T_L is a $N_y \times N_y$. A left-multiplied ETM can be obtained by writing the posterior covariance matrix as

$$C_{\mathrm{Y}}^a = \left[(C_{\mathrm{Y}}^f)^{-1} + H^{\mathrm{T}} C_{\mathrm{D}}^{-1} H\right]^{-1}. \tag{2.106}$$

The equivalence between Eq. 2.92 and Eq. 2.106 can be shown by using the matrix inversion lemmas as described in [135, Chap. 7]. We can rewrite Eq. 2.106 as

$$C_{\mathrm{Y}}^a = \left[I_{N_y} + C_{\mathrm{Y}}^f H^{\mathrm{T}} C_{\mathrm{D}}^{-1} H\right]^{-1} C_{\mathrm{Y}}^f. \tag{2.107}$$

Hence, the square root of C_{Y}^a becomes

$$(C_{\mathrm{Y}}^a)^{1/2} = \left\{\left[I_{N_y} + C_{\mathrm{Y}}^f H^{\mathrm{T}} C_{\mathrm{D}}^{-1} H\right]^{-1} C_{\mathrm{Y}}^f\right\}^{1/2}. \tag{2.108}$$

Replacing the covariance matrix C_{Y}^a by its ensemble approximation, $\widetilde{C}_{\mathrm{Y}}^a$, given by

$$\widetilde{C}_{\mathrm{Y}}^a = \frac{1}{N_e - 1} \Delta Y^a (\Delta Y^a)^{\mathrm{T}}, \tag{2.109}$$

and considering the general equation, $\widetilde{C}_{\mathrm{Y}}^a = (\widetilde{C}_{\mathrm{Y}}^a)^{1/2} (\widetilde{C}_{\mathrm{Y}}^a)^{\mathrm{T}/2}$, we can write

$$(\widetilde{C}_{\mathrm{Y}}^{a})^{1/2} = \frac{1}{\sqrt{N_e - 1}} \Delta Y^{a}. \tag{2.110}$$

Using Eq. 2.110 in Eq. 2.108 and following the same procedure for the square root of C_{Y}^{f}, we obtain

$$\Delta Y^{a} = \left[I_{N_y} + \widetilde{C}_{\mathrm{Y}}^{f} H^{\mathrm{T}} C_{\mathrm{D}}^{-1} H \right]^{-1/2} \Delta Y^{f}. \tag{2.111}$$

Thus, we define the ETM as

$$T_L \equiv \left[I_{N_y} + \widetilde{C}_{\mathrm{Y}}^{f} H^{\mathrm{T}} C_{\mathrm{D}}^{-1} H \right]^{-1/2}. \tag{2.112}$$

However, this scheme requires the square root of a $N_y \times N_y$ matrix, which is not feasible for typical reservoir history-matching applications. Whitaker and Hamill [186] introduced a left-multiplied EnSRF scheme which does not require the explicit computation of a $N_y \times N_y$ matrix. In this scheme, we define a modified Kalman gain matrix, $\widetilde{K}^{\star}$, such that

$$\Delta Y^{a} = \left[I_{N_y} - \widetilde{K}^{\star} H \right] \Delta Y^{f} \tag{2.113}$$

exactly satisfies the analysis covariance equation (2.88). From Eq. 2.113, we have

$$\Delta Y^{a} (\Delta Y^{a})^{\mathrm{T}} = \left[I_{N_y} - \widetilde{K}^{\star} H \right] \Delta Y^{f} (\Delta Y^{f})^{\mathrm{T}} \left[I_{N_y} - \widetilde{K}^{\star} H \right]^{\mathrm{T}}, \tag{2.114}$$

and using $\widetilde{C}_{\mathrm{Y}}^{a} = \frac{1}{N_e - 1} \Delta Y^{a} (\Delta Y^{a})^{\mathrm{T}}$ and $\widetilde{C}_{\mathrm{Y}}^{f} = \frac{1}{N_e - 1} \Delta Y^{f} (\Delta Y^{f})^{\mathrm{T}}$, we obtain

$$\widetilde{C}_{\mathrm{Y}}^{a} = \left[I_{N_y} - \widetilde{K}^{\star} H \right] \widetilde{C}_{\mathrm{Y}}^{f} \left[I_{N_y} - \widetilde{K}^{\star} H \right]^{\mathrm{T}} = \left[I_{N_y} - \widetilde{K} H \right] \widetilde{C}_{\mathrm{Y}}^{f}, \tag{2.115}$$

where the last equality came from Eq. 2.88. This procedure leads to a nonlinear equation whose solution was first presented by Andrews [8], and it is given by

$$\widetilde{K}^{\star} = \widetilde{C}_{\mathrm{Y}}^{f} H^{\mathrm{T}} \left[H \widetilde{C}_{\mathrm{Y}}^{f} H^{\mathrm{T}} + C_{\mathrm{D}} \right]^{-\mathrm{T}/2} \left[\left(H \widetilde{C}_{\mathrm{Y}}^{f} H^{\mathrm{T}} + C_{\mathrm{D}} \right)^{1/2} + C_{\mathrm{D}}^{1/2} \right]^{-1}. \tag{2.116}$$

The interesting feature of this equation is that it requires square roots of $N_n \times N_n$ matrices only. For individual observations, the matrices C_{DD}^{f} and C_{D} reduce to scalars, and the computation of $\widetilde{K}^{\star}$ can be simplified. Denoting these scalars by

$$H \widetilde{C}_{\mathrm{Y}}^{f} H^{\mathrm{T}} = \widetilde{C}_{\mathrm{DD}}^{f} = \sigma_{dd}^{2} \tag{2.117}$$

and

$$C_{\mathrm{D}} = \sigma_{d}^{2}, \tag{2.118}$$

we can rewrite Eq. 2.116 as

$$\widetilde{K}^{\star} = \widetilde{C}_{\mathrm{Y}}^{f} H^{\mathrm{T}} \left(\frac{1}{\sqrt{\sigma_{dd}^{2} + \sigma_{d}^{2}}} \right) \times \left(\frac{1}{\sqrt{\sigma_{dd}^{2} + \sigma_{d}^{2}} + \sigma_{d}} \right), \tag{2.119}$$

and write the original Kalman gain as

$$\widetilde{K} = \widetilde{C}_{\mathrm{Y}}^{f} H^{\mathrm{T}} \left(\frac{1}{\sigma_{dd}^{2} + \sigma_{d}^{2}} \right). \tag{2.120}$$

Whitaker and Hamill [186] found a scalar, α, such that we can write

$$\widetilde{K}^{\star} = \alpha \widetilde{K}. \tag{2.121}$$

From Eqs. 2.119 and 2.120, we find that α is given by

$$\alpha = \frac{\sqrt{\sigma_{dd}^2 + \sigma_d^2}}{\sqrt{\sigma_{dd}^2 + \sigma_d^2} + \sigma_d}. \tag{2.122}$$

From Eq. 2.122, we note that α is a positive number less than one. It would be one for a noise-free measurement. Although here we use a different notation, the coefficient α from Eq. 2.122 is the same coefficient given by Eq. 13 of Whitaker and Hamill [186]. According to Whitaker and Hamill [186], if the measurement errors are uncorrelated, i.e., C_{D} is diagonal, then the observations may be assimilated one at a time. Therefore, we can assimilate data using $\widetilde{K}^{\star} = \alpha \widetilde{K}$ and the EnSRF requires no more computation than the traditional EnKF.

2.7 Iterative Forms of EnKF

Even though the EnKF was originally proposed as an alternative to the extended Kalman filter for applications in nonlinear dynamical systems, the update step in the EnKF is still linear. This linear update may result in a suboptimal performance for highly nonlinear problems. In these situations, EnKF may fail to provide reasonable data matches. This fact motivated the development of several iterative forms of EnKF in the last few years. In this section, we present some of iterative methods which were proposed in the context of reservoir history matching. In Chapter 9, we present a comparative study of these methods.

2.7.1 Half-iteration EnKF

When applied to reservoir history-matching problems, EnKF requires us to update a combined parameter-state vector, which includes model parameters and primary variables of the reservoir simulator. Hence, there is an underlying assumption of statistical consistency between the updated vectors of model parameters and states, where "consistency" is defined in [167]. However, because the reservoir simulator equations are highly nonlinear, this assumption can be invalid and the results obtained by EnKF can be unreliable. A simple procedure to avoid inconsistency is

to rerun the simulator with the latest ensemble of model parameters from time zero after each data assimilation time-step. This procedure is referred to as half-iteration EnKF (HI-EnKF) [182]. To improve computational efficiency, Wang et al. [182] proposed rerunning the ensemble from time zero only if the average relative change in the ensemble mean model is larger than a threshold value.

2.7.2 Lorentzen-Nævdal iterative EnKF

Lorentzen and Nævdal [112] introduced an iterative procedure to improve the estimates obtained by EnKF for highly nonlinear problems. Here, we refer to this method as LN-IEnKF. Algorithm 2.2 presents the LN-IEnKF procedure.

In this procedure, the state vector at the ℓth iteration is updated using Eq. 2.123, which corresponds to the EnKF analysis equation (Eq. 2.49) with $d_j^{n,f}$ replaced by $d_j^{n,\ell}$, and $\widetilde{C}_{\mathrm{YD}}^{n,\ell}$ and $\widetilde{C}_{\mathrm{DD}}^{n,\ell}$ computed based on $d_j^{n,\ell}$. In Eq. 2.124, $(C_{\mathrm{Y}}^{n,f})^{-1}$ is the inverse of the forecast state covariance matrix, which is fixed during the iterations. However, Lorentzen and Nævdal [112] do not mention how they compute $(C_{\mathrm{Y}}^{n,f})^{-1}$. For reservoir history-matching problems, the only feasible option is to approximate $C_{\mathrm{Y}}^{n,f}$ based on the forecast ensemble and compute its inverse by SVD.

The LN-IEnKF method requires one to compute the predicted data vector, $d_j^{n,\ell}$, for a given updated state vector, $y_j^{n,\ell}$, i.e., $d_j^{n,\ell} = h(y_j^{n,\ell})$. For a typical reservoir history-matching application, the observations correspond to data at wells; e.g., well water-cut or bottomhole pressure data. In this case, $h(y_j^{n,\ell})$ can be computed using Peaceman's equation [136]. Note that we use $h(\cdot)$ to denote the nonlinear relationship between the state vector and the predicted data, which is different from our previous notation where we used $g(\cdot)$ to denote the nonlinear relationship between the vector of model parameters and the predicted data.

2.7.3 Krymskaya-Hanea-Verlaan iterative EnKF

Krymskaya et al. [98] also proposed an iterative form of EnKF for reservoir

Algorithm 2.2 LN-IEnKF

1. At the nth data assimilation time-step, update the forecast state vector, $y_j^{n,f}$, using the standard EnKF equation (Eq. 2.49).

2. Set $\ell = 1$ and $y_j^{n,1} = y_j^{n,a}$, for $j = 1, 2, \ldots, N_e$.

3. Repeat

   (a) For $j = 1$ to N_e:

   - Compute the predicted data $d_j^{n,\ell} = h(y_j^{n,\ell})$, where $h(y_j^{n,\ell})$ denotes a nonlinear function of the state vector $y_j^{n,\ell}$.
   - Compute

$$y_j^{n,\ell+1} = y_j^{n,\ell} + \widetilde{C}_{\mathrm{YD}}^{n,\ell} \left(\widetilde{C}_{\mathrm{DD}}^{n,\ell} + C_{\mathrm{D}}^{n} \right)^{-1} \left(d_{\mathrm{uc},j}^{n} - d_j^{n,\ell} \right), \tag{2.123}$$

   - Compute

$$\begin{aligned} O_j^{n,\ell+1} &= \left(y_j^{n,\ell+1} - y_j^{n,f} \right)^{\mathrm{T}} (C_{\mathrm{Y}}^{n,f})^{-1} \left(y_j^{n,\ell+1} - y_j^{n,f} \right) \\ &+ \left(h(y_j^{n,\ell+1}) - d_{\mathrm{uc},j}^{n} \right)^{\mathrm{T}} (C_{\mathrm{D}}^{n})^{-1} \left(h(y_j^{n,\ell+1}) - d_{\mathrm{uc},j}^{n} \right) \end{aligned} \tag{2.124}$$

   - If $(O_j^{n,\ell+1} \geq O_j^{n,\ell})$ set $y_j^{n,\ell+1} = y_j^{n,\ell}$, i.e., do not update the state.

   end (for).

   (b) Set $\ell = \ell + 1$.

   until $(O_j^{n,\ell+1} \geq O_j^{n,\ell}$, for $j = 1, 2, \ldots, N_e)$, i.e., no states are updated; or ($\ell >$ maximum number of iterations).

history-matching problems. Here, we refer to this method as KHV-IEnKF. In this method, we assimilate data with standard EnKF and use the final ensemble mean of model parameters as the "prior mean" for a subsequent data assimilation. Algorithm 2.3 presents the KHV-IEnKF procedure:

Algorithm 2.3 KHV-IEnKF

1. Initialize $\overline{m}^0 = m_{\text{pr}}$ and $\ell = 0$.
2. Generate the ensemble of model parameters by sampling $m_j^\ell \sim \mathcal{N}(\overline{m}^\ell, C_{\text{M}})$, for $j = 1, 2, \cdots, N_e$.
3. Assimilate data using EnKF and set $m_j^{\ell+1} = m_j^a$, for $j = 1, 2, \cdots, N_e$, where m_j^a is the analyzed vector of model parameters.
4. Rerun the ensemble from time zero and compute the average objective function: $\overline{O}^{\ell+1} = \frac{1}{N_e}\sum_{j=1}^{N_e} O(m_j^{\ell+1})$, where $O(m_j^{\ell+1})$ is computed using Eq. 2.3.
5. If $\left(\left|\frac{\overline{O}^{\ell+1} - \overline{O}^\ell}{\overline{O}^\ell + 10^{-8}}\right| < \epsilon_o\right)$ or (ℓ = maximum number of iterations) then:
   - Stop data assimilation.

   Else:
   - Compute the ensemble mean: $\overline{m}^{\ell+1} = \frac{1}{N_e}\sum_{j=1}^{N_e} m_j^{\ell+1}$.
   - Set $\ell = \ell + 1$.
   - Return to step 2.

   end (if).

In [98], the authors use the "confirming step" [184] option to avoid non-physical values of pressure and saturation when restarting simulations. However, this procedure is inconsistent for the linear-Gaussian case [191]. Unfortunately, the KHV-IEnKF is also inconsistent for the linear-Gaussian case. By simple inspection of the algorithm, we can conclude that KHV-IEnKF will not provide a correct sampling of the posterior pdf for the linear-Gaussian case because we use the posterior mean from the previous iteration as the "prior mean" for the next iteration, which is clearly statistically incorrect.

2.7.4 Ensemble randomized maximum likelihood

The ensemble randomized maximum likelihood (EnRML) method was introduced by Gu and Oliver [70] as an iterative EnKF. EnRML uses a Gauss-Newton update equation with an "average sensitivity matrix" estimated by the ensemble. In the EnRML, each model of the ensemble is updated using

$$m_j^{\ell+1} = m_j^{\ell} + \beta_\ell \delta m_j^{\ell+1}, \tag{2.125}$$

with

$$\delta m_j^{\ell+1} = m_j^f - m_j^{\ell} - \widetilde{C}_{\mathrm{M}}^f \widetilde{G}_\ell^{\mathrm{T}} \left(\widetilde{G}_\ell \widetilde{C}_{\mathrm{M}}^f \widetilde{G}_\ell^{\mathrm{T}} + C_{\mathrm{D}} \right)^{-1} \left[d_j^{\ell} - d_{\mathrm{uc},j} - \widetilde{G}_\ell \left(m_j^{\ell} - m_j^f \right) \right]. \tag{2.126}$$

In the above equations, ℓ denotes the iteration index and β_ℓ denotes the step size. During the iterative process, $\widetilde{C}_{\mathrm{M}}^f$ is fixed and estimated based on the forecast ensemble, i.e.,

$$\widetilde{C}_{\mathrm{M}}^f = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(m_j^f - \overline{m}^f \right) \left(m_j^f - \overline{m}^f \right)^{\mathrm{T}} = \frac{\Delta M^f (\Delta M^f)^{\mathrm{T}}}{N_e - 1}. \tag{2.127}$$

The average sensitivity matrix, $\widetilde{G}_\ell$, is computed using

$$\widetilde{G}_\ell = \Delta D^{\ell} \left(\Delta M^{\ell} \right)^{+}, \tag{2.128}$$

where the superscript "+" denotes the pseudo-inverse of ΔM^{ℓ} computed by SVD. $\Delta D^{\ell} = D^{\ell} - \overline{D}^{\ell}$, where D^{ℓ} is the matrix with the ensemble of predicted data at the ℓth iteration, i.e., the jth column of D^{ℓ} corresponds to the predicted data from the jth ensemble member. $\overline{D}^{\ell}$ is the matrix with all columns equal to $\overline{d}^{\ell}$, which represents the average of all columns of D^{ℓ}. Similarly, $\Delta M^{\ell} = M^{\ell} - \overline{M}^{\ell}$, where the

jth column of M^ℓ contains the vector of model parameters corresponding to the jth ensemble member at the ℓth iteration. $\overline{M}^\ell$ is the matrix with all columns equal to $\overline{m}^\ell$, which represents the average of all columns of M^ℓ.

The EnRML procedure is presented in Algorithm 2.4. Apparently, the last termination criterion in Algorithm 2.4, i.e., maximum number of step size cuts, was not used in Gu and Oliver [70]. In fact, [70] gives very little information about the line-search procedure. In our implementation, we use an initial step size $\beta_0 = 1$, a step size increasing/reduction factor of $\alpha = 2$, and a maximum number of step size cuts equal to five.

In the EnRML there is no guarantee that the search direction is downhill. When the search direction obtained is not downhill, the iterative procedure fails to improve the data matches. Moreover, because all models are updated using the same average sensitivity, similar to the standard EnKF, EnRML is not expected to sample multiple modes of the posterior distribution. In the original EnRML, during an iteration, all ensemble members are updated with the same step size. Wang et al. [182] suggested computing a different step size for each ensemble member. Although this procedure seems to be more efficient, the results presented in [182] do not show significant differences between the two line-search procedures.

2.8 Ensemble Smoother

The ensemble smoother (ES) was proposed by van Leeuwen and Evensen [174]. Unlike EnKF, ES does not assimilate data sequentially in time. Instead, ES computes a global update including all data available. Because ES assimilates all data simultaneously, we only need to consider the parameter estimation problem. Other than that, ES formulation is similar to EnKF. For ES, we write the analyzed vector of model parameters, m^a, as

$$m_j^a = m_j^f + \widetilde{C}_{\mathrm{MD}}^f \left(\widetilde{C}_{\mathrm{DD}}^f + C_{\mathrm{D}}\right)^{-1} \left(d_{\mathrm{uc},j} - d_j^f\right), \tag{2.129}$$

Algorithm 2.4 EnRML

1. Run the ensemble from time zero until the next data assimilation time-step.
2. Initialize: $\ell = 0$, $N_{\text{cuts}} = 0$, β_0 and $m_j^0 = m_j^f$ for $j = 1, 2, \ldots, N_e$.
3. Compute $\widetilde{G}_\ell$ (Eq. 2.128).
4. For $j = 1$ to N_e:
   (a) Compute $m_j^{\ell+1}$ (Eqs. 2.125 and 2.126).
   (b) Rerun ensemble from time zero.
   (c) Compute $O_{d,j}^{\ell+1} = 0.5 \left(d_j^{\ell+1} - d_{\text{uc},j}\right)^{\text{T}} C_{\text{D}}^{-1} \left(d_j^{\ell+1} - d_{\text{uc},j}\right)$.

   end (for).
5. Compute $\overline{O}_d^{\ell+1} = \frac{1}{N_e} \sum_{j=1}^{N_e} O_{d,j}^{\ell+1}$.
6. If $\overline{O}_d^{\ell+1} < \overline{O}_d^{\ell}$ then:
   (a) Accept the step and increase the step size for the next iteration, $\beta_{\ell+1} = \alpha\beta_\ell$.
   (b) Set $N_{\text{cuts}} = 0$.
   (c) Increase the iteration index, $\ell = \ell + 1$.

   Else:
   (a) Reduce the step size, $\beta_\ell = \beta_\ell / \alpha$.
   (b) Set $N_{\text{cuts}} = N_{\text{cuts}} + 1$.
   (c) If $N_{\text{cuts}} \leq N_{\text{cuts,max}}$ then return to step 4.

   end (if).
7. Check termination criteria.
8. If any one of the termination criteria is satisfied, then go to the next data assimilation time-step. Otherwise, return to the step 3.

The termination criteria include the following:

- $\max |m_{i,j}^{\ell+1} - m_{i,j}^{\ell}| < \epsilon_m$ for $i = 1, 2, \ldots, N_m$ and $j = 1, 2, \ldots, N_e$.
- $\left| \frac{\overline{O}_d^{\ell+1} - \overline{O}_d^{\ell}}{\overline{O}_d^{\ell}} \right| < \epsilon_o$.
- Maximum number of iterations.
- Maximum number of step size cuts ($N_{\text{cuts,max}}$).

for $j = 1, 2, \cdots N_e$. The notation is similar to the one presented for EnKF. $\widetilde{C}^f_{\mathrm{MD}}$ is the cross-covariance matrix between the prior vector of model parameters, m^f, and the vector of predicted data, d^f; $\widetilde{C}^f_{\mathrm{DD}}$ is the $N_d \times N_d$ auto-covariance matrix of predicted data. Note that we use N_d to denote the total number of measurements assimilated, which is different from our previous notation for EnKF, where we used N_n to denote the number of measurements at the nth data assimilation time-step. Similar to the EnKF case, $d_{\mathrm{uc}} \sim \mathcal{N}(d_{\mathrm{obs}}, C_{\mathrm{D}})$ where d_{obs} denotes the N_d-dimensional vector of observed data and C_{D} denotes the $N_d \times N_d$ covariance matrix of observed data measurement errors.

2.9 Normalized Objective Function

According to Tarantola [164, p. 74] and discussed in more detail in the history matching context in [193, 59], for the linear case, the minimum of $2O(m)$ at the MAP estimate has a χ^2 distribution with N_d degrees of freedom. For N_d large, this χ^2 distribution can be approximated by a Gaussian distribution with mean equal to N_d and variance equal to $2N_d$. Assuming that any legitimate realization of this χ^2 distribution should be within five standard deviations of the mean implies that the final objective function at the MAP estimate should satisfy

$$1 - 5\sqrt{\frac{2}{N_d}} \leq \frac{2O(m_{\mathrm{MAP}})}{N_d} \leq 1 + 5\sqrt{\frac{2}{N_d}}. \tag{2.130}$$

Although this criterion is strictly valid only for the linear case, we expect that it approximately holds for the nonlinear case. According to Oliver et al. [135], failure to satisfy Eq. 2.130 may indicate convergence to a local minimum, in which case, we should question whether a reasonable MAP estimate was obtained. Based on Eq. 2.130, Gao et al. [58] defined the normalized objective function at the MAP estimate as

$$O_N(m_{\text{MAP}}) \equiv \frac{2O(m_{\text{MAP}})}{N_d}. \tag{2.131}$$

According to Eqs. 2.130 and 2.131, as the number of data points increases, we expect $O_N(m_{\text{MAP}}) \approx 1$.

When we use RML for obtaining samples of the posterior pdf, we minimize the modified objective function given by Eq. 2.21. According to Oliver et al. [135, p. 149], because we use m_{uc} and d_{uc} instead of m_{pr} and d_{obs} in Eq. 2.21, the criterion of Eq. 2.130 should be modified to

$$1 - 5\sqrt{\frac{2}{N_d}} \leq \frac{O(m_{\text{c}})}{N_d} \leq 1 + 5\sqrt{\frac{2}{N_d}}, \tag{2.132}$$

where, m_{c} denotes a conditional realization obtained by RML. The same argument is valid for EnKF. In fact, as shown in [143], the EnKF analysis equation can be derived from RML so that both methods are equivalent for the linear-Gaussian case when the size of the ensemble used in the EnKF goes to infinity. Hence, it is convenient to re-define the normalized objective function for a RML or EnKF sample as

$$O_N(m_{\text{c}}) \equiv \frac{O(m_{\text{c}})}{N_d}. \tag{2.133}$$

2.10 Normalized Variance

For the linear-Gaussian case, the posterior covariance, C_{MAP}, is given by Eq. 2.14. The last equality in Eq. 2.14 shows that C_{MAP} is given by the prior, C_{M}, minus a matrix ΔC_{M}, i.e,

$$C_{\text{MAP}} = C_{\text{M}} - \Delta C_{\text{M}}, \tag{2.134}$$

where

$$\Delta C_{\mathrm{M}} = C_{\mathrm{M}} G^{\mathrm{T}} \left(G C_{\mathrm{M}} G^{\mathrm{T}} + C_{\mathrm{D}} \right)^{-1} G C_{\mathrm{M}}. \tag{2.135}$$

Because ΔC_{M} is a positive semidefinite matrix [135, Chap. 9], its diagonal entries are non-negative. Hence, the posterior variances, i.e., the diagonal entries of C_{MAP} are less than or equal to the prior variances. Therefore, dividing a posterior variance by the corresponding prior variance results in a number less or equal to one, which can be interpreted as an approximate measure of reduction in uncertainty [135, Chap. 9]. However, it is clear that the normalized variances ignore the change in uncertainty due to the change in the covariances [135]. Nevertheless, in this dissertation, we use the sum of normalized variances (SNV) as an indication of reduction in uncertainty due to data assimilation. We compute the SNV using

$$\mathrm{SNV} = \sum_{i=1}^{N_m} \frac{\mathrm{var}[m_i^a]}{\mathrm{var}[m_i^f]}, \tag{2.136}$$

with m_i^a and m_i^f denoting the values of the ith model parameter after and before data assimilation.

CHAPTER 3

COVARIANCE LOCALIZATION

Covariance localization is a standard procedure for reducing the spurious correlations caused by sampling errors in the EnKF. Another advantage of covariance localization is that it increases the degrees of freedom available to assimilate data. Most of the localization procedures are distance-dependent, i.e., they assume that only measurements positioned within a certain distance from a gridblock should affect the analysis in that gridblock. Covariance localization is done by forming the Schur (or Hadamard) product between a correlation matrix with compact support and the covariance of the forecast state vector. Hence, the most important decision required for distance-based covariance localization is the choice of the correlation length to build the correlation matrix. Throughout, we refer to this correlation length as critical length. This chapter presents a procedure that combines the correlation lengths from the prior geological model and the region of data sensitivity to define the critical lengths. In addition, this chapter presents the results of a real field application of EnKF with the proposed covariance localization procedure. The results presented in this chapter were published in Emerick and Reynolds [36] and Emerick and Reynolds [37].

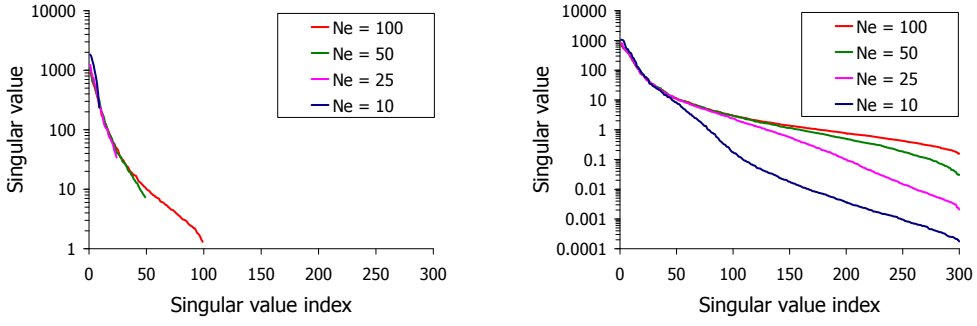
3.1 EnKF Analysis with Covariance Localization

For covariance localization, we replace the forecast covariance matrix, $\widetilde{C}_{\mathrm{Y}}^{n,f}$, by the Schur product between a correlation matrix, ρ^n, and $\widetilde{C}_{\mathrm{Y}}^{n,f}$, i.e.,

$$\widetilde{C}_{\mathrm{Y}_\rho}^{n,f} = \rho^n \circ \widetilde{C}_{\mathrm{Y}}^{n,f}, \tag{3.1}$$

where "$\circ$" denotes the Schur product, which is an element-wise product between ρ^n and $\widetilde{C}_{\mathrm{Y}}^{n,f}$.

By computing eigenvalues of estimated covariance matrices, Hamill et al. [71] showed that the covariance matrix estimated by an ensemble becomes full rank when this Schur product is employed. This is a consequence of the Schur Product Theorem [79]: *If the matrix A is positive-definite and the matrix B is positive semi-definite with all of its main diagonal entries positive, then the product $A \circ B$ is positive-definite.* Because the sampled covariance matrix is positive semi-definite with positive values on the diagonal, Schur's Product Theorem guarantees that the Schur product of the sampled covariance with a positive-definite correlation matrix will result in a positive-definite matrix. As a consequence, localization allows the model update to be obtained from a much larger space than the one spanned by the ensemble. Fig. 3.1 illustrates this fact by showing the singular values of a 300×300 covariance matrix estimated from ensembles before and after the Schur product with a positive-definite correlation matrix for different ensemble sizes. This figure shows that the number of nonzero singular values of the ensemble covariance matrices is limited by the size of the ensemble. After the Schur product, the matrices become full rank, i.e., there are 300 nonzero singular values.



(a) Before Schur product **(b)** After Schur product

Figure 3.1: Singular values of covariance matrices estimated by ensembles before and after the Schur product with a positive-definite correlation matrix.

For covariance localization, we replace the EnKF analysis (Eq. 2.39) by

$$y_j^{n,a} = y_j^{n,f} + (\rho^n \circ \widetilde{C}_{\mathrm{Y}}^{n,f}) H_n^{\mathrm{T}} \left[C_{\mathrm{D}}^n + H_n (\rho^n \circ \widetilde{C}_{\mathrm{Y}}^{n,f}) H_n^{\mathrm{T}} \right]^{-1} \left(d_{\mathrm{uc},j}^n - d_j^{n,f} \right). \tag{3.2}$$

Similarly to $\widetilde{C}_{\mathrm{Y}}^{n,f}$, the correlation matrix ρ^n also can be partitioned as

$$\rho^n = \begin{bmatrix} \rho_{\mathrm{MM}}^n & \rho_{\mathrm{MP}}^n & \rho_{\mathrm{MD}}^n \\ \rho_{\mathrm{PM}}^n & \rho_{\mathrm{PP}}^n & \rho_{\mathrm{PD}}^n \\ \rho_{\mathrm{DM}}^n & \rho_{\mathrm{DP}}^n & \rho_{\mathrm{DD}}^n \end{bmatrix}. \tag{3.3}$$

It follows that

$$(\rho^n \circ \widetilde{C}_{\mathrm{Y}}^{n,f}) H_n^{\mathrm{T}} = \begin{bmatrix} \rho_{\mathrm{MD}}^n \circ \widetilde{C}_{\mathrm{MD}}^{n,f} \\ \rho_{\mathrm{PD}}^n \circ \widetilde{C}_{\mathrm{PD}}^{n,f} \\ \rho_{\mathrm{DD}}^n \circ \widetilde{C}_{\mathrm{DD}}^{n,f} \end{bmatrix} = \rho_{\mathrm{YD}}^n \circ \widetilde{C}_{\mathrm{YD}}^{n,f}, \tag{3.4}$$

and

$$H_n (\rho^n \circ \widetilde{C}_{\mathrm{Y}}^{n,f}) H_n^{\mathrm{T}} = \rho_{\mathrm{DD}}^n \circ \widetilde{C}_{\mathrm{DD}}^{n,f}. \tag{3.5}$$

Using Eqs. 3.4 and 3.5, the EnKF analysis equation with covariance localization (3.2) becomes

$$y_j^{n,a} = y_j^{n,f} + \left(\rho_{\mathrm{YD}}^n \circ \widetilde{C}_{\mathrm{YD}}^{n,f} \right) \left(C_{\mathrm{D}}^n + \rho_{\mathrm{DD}}^n \circ \widetilde{C}_{\mathrm{DD}}^{n,f} \right)^{-1} \left(d_{\mathrm{uc},j}^n - d_j^{n,f} \right). \tag{3.6}$$

Eq. 3.6 indicates that to perform localization, we need to compute only the submatrices, ρ_{MD}^n, ρ_{PD}^n and ρ_{DD}^n. Although we have made a distinction between ρ_{MD}^n, ρ_{PD}^n and ρ_{DD}^n, in the results considered in this dissertation, we use correlations which are distance-dependent only; and the entries of these three localization matrices are calculated using the same correlation function and the same critical length.

Alternatively, we can apply the Schur product directly to the Kalman gain. We refer to this procedure as Kalman gain localization. In this case, we write the EnKF analysis equation as

$$y_j^{n,a} = y_j^{n,f} + \left(\rho_{\mathrm{YD}}^n \circ \widetilde{K}_n\right)\left(d_{\mathrm{uc},j}^n - d_j^{n,f}\right), \tag{3.7}$$

where, $\widetilde{K}_n$ is the regular Kalman gain given by Eq. 2.40. Note that the subspace inversion procedure described in Section 2.5.2 does not allow use of covariance localization because the matrix $\widetilde{C}_{\mathrm{DD}}^{n,f}$ is not formed. In this case, Kalman gain localization can be used.

Eq. 3.6 shows that only the last N_n columns of $\rho^n \circ \widetilde{C}_{\mathrm{Y}}^{n,f}$ directly affect the analysis step, and these last N_n columns are given by $\rho_{\mathrm{YD}}^n \circ \widetilde{C}_{\mathrm{YD}}^{n,f}$; see Eq. 3.4. Ideally, we wish the N_n columns of $\rho_{\mathrm{YD}}^n \circ \widetilde{C}_{\mathrm{YD}}^{n,f}$ to be linearly independent because the updated state vector is a linear combination of the columns of $\rho_{\mathrm{YD}}^n \circ \widetilde{C}_{\mathrm{YD}}^{n,f}$. Assuming that $N_y > N_n$, the maximum rank of $\rho_{\mathrm{YD}}^n \circ \widetilde{C}_{\mathrm{YD}}^{n,f}$ is N_n. When the rank of $\rho_{\mathrm{YD}}^n \circ \widetilde{C}_{\mathrm{YD}}^{n,f}$ is equal to N_n, we obtain the maximum number of degrees of freedom available to assimilate the N_n observations. The Schur product $\rho^n \circ \widetilde{C}_{\mathrm{Y}}^{n,f}$ is positive-definite as long as we employ a positive-definite matrix ρ^n. Unfortunately, we cannot guarantee that the composite localization matrix defined in Eq. 3.3 is positive-definite. However, even if the last block column of ρ^n does not have rank N_n, it is plausible that $\rho_{\mathrm{YD}}^n \circ \widetilde{C}_{\mathrm{YD}}^{n,f}$ will have rank N_n.

Chen and Oliver [22] concluded that for reservoir problems, the cross-covariance between data and state variables, $\widetilde{C}_{\mathrm{PD}}^{n,f}$, may be very different from the cross-covariance between data and model parameters, $\widetilde{C}_{\mathrm{MD}}^{n,f}$. Thus, one should apply different localization for $\widetilde{C}_{\mathrm{MD}}^{n,f}$ and $\widetilde{C}_{\mathrm{PD}}^{n,f}$. Because the sensitivity of predicted data to model parameters, say gridblock permeabilities, is different from the sensitivity of data to primary variables, say pressure and fluid saturations, the use of different localization functions seems preferable. However, unless we can accurately calculate cross-covariances,

which is not computationally feasible, we would not be able to design proper localization functions. Thus, in the localization procedure introduced in this chapter, we apply the same localization function to modify the two cross-covariances, $\widetilde{C}^{n,f}_{\mathrm{MD}}$ and $\widetilde{C}^{n,f}_{\mathrm{PD}}$. For the auto-covariance of predicted data, $\widetilde{C}^{n,f}_{\mathrm{DD}}$, we also apply the same localization function. It is important to observe that the motivation for including primary variables in the state vector in the EnKF is to avoid the need to run the simulator from time zero before each data assimilation time-step. In this case, there is an underlying assumption that the ensemble of updated primary variables obtained with EnKF are statistically consistent with the ensemble of primary variables that would be obtained by running the reservoir simulator from time zero using the ensemble of updated vectors of model parameters as input. Once localization is applied, we can not guarantee this consistency even for the linear-Gaussian case. Nevertheless, the computational results presented in this chapter indicate that the proposed localization procedure does not increase inconsistency problems. In fact, the results indicate that localization improves the data matches obtained by EnKF.

3.1.1 Correlation matrix for localization

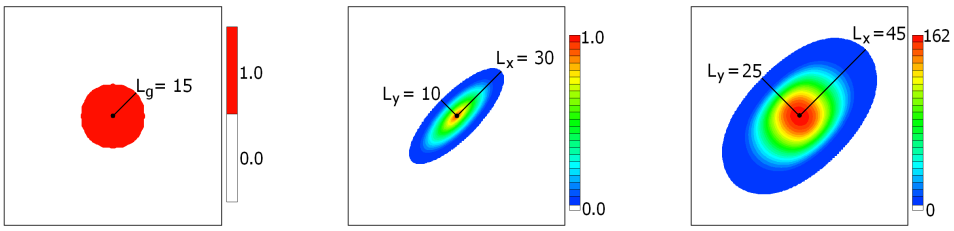
The most important decision to be made in the correlation matrix for localization is the choice of the critical length(s). Based on sensitivities [188, 105], we expect that each type of data has a different region of influence. For example, bottomhole pressure data at a flowing well are strongly sensitive only to gridblock permeabilities and porosities fairly close to the well, but water-cut data at a producing well are sensitive to porosities and permeabilities along streamlines connecting a producing well to a water injection well. It is intuitive that we should consider the region of influence of data when selecting the critical length(s) for localization. Using a localization region significantly smaller than the actual region of influence may lead to unreasonably large local changes in model parameters. Besides that, the localization region is expected to be affected by the geological structure, e.g., the covariance

function for a prior probability density function for the vector of model parameters. Applying localization to restrict changes in the state vector to a region smaller than the region based on the correlation length(s) inherent in the geological model may lead to model changes that destroy the geological realism of the EnKF estimates. Therefore, the appropriate choice of the critical length(s) should be based both on the underlying correlation length(s) of the geological model and data sensitivity.

In order to analyze our conjecture on the choice of the critical length, we consider the linear case. For the linear case, the covariance $C_{\mathrm{MD}}^{n,f}$ in the analysis equation reduces to the product $C_{\mathrm{M}}^{n,f}G_n^{\mathrm{T}}$, where $C_{\mathrm{M}}^{n,f}$ denotes the forecast covariance matrix of model parameters and G_n^{T} denotes the transpose of the sensitivity matrix. Assume that for a particular well datum, nonzero sensitivities correspond to a circular region with radius L_g (Fig. 3.2a) and the model covariance at this location can be described by the correlation lengths L_x and L_y (Fig. 3.2b). For this case, the resulting correlation lengths of the product $C_{\mathrm{M}}^{n,f}G_n^{\mathrm{T}}$ are the sums $L_g + L_x$ and $L_g + L_y$, as illustrated in Fig. 3.2c. Assuming that instead of computing the product $C_{\mathrm{M}}^{n,f}G_n^{\mathrm{T}} = C_{\mathrm{MD}}^{n,f}$, we use the ensemble approximation $\widetilde{C}_{\mathrm{MD}}^{n,f}$. For this case, if we apply localization to $\widetilde{C}_{\mathrm{MD}}^{n,f}$, there is no reason to choose critical lengths larger than the prior principal critical lengths plus the length (radius) of the region of sensitivity, i.e., $L_g + L_x$ and $L_g + L_y$. If the distance between a model parameter and a datum is greater than these critical lengths, we know that the covariance must be zero. On the other hand, if we choose critical lengths much smaller than L_g+L_x and L_g+L_y, we may eliminate actual nonzero covariances, which is clearly undesirable. Thus, using critical lengths equal to the sum of prior principal correlation lengths and the sensitivity "range" is a reasonable and simple choice. Moreover, this choice is easy to implement in practice. In the linear case, $C_{\mathrm{DD}}^{n,f} = G_n C_{\mathrm{M}}^{n,f} G_n^{\mathrm{T}}$, so using this same procedure to approximate the critical lengths for ρ_{DD}^{n} is also reasonable. Although our motivating assumption that the sensitivity region corresponding to a particular datum is a circular region

centered on the datum is rarely realistic, the results presented in this chapter show that the localization procedure based on this assumption effectively eliminates the detrimental results commonly associated with sampling error when applying the EnKF.

One might argue that because covariances are sequentially updated at each EnKF analysis step, basing critical correlation lengths on the covariance of the prior geological model is inappropriate because at each data assimilation time-step, the ensemble of forecasts implicitly represents the prior covariance for the subsequent data assimilation (analysis) step. Thus, the critical correlations lengths used to update model parameters at the data assimilation time-step t_n should depend on the critical lengths inherent in the analyzed covariance matrix at time t_{n-1} plus the sensitivity matrix at time-step t_n. Nevertheless, we argue that the correlation lengths inherent in the initial ensemble of vectors of model parameters should be essentially maintained after each data assimilation time-step assuming of course that the initial model is based on a reasonable geological description. Even though we expect the initial critical correlation lengths in the model will be roughly preserved, it is clear that the initial prior covariance matrix for model parameters multiplied by the transpose of the sensitivity matrix at time-step t_n may be numerically very different than the posterior covariance for model parameters at time-step t_{n-1} multiplied by the transpose of the sensitivity matrix at time-step t_n. The last statement is illustrated by the 1D immiscible flow example in the recent paper of Chen and Oliver [22]. Nevertheless, we argue that the incorporation of the principal correlation lengths inherent in the initial ensemble of vectors of model parameters in conjunction with sensitivity ranges at each data assimilation time-step provides a reasonable localization procedure even though this means that any change in the localization critical lengths with time must be accounted for entirely by changes in the ranges of sensitivities with time.

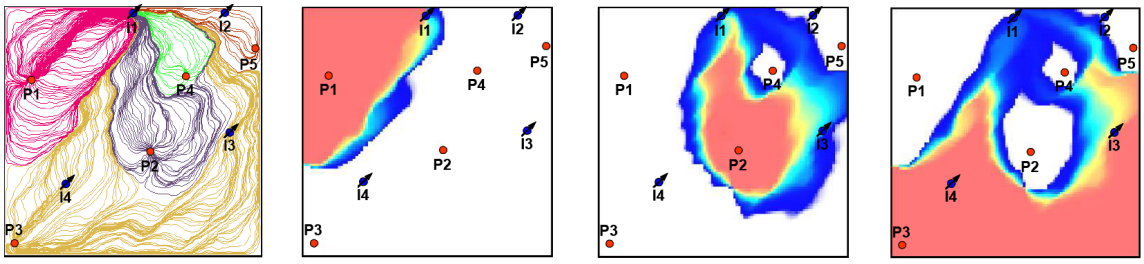


(a) Sensitivity (G_n) **(b)** Prior covariance ($C_{\mathrm{M}}^{n,f}$) **(c)** Product $C_{\mathrm{M}}^{n,f}G_n^{\mathrm{T}}$

Figure 3.2: Example illustrating that the ranges of the product $C_{\mathrm{M}}^{n,f}G_n^{\mathrm{T}}$ are equal to the sum of the radius of influence of the sensitivity matrix and the correlation lengths of the prior covariance.

In general, sensitivities are not readily computed with commercial reservoir simulators, and even when they are available, their computation is expensive. Therefore, computing sensitivities simply for the purpose of defining critical lengths for localization is not an attractive option. Arroyo-Negrete et al. [10] and Devegowda et al. [31] proposed a localization procedure that used streamlines to quantify the region of influence of model parameters on the observed data. The main attraction of their idea is that using streamlines, one can identify the region of influence of well data with a relatively low computational cost. In a recent paper, Watanabe and Datta-Gupta [183] modified the procedure proposed in [10] by using phase-streamlines instead of total-streamlines to obtain the localization region. By comparing phase-streamlines with cross-covariances between permeability and water and gas-oil ratio data estimated by an ensemble of 2,000 models, Watanabe and Datta-Gupta [183] concluded that phase-streamlines approximates reasonably well the regions of high cross-covariances. Watanabe and Datta-Gupta [183] used this result to justify the use of phase-streamlines for covariance localization. However, the ensemble of permeability used in [183] was uncorrelated. Therefore, for this case, basing the localization only on the phase-streamlines, which roughly represents the region of data sensitivity, is reasonable. In fact, this result does not contradict our conjecture that localization needs also to consider the correlation lengths in the prior geological model.

Damiani [28] proposed a method based on calculation of a pseudo-tracer concentration to define drainage areas (or irrigated areas in case of injection) associated with each well. The procedure assumes the injection of a pseudo-tracer which is fully miscible with the reservoir fluid. The computations are performed as a post-processing of results of a reservoir simulation using the velocity fields stored during the simulation. As shown by Damiani [28], the drainage areas obtained with this procedure are equivalent to the ones obtained by tracing the streamlines (Fig. 3.3).



(a) Streamlines **(b)** Pseudo-tracer concentration. P1 **(c)** Pseudo-tracer concentration. P2 **(d)** Pseudo-tracer concentration. P3

Figure 3.3: Comparison between drainage areas obtained by streamlines and pseudo-tracer [28].

We propose the following approach for covariance localization in reservoir simulation applications: at each data assimilation time-step, compute the region of influence of each well based on the ensemble mean model; the region of influence can be calculated either by using streamlines or by using the procedure proposed by Damiani [28]. Next, compute the area for each region and then calculate the radius of the circle with the same area. Alternatively, instead of using a circular region, we can generate an ellipse centered on each datum location that "best" represents the drainage region in a least squares sense. Then, the radii of influence are added to the principal correlation lengths of the prior geological model to obtain the final critical lengths that will be applied for localization using the fifth-order compact correlation function proposed by Gaspari and Cohn [61]

$$\rho = \begin{cases} -\frac{1}{4}(\frac{h}{L})^5 + \frac{1}{2}(\frac{h}{L})^4 + \frac{5}{8}(\frac{h}{L})^3 - \frac{5}{3}(\frac{h}{L})^2 + 1 & 0 \leq h \leq L \\ \frac{1}{12}(\frac{h}{L})^5 - \frac{1}{2}(\frac{h}{L})^4 + \frac{5}{8}(\frac{h}{L})^3 + \frac{5}{3}(\frac{h}{L})^2 - 5(\frac{h}{L}) + 4 - \frac{2}{3}(\frac{h}{L})^{-1} & L \leq h \leq 2L \\ 0 & h > 2L, \end{cases} \tag{3.8}$$

where L is the critical length and h is the Euclidean distance between any grid point and an observation location. Note that ρ is zero for $h \geq 2L$, and when $h = L$, $\rho \approx 0.21$. Fig. 3.4 illustrates the shape of the Gaspari-Cohn correlation function.

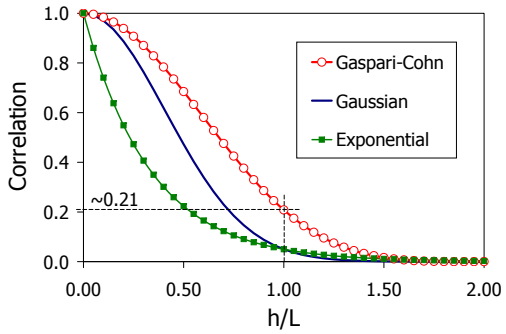


Figure 3.4: Correlation functions.

This localization procedure is applied on a layer-by-layer basis, i.e., each layer may have different critical lengths. When using ellipses, as opposed to simply circles to represent the drainage regions, we compute the lengths of the axes of each ellipse, L_1 and L_2, by minimizing a least-squares function. In our implementation, this minimization is carried out using the downhill simplex method [139, Chap. 10]. The least-squares function is defined by

$$F(L_1, L_2) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (c_{ij} - e(h))^2 \tag{3.9}$$

where N_x and N_y are the number of gridblocks in the x- and y-directions; c_{ij} is the pseudo-tracer concentration in the gridblock with index (i, j). The values of pseudo-tracer concentration vary between zero and one, but in most of the drainage region they are either one or very close to one. Values lower than one are typically observed at the edges of the drainage region. In Eq. 3.9, $e(h)$ is the "ellipse" function, defined by

$$e(h) = \begin{cases} 0, & h > 1 \\ \\ 1, & h \leq 1 \end{cases}, \tag{3.10}$$

where

$$h = \sqrt{\left(\frac{\delta_{x'}}{L_1}\right)^2 + \left(\frac{\delta_{y'}}{L_2}\right)^2}. \tag{3.11}$$

In Eq. 3.11, L_1 and L_2 are the axes of the ellipse rotated by the angle θ (Fig. 3.5); $\delta_{x'}$ and $\delta_{y'}$ are the distances from the well location in the rotated directions x' and y', respectively. First, we compute the distances in the x- and y-directions and then compute the rotated distances using

$$\begin{bmatrix} \delta_{x'} \\ \delta_{y'} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix}. \tag{3.12}$$

In computing the axes of each ellipse, we assume that the rotation angle, θ, is the same one defined by the prior geological model so that the axes, L_1 and L_2, can be directly added to the prior correlation lengths to define the final localization region. Fig. 3.6 illustrates two "ellipses" obtained for the drainage region shown in Fig. 3.6a assuming $\theta = 0$ (Fig. 3.6b) and $\theta = 45^{\text{o}}$ (Fig. 3.6c).

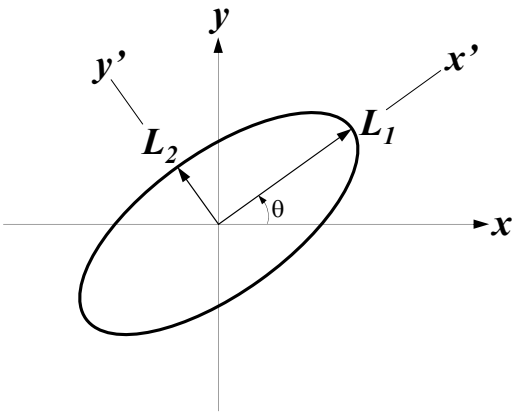
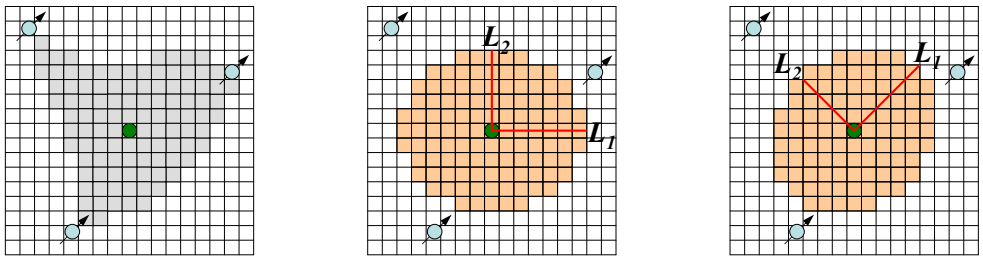


Figure 3.5: Ellipse defining anisotropic directions in the prior model.



(a) Drainage region

(b) Ellipse matched for $\theta = 0$

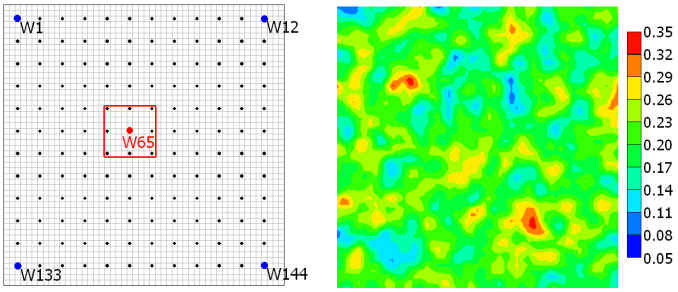
(c) Ellipse matched for $\theta = 45^o$

Figure 3.6: Illustration of ellipses generated from a drainage region with different anisotropic directions in the prior model.

3.2 Examples

3.2.1 Example 1 – Linear model

The first example is a 2D linear model on a 50×50 uniform grid. The model parameters are gridblock porosities. The measured data correspond to the average porosity of the 11×11 gridblocks centered on each well position as illustrated in Fig. 3.7a. The total number of wells is 144. At each data assimilation time-step, we assimilate the datum from one well, such that the first datum assimilated corresponds to well number 1, the second to well 2 and so on, until the assimilation of the datum from the 144th well. To generate the measurements, we added to the well data predicted by the true model Gaussian random noise with zero mean and standard deviation equal to 3% of the true data. We assumed that the measurement errors are uncorrelated, so that the covariance matrix C_D is diagonal. The true porosity field was generated using an isotropic exponential covariance function with a correlation length of 5 gridblocks, i.e., equal to $5\Delta x$ where Δx is the width of a gridblock. The prior mean porosity of every gridblock is 0.20 and the prior variance is 0.05^2. Fig. 3.7b shows the true porosity field.



(a) Grid with well locations

(b) True porosity field

Figure 3.7: Well locations and true model. In (a), the black dots represent the well locations and the red square illustrates the size of the sensitivity region for well W65.

The initial ensemble was generated using the same covariance function and prior mean specified above. We use EnKF to assimilate data with two ensemble sizes,

25 and 50, with and without covariance localization. For covariance localization, we consider three cases: (1) applying the Gaspari-Cohn correlation function with a critical length of 5 gridblocks, which is equal to the underlying correlation length of the prior geostatistical model from which the initial ensemble and the true model were generated. Note that 5 gridblocks also approximately corresponds to the distance of influence of the sensitivity; (2) the critical length of 10 gridblocks, which corresponds to the correlation length of the prior model plus the distance of influence of the sensitivity (our recommended procedure); (3) critical length of 15 gridblocks.

Figs. 3.8 and 3.9 show the ensemble mean after data assimilation for ensemble sizes of 25 and 50, respectively. For comparison, the MAP estimate and the EnKF ensemble mean obtained with an ensemble of 1,000 models are also presented. Note that with this large ensemble, the EnKF ensemble mean and the MAP estimate share many features. Also, note that for this case, the MAP estimate corresponds to the correct posterior mean and the ensemble mean from EnKF will converge to the MAP estimate when $N_e \to \infty$. Compared to the MAP estimate, the ensemble mean from EnKF with $N_e = 25$ and $N_e = 50$ without localization are very rough. The ensemble means from the localization cases, on the other hand, are smoother and closer to the MAP estimate. Each of the three localization procedures yields an estimate (ensemble mean) of the porosity field that shows the main features of the MAP estimate, even for an ensemble of 25 models.

In order to have statistically more significant results, we repeated the data assimilation with 10 different initial ensembles. Table 3.1 presents the mean and standard deviation of the objective function for each case. The objective functions were evaluated based on the ensemble mean model. For this example, we have $N_d = 144$ observed data points. Hence, according to the quality criterion of Eq. 2.130, the normalized objective function (Eq. 2.131) should be less than 1.589. According to Table 3.1, most of the EnKF results do not satisfy this quality criterion. Except

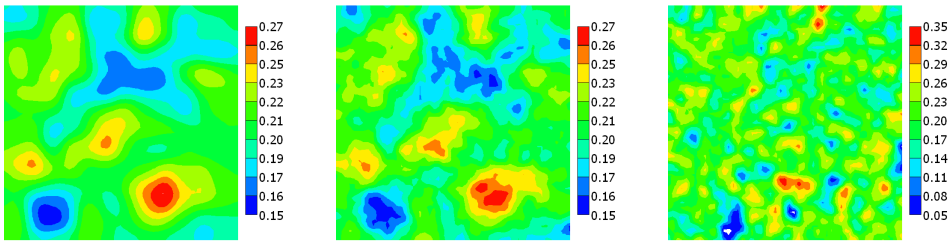
Table 3.1: Objective function. For EnKF cases with ensembles of 25 and 50 models, this table presents the mean and standard deviation for 10 different initial ensembles.

Case	$O_m(m)$		$O_d(m)$		$O_N = \frac{2O(m)}{N_d}$	
	mean	std. dev.	mean	std. dev.	mean	std. dev.
$N_e = 25$, no local.	1310.4	318.4	228.0	71.7	21.366	4.768
$N_e = 25$, local. 1	105.3	5.1	46.8	2.7	2.112	0.084
$N_e = 25$, local. 2	187.8	11.5	24.5	2.6	2.944	0.154
$N_e = 25$, local. 3	293.8	20.8	24.5	3.9	4.412	0.268
$N_e = 50$, no local.	1359.9	375.6	102.2	29.7	20.308	5.532
$N_e = 50$, local. 1	56.7	2.0	44.7	2.8	1.408	0.054
$N_e = 50$, local. 2	98.4	3.5	20.4	1.8	1.650	0.060
$N_e = 50$, local. 3	150.1	5.6	17.2	1.8	2.324	0.084
MAP	19.8	–	7.6	–	0.380	–
$N_e = 1{,}000$	82.1	–	9.0	–	1.265	–

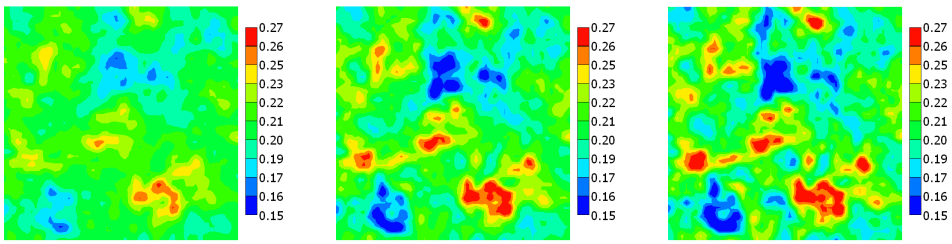
for the ensemble of 1,000 models, EnKF cases without localization gave very high values of the objective function. Cases with localization, on the other hand, resulted in values of the objective function that come close to satisfying the quality criterion. Besides the normalized objective function, in Table 3.1, we present the mean values of the two parts of the objective function, O_m and O_d. EnKF cases without localization resulted in large values of O_m, meaning very rough models. Besides that, these cases resulted in larger values of O_d, indicating poor data matches. Localization improved both parts of the objective function. Note that as the specified critical length increased, EnKF resulted in rougher models, but with better data matches.

Figs. 3.10 and 3.11 present the final posterior covariance between the center gridblock and gridblocks on a horizontal line passing through the center gridblock. For EnKF cases, the posterior covariances were estimated using the standard approximation, i.e.,

$$\widetilde{C}^a_{\mathrm{M}} = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (m_j^a - \overline{m}^a)(m_j^a - \overline{m}^a)^{\mathrm{T}}. \tag{3.13}$$

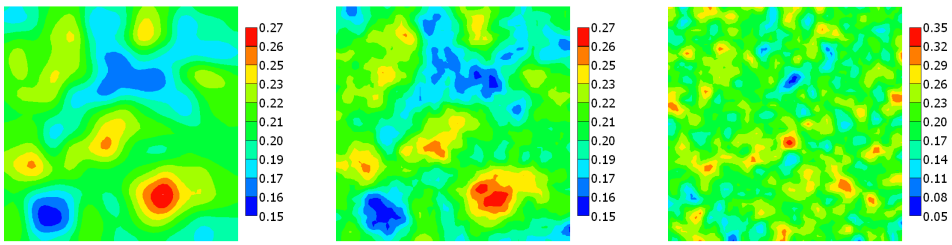


(a) MAP **(b)** $N_e = 1{,}000$ **(c)** No localization

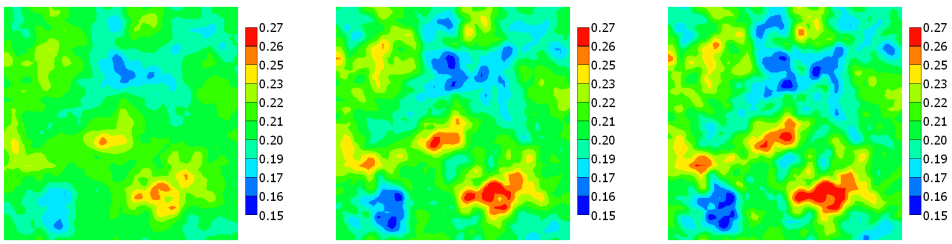


(d) Localization case 1 (5 gridblocks) **(e)** Localization case 2 (10 gridblocks) **(f)** Localization case 3 (15 gridblocks)

Figure 3.8: Ensemble mean after data assimilation. Ensemble size 25.



(a) MAP **(b)** $N_e = 1{,}000$ **(c)** No localization



(d) Localization case 1 (5 gridblocks) **(e)** Localization case 2 (10 gridblocks) **(f)** Localization case 3 (15 gridblocks)

Figure 3.9: Ensemble mean after data assimilation. Ensemble size 50.

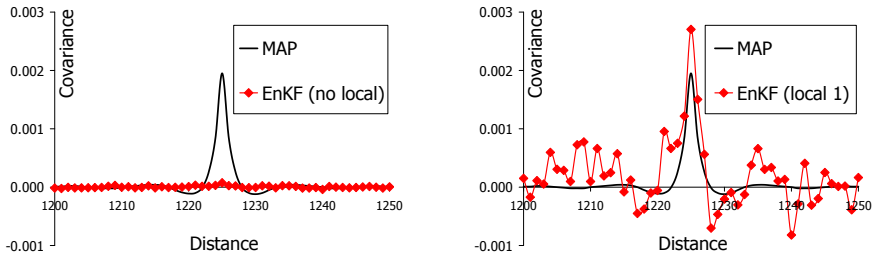
Note that these posterior covariances correspond to the final ensemble of models, and they are not multiplied by the localization correlation matrices. For comparison, we computed the correct posterior covariance using Eq. 2.14. According to Figs. 3.10a and 3.11a, EnKF without localization clearly results in a significant underestimation

of the posterior covariances, i.e., EnKF overestimates the reduction in uncertainty obtained by the assimilation of data. Note that for the ensemble of 25 models, the posterior covariances are very close to zero, which indicates that the final ensemble has almost collapsed to a single model. Localization resulted in better estimates of the posterior covariances. Note that localization case 2 (our recommended procedure) gave posterior covariances very close to the correct ones in the vicinity of the central gridblock. A smaller critical length for localization tends to result in an overestimation of the posterior covariance while a larger length tends to give an underestimation of the covariances. As observed by other investigators [80, 71, 57], as we increase the size of the ensemble, we may need to increase the critical length in the localization function in order to obtain optimal results. We observe this behavior in this example; i.e., using a critical length of 15 gridblocks, we approximate very well the posterior covariance for the ensemble size of 50, while for the ensemble size of 25, we underestimate it. However, even for an ensemble of 50 models, a critical length defined by prior plus sensitivity (10 gridblocks) provides a good approximation of the posterior covariance.

Furrer and Bengtsson [57] derived a relationship between a nearly optimal localization function and the true covariance. In this paper, they compared the true covariance matrix, C_{M}, with a covariance matrix $\widetilde{C}_{\mathrm{M}}$ generated from an ensemble using the expectation of the norm of the difference between these two matrices defined by

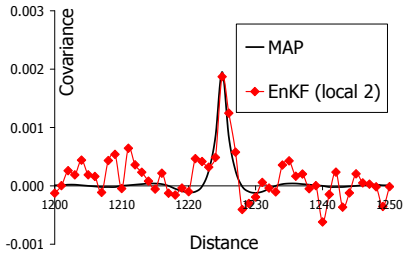
$$\mathrm{E}\left[\|C_{\mathrm{M}} - \widetilde{C}_{\mathrm{M}}\|^2\right] = \mathrm{E}\left[\mathrm{trace}\left((C_{\mathrm{M}} - \widetilde{C}_{\mathrm{M}})^2\right)\right]. \tag{3.14}$$

According to this definition, the smaller the norm of the difference, the closer $\widetilde{C}_{\mathrm{M}}$ is to the true covariance. Using the same metric, we computed the norm of the prior and posterior covariances with and without localization. Table 3.2 presents the average values of the expectations of the norms based on 10 different ensembles

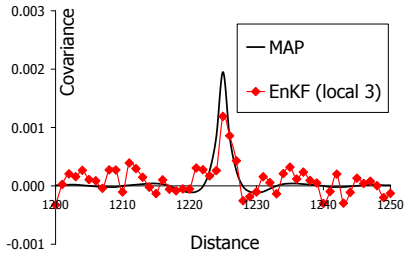


(a) No localization

(b) Localization case 1 (5 gridblocks)

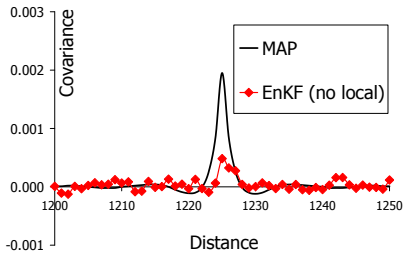


(c) Localization case 2 (10 gridblocks)

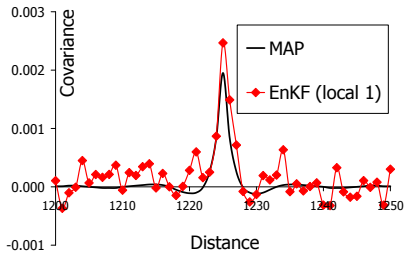


(d) Localization case 3 (15 gridblocks)

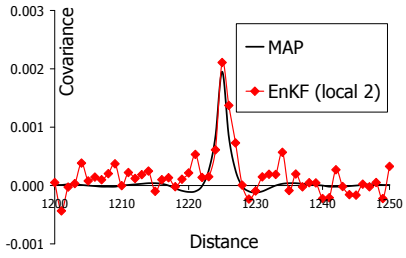
Figure 3.10: Posterior covariances for the center gridblock of the model. Ensemble of 25 models.



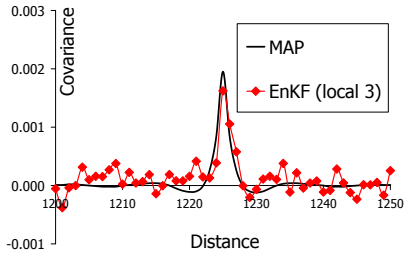
(a) No localization



(b) Localization case 1 (5 gridblocks)



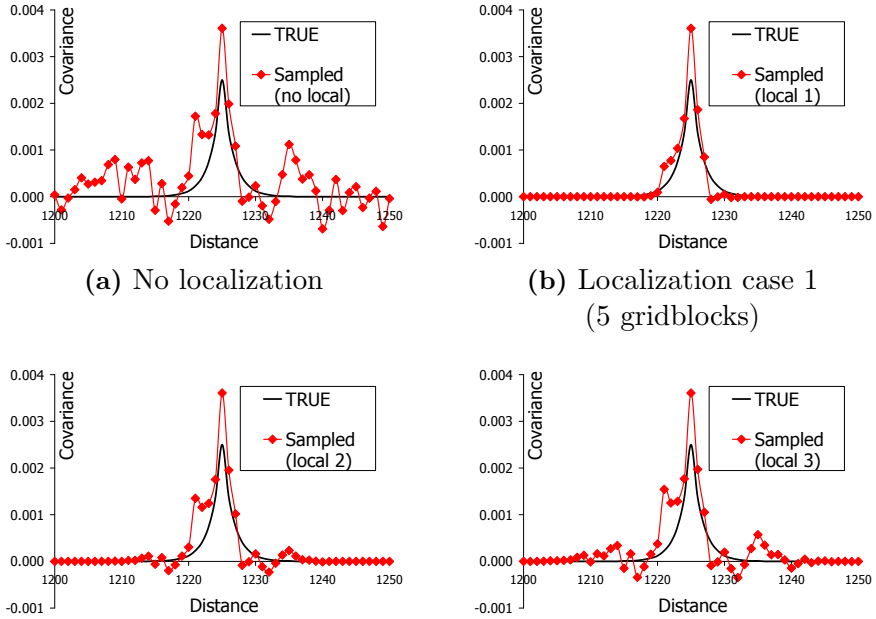
(c) Localization case 2 (10 gridblocks)



(d) Localization case 3 (15 gridblocks)

Figure 3.11: Posterior covariances for the center gridblock of the model. Ensemble of 50 models.

considering the prior, C_{M}^{f}, and the posterior, C_{M}^{a}, covariance matrices. The results in Table 3.2 indicate that the best approximation of the true prior covariance matrix is for the case with localization using a critical length of 5 gridblocks (localization case 1), because it yields the lowest value for the norm $\|C_{\mathrm{M}}^{f} - \widetilde{C}_{\mathrm{M}}^{f}\|^2$. Fig. 3.12 presents the prior covariances for the first ensemble between the center gridblock of the model and gridblocks in a horizontal line passing through the center gridblock. The results presented in this figure agree with the results of Table 3.2, i.e., the best approximation of the prior covariance is for the case with localization with 5 gridblocks.



(a) No localization

(b) Localization case 1 (5 gridblocks)

(c) Localization case 2 (10 gridblocks)

(d) Localization case 3 (15 gridblocks)

Figure 3.12: Prior covariances. Ensemble of 25 models.

Using a matrix norm for comparing the posterior covariances may not be definitive. For example, consider the results of Fig. 3.10. If we define the best result based on the norm of the difference (third column of Table 3.2), as we did for the prior, the best approximation would be the case without localization. However, as illustrated in Fig. 3.10a, the posterior covariances for the case without localization are almost zero everywhere. As the true values are also close to zero except in a small

region around the central gridblock, at most gridblocks, EnKF without localizaiton gives a good approximation of the true covariance. In fact, however, the EnKF-generated ensemble almost collapsed and we obtained a poor data match as well as an overly rough approximation of the MAP estimate. The posterior covariances obtained with localization (Figs. 3.10b, 3.10c and 3.10d) give better approximations of the correct covariances near the central gridblock, even though they are noisy. Thus, the matrix norm of Eq. 3.14 may not be a good metric for comparing the posterior covariances. The results of Fig. 3.10 and Table 3.1 suggest that the best results were obtained by using a critical correlation length equal to the prior correlation length plus the sensitivity range.

In fact, in order to obtain a good estimate of the posterior mean and posterior covariance with EnKF, we need to have good approximations for the products $C_{\mathrm{M}}^{f}G^{\mathrm{T}}$ and $GC_{\mathrm{M}}^{f}G^{\mathrm{T}}$. If an exact representation of the sensitivity matrix G is available, we can improve the estimate of $C_{\mathrm{M}}^{f}G^{\mathrm{T}}$ by improving the estimate of the prior covariance C_{M}^{f}, which, for example 1, could be done by selecting the prior correlation length (5 gridblocks) for localization. However, this is not true whenever we use sample approximations. In this case, the spurious correlations are distributed over the product $C_{\mathrm{M}}^{f}G^{\mathrm{T}}$. Therefore, a proper selection of correlation length(s) for localization should consider both the principal correlation ranges of the initial prior covariance and the range of sensitivities reflected in the sensitivity matrix.

Table 3.2: Expectation of the norm of the difference between the true and sampled covariances.

Case	$\mathrm{E}\left[\|C_{\mathrm{M}}^{f}-\widetilde{C}_{\mathrm{M}}^{f}\|^{2}\right]$	$\mathrm{E}\left[\|C_{\mathrm{M}}^{a}-\widetilde{C}_{\mathrm{M}}^{a}\|^{2}\right]$
$N_e = 25$, no local.	23.61	4.67
$N_e = 25$, local. 1	3.82	17.93
$N_e = 25$, local. 2	6.24	13.33
$N_e = 25$, local. 3	8.46	9.39

3.2.2 Example 2 – Single phase pressure data

The second example corresponds to a single-phase flow in a one-dimensional horizontal reservoir partitioned into 300 gridblocks for the purpose of numerical simulation. Only one well is completed in the reservoir, a producing well in the 300th gridblock as illustrated in Fig. 3.13. The porosity is assumed constant and equal to 0.2. The well is operated at a specified oil rate, and the observed data are bottomhole pressures. The production history consists of 246 days with alternating periods of flow and build-up as illustrated in Fig. 3.14. Independent Gaussian random noise with zero mean and standard deviation of 5 psi was added to the data generated from the simulator with the true permeability field in order to generate the observed data for history matching. We assumed that the measurement errors are uncorrelated, so that the covariance matrix C_{D} is diagonal.

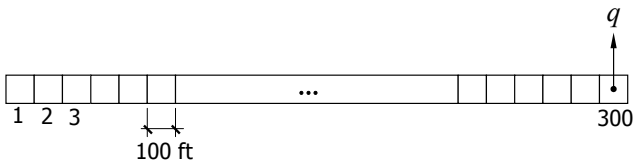


Figure 3.13: Grid of example 2.

The model parameters are the gridblock log-permeabilities. We generated the true model and the initial ensemble of log-permeability using an exponential correlation function with prior mean of 5.0, variance of 1.0 and correlation length of

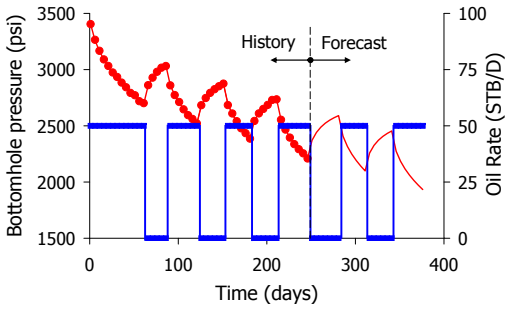


Figure 3.14: Well data. Red dots are the history (bottomhole pressure measurements), solid red curve is the bottomhole pressure and solid blue curve is the oil flow rate obtained by running the true model.

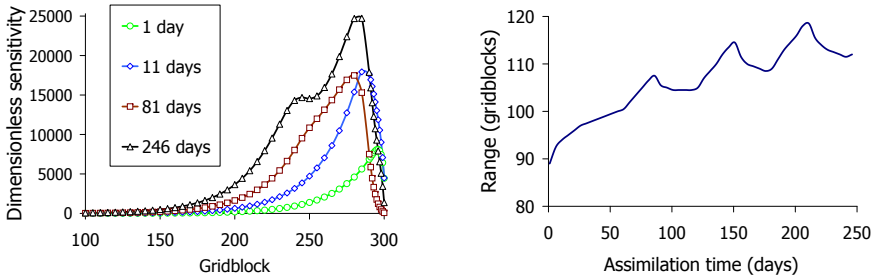
50 gridblocks (5,000 ft). We assimilate data using EnKF with and without covariance localization. For covariance localization, we use the Gaspari-Cohn correlation function with two different critical lengths: (1) 50 gridblocks, which corresponds to the prior correlation length of the true model; (2) 50 gridblocks (from prior correlation) plus the "range" of the dimensionless sensitivity at the MAP estimate. The dimensionless sensitivity matrix, G_{D}, is defined by [192]

$$G_{\mathrm{D}} = C_{\mathrm{D}}^{-1/2} G C_{\mathrm{M}}^{1/2}, \tag{3.15}$$

where G is the sensitivity matrix evaluated at some particular m, $C_{\mathrm{M}}^{1/2}$ denotes the square root of the prior model covariance matrix and $C_{\mathrm{D}}^{-1/2}$ denotes the inverse of the square root of data error covariance matrix. We define the "range" of the dimensionless sensitivity at each data assimilation time as the distance from the well to the closest gridblock where the sensitivity of predicted pressure to the gridblock log-permeability is less than 0.1% of the maximum sensitivity. Fig. 3.15a shows the dimensionless sensitivity obtained for the MAP estimate at different times. The sensitivities were calculated using the adjoint method. As we can see in Fig. 3.15a, the range of this sensitivity increases as the time increases. Fig. 3.15b shows the ranges obtained based on the dimensionless sensitivity in terms of the number of gridblocks. Interestingly, the sensitivity range during a build-up period is slightly higher than the range during the preceding flow period.

Fig. 3.16a shows the true model and the MAP estimate. The observed data can resolve quite well the model parameters (log-permeability) up to approximately 100 gridblocks from the producing well. At distances more than 150 gridblocks away from this well, the estimated log-permeabilities are equal to the prior mean. Fig. 3.16b compares the MAP estimate and the ensemble mean from EnKF with $N_e = 999$. For this ensemble size, EnKF provides a good estimate of the MAP.

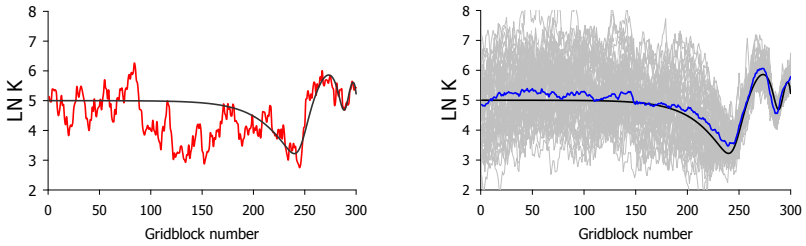
Figs. 3.17 and 3.18 present the final ensemble mean from EnKF with ensemble



(a) Dimensionless sensitivity obtained for the MAP

(b) Sensitivity range used for covariance localization

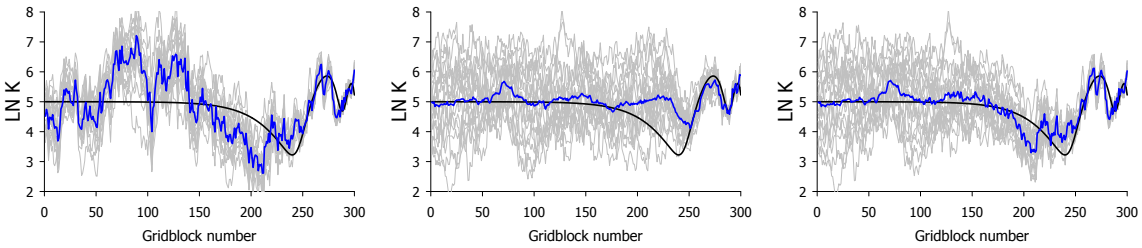
Figure 3.15: Sensitivity range used for covariance localization.



(a) True model and MAP

(b) EnKF, $N_e = 999$

Figure 3.16: Log-permeability fields. In red, the true model; in black, the MAP estimate; in blue, the ensemble mean and in gray the final models from EnKF with $N_e = 999$.

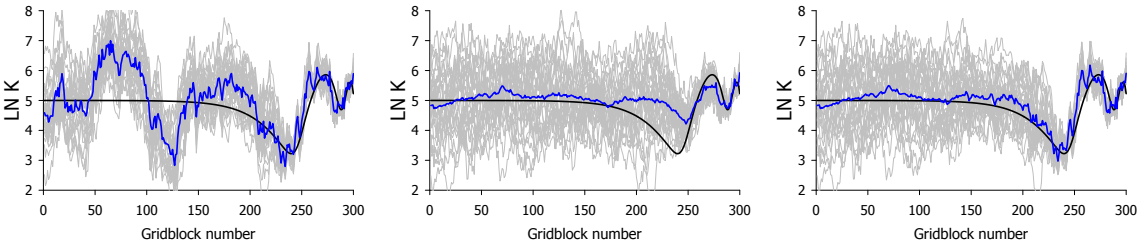


(a) No localization

(b) Local. 1 (prior)

(c) Local. 2 (prior + sens.)

Figure 3.17: Log-permeability fields. In black, the MAP estimate; in blue, the ensemble mean and in gray the final models from EnKF cases. $N_e = 25$.



(a) No localization **(b)** Local. 1 (prior) **(c)** Local. 2 (prior + sens.)

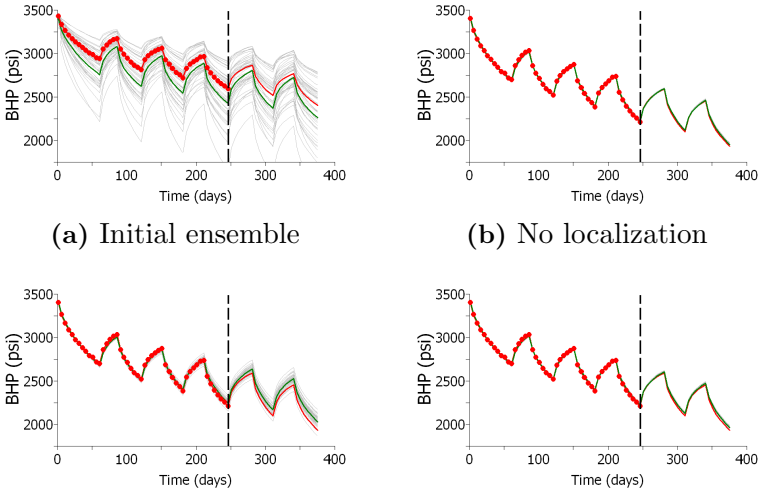
Figure 3.18: Log-permeability fields. In black, the MAP estimate; in blue, the ensemble mean and in gray the final models from EnKF cases. $N_e = 50$.

sizes of 25 and 50, respectively. An ensemble size of 50 is relatively large if we consider that the number of model parameters is only 300. For cases without localization, EnKF gave very rough estimates of the log-permeability field with values far from the prior mean at locations farther than 150 gridblocks from the producing well. A critical length of 50 gridblocks for localization is clearly too small, while a critical length defined by the sensitivity range plus the prior correlation length gave an ensemble mean which is in reasonable agreement with the MAP estimate. Note that for the ensemble of 50 models, by doing localization, we obtained an estimate comparable to the one obtained using an ensemble of 999 models (compare Figs. 3.16b and 3.18c).

Fig. 3.19 presents the bottomhole pressure obtained from the true model and from EnKF cases with $N_e = 50$. For EnKF the results presented in Fig. 3.19 were obtained by running the simulator from time zero using the final ensemble. In all cases, a good data-match was obtained. However, for the case with localization using the critical length of 50 gridblocks, the forecasts ($t > 246$ days) deviate slightly from the truth because this critical length restricted the model changes to a too small region around the well.

3.2.3 Example 3 – Synthetic reservoir

The third example corresponds to a 2D reservoir model on a 60×60 uniform



(a) Initial ensemble **(b)** No localization

(c) Local. 1 (prior) **(d)** Local. 2 (prior + sens.)

Figure 3.19: Bottomhole pressure (history and forecast) in psi. Red dots are the history; red solid curve represents the true model; solid green curve is the ensemble mean and solid gray curves represent all models of the ensemble. The dashed vertical black line indicates the end of the history and beginning of forecast. $N_e = 50$.

grid. The model parameters are gridblock log-permeabilities. The true model was generated from an anisotropic exponential correlation function with major correlation length of 3,000 ft (which corresponds to the width of 30 gridblocks) and minor correlation length of 1,000 ft oriented at 45°. The prior mean is 5.0 and the prior variance is 1.0 for all gridblock log-permeabilities. Fig. 3.20 shows the true permeability field used as the reference to generate the observed data. The observed data corresponds to oil and water production rates and water injection rates. To generate the observed data, we added to the well data predicted by the true model Gaussian random noise with zero mean and standard deviation equal to 5% of the true data. We assumed that the measurement errors are uncorrelated, so that the covariance matrix C_{D} is diagonal. For this problem, 3 producing wells, P1, P2, P3, and one injection well, I1, start producing or injecting water at the beginning of the simulation time. Another well, P4, starts production at 6,000 days and the other two wells, P5 and P6, start production after 9,000 days. The total production time is 12,000 days. The observed data are oil and water production rates and water

injection rates. During simulations, all wells are controlled by specified bottomhole pressure, 1,500 psi for the producers and 4,000 psi for the water injector well. This situation in which we introduce new measurements at different locations after some period of data assimilation is potentially unfavorable for EnKF. We expect that spurious correlations will result in changes in model parameters pertaining to gridblocks far from the observation locations. When a measurement at a new location is introduced, the variability in the model parameters at this gridblock and nearby gridblocks may be too small to allow for the proper changes necessary to assimilate these measurements. Thus, this is an example in which localization should be beneficial and it is also a situation that occurs very often in practice.

The EnKF was applied using an ensemble of 100 models considering four cases: (1) no localization; (2) localization based on the drainage areas defined by the pseudo-tracer concentration calculated from the method proposed by Damiani [28]; (3) localization using the Gaspari-Cohn correlation function with critical length equal to the prior model correlation lengths (3,000 ft in the principal direction and 1,000 ft in the orthogonal direction with anisotropy angle of 45°); (4) localization using the Gaspari-Cohn correlation function with critical lengths for the localization function equal to the prior model principal correlation lengths plus the radius of the circle with area equal to the drainage region calculated with the pseudo-tracer (our recommended procedure). For comparisons, the results from EnKF with an ensemble of 1,000 models without localization and the MAP estimate are also presented. For the EnKF case 2, the values of the pseudo-tracer concentration were used as the entries of the localization matrix, ρ^n. These values are close to one inside the drainage area and zero outside this area. Fig. 3.21 shows the estimated drainage areas at 12,000 days based on the true model. These areas change with time; for example, a big change is observed when a new producing well is opened. The drainage areas used in the EnKF process were calculated based on the ensemble mean model, so that a

different localization was applied at each data assimilation time. Fig. 3.22 illustrates the Gaspari-Cohn correlation function values using the prior model's correlation lengths (case 3) and the prior correlation lengths plus drainage radius at 12,000 days (case 4), respectively.

Figs. 3.23 and 3.24. compare the sensitivities and drainage areas for wells P1 and P4. In these figures, both sensitivities and drainage areas were calculated based on the true permeability field. The sensitivities were calculated using the adjoint method. Clearly, the drainage areas do not correspond exactly to the sensitivities, but a well's drainage area describes fairly well the region around the well with high sensitivity values, which we believe is sufficient for the purpose of doing localization. In Figs. 3.23 and 3.24, we also observe high sensitivity at other well locations. These high sensitivities indicate that oil and water rate at a particular well are also sensitive to log-permeability at gridblocks where other wells are completed. For this example, as producing wells operate at a fixed bottomhole pressure, increasing the permeability in a gridblock containing a well increases the flow rates at this well, and thus by material balance, causes a decrease in the flow rates at other producing wells. With the proposed localization procedure, depicted in Fig. 3.22b, the permeability at most other gridblocks containing producing wells will not be updated by assimilating oil and water flow rate data at well P2. Although strictly speaking the failure to update permeabilities near other producing wells is not correct, increasing the critical lengths in our localization procedure so that these permeabilities would be updated would effectively eliminate localization and thus reintroduce the problem of spurious correlations.

Fig. 3.25 presents the ensemble mean permeability fields for all EnKF cases considered. Note that the permeability field obtained by EnKF without localization and $N_e = 100$ displays several regions with overshooting an undershooting of the permeability. In some gridblocks, permeability values greater than 35,000 mD were

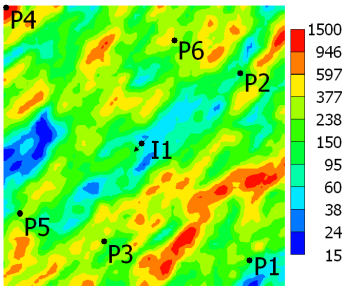


Figure 3.20: True permeability field (mD).

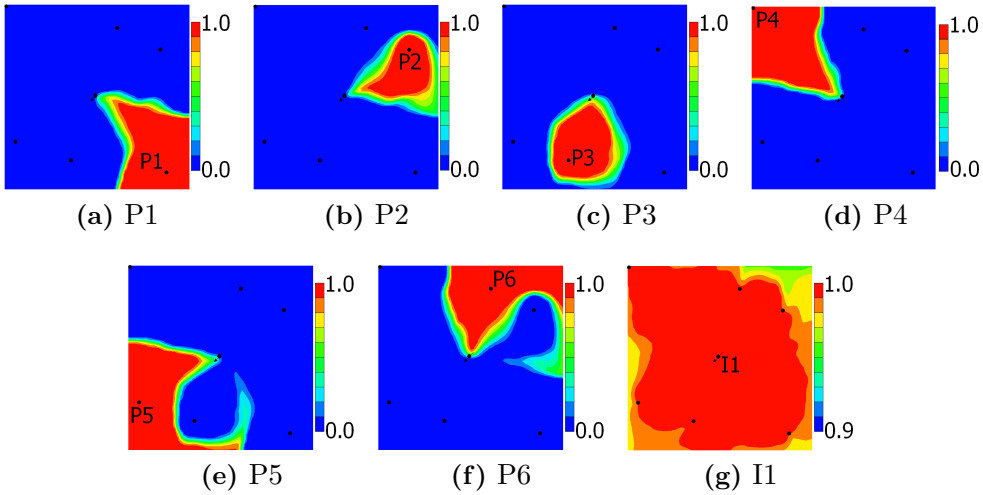
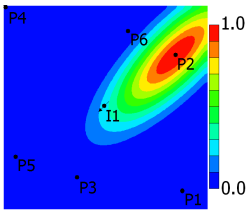


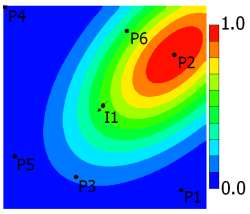
Figure 3.21: Drainage areas obtained using the pseudo-tracer concentration at 12,000 days for the true model.

obtained. Localization improved the estimate of the permeability field and gave smoother models. Figs. 3.26–3.31 present the results in term of the predicted oil and water rates for wells P1 and P4. Results shown in these figures were obtained by running the final ensembles from time zero. EnKF with $N_e = 100$ without localization did not give a reasonable data match of either oil or water rate at either well. Besides that, all models of the final ensemble gave almost the same prediction indicating a significant loss of variability in the ensemble.

Table 3.3 shows the average objective function obtained by each case. For this example, we have 651 observed data points. Thus, according to the quality criterion of Eq. 2.132, the final normalized objective function should not exceed 1.277. However, according to Table 3.3 none of the cases, including the MAP estimate itself,

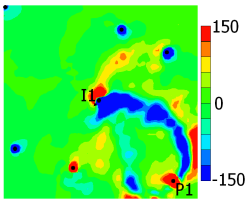


(a) Case 3 (prior)

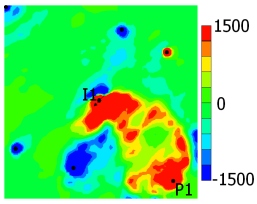


(b) Case 4 (prior + drainage radius)

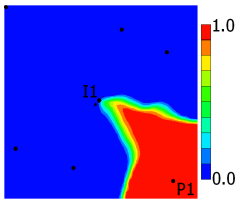
Figure 3.22: Correlations values for localization in cases 3 and 4 at well P2. For case 4 this figure illustrates the localization at 12,000 days.



(a) Oil rate

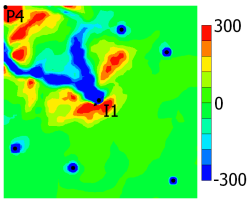


(b) Water rate

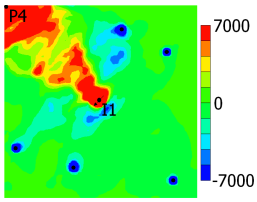


(c) Drainage area

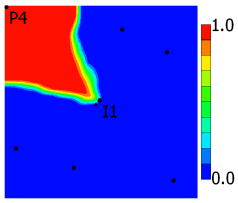
Figure 3.23: Comparison between sensitivities of oil and water rates with respect to log-permeability and drainage area for well P1 at 12,000 days.



(a) Oil rate



(b) Water rate



(c) Drainage area

Figure 3.24: Comparison between sensitivities of oil and water rates with respect to log-permeability and drainage area for well P4 at 12,000 days.

satisfy this quality criterion, which is strictly valid only for the linear case. The EnKF results violate the 1.277 quality criterion value by more than a factor of ten in all cases, and the violation is mainly because of the data mismatch part of the objective function, O_d. Except for the case of localization using only the prior correlation length, localization improved the data match compared to the no localization case. The third localization case (prior principal correlation lengths plus sensitivity radius), performed better in terms of data match than all other EnKF cases, including the case with $N_e = 1{,}000$.

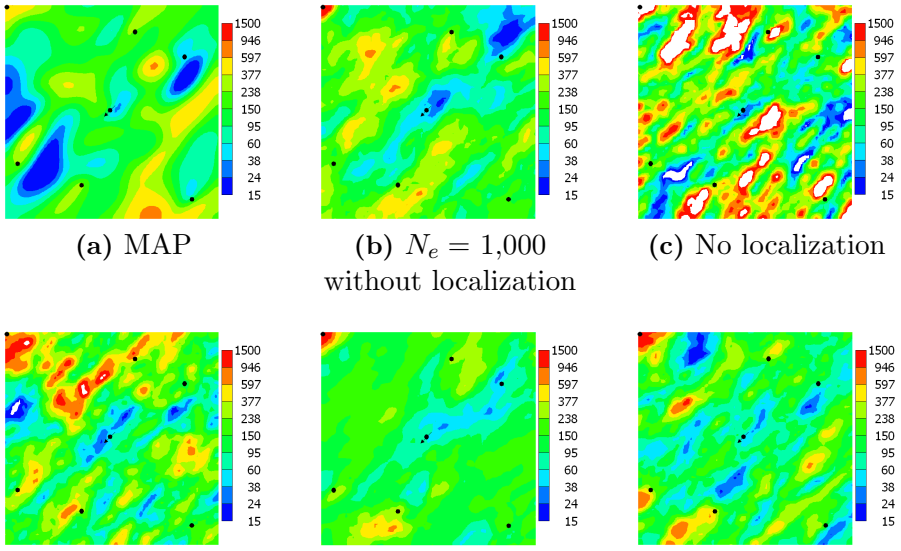
From the results of Table 3.3, it is clear that localization using only the

drainage areas gave inferior results to those obtained from EnKF with critical lengths based on the sum of the prior principal correlation lengths and the sensitivity range estimated from the drainage areas. Assuming that the drainage areas are somewhat related to the sensitivities, when we keep only the drainage areas for localization, we are selecting too small a region, unless all model parameters are completely uncorrelated in the prior model. A second argument for using a correlation matrix instead of drainage areas is that the value of the correlation function smoothly decreases to zero as the distance from the well location increases; this smooth variation tends to produce smoother model changes. The drainage areas, on the other hand, have sharp boundaries, and as we move across a boundary, the concentration calculated from the pseudo-tracer changes abruptly from one to zero, which can even introduce discontinuities in the estimated rock properties. A third argument for using a correlation function is that if we take sufficient care to ensure the correlation function ρ is positive definite, then we can obtain the maximum possible degrees of freedom when assimilating data.

It is important to observe that in this example, the drainage areas were calculated based on the ensemble mean model. Arroyo-Negrete et al. [10] and Devegowda et al. [31], on the other hand, have used the region of influence defined by stacking regions calculated for each model of the ensemble. Considering that each model may have different streamlines, the final region of influence based on their procedure may be larger than the one that would be obtained from the mean model. This fact could possibly alleviate problems that can arise by not considering the correlation lengths of the prior geological model when applying localization.

Table 3.3: Average objective function.

Case	$O_m(m)$	$O_d(m)$	$O_N = \frac{O(m)}{N_d}$
(1) No localization	4,833	11,467	25.04
(2) Local. using drainage area	2,613	10,803	20.61
(3) Local. using prior corr.	1,584	13,844	23.70
(4) Local. using prior plus drainage radius	2,016	5,955	12.25
(5) No local. $N_e = 1{,}000$	1,912	8,988	16.75
(6) MAP	73	1,369	4.43 [1]



(a) MAP

(b) $N_e = 1{,}000$ without localization

(c) No localization

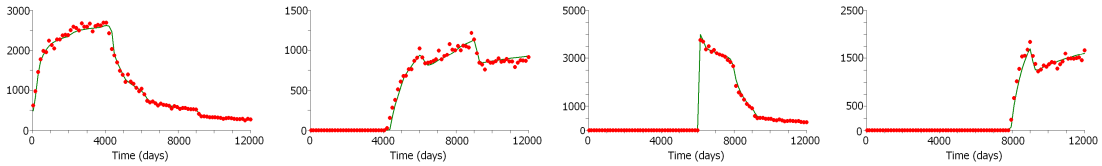
(d) Localization using drainage areas

(e) Localization using prior correlation length

(f) Localization using prior plus drainage area radius

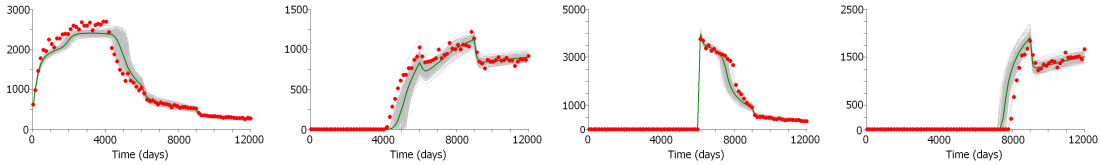
Figure 3.25: Final mean permeability fields (mD). Blank areas represent overshooting or undershooting on the permeability values. In cases (c)–(f), $N_e = 100$.

[1] For the MAP estimate, we use $O_N(m_{\text{MAP}}) = 2O(m_{\text{MAP}})/N_d$.



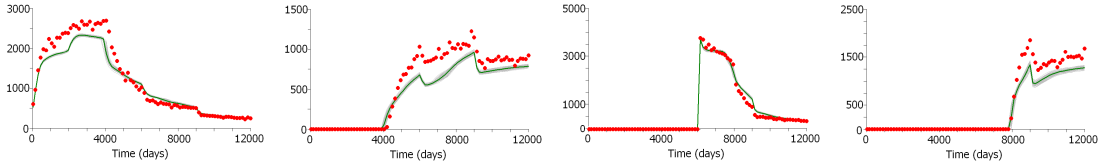
(a) P1 – Oil rate **(b)** P1 – Water rate **(c)** P4 – Oil rate **(d)** P4 – Water rate

Figure 3.26: Predicted well data from the MAP estimate. Red dots are the history, solid green curve is the MAP estimate.



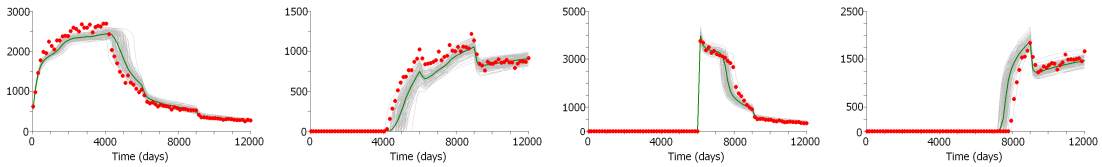
(a) P1 – Oil rate **(b)** P1 – Water rate **(c)** P4 – Oil rate **(d)** P4 – Water rate

Figure 3.27: Predicted well data running final ensemble from time zero (stb/day). N_e = 1,000 without localization. Red dots are the history; solid green curve is the ensemble mean and solid gray curves represent all models of the ensemble.



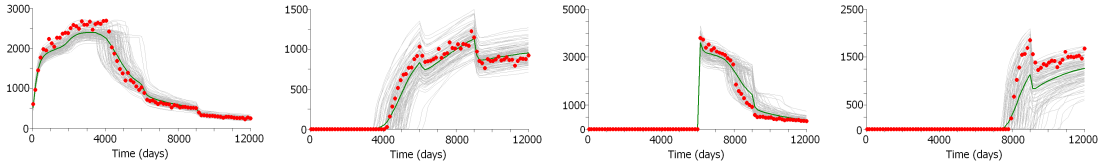
(a) P1 – Oil rate **(b)** P1 – Water rate **(c)** P4 – Oil rate **(d)** P4 – Water rate

Figure 3.28: Predicted well data running final ensemble from time zero (stb/day). No localization. N_e = 100. The colors in this figure have the same meaning as in Fig. 3.27.



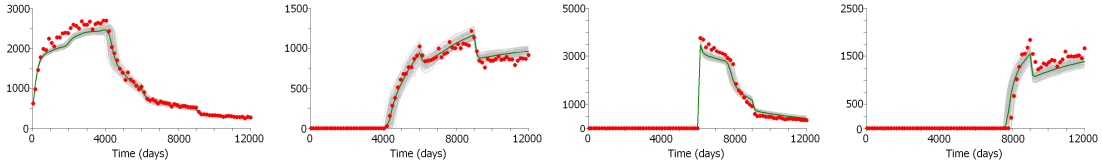
(a) P1 – Oil rate **(b)** P1 – Water rate **(c)** P4 – Oil rate **(d)** P4 – Water rate

Figure 3.29: Predicted well data running final ensemble from time zero (stb/day). Localization using drainage areas. N_e = 100. The colors in this figure have the same meaning as in Fig. 3.27.



(a) P1 – Oil rate **(b)** P1 – Water rate **(c)** P4 – Oil rate **(d)** P4 – Water rate

Figure 3.30: Predicted well data running final ensemble from time zero (stb/day). Localization using prior correlation length. $N_e = 100$. The colors in this figure have the same meaning as in Fig. 3.27.



(a) P1 – Oil rate **(b)** P1 – Water rate **(c)** P4 – Oil rate **(d)** P4 – Water rate

Figure 3.31: Predicted well data running final ensemble from time zero (stb/day). Localization using prior plus drainage area radius. $N_e = 100$. The colors in this figure have the same meaning as in Fig. 3.27.

3.3 Field Case

3.3.1 Reservoir model

The field case corresponds to a turbidite heavy-oil field in the Campos Basin in water depths between 600 and 1,100 meters. To this date, this is one of the largest producing fields in Brazil. Reservoir facies consist of amalgamated graded beds of poorly consolidated sandstone with very low silt and clay content and high porosity and permeability [89]. In this field, there are two turbidite systems with depositional directions as illustrated in Fig. 3.32. According to Lorenzatto et al. [115], "because of the reservoir continuity, the relative scarcity of gas, the favorable characteristics of the relative permeability curves and the deepwater location, the recovery method employed in this field is water injection in the oil zone." More details about this field can be found in Bruhn et al. [17], Lorenzatto et al. [115] and Oliveira et al. [129].

The engineers of the asset team modified the permeability field of the reservoir simulation model to match the production history. In this manual history-matching

process, the reservoir simulator was run with each producing well constrained to its historical oil rate and each injector well constrained to its historical water injection rate. For most of the time, wells produce with the original gas-oil ratio and the main objective of the history matching is to match water rate at producing wells. Although this manually history-matched model gives a reasonable data match of water production rate, the permeability field is very artificial, with several "patches" and a few artificial high permeability streaks connecting injectors to producing wells, as illustrated in Fig. 3.33.

The original reservoir simulation model has $200 \times 240 \times 14$ gridblocks. Currently, there are 136 wells operating in this field. However, we decided to limit the scope of this study to only the upper production zone of the field, which corresponds to turbidite system 2 (Fig. 3.32). In this zone there are 17 producing wells and 12 water injectors. We extracted a submodel with $165 \times 86 \times 4$ gridblocks where 20,258 are active. This zone is hydraulically isolated from the rest of the field. However, there are two wells (P-147 and P-150) with commingled production from both the upper and the lower zones. For these two wells, we divided the historical oil production rates between the two zones based on the layer rates obtained from the manually history-matched model. We included the production rate of these two wells in the submodel, but we did not match data (water production rate) from these two wells. We considered only the turbidite system 2 because the primary objectives of this application were to illustrate that a proper implementation of EnKF could yield a reasonable history matching that was consistent with known geology and could "automatically" produce a data match as good as or superior to the one obtained from a tedious manual history matching exercise. Secondly, we wanted to compare EnKF with and without localization and consider the usefulness of HI-EnKF. As we envisioned that we would have to make several runs to do all the comparisons, we wanted to limit the size of the problem, which was easy to do since the two turbidite

systems are hydraulically isolated.

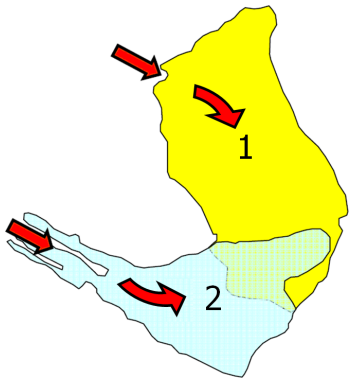


Figure 3.32: Turbidite systems 1 and 2. The red arrows indicate the main depositional direction of each system [129].

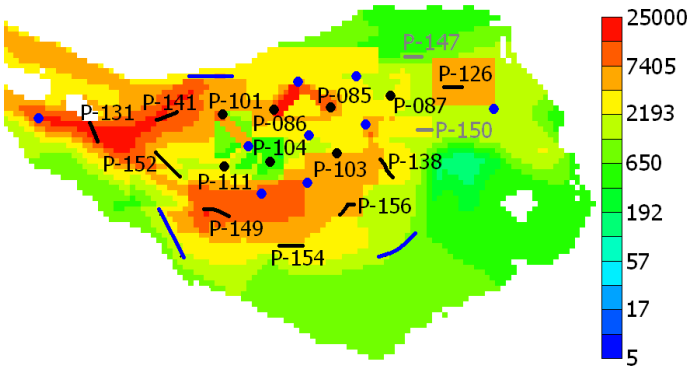


Figure 3.33: Permeability field (mD) of the manually history-matched model (layer 4). This figure indicates the names of the producing wells with a history of water production. The producing wells are shown in black and the water injector wells in blue. Dots represent vertical wells and lines represent horizontal wells. Data from producers P-147 and P-150 (in gray) were not history-matched.

3.3.2 History matching with EnKF

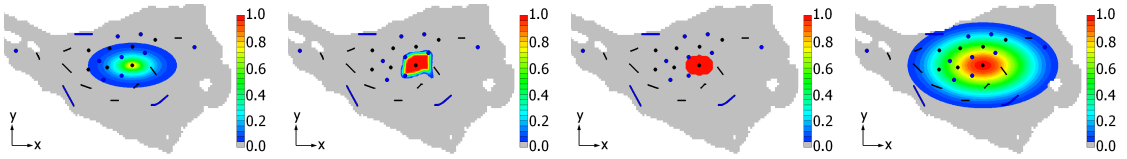
Typically, an ensemble size of the order of 100 has been used in reservoir applications [74, 14, 156]. However, there is no guarantee that 100 realizations are always adequate to avoid problems with insufficient degrees of freedom and spurious correlations, especially when we do not apply covariance localization. Here, we chose an ensemble size of 200 for applying EnKF with and without covariance localization. Although we did not experiment different ensemble sizes, 200 seemed to be a rea-

sonable choice for comparing EnKF with and without covariance localization in this field case. As will be shown, with this ensemble size, we were able to match data with and without localization, but we still observed problems related to spurious correlations in EnKF without localization, namely, significant loss of variance in the final ensemble and a rough final ensemble mean.

In order to make the results comparable to the manually history-matched model, we selected as model parameters for history matching only the gridblock horizontal log-permeabilities, which are the same parameters used to perform the manual history matching. The vertical permeability was calculated using a fixed anisotropic ratio, $k_v/k_h = 0.3$. This ratio is the same as in the manually history-matched model. Other reservoir properties are kept as the same as the manually history-matched model. The initial ensemble of log-permeability fields was obtained using sequential Gaussian simulation [30, Chap. V] conditioned to interpreted pressure transient data using an approximate covariance function. In this case, we simply assigned the values of log-permeabilities obtained from pressure transient analysis at 13 wells to the gridblocks corresponding to each well location in order to define the "hard data" for sequential Gaussian simulation. Based on these data, we expect very high values of horizontal permeability in this field, including permeabilities around 20,000 mD. For turbidite system 2, the geologists of this field expect a higher correlation length in the x-direction, which approximately corresponds to the depositional direction. As we do not have a geological model for the permeabilities in this field, we chose the prior correlation lengths by visual inspection of the manually history-matched model. We chose an anisotropic exponential correlation function with ranges of 2,000 meters (corresponding to the size of 20 gridblocks) in the x-direction, and 1,000 meters (10 gridblocks) in the y-direction. Fig. 3.34a illustrates the correlation function used for building the initial ensemble. The choice of the prior correlation lengths impacts the generation of the initial ensemble as well as the covariance localization

procedure. Therefore, it is desirable to define the prior model based on a realistic geological model. Unfortunately, in the present field application, we do not have a prior geological model that allows stochastic generation of plausible permeability fields. Nevertheless, we obtained good results for this field case with our choice of correlation lengths.

For covariance localization, we defined the critical lengths as the prior correlation lengths plus the lengths of the axes of the ellipses centered at the well locations that "best" represent the drainage region (in a least squares sense) of each well calculated with the pseudo-tracer approach proposed by Damiani [28]. Fig. 3.34 shows the drainage area, the ellipse obtained by matching the drainage area and the final localization region for well P-103 at the last data assimilation time-step. Note that as the drainage areas change with time, the localization regions also change with time. Also note that as we assumed a long correlation length in the prior model, the final localization region is relatively large, as illustrated in Fig. 3.34d. If we had chosen the localization based only on the sensitivity region, as in Arroyo-Negrete et al. [10], the localization would be considerably smaller as illustrated in Fig. 3.34b.



(a) Prior correlation (exponential correlation function) **(b)** Drainage area **(c)** Ellipse **(d)** Localization region (Gaspari-Cohn correlation function)

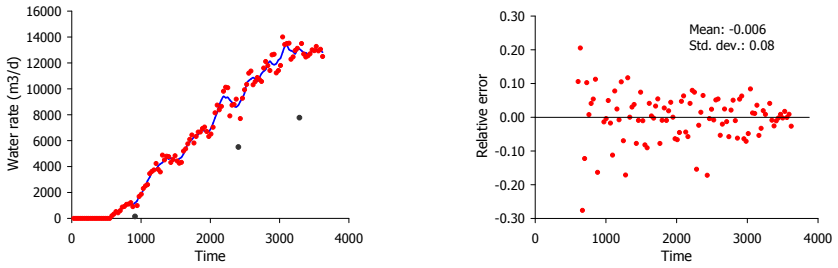
Figure 3.34: Localization for well P-103. (a) Prior correlation; (b) drainage areas at 2,770 days (last data assimilation time); (c) ellipse matched from the drainage area and (d) the corresponding correlation values used for localization (layer 4).

Recall that the data used for history matching is the water production rate at the labeled producing wells shown in Fig. 3.33. The total number of producing wells with a water production history is 14 and the total production time is 3,621 days. From the total production period, we assimilated data only for the first 2,770 days.

The manually history-matched model was also obtained considering only the first 2,770 days of history. The remaining 851 days of history were used for comparing the predictions from the history-matched models.

The measurement errors were assumed to be Gaussian with zero mean and standard deviation estimated as 10% of the water production rate for each well. The measurement errors were also assumed to be uncorrelated in time and space. We estimated the standard deviation of the measurement errors by smoothing the observed data using a moving average and subtracting the observations from the smoothed data. Then, we computed the relative errors by dividing each data difference by the value corresponding to the smoothed data. After that, we simply computed the standard deviation of the relative errors. As we have measurements of water rate every 30 days, we chose the length of the averaging window to be 120 days (five data points centered at each datum). During the smoothing process, we removed outliers and did not smooth across sharp changes in the production data. We applied this procedure for each well. The final average standard deviation of relative measurement errors was 10%. Fig. 3.35a shows the field water production rate (all wells) and the smoothed data obtained by using the averaging window. In this figure, the gray dots correspond to data not used in the smoothing process. Fig. 3.35b shows the corresponding relative errors indicating that, for the total field production, the estimated standard deviation is 8%. Even though we have data with a frequency of 30 days, we assimilate data every 90 days with EnKF. Note that we also include zero water production rate data as observed data in the history-matching process. In this case, we assume a constant measurement error of 1 std m^3/d.

In the simulation model, there are 136 gridblocks with local refinements. These refined grids are located around seven producing wells. The permeabilities of these refined grids are the same as for the coarse gridblock that contains the refined grid. During data assimilation with EnKF, we also update the pressure and

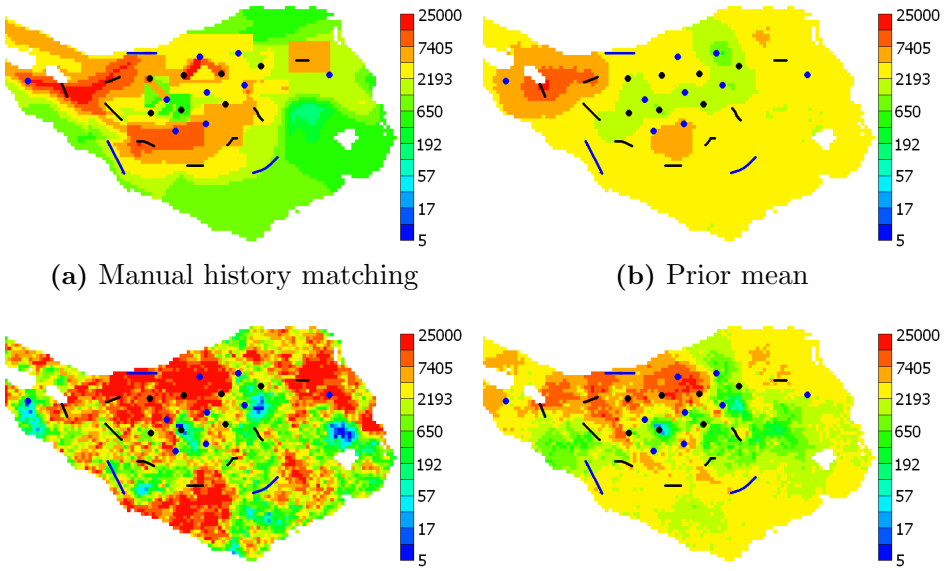


(a) Water production rate **(b)** Relative errors

Figure 3.35: (a) Field water production (all wells) in std m^3/day. Red dots are the observed data points and the solid blue curve is the smoothed data obtained with a moving averaging window. The grey dots were not used in the smoothing process. (b) Relative errors.

saturations at these refined grids in order to be able to restart the simulations after each data assimilation time-step.

Fig. 3.36 presents the initial ensemble mean permeability field and the mean permeability fields obtained by data assimilation using EnKF with and without covariance localization. For comparison, the manually history-matched permeability field is also included in this figure. According to the results in Fig. 3.36, the resulting mean permeability field from EnKF without localization has several regions with high permeability (25,000 mD). On the other hand, EnKF with covariance localization obtained a smoother ensemble mean permeability field, and the main changes were concentrated in regions near wells. Note that in the south and southeast part of the field, where there are no producing wells, EnKF with covariance localization did not significantly change the values of permeabilities, i.e., these values are close to those from the prior mean. This indicates that the effect of spurious correlations in EnKF was reduced by the application of covariance localization. Also note that the estimated permeability field obtained from EnKF with localization shows none of the high permeability streaks connecting injector/producer pairs that appeared in the manually history-matched model. Fig. 3.37 presents the standard deviation of log-permeability for the initial ensemble and after data assimilation using EnKF with and without covariance localization. As illustrated in Fig. 3.37, EnKF without

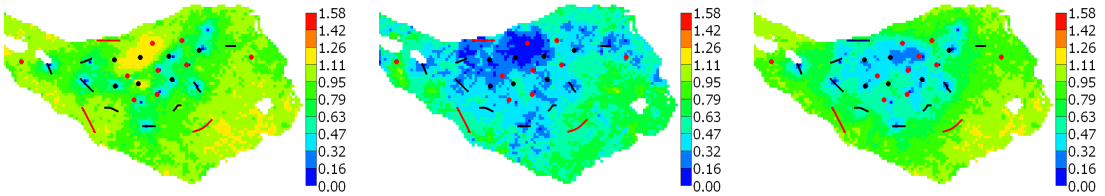


(a) Manual history matching **(b)** Prior mean

(c) Posterior mean. EnKF **(d)** Posterior mean. EnKF with localization

Figure 3.36: Permeability fields (mD) for layer 4. Producing wells are shown in black and water injector wells are shown in blue.

localization resulted in a significant reduction in the ensemble standard deviations, including at locations far from wells, while EnKF with covariance localization resulted in higher values of standard deviation, i.e., there is higher variability in the final ensemble after data assimilation. Note that avoiding excessive loss of variance in the ensemble is important for assimilating later data and for obtaining reasonable estimations of the uncertainty in the predictions.



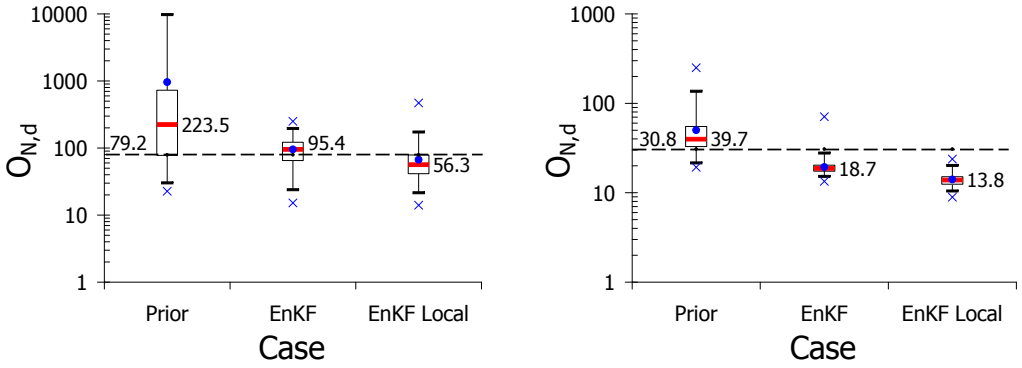
(a) Initial **(b)** EnKF **(c)** EnKF with localization

Figure 3.37: Standard deviation of log-permeability (layer 4).

In order to compare the final data matches, we reran the final models from time zero (initial reservoir condition) and computed the normalized data mismatch

objective function, $O_{N,d} = O_d(m)/N_d$. Fig. 3.38a shows the box plots of $O_{N,d}$ for the initial ensemble, EnKF and EnKF with covariance localization. Note that these values of $O_{N,d}$ consider only the history used for data assimilation (2,770 days). For comparison, we also include the $O_{N,d}$ obtained by the manually history-matched model (horizontal dashed line in Fig. 3.38a). All box plots presented in this dissertation follow the description presented in Fig. 3.39. According to the results in Fig. 3.38a, covariance localization gave better data matches than the standard EnKF. The data matches obtained by EnKF with covariance localization are also slightly better than those obtained with the manually history-matched model. In order to compare the forecasts, we also computed the $O_{N,d}$ considering only the forecast period (851 days) and the results are presented in Fig. 3.38b. According to the results in Fig. 3.38b, both EnKF and EnKF with covariance localization gave better predictions than were obtained with the manually history-matched model. EnKF with covariance localization also gave better predictions than standard EnKF. Figs. 3.40–3.46 present the predicted water production rate for seven wells. Fig. 3.47 shows the total predicted water production rate (all wells). In these figures, we include both the history for data assimilation period (2,770 days) and the forecast period (851 days). These figures illustrate that EnKF with covariance localization resulted in the best data matches and most accurate "future" predictions. Although the overall data matches obtained for each well seem reasonable, for wells P-85 and P-101, the breakthrough times are not matched well even when EnKF with covariance localization is used. We believe that most of the difficulty in matching breakthrough time arises from the fact that the initial ensemble does not give a set of breakthrough times that spans the actual breakthrough time adequately (Figs. 3.40b and 3.41b), i.e., most of the ensemble members in the initial ensemble predict the beginning of water production after the observed breakthrough time. As shown by Reynolds et al. [143], EnKF is similar to applying Gauss-Newton sequentially, with a full step and

replacement of the sensitivity matrix by an average sensitivity matrix obtained from the ensemble. Therefore, if none of the ensemble members produce water before the observed breakthrough time, the estimated sensitivity of water rate at times near breakthrough will be zero. If sensitivity is zero, EnKF cannot adjust parameters to match the water breakthrough. Even with imperfect matches of some breakthrough times, the overall data matches are reasonably good; thus, we did not choose to incur the additional computational time necessary to obtain a better initial ensemble.



(a) History (2,770 days) **(b)** Forecast (851 days)

Figure 3.38: Box plots of the normalized data mismatch objective function. The horizontal dashed lines correspond to the $O_{N,d}$ obtained by the manually history-matched model. The numbers next to each box plot correspond to the values of the median of $O_{N,d}$ and the numbers in the beginning of the horizontal dashed lines correspond to the values of $O_{N,d}$ obtained by the manually history-matched model.

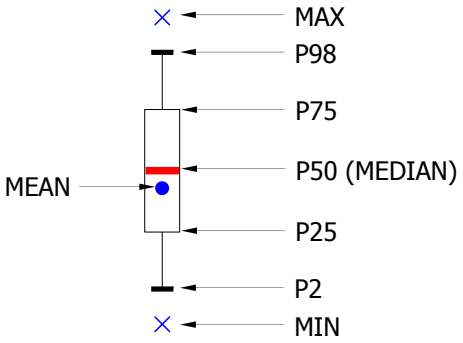
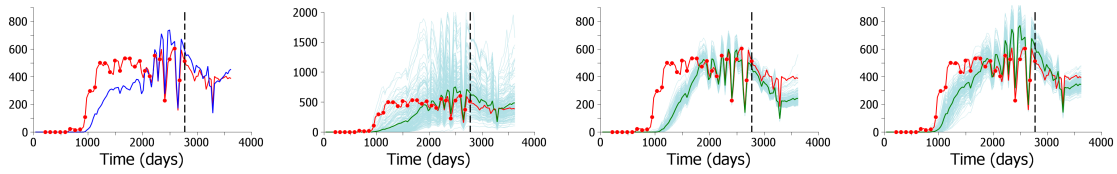
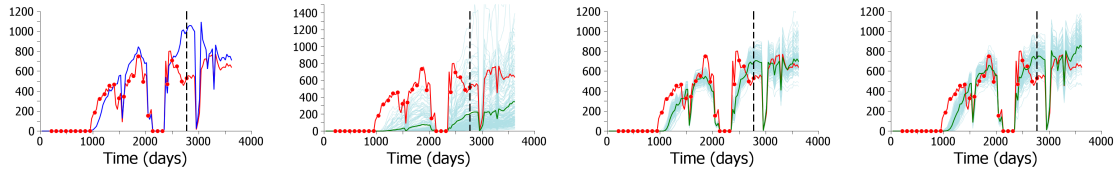


Figure 3.39: Box plot description.



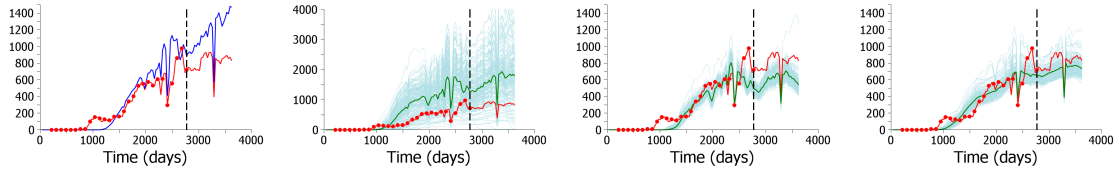
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** EnKF with localization

Figure 3.40: Well P-85. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). Red dots are the data points used for data assimilation; solid red curve is the production history; solid blue curve is the prediction from the manually history-matched model; solid green curves are the mean predictions from the ensembles; solid light blue curves represent predictions from all models of the ensembles. The dashed vertical black line indicates the end of the history and beginning of the forecast period.



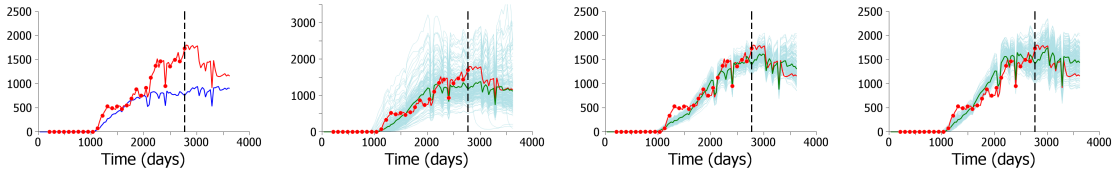
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** EnKF with localization

Figure 3.41: Well P-101. Predicted water production rate obtained by running final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 3.40.



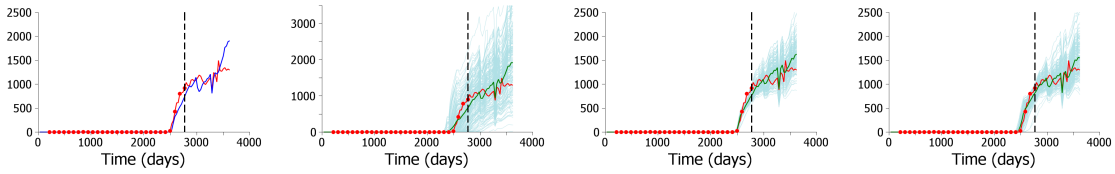
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** EnKF with localization

Figure 3.42: Well P-104. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 3.40.



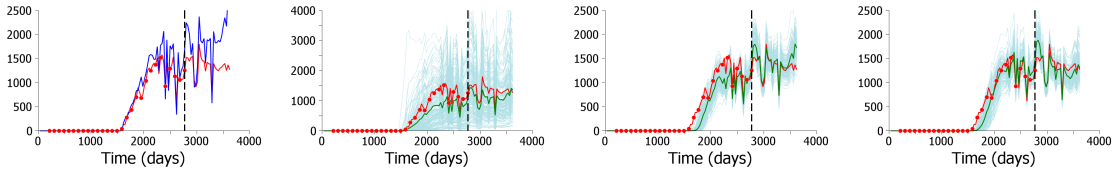
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** EnKF with localization

Figure 3.43: Well P-111. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 3.40.



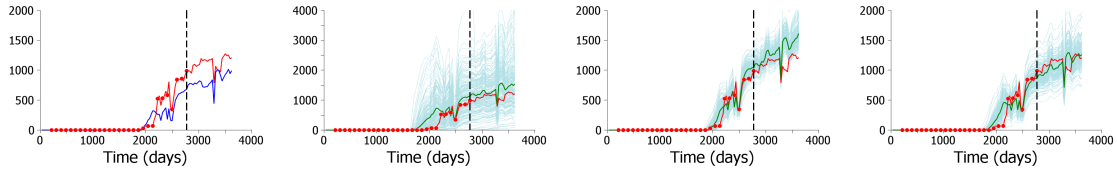
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** EnKF with localization

Figure 3.44: Well P-141. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 3.40.



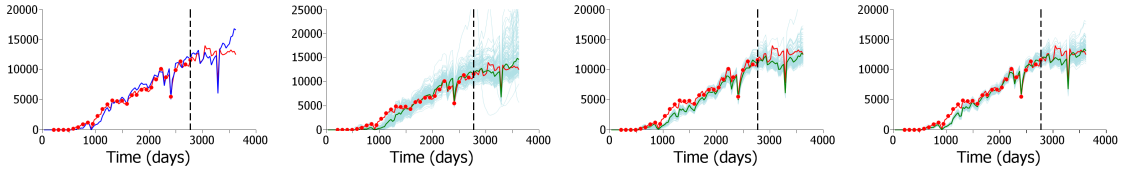
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** EnKF with localization

Figure 3.45: Well P-149. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 3.40.



(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** EnKF with localization

Figure 3.46: Well P-152. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 3.40.



(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** EnKF with localization

Figure 3.47: Field water production (all wells). Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 3.40.

3.3.3 Half-iteration EnKF

Here, we present the results of history matching the field case using HI-EnKF with covariance localization considering two cases: (1) rerunning the ensemble every data assimilation time-step and (2) rerunning the ensemble if the average relative change in the updated ensemble mean is larger than 1%. In this case, we rerun the ensemble from time zero if any one the following three the conditions is satisfied:

$$\frac{1}{N_g}\sum_{i=1}^{N_g}\left|\frac{\overline{\ln k}_i^a - \overline{\ln k}_i^f}{\overline{\ln k}_i^f}\right| \geq 0.01, \tag{3.16}$$

or

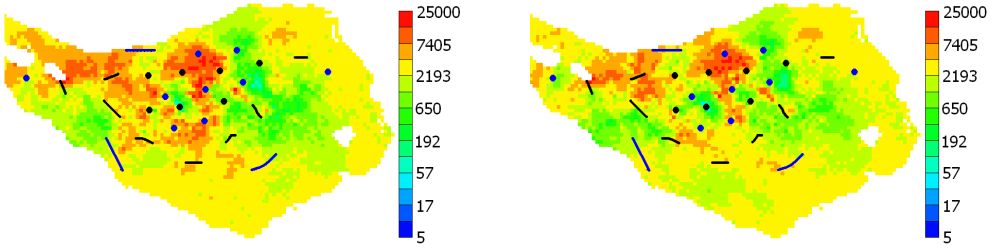
$$\frac{1}{N_g}\sum_{i=1}^{N_g}\left|\frac{\overline{p}_i^a - \overline{p}_i^f}{\overline{p}_i^f}\right| \geq 0.01, \tag{3.17}$$

or

$$\frac{1}{N_g}\sum_{i=1}^{N_g}\left|\frac{\overline{S_{\mathrm{o}i}}^{a}-\overline{S_{\mathrm{o}i}}^{f}}{\overline{S_{\mathrm{o}i}}^{f}}\right|\geq 0.01. \tag{3.18}$$

where N_g is the number of active gridblocks, $\overline{\ln k}$, $\overline{p}$ and $\overline{S_{\mathrm{o}}}$ are the ensemble mean vectors of gridblock log-permeabilities, pressures and oil saturations, respectively.

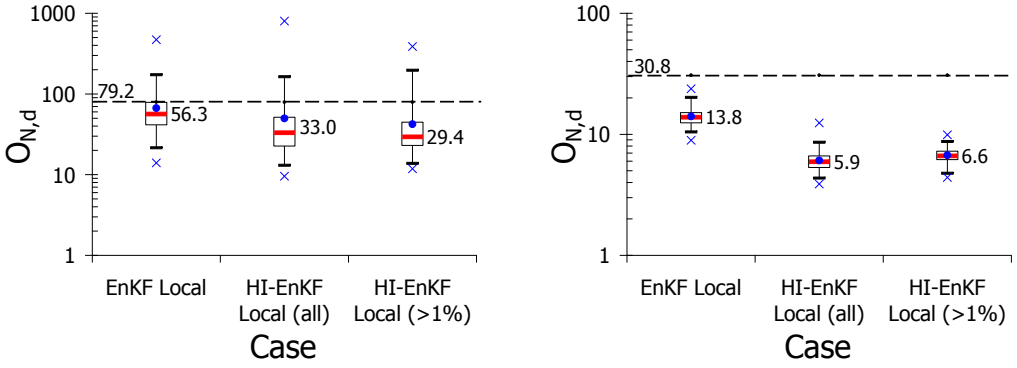
Fig. 3.48 presents the ensemble mean permeability fields obtained by data assimilation using HI-EnKF with covariance localization for the two cases considered. As illustrated in this figure, there are no significant differences between the two permeability fields. In addition, they are both very similar to the mean permeability field obtained using EnKF with localization without half-iteration (Fig. 3.36c). Fig. 3.49 shows the box plots of $O_{N,d}$ obtained by the two HI-EnKF cases considering the history and forecast periods. For visual comparison, we also include the box plot of $O_{N,d}$ obtained by EnKF with covariance localization and the $O_{N,d}$ obtained by the manually history-matched model in Fig. 3.49. Comparing the two HI-EnKF procedures, we observe that the quality of the final data matches and predictions are basically the same, but they are considerably better than the ones obtained using EnKF with localization without half-iteration and the manually history-matched model. Fig. 3.50 shows the total predicted water production rate (all wells) for the two HI-EnKF cases. This figure also illustrates the similarity in the results obtained by the two HI-EnKF procedures. The major difference between these two cases is the computational cost. In the first case, we reran from time zero after each data assimilation time-step, but in the second case, we reran from time zero only five times. The total number of data assimilations time-steps is 29. The data assimilation time-steps in which the ensemble was rerun from time zero are indicated in the Fig. 3.50b as blue circles. Note that four of the five data assimilations that resulted in rerunning from time zero according to the criteria of Eqs. 3.16, 3.17 and 3.18, occurred at the beginning of the water production; the fifth rerun was at the last data assimilation time-step.



(a) Posterior mean. HI-EnKF rerunning all data assimilation time-steps

(b) Posterior mean. HI-EnKF rerunning if average change in the ensemble mean is larger than 1%

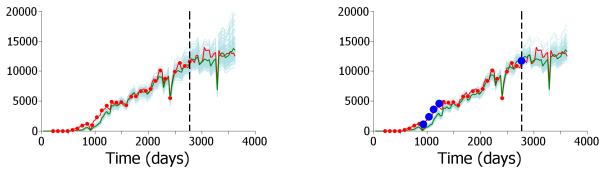
Figure 3.48: Permeability fields for layer 4 (mD). Producing wells are shown in black and water injector wells are shown in blue.



(a) History (2,770 days)

(b) Forecast (851 days)

Figure 3.49: Box plots of the normalized data mismatch objective function. The numbers next to each box plot correspond to the value of the median of $O_{N,d}$. The horizontal dashed line corresponds to the manually history-matched model.



(a) HI-EnKF rerunning after each data assimilation time-steps

(b) HI-EnKF rerunning if average change in the ensemble mean is larger than 1%

Figure 3.50: Field water production (all wells). Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The blue circles in (b) indicate the data assimilation time-steps with rerun from time zero. The colors in this figure have the same meaning as in Fig. 3.40.

CHAPTER 4

STRATEGIES TO REDUCE LOSS OF VARIANCE DUE TO SAMPLING ERRORS IN THE ENSEMBLE KALMAN FILTER

Besides covariance localization, other methods for reducing the negative effects of a small ensemble have been proposed in the literature. This chapter presents a comparative evaluation of some of those methods, namely, distance-based covariance localization, non-distance dependent localization and covariance inflation. We also consider the ensemble square root filter (EnSRF), which typically exhibits a smaller loss of variance than occurs with EnKF.

4.1 Non-distance Dependent Covariance Localization

4.1.1 Furrer and Bengtsson taper

Furrer and Bengtsson [57] showed that it is possible to minimize term-by-term the norm of the difference between the true forecast covariance matrix and the localized estimate ignoring the positive-definiteness constraint. Based on this procedure, Furrer and Bengtsson [57] obtained the following expression to compute the components of the correlation matrix for localization,

$$\rho_{ij} = \frac{c_{ij}^2}{c_{ij}^2 + (c_{ij}^2 + c_{ii}c_{jj})/N_e}, \tag{4.1}$$

where c_{ij} is the forecast covariance between the ith and and jth variables. In practice, c_{ij} is the (i,j) entry of the matrix $\widetilde{C}_{\mathrm{YD}}^{n,f}$, which corresponds to the cross-covariance between the ith element of the forecast augmented state vector, $y^{n,f}$ (Eq. 2.36), and the jth element of the predicted data vector $d^{n,f}$. In [57], the authors suggest that

sparseness can be introduced by zeroing "small" values of c_{ij}. Although the definition of "small" is problem-dependent, a reasonable choice is set $\rho_{ij} = 0$ if

$$|c_{ij}| < \epsilon\sqrt{c_{ii}c_{jj}}, \tag{4.2}$$

with $\epsilon = 10^{-2}$ to 10^{-3}.

4.1.2 Hierarchical ensemble filter

Anderson [6] also presented a non-distance dependent localization procedure, which is referred to as hierarchical ensemble filter. This procedure computes confidence factors, α's, to "localize" the Kalman gain based on groups of N_h independent ensembles. The confidence factor, α_i, corresponding to the ith entry in the Kalman gain matrix, $[\widetilde{K}]_i$, is defined as the minimizer of

$$\sqrt{\sum_{j=1}^{N_h}\sum_{l=1,l\neq j}^{N_h}\left(\alpha_i[\widetilde{K}_l]_i - [\widetilde{K}_j]_i\right)^2}, \tag{4.3}$$

which leads to

$$\alpha_i = \max\left\{\frac{1}{N_h-1}\left[\frac{\left(\sum_{j=1}^{N_h}[\widetilde{K}_j]_i\right)^2}{\sum_{j=1}^{N_h}[\widetilde{K}_j]_i^2} - 1\right], 0\right\}. \tag{4.4}$$

In the above equation, a confidence factor α_i is truncated to zero if the α_i computed by minimizing Eq. 4.3 is negative, as proposed in [6]. The "localized" Kalman gain matrix, $\widetilde{K}_\alpha$, for each of the N_h ensembles, is computed by multiplying each Kalman gain matrix with a matrix of confidence factors, A, in an element-wise manner

$$\widetilde{K}_{j,\alpha} = A \circ \widetilde{K}_j, \text{ for } j = 1, 2, \cdots N_h. \tag{4.5}$$

One difficulty with this approach is that it requires assimilation of data with N_h ensembles simultaneously, which increases the computational cost of the process.

Zhang and Oliver [194] proposed that instead of N_h ensembles of N_e members, we resample a single ensemble with replacement to generate N_b bootstrapped ensembles. Then, the confidence factors can be computed with same expression (4.4) using N_b instead of N_h. In this procedure, the N_b bootstrapped ensembles are used only to compute the confidence factors; the data assimilation is done with the original ensemble, i.e., before resampling.

4.2 Covariance Inflation

Another procedure that has appeared in the literature to compensate for the underestimation of posterior variances in the EnKF is covariance inflation [7]. In covariance inflation, we replace the forecast ensemble by

$$y_j^{n,f} = \gamma(y_j^{n,f} - \overline{y}^{n,f}) + \overline{y}^{n,f}, \text{ for } j = 1, 2, \cdots, N_e, \tag{4.6}$$

where γ is the inflation factor and $\overline{y}^{n,f}$ is the forecast ensemble mean state vector at the nth data assimilation step. Note that Eq. 4.6 increases the variance but does not change the mean of $y^{n,f}$.

Typically the inflation factor is a number slightly greater than one. The optimal inflation factor is problem dependent. One possible advantage of covariance inflation over localization is that, because we multiply the whole state by a constant factor, we do not artificially change the correlation coefficient between each pair of variables, only the variance is changed. For distance-based localization on the other hand, this is not true. In fact, depending on the choice of the critical length, we may even destroy the geological realism of the estimates.

4.2.1 Adaptive inflation

Evensen [47] proposed an adaptive covariance inflation procedure which can be described as follows. For each data assimilation step, we generate the additional $N_r \times N_e$ matrix B^f with random normally distributed numbers, i.e., samples from

$\mathcal{N}(0,1)$. Then, we update B using the standard analysis equation of EnKF. However, as the columns of B^f are uncorrelated with the data measurements, the variance among the N_e columns of B^f should be preserved after analysis, i.e., we should have $\text{var}[b_i^a] = \text{var}[b_i^f] = 1.0$, for $i = 1, 2, \ldots, N_r$; where b_i^a and b_i^f denote the ith row of B^a and B^f respectively. However, because of sampling errors, this is not true and we observe some reduction on $\text{var}[b_i^a]$. Thus, we seek for an inflation factor to correct this variance reduction. This inflation factor is given by,

$$\gamma = \frac{1}{\frac{1}{N_r}\sum_{i=1}^{N_r}\sigma_i}, \tag{4.7}$$

where

$$\sigma_i = \sqrt{\text{var}[b_i^a]} \tag{4.8}$$

is the standard deviation for the ith row of B^a. The same inflation factor, γ, is applied in the whole state. Evensen [47] does not mention the size of N_r that one should select to compute the inflation factor. Here, we set N_r equal to the size of the state vector, N_y.

4.2.2 Adaptive inflation for distance-based localization

Although covariance localization reduces the underestimation of posterior variances, we still observe some underestimation, especially when the localization region is large. In fact, for distance-based localization, we have a trade-off: reducing the critical length tends to increase the posterior variances, but it removes true long distance correlations. Increasing the critical length, we keep long distance correlations, but we may observe some underestimation in the posterior variances. This suggests that one could apply both localization and inflation simultaneously. However, when we apply localization only states within a pre-defined distance are updated. Therefore, if we also apply inflation, we may inflate states that were not

even updated during the data assimilation, which is not desirable. This suggests that one should apply inflation only in the region corresponding to entries of $y_j^{n,f}$ influenced by the measurements, i.e., the localization region.

Considering that, we introduced a modification of the covariance inflation procedure proposed by Evensen [47] that is suitable for application in cases in which a distance-based localization procedure is also applied. Algorithm 4.1 presents the proposed local-inflation procedure.

In this procedure, we first compute the inflation factors, apply inflation to the forecast state and then, we perform the analysis of the state with covariance localization. One drawback of this approach is that we change the correlation coefficient between each pair of variables because we do not multiply the whole state by the same inflation factor. Another problem is the extra computational time required to perform the N_r iterations of the EnKF analysis routine. However, some computation time can be saved if we consider that part of the EnKF analysis equation does not change from iteration to iteration. In fact, it is easy to show that the EnKF analysis with localization can be written as

$$B^a = B^f + \rho_{\mathrm{BD}} \circ \left[\Delta B^f \left(\Delta D^{n,f}\right)^{\mathrm{T}}\right] \frac{1}{N_e - 1} \left[C_{\mathrm{D}}^n + \left(\rho_{\mathrm{DD}} \circ \widetilde{C}_{\mathrm{DD}}^{n,f}\right)\right]^{-1} \left(D_{\mathrm{uc}}^n - D^{n,f}\right), \tag{4.15}$$

or simply,

$$B^a = B^f + \rho_{\mathrm{BD}} \circ \left[\Delta B^f \left(\Delta D^{n,f}\right)^{\mathrm{T}}\right] X, \tag{4.16}$$

where

$$X \equiv \frac{1}{N_e - 1} \left[C_{\mathrm{D}}^n + \left(\rho_{\mathrm{DD}} \circ \widetilde{C}_{\mathrm{DD}}^{n,f}\right)\right]^{-1} \left(D_{\mathrm{uc}}^n - D^{n,f}\right). \tag{4.17}$$

Note that the $N_n \times N_e$ matrix X does not change from iteration to iteration, and

Algorithm 4.1 Adaptive inflation for distance-based localization

1. For $\ell = 1$ to N_r:

   (a) Define a $N_g \times N_e$ matrix B^f, where each entry of B^f is generated as an independent standard normal deviate. N_g denotes the number of active gridblocks of the model. Each column of the matrix B^f correspond to an ensemble member and each row of B^f corresponds to a gridblock location.

   (b) Correct the matrix B^f such that each row has exactly zero mean and variance equal to one. This can be done by replacing the entries of B^f according to

   $$b_{ij}^{f,\text{corr}} = \frac{b_{ij}^f - \overline{b}_i^f}{\sigma_i^f}, \text{ for } i = 1, 2, \cdots N_g, \tag{4.9}$$

   where

   $$\overline{b}_i^f = \frac{1}{N_e} \sum_{j=1}^{N_e} b_{ij}^f \tag{4.10}$$

   and

   $$\sigma_i^f = \sqrt{\frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(b_{ij}^f - \overline{b}_i^f \right)}. \tag{4.11}$$

   (c) Compute B^a using the EnKF analysis equation with covariance localization (Eq. 4.15).

   (d) Compute:

   $$\sigma_{i,\ell}^a = \sqrt{\frac{1}{N_e - 1} \sum_{j=1}^{N_e} \left(b_{ij}^a - \overline{b}_i^a \right)}, \text{ for } i = 1, 2, \cdots N_g. \tag{4.12}$$

   end (for).

2. Compute:

   $$\overline{\sigma}_i^a = \frac{1}{N_r} \sum_{\ell=1}^{N_r} \sigma_{i,\ell}^a, \text{ for } i = 1, 2, \cdots N_g. \tag{4.13}$$

3. Compute the N_g-dimensional vector of inflation factors, where each entry, γ_i, is given by

   $$\gamma_i = \frac{1}{\overline{\sigma}_i^a}. \tag{4.14}$$

4. Apply the inflation factor vector to the forecast state considering that each gridblock location has a different inflation factor.

thus it needs be computed only once. In the above equations, $\Delta B^f = B^f - \overline{B}^f$ where $\overline{B}^f$ denotes the $N_g \times N_e$ matrix with all columns equal to the ensemble mean and ρ_{BD} denotes the $N_g \times N_n$ correlation matrix for localization. This matrix is the same one employed to update model parameters. $D^{n,f}$ denotes the $N_n \times N_e$ matrix of predicted data, $\Delta D^{n,f} = D^{n,f} - \overline{D}^{n,f}$, and D^n_{uc} denotes the $N_n \times N_e$ matrix of perturbed observations.

Another aspect of this method is the number of iterations, N_r, required to estimate an appropriate inflation factor. Instead of defining a fixed number of iterations, we define the maximum number of iterations and check to see if there is convergence in the inflation factor. We use the following convergence criterion

$$\max_{1 \le i \le N_g} \left| \frac{\gamma_i^\ell - \gamma_i^{\ell-1}}{\gamma_i^{\ell-1}} \right| \le 10^{-4}. \tag{4.18}$$

The maximum allowed number of iterations is 1,000.

4.3 Ensemble Square Root Filter

The perturbed observation scheme used in the standard EnKF is often identified as another source of sampling errors [186]. In the square root filters [170], the ensemble is updated using deterministic formulations, which is believed to reduce sampling errors, resulting in smaller underestimation of posterior variances [186]. In the results presented in the next section, we use the left-multiplied EnSRF scheme proposed by Whitaker and Hamill [186]. The reason we chose this scheme is because it allows covariance localization, which can be done by simply replacing $\widetilde{C}_{\text{Y}}^{n,f}$ in Eq. 2.116 by the Schur product $\rho^n \circ \widetilde{C}_{\text{Y}}^{n,f}$. Eq. 2.116 with covariance localization can be written as

$$\widetilde{K}^{\star}_{n,\rho} = \rho^{n}_{\mathrm{YD}} \circ \widetilde{C}^{n,f}_{\mathrm{YD}} \left[\rho^{n}_{\mathrm{DD}} \circ \widetilde{C}^{n,f}_{\mathrm{DD}} + C^{n}_{\mathrm{D}} \right]^{\mathrm{T}/2} \left[\left(\rho^{n}_{\mathrm{DD}} \circ \widetilde{C}^{n,f}_{\mathrm{DD}} + C^{n}_{\mathrm{D}} \right)^{1/2} + (C^{n}_{\mathrm{D}})^{1/2} \right]^{-1}. \tag{4.19}$$

As discussed in Section 2.6.2, for the case of uncorrelated measurement errors, we can assimilate one datum at time, which simplifies the calculation of $\widetilde{K}^{\star}$. In this case, we can localize the Kalman gain matrix, i.e.,

$$\widetilde{K}^{\star}_{n,\rho} = \alpha \left(\rho^{n}_{\mathrm{YD}} \circ \widetilde{K}_{n} \right), \tag{4.20}$$

where α is the scalar given by Eq. 2.122. In the results presented in the next section, we assimilate one data at a time with EnSRF.

4.4 Examples

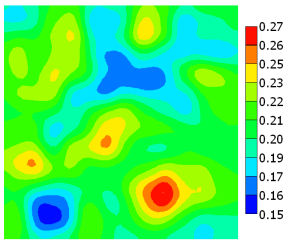
4.4.1 Example 1 – Linear model

The first example is the same linear case considered in Chapter 3 (Section 3.2.1). Recall that the model parameters are gridblock porosities and the measured data correspond to the average porosity of the 11×11 gridblocks centered on each well position as illustrated in Fig. 3.7a. The total number of data points is 144 and we assimilate one datum each time-step. We use an ensemble size of 50 and consider the following cases: (1) standard EnKF; (2) EnKF with adaptive inflation [47]; (3) EnKF with distance-based localization; (4) EnKF with distance-based localization and adaptive inflation; (5) EnKF with Furrer and Bengtsson non-distance dependent localization; (6) bootstrap hierarchical EnKF [194]; (7) left-multiplied EnSRF [186]; (8) left-multiplied EnSRF with distance-based localization. The cases with distance-based localization use the Gaspari-Cohn correlation function with critical length equal 10 gridblocks, which corresponds to the correlation length of the prior model plus the distance of influence of the data (sensitivity region).

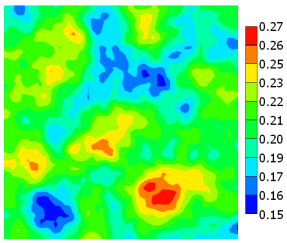
Fig. 4.1 shows the ensemble mean after data assimilation for all cases considered. For comparison, the maximum a posteriori (MAP) estimate and the EnKF ensemble mean obtained with an ensemble of 1,000 models are also presented. According to the results in this figure, EnKF (Fig. 4.1c) and EnSRF (Fig. 4.1i) obtained very rough ensemble means. The application of adaptive inflation (Fig. 4.1d) did not improve the results in terms of the ensemble mean. Distance-based localization, on the other hand, clearly resulted in significant improvements in the ensemble mean (Figs. 4.1e, 4.1f and 4.1j). The non-distance dependent localization cases (Figs. 4.1g and 4.1h) did not improve significantly the results in terms of the ensemble mean.

Table 4.1 presents the values of the sum of normalized variances (SNV) for all cases. EnKF with $N_e = 1{,}000$ gave a value of SNV 8.4% lower than the value obtained with the MAP. Note that the SNV obtained for the MAP corresponds to the correct value. Standard EnKF (case 1) resulted in a severe underestimation of the SNV, i.e., EnKF greatly overestimated the reduction in uncertainty due to data assimilation. For this case, the final ensemble almost collapsed to a single model. Inflation (case 2), non-distance dependent localization (cases 5 and 6) and EnSRF (case 7) approximately doubled the SNV obtained with EnKF, but they still give a severe underestimation of SNV. Significant improvements were obtained only for the cases with distance-based localization. EnKF with localization (3) results in a SNV 13% less than the correct value, while for EnSRF with localization (case 8), the SNV is only 7.6% less than the correct value, and is closer to the correct SNV than the result obtained using EnKF with $N_e = 1{,}000$. The combined application of localization and inflation (case 4) resulted in a very small overestimation of the SNV (2.3% greater than the correct value).

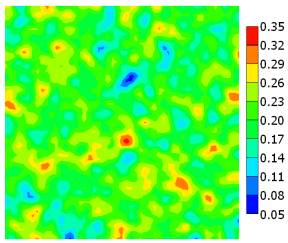
Table 4.2 presents values of the objective function obtained for each case. For this example, we have 144 observed data points, which results in a value of 1.589 for the upper limit of the data-match quality criterion (Eq. 2.130). According to



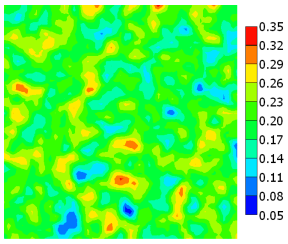
(a) MAP



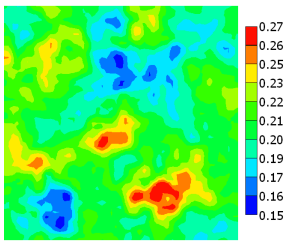
(b) EnKF ($N_e = 1{,}000$)



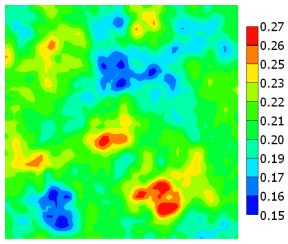
(c) Case 1: EnKF



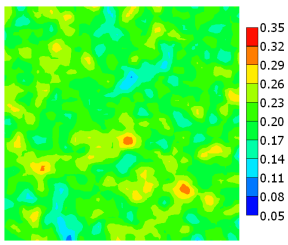
(d) Case 2: EnKF with adaptive inflation



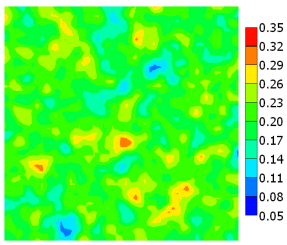
(e) Case 3: EnKF with distance-based localization



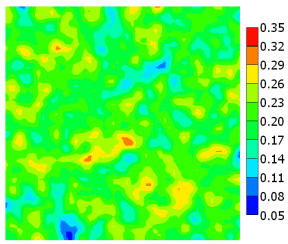
(f) Case 4: EnKF with distance-based localization and inflation



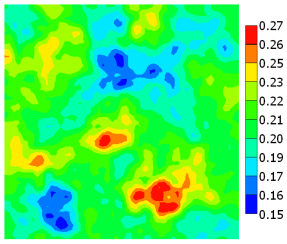
(g) Case 5: EnKF + Furrer and Bengtsson



(h) Case 6: Bootstrap hierarchical EnKF



(i) Case 7: EnSRF



(j) Case 8: EnSRF with distance-based localization

Figure 4.1: Ensemble mean after data assimilations. Note that some figures have different color scale.

Table 4.1: Sum of normalized variances of porosity.

Case	SNV
MAP	0.764
EnKF with $N_e = 1{,}000$	0.697
Case 1: EnKF	0.079
Case 2: EnKF with inflation	0.188
Case 3: EnKF with dist. local.	0.664
Case 4: EnKF with dist. local. and infl.	0.794
Case 5: EnKF with F-B non-dist. local.	0.150
Case 6: Bootstrap hierarchical EnKF	0.147
Case 7: EnSRF	0.176
Case 8: EnSRF with dist. local.	0.706

Table 4.2, except for the MAP estimate and EnKF with $N_e = 1{,}000$, none of the cases satisfy this quality criterion. However, the cases with distance-based localization (cases 3, 4 and 8) resulted in values of the objective function which come very close to satisfying the quality criterion. All cases without distance-based localization resulted in high values of O_m, indicating rough ensemble means, and high values of O_d, indicating poor data matches.

Fig. 4.2 presents the inflation factors obtained for case 2. In order to illustrate the effect of ensemble size, Fig. 4.2 also includes the inflation factors obtained for an ensemble of 25 models. According to the results in Fig. 4.2, increasing the size of the ensemble reduces the inflation factors, which is consistent with the conjecture that larger ensembles need less inflation. However, for both ensembles sizes, the values of the inflation factor shows a very oscillatory behavior over the data assimilation steps. Fig. 4.3 presents the inflation factors obtained for the 10th well data using the combined application of localization and inflation. The values of the inflation factors show a behavior similar to the localization used, i.e., the inflation factor is higher closer to the well location and decreases to one, outside the localization region. Also, increasing the size of the ensemble, reduced the inflation factors.

Table 4.2: Average objective function.

Case	$O_m(m)$	$O_d(m)$	$O_N = \frac{2O(m)}{N_d}$
MAP	19.8	7.6	0.380
EnKF with $N_e = 1{,}000$	82.1	9.0	1.265
Case 1: EnKF	1483.4	97.8	21.961
Case 2: EnKF with inflation	1771.4	85.3	25.788
Case 3: EnKF with dist. local.	100.0	20.2	1.668
Case 4: EnKF with dist. local. and infl.	103.3	18.8	1.696
Case 5: EnKF with F-B non-dist. local.	920.3	61.6	13.632
Case 6: Bootstrap hierarchical EnKF	1131.9	62.6	16.590
Case 7: EnSRF	995.8	55.1	14.597
Case 8: EnSRF with dist. local.	94.3	20.2	1.591

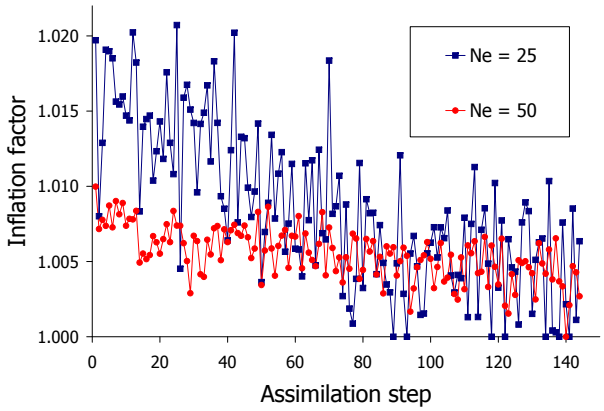
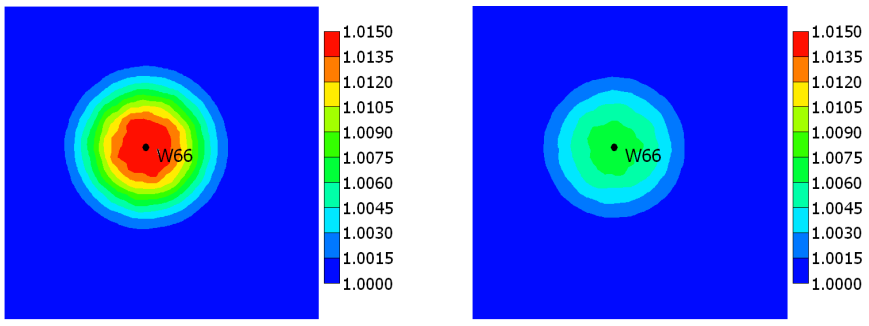


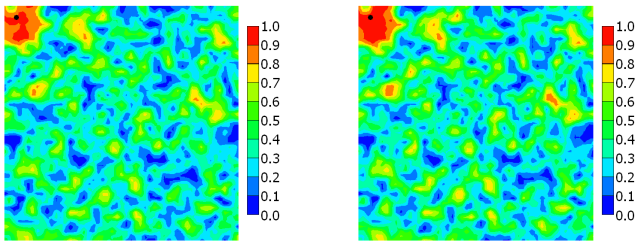
Figure 4.2: Inflation factors (case 2).



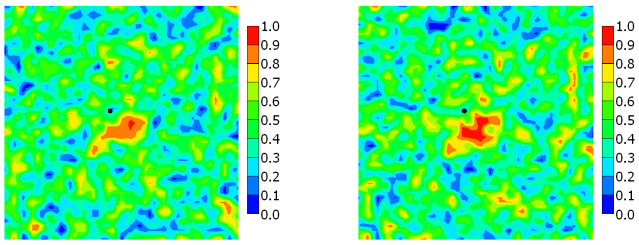
(a) $N_e = 25$ **(b)** $N_e = 50$

Figure 4.3: Inflation factor for well 66.

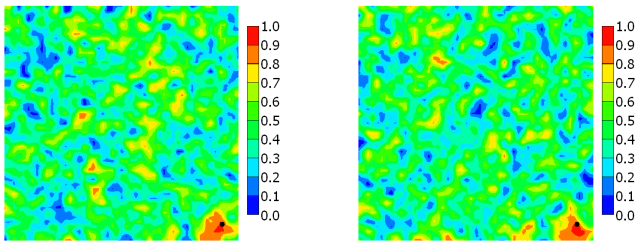
Fig. 4.4 presents the localization values obtained for the non-distance dependent localization methods (cases 5 and 6). This figure shows that the localization values are closer to one around the well location, but they do not decrease monotonically to zero as we increase the distance from the well. In fact, the average value of the localization coefficients far away from the well is around 0.5. Besides that, the localization values are very noisy. As a result, both procedures failed to eliminate spurious correlations.



(a) Case 5 – Well 1 **(b)** Case 6 – Well 1



(c) Case 5 – Well 66 **(d)** Case 6 – Well 66



(e) Case 5 – Well 144 **(f)** Case 6 – Well 144

Figure 4.4: Localization values obtained using the non-distance dependent localization procedures.

4.4.2 Example 2 – Synthetic reservoir

The second example corresponds to 2D synthetic reservoir on a 60×60 uniform grid. The model parameters are gridblock log-permeabilities. The true model

was generated from an anisotropic exponential correlation function with major correlation length of 2,400 ft (which corresponds to the width of 12 gridblocks) and minor correlation length of 1,200 ft; the direction of maximum correlation length is oriented at 45° to the horizontal direction. The prior mean is 5.0 and the prior variance is 1.0 for all gridblock log-permeabilities. Fig. 4.5 shows the true permeability field used as the reference to generate the observed data. For this reservoir, there are 16 producing wells and 9 water injectors placed in five-spot patterns. During simulations, the producing wells operate at a specified maximum liquid rate. However, a minimum bottomhole pressure constraint of 1,000 psi is also imposed. The observed data at the producers are oil and water flow rates and bottomhole pressures. The water injectors operate at constant water injection rate and the measurements are bottomhole pressures. For the flow rate data, random normally distributed noise with zero mean and standard deviation corresponding to 5% of the true data was added to the true data to define the the random variables that represent the measurements. For pressure data, a constant measurement error of 10 psi was considered. This example was designed to have a large number of measurements, which is the typical situation where problems related to sampling errors and limited degrees of freedom are more severe.

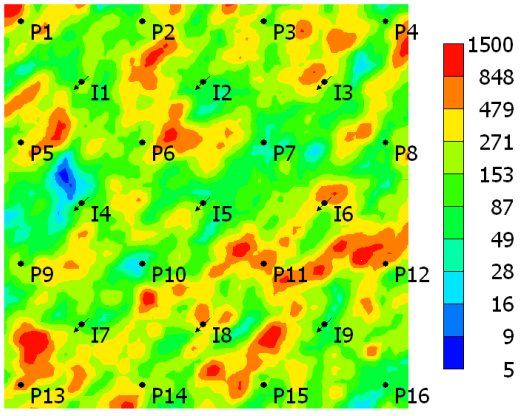


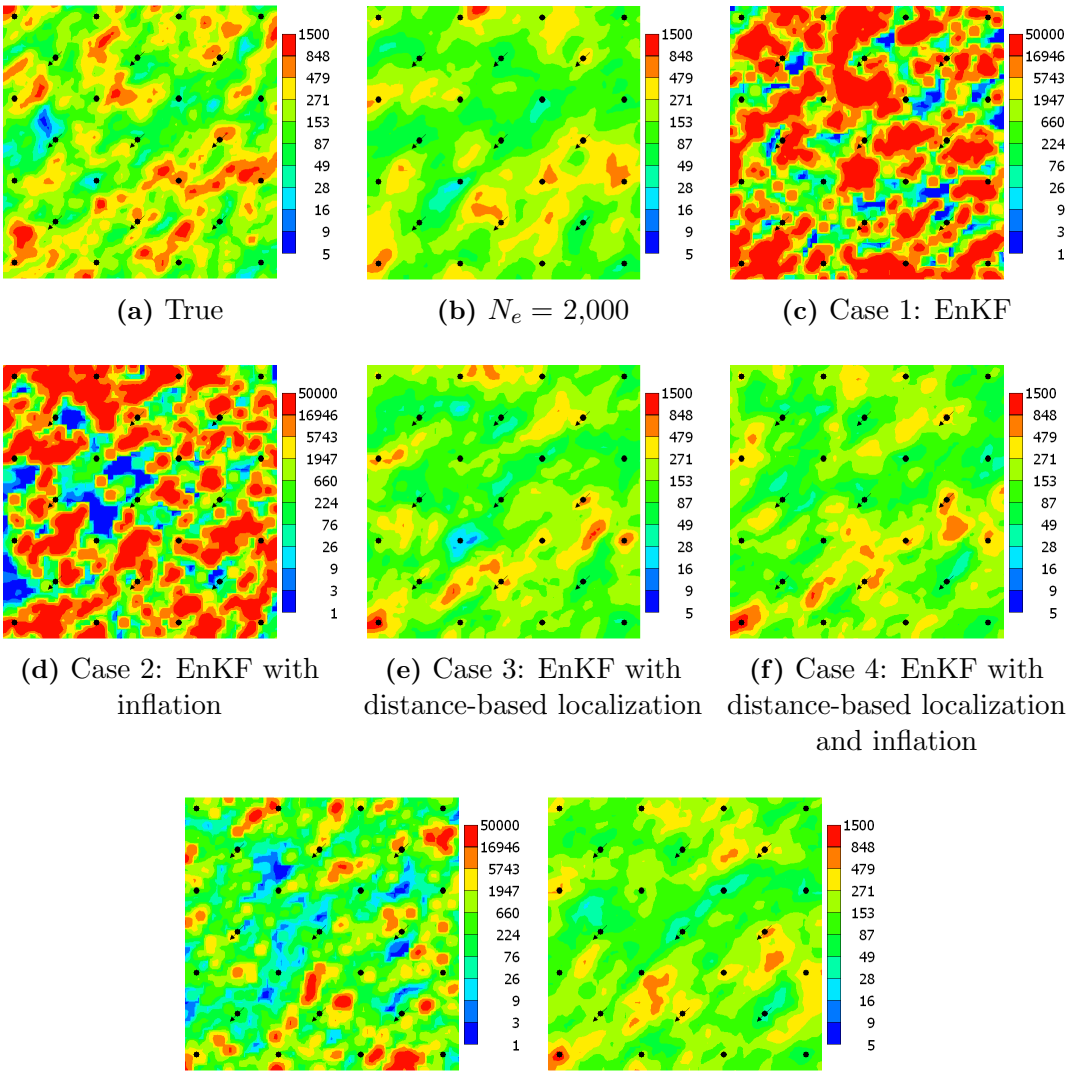
Figure 4.5: True permeability field.

The initial ensemble was generated using the same covariance function and

prior mean specified above. Data assimilations were performed using an ensemble of 100 models considering the following cases: (1) EnKF; (2) EnKF with adaptive inflation; (3) EnKF with distance-based localization; (4) EnKF with distance-based localization and inflation; (5) EnSRF; (6) EnSRF with distance-based localization. Cases with localization use the procedure described in Chapter 3.

Fig. 4.6 shows the ensemble mean permeability fields obtained after data assimilations. For comparison, the true model and the ensemble mean from EnKF with N_e = 2,000 models are also included in this figure. The ensemble mean permeability field obtained by EnKF (case 1) and EnSRF (case 5) are very rough and display several regions with permeability equal to 50,000 mD, which corresponds to the truncation value we use during data assimilations. Note that the color scale is not the same for all the cases. Inflation (case 2) did not improve the final ensemble mean permeability field and we still observe several regions of overshooting. Distance-based localization (cases 3, 4 and 6), on the other hand, gave smoother ensemble means, which resembles geologically the true model and the N_e = 2,000 result.

Table 4.3 presents the values of SNV for all cases. Compared to the ensemble of 2,000 models, EnKF (case 1) and EnSRF (case 5) clearly result in severe underestimation of the posterior variances. The application of covariance inflation (case 2) did not improve the results. EnKF with localization (case 3) gave a significant increase in the SNV, although there is still an underestimation. EnSRF with localization (case 6) gave a higher value of SNV compared to case 1, but still 32% less than the value obtained with N_e = 2,000. Localization with inflation (case 4) gave SNV 50% higher than EnKF with N_e = 2,000. Although EnKF with N_e = 2,000 does not correspond to the correct value of SNV, it seems that case 4 overestimated the SNV.



(a) True **(b)** $N_e = 2{,}000$ **(c)** Case 1: EnKF

(d) Case 2: EnKF with inflation **(e)** Case 3: EnKF with distance-based localization **(f)** Case 4: EnKF with distance-based localization and inflation

(g) Case 5: EnSRF **(h)** Case 6: EnSRF with distance-based localization

Figure 4.6: Final average permeability fields.

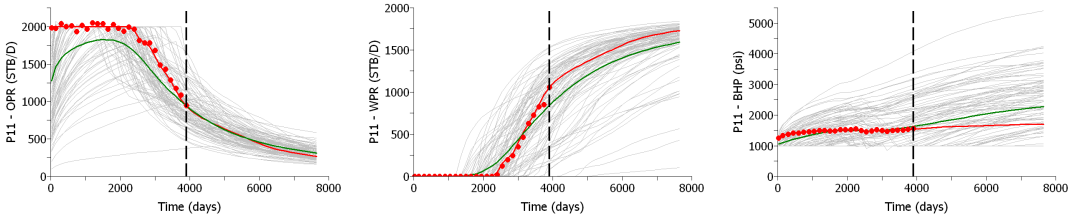
Table 4.3: Sum of normalized variance of log-permeability.

Case	SNV
N_e = 2,000	0.5631
Case 1: EnKF	0.0003
Case 2: EnKF with inflation	0.0005
Case 3: EnKF with dist. local.	0.3411
Case 4: EnKF with dist. local. and infl.	0.8499
Case 5: EnSRF	0.0022
Case 6: EnSRF with dist. local.	0.3829

Table 4.4 shows the average objective function for each case. For this example, we have N_d = 1,352 observed data points. Thus, according to the quality criterion of Eq. 2.132, the final normalized objective functions should not exceed 1.192. However, according to results in Table 4.4, none of the cases satisfy this quality criterion. EnKF with N_e = 2,000 violates the quality criterion value by a factor of 17. EnKF (case 1) and EnSRF (case 5) gave a very high value of O_m, indicating very rough models, and very high values of the data mismatch, O_d. In fact, for these cases, we obtained ensemble collapse. Inflation (case 2) did not improve the results. Distance-based localization, on the other hand, gave smoother models and significantly better data matches. EnSRF with localization (case 6) gave better results than EnKF with localization (case 3). The combined application of covariance localization and inflation resulted in higher values of the objective function than EnKF with covariance localization only. A possible explanation for the worse data matches in this case is that inflation might have increased inconsistency between the updated model parameters and the updated primary variables during the data assimilation. Figs. 4.7–4.14 show the predicted data for well P11. In these figures, besides the historical period, we also include 10 years of forecast. The results presented in these figures illustrate that reasonable data matches were obtained only by the cases with covariance localization and the case with N_e = 2,000.

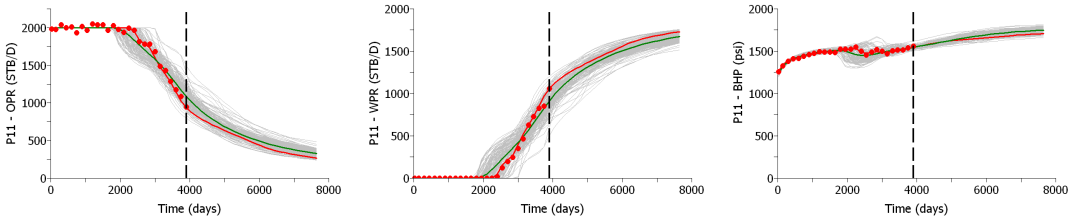
Table 4.4: Average objective function.

Case	$O_m(m)$	$O_d(m)$	$O_N = \frac{O(m)}{N_d}$
N_e = 2,000	1,709	25,891	20.42
Case 1: EnKF	49,547	27,198,550	20,153.92
Case 2: EnKF with inflation	47,381	17,134,080	12,708.23
Case 3: EnKF with dist. local.	1,416	41,274	31.58
Case 4: EnKF with dist. local. and infl.	2,705	157,900	118.79
Case 5: EnSRF	14,727	3,324,514	2,469.86
Case 6: EnSRF with dist. local	1,413	23,981	18.79



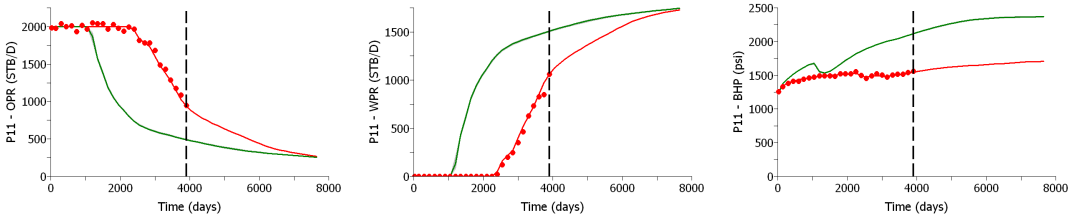
(a) Oil rate **(b)** Water rate **(c)** Bottomhole pressure

Figure 4.7: Prior ensemble. Predicted data for well P11. Red dots are the history; solid red line is the true model; solid green curve is the ensemble mean and solid gray curves represent all models of the ensemble.



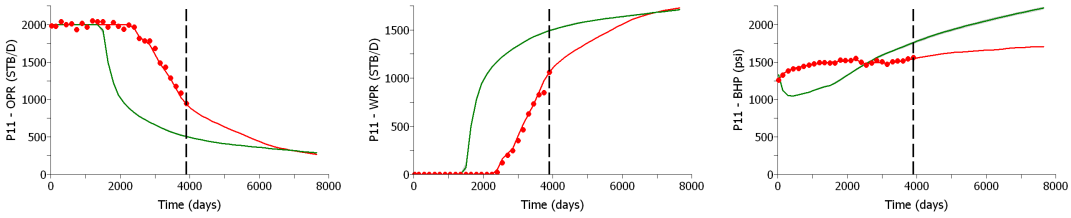
(a) Oil rate **(b)** Water rate **(c)** Bottomhole pressure

Figure 4.8: EnKF with N_e = 2,000. Predicted data for well P11 running final ensemble from time zero. The colors in this figure have the same meaning as in Fig. 4.7.



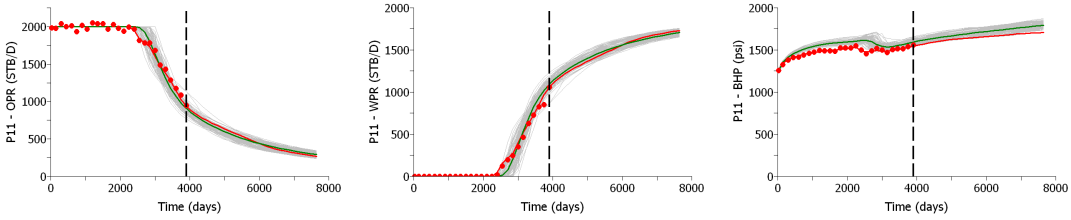
(a) Oil rate **(b)** Water rate **(c)** Bottomhole pressure

Figure 4.9: Case 1: EnKF. Predicted data for well P11 running final ensemble from time zero. The colors in this figure have the same meaning as in Fig. 4.7.



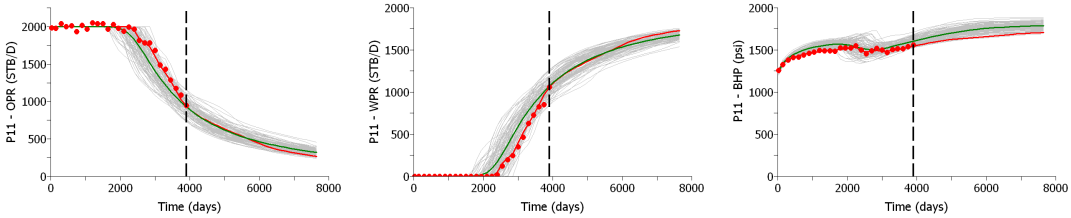
(a) Oil rate **(b)** Water rate **(c)** Bottomhole pressure

Figure 4.10: Case 2: EnKF with adaptive inflation. Predicted data for well P11 running final ensemble from time zero. The colors in this figure have the same meaning as in Fig. 4.7.



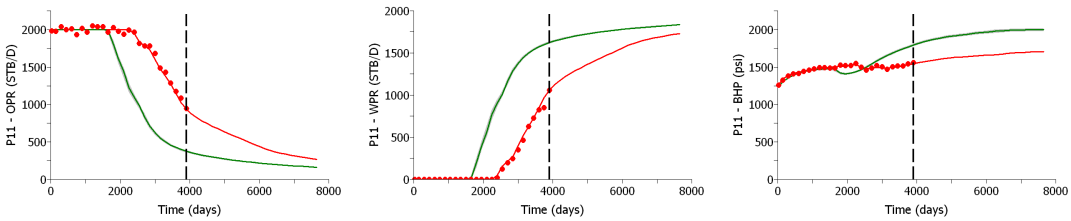
(a) Oil rate **(b)** Water rate **(c)** Bottomhole pressure

Figure 4.11: Case 3: EnKF with distance-based localization. Predicted data for well P11 running final ensemble from time zero. The colors in this figure have the same meaning as in Fig. 4.7.



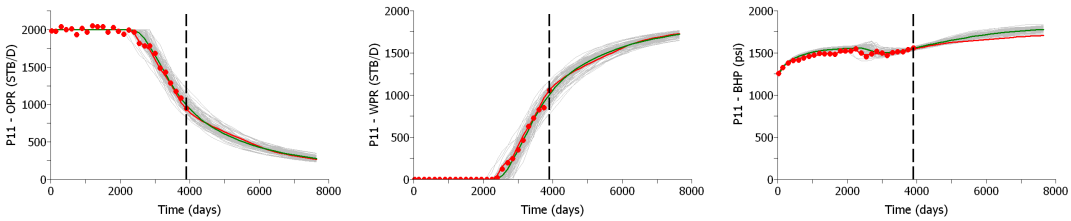
(a) Oil rate **(b)** Water rate **(c)** Bottomhole pressure

Figure 4.12: Case 4: EnKF with distance-based localization and inflation. Predicted data for well P11 running final ensemble from time zero. The colors in this figure have the same meaning as in Fig. 4.7.



(a) Oil rate **(b)** Water rate **(c)** Bottomhole pressure

Figure 4.13: Case 5: EnSRF. Predicted data for well P11 running final ensemble from time zero. The colors in this figure have the same meaning as in Fig. 4.7.



(a) Oil rate **(b)** Water rate **(c)** Bottomhole pressure

Figure 4.14: Case 6: EnSRF with distance-based localization. Predicted data for well P11 running final ensemble from time zero. The colors in this figure have the same meaning as in Fig. 4.7.

CHAPTER 5

COMBINING ENSEMBLE KALMAN FILTER AND MARKOV CHAIN MONTE CARLO

One problem that often occurs when EnKF is applied to reservoir history-matching problems is that the values of the objective function corresponding to the models of the final ensemble are relatively high, especially when compared to gradient-based history matching. In fact, for most of the reservoir history-matching examples considered in the previous two chapters, we observed that the final values of the normalized objective function obtained by EnKF cases did not come close to satisfying the criterion of Eq. 2.132. Large values of the objective function occur because of poor data matches, or rough models, or both. However the statistical interpretation of high values of the objective function is perhaps more important. A model which gives a high value of the objective function corresponds to a sample of a low probability region of the posterior pdf, which is clearly undesirable, especially if the objective is to generate a good sampling of the posterior pdf in order to have a reliable characterization of uncertainty.

This chapter presents a procedure which combines EnKF and MCMC for the purpose of improving the final data matches and obtaining a more accurate sampling of the posterior pdf for reservoir model parameters. The results presented in this chapter were published in Emerick and Reynolds [38].

5.1 EnKF-MCMC

5.1.1 Proposing realizations from the square root

For multivariate Gaussian variables, it is possible to generate realizations from a square root of the covariance matrix. Based on that fact, Emerick and Reynolds [35] proposed computing the square root of the posterior covariance based on the final ensemble obtained from EnKF and using this square root to generate independent conditional realizations. The outline of the procedure is as follows: (i) Generate an ensemble of vectors of model parameters by assimilating data with EnKF. (ii) Apply SVD to estimate the square root of the covariance matrix approximated from this ensemble of model parameters. (iii) Use this square root and the ensemble mean to propose new states in the Markov chain in the implementation of MCMC. Note that this mechanism for proposing transitions in the Markov chain means that the proposed states in the chain are generated from a Gaussian pdf with a mean equal to the ensemble mean and a covariance matrix given by the approximate covariance matrix estimated from the final EnKF ensemble of model parameters. In a synthetic reservoir problem, the EnKF-MCMC method suggested in [35] gave better data matches and a better characterization of uncertainty in reservoir performance predictions than were obtained by EnKF even though it is clear that the Markov chains generated were not sufficiently long to complete the burn-in period.

There are two main problems associated with the EnKF-MCMC algorithm described in the previous paragraph. (a) First, applying the Metropolis-Hastings criterion for determining whether to accept a proposed transition in the chain requires running the reservoir simulator to evaluate the likelihood part of the posterior pdf. Thus, for a realistically-sized simulation model, it is not computationally feasible to consider more than a few thousand proposals at the very most. Because of the high computational cost, Emerick and Reynolds [35] considered only chains generated using 1,000 proposed transitions and thus, as mentioned above, MCMC was termi-

nated while it was still in the burn-in period. (b) Secondly, because a new state (new vector of model parameters) is proposed from the square root of the approximate covariance matrix in step (iii), any proposed state is a vector in the subspace spanned by the final ensemble of vectors of model parameters obtained with EnKF. Thus, it is doubtful that we can generate an accurate characteristic of the posterior pdf for any predicted outcome, e.g., cumulative water production at some future time or the predicted oil rate at a well, if states in the chain are restricted to this subspace. This difficulty in correctly characterizing uncertainty is to be expected because EnKF is a stochastic process, and the mean and covariance for the vector of model parameters estimated from the final EnKF ensemble depends on the initial ensemble.

In this chapter, we extend the basic ideas of Emerick and Reynolds [35] in a way that eliminates problem (a) of the preceding paragraph and at least ameliorates the negative effects that arise from problem (b). In the modified algorithm presented here, after assimilating data with EnKF, we run the simulator with each member of the ensemble of vectors of model parameters from time zero to compute the corresponding primary simulator variables associated with each realization of model parameters. Then, we compute the square root of the covariance of the resulting state vector which includes both model parameters and primary variables. Next, somewhat similar to the original Emerick and Reynolds [35] EnKF-MCMC algorithm outlined above, the ensemble mean and the square root of this covariance matrix are used to propose new states in the Markov chain. As the proposed states include values of the simulation primary variables, the predicted well data associated with the proposed state can be calculated by a direct application of Peaceman's equation [136]. With these predicted data, the likelihood part of the objective function can be evaluated without running the simulator and so the Metropolis-Hastings probability of accepting the proposal as the next state in the Markov chain can be calculated without running the simulator. Thus, the computational cost of running

a very long chain is reasonable. This eliminates problem (a) of the original EnKF-MCMC algorithm proposed in [35]. To reduce the effects of problem (b) and further improve the method, three more modifications are introduced. First, we apply EnKF multiple times, where each time we start with a different initial ensemble of state vectors. Secondly, from each final ensemble, we generate N_e Markov chains, each chain starting with a different initial state, and simply select the final state in each chain as a sample of the posterior pdf. If EnKF is applied N_f times, then $N_s = N_f \times N_e$ samples are obtained. Thirdly, we run the reservoir simulation with each vector of model parameters in the combined ensemble of N_s samples and resample based on the actual values of the normalized objective function.

From this point, we refer to this method simply as EnKF-MCMC. The description of the EnKF-MCMC follows: After data assimilation with EnKF, we rerun the final ensemble of vectors of model parameters from time zero (initial reservoir condition) to compute the corresponding simulator primary variables. Then, we build the matrix

$$\Delta Y = Y - \overline{Y} = \begin{bmatrix} \Delta M \\ \Delta P^1 \\ \Delta P^2 \\ \vdots \\ \Delta P^{N_t} \end{bmatrix}. \tag{5.1}$$

Here, $\Delta M = M - \overline{M}$, where M is the $N_m \times N_e$ matrix with the jth column equal to the jth realization of the vector of model parameters obtained by data assimilation with EnKF, and $\overline{M}$ is the $N_m \times N_e$ matrix with all columns equal to $\overline{m}$, which represents the average of all columns of M, i.e., the ensemble mean. Similarly, $\Delta P^n = P^n - \overline{P}^n$, where P^n is the $N_p \times N_e$ matrix with its jth column equal to the jth vector of reservoir simulator primary variables at the nth data assimilation step.

Each column of $\overline{P}^n$ is equal to $\overline{p}^n$, the average of the columns of P^n. N_t denotes the total number of data assimilation time-steps. We include primary variables in ΔY so that we can compute predicted well data corresponding to observed data using Peaceman's equation [136] without running the reservoir simulator when applying the MCMC algorithm. Thus, we need to include in P^n only the primary simulation variables at gridblock locations in which there are wells completed, and by doing so, we reduce significantly the size of the matrix ΔY. An alternative approach would be to include in the matrix ΔY only the predicted data obtained by running the ensemble from time zero. In this case, one could estimate directly the well data without applying Peaceman's equation. While this procedure would be much easier to implement, we believe that our approach tends to give better predictions because the relation between model parameters and primary reservoir simulator variables tends to be more linear than the relation between model parameters and predicted well data.

Let $N_y = N_m + N_t \times N_p$ denote the number of rows of the matrix ΔY. Assuming that the final ensemble describes a Gaussian distribution, we can propose independent conditional realizations, $\widehat{y}$, using

$$\widehat{y} = \overline{y} + \widetilde{C}_{\mathrm{Y}}^{1/2} z. \tag{5.2}$$

Here, $\overline{y}$ is the average of the columns of Y; z is a N_y-dimensional column vector of random independent standard normal deviates, i.e., $z \sim \mathcal{N}(0, I_{N_y})$; and $\widetilde{C}_{\mathrm{Y}}^{1/2}$ is the square root of the posterior covariance matrix, which is estimated using the final ensemble from EnKF. To compute $\widetilde{C}_{\mathrm{Y}}^{1/2}$, we first calculate the SVD of ΔY, which can be written as

$$\Delta Y = U \Lambda V^{\mathrm{T}}, \tag{5.3}$$

where U is the $N_y \times N_y$ orthogonal matrix with its jth column equal to the jth left

singular vector; Λ is the $N_y \times N_e$ matrix with all nonzero entries corresponding to singular values, and V is the $N_e \times N_e$ orthogonal matrix of right singular vectors. Considering only the N_r nonzero singular values of ΔY, we can write the posterior covariance matrix as

$$\widetilde{C}_{\mathrm{Y}} = \frac{\Delta Y \Delta Y^{\mathrm{T}}}{N_e - 1} = \frac{(U_r \Lambda_r)(U_r \Lambda_r)^{\mathrm{T}}}{N_e - 1} = \widetilde{C}_{\mathrm{Y}}^{1/2} \widetilde{C}_{\mathrm{Y}}^{\mathrm{T}/2}, \tag{5.4}$$

where U_r is the $N_y \times N_r$ matrix of left singular vectors corresponding to nonzero singular values, and Λ_r is the $N_r \times N_r$ diagonal matrix with nonzero singular values. From Eq. 5.4, one can define the square root of the posterior covariance as

$$\widetilde{C}_{\mathrm{Y}}^{1/2} = \frac{U_r \Lambda_r}{\sqrt{N_e - 1}}, \tag{5.5}$$

which is a $N_y \times N_r$ matrix and thus not square. Instead, we define the square root of $\widetilde{C}_{\mathrm{Y}}$ as

$$\widetilde{C}_{\mathrm{Y}}^{1/2} = \frac{U_r \Lambda_r U_r^{\mathrm{T}}}{\sqrt{N_e - 1}}, \tag{5.6}$$

which is a $N_y \times N_y$ symmetric matrix. Because $U_r^{\mathrm{T}} U_r = I_{N_r}$, where I_{N_r} denotes the $N_r \times N_r$ identity matrix, $\widetilde{C}_{\mathrm{Y}}^{1/2}$ given by Eq. 5.6 is still a square root of $\widetilde{C}_{\mathrm{Y}}$.

In realistic reservoir history-matching applications, we have $N_e \ll N_y$ and thus $N_r \leq N_e$. Typically, we have $N_r = N_e - 1$. Note that N_y may be very large. However, we do not need to compute or store the $N_y \times N_y$ matrix $\widetilde{C}_{\mathrm{Y}}^{1/2}$. Instead, we store only the $N_y \times N_r$ matrix U_r and the N_r diagonal elements of Λ_r corresponding to the nonzero singular values of ΔY. Then, we compute the N_r-dimensional vector x by

$$x = \Lambda_r U_r^{\mathrm{T}} z \tag{5.7}$$

and compute the new proposal for the Markov chain using

$$\widehat{y} = \overline{y} + \frac{1}{\sqrt{N_e - 1}} U_r x, \tag{5.8}$$

which involves the product of a $N_y \times N_r$ matrix and a N_r-dimensional vector.

We use Eq. 5.8 to propose states for the MCMC method. From Eq. 5.8, we note that any state proposed during MCMC is in the subspace spanned by the N_r left singular vectors of ΔY contained in U_r. After proposing the new state $\widehat{y}$, we effectively have a new model and the corresponding values of primary simulation variables at gridblocks that contain wells at all data assimilation time-steps. At this point, we have made the underlying assumption that those primary variables are statistically consistent with the model parameters. A similar assumption is present in the EnKF method. Once we compute the vector $\widehat{y}$, we can easily compute the predicted well data by the standard application of Peaceman's equation. For example, for a well completed in a single gridblock, we could compute the predicted oil rate, q_o, at the nth time-step using

$$q_{\mathrm{o}}^n = WI \frac{k_{\mathrm{ro}}(S_{\mathrm{o}}^n)}{B_{\mathrm{o}}(p^n)\mu_{\mathrm{o}}(p^n)} \left(p^n - p_{\mathrm{wf}}^n\right), \tag{5.9}$$

where WI is the well index computed with Peaceman's equivalent radius; $k_{\mathrm{ro}}(S_{\mathrm{o}}^n)$ is the oil relative permeability, which is a function of the gridblock oil saturation, S_{o}^n; $B_{\mathrm{o}}(p^n)$ is the oil formation volume factor and $\mu_{\mathrm{o}}(p^n)$ is the oil viscosity, which are both functions of the gridblock pressure, p^n; p_{wf}^n is the specified well bottomhole pressure. The values of p^n and S_{o}^n are components of the state vector $\widehat{y}$, computed using the square root scheme discussed before.

As discussed in Chapter 2 (Section 2.2), for reservoir history-matching problems, the application of MCMC with global perturbations does not seem to be feasible because the acceptance rate is too small to sample the posterior pdf in a reasonable number of proposals. Therefore, it is far more efficient to generate chains using local perturbations. Here, we use the local perturbation procedure described in Section 2.2

(Eq. 2.18). Unfortunately, in a Markov chain built using local perturbations, most of the states are correlated, an thus we obtain only a few "independent" samples of the posterior pdf. One option to ameliorate this problem is to run an extremely long chain to try to ensure that enough independent samples are generated in this chain. Here, we consider an alternative approach. To obtain N_e new samples, we run N_e chains starting from different models. We run each chain long enough to pass through the burn-in period and select only one model as a sample, the last accepted state in the chain. Because we start each chain from an independently proposed initial state, we obtain N_e independent samples using this procedure. Note that it is more efficient to run several small chains than one single very long chain because we can run the chains in parallel on multiple computers, reducing the time required to generate the new samples.

5.1.2 Comments on underestimation of posterior covariance

The proposed procedure uses the final ensemble from EnKF to estimate the posterior covariance of model parameters. As we use the square root of this covariance matrix, any model proposed during MCMC is in the subspace spanned by the vectors of model parameters corresponding to the final ensemble from EnKF. Thus, if the final ensemble from EnKF does not span the whole pdf, MCMC will not provide samples which span the whole pdf either. This is an important limitation because EnKF tends to underestimate posterior variances of reservoir model parameters after data assimilation [172, 57, 6]. Covariance localization usually reduces this problem, but we still expect some underestimation.

As will be shown in the example section, if we apply EnKF for the same problem with different initial ensembles, we obtain very different distributions of predicted cumulative oil and water production, which indicates that the resulting models from EnKF are not samples of the same pdf, or, at least, they are samples from different parts of the same pdf. Therefore, uncertainty is not reasonably char-

acterized by a single ensemble of 100 models using EnKF. The same conclusion was obtained by Lorentzen et al. [113] using the PUNQ-S3 case [55]. Thulin et al. [168] also concluded that with only a single EnKF run, there is no control over the Monte Carlo error in the estimate of the conditional pdf. They proposed performing various data assimilations with EnKF and using as the final distribution of predictions, the average cdf from all data assimilations. Although this procedure may improve the estimate of the conditional cdf, there is still a problem related to the poor data matches obtained by EnKF. As shown later, applying EnKF-MCMC results in better data matches and predictions, but the resulting distributions are still different. In order to alleviate this problem, we propose repeating EnKF-MCMC with different initial ensembles so that we expect to generate samples of different parts of the posterior pdf.

5.1.3 Resampling based on the normalized objective function

As discussed in the previous section, in order to improve sampling, we apply EnKF-MCMC with different initial ensembles. Then, we can assume that all the resulting models of all EnKF-MCMC runs provide the final sampling of the posterior pdf. However, because our MCMC procedure is based on an approximate likelihood, we cannot expect to sample the posterior pdf correctly, although we do expect to improve the sampling compared to EnKF. In practice, we observe that some models resulting from EnKF-MCMC will not provide reasonable data matches after we run the reservoir simulator. As a result, we observe unreasonably high variance in the distributions of reservoir predictions (e.g., a high variance in predicted cumulative oil production). Hence, in order to have a reasonable characterization of uncertainty, we need to eliminate models with unacceptable data matches. From a Bayesian viewpoint, we wish to eliminate the samples from negligible probability regions of the posterior pdf.

The obvious way to remove these low probability samples is by using impor-

tance sampling (IS) [120, 33, 34]. IS is the basis of particle filters [33, 34] where the samples (particles) are re-weighted based on the likelihood (or the target pdf) evaluated at the sample. In this case, it is straightforward to show that each sample get an importance weight w_j given by

$$w_j = \frac{\pi(m_j)}{\sum_{k=1}^{N_s} \pi(m_k)} = \frac{\exp(-O(m_j))}{\sum_{k=1}^{N_s} \exp(-O(m_k))}, \text{ for } j = 1, 2, \cdots N_s, \tag{5.10}$$

where $O(m)$ is the objective function given by Eq. 2.3. However, when the objective function values of the samples are unevenly distributed, computing weights using Eq. 5.10 leads to a situation in which most of the samples have weights essentially equal to zero and only very few samples have non-negligible weights. In the particle filter literature, this phenomenon is known as particle degeneracy [34]. In the extreme case, the sampling collapses to a single particle. In practice, we observe that re-weighting the history-matched models based on the likelihood (or even the on target pdf) results in sampling collapse. This problem is well known in the literature, and it is often attributed to the so-called "curse-of-dimensionality" [12], which states that the sample size needs to increase exponentially with the dimensionality of the problem to avoid collapse of the particles.

Here, we propose an approximate procedure inspired by IS and particle filters. We re-weight the samples based on the target pdf evaluated at the normalized objective function, i.e., we compute the weights using

$$w_j = \frac{\exp(-O_N(m_j))}{\sum_{k=1}^{N_s} \exp(-O_N(m_k))}, \text{ for } j = 1, 2, \cdots N_s, \tag{5.11}$$

where $O_N(m) = O(m)/N_d$ is the normalized objective function. Re-weighting based on the normalized objective function does not represent a rigorous resampling procedure because one no longer can prove that the resulting weights results into a correct

sampling of the posterior pdf. However, this procedure approximately does what we wish. First, normalizing the objective function by the number of data points significantly reduces the sampling collapse problem. Secondly, with this procedure, we are still able to remove models which result in unreasonable data matches and increase the importance of more probable models when generating distributions of reservoir predictions. We use the normalized objective function because the values of $O_N(m)$ for models that result in a good history match are typically of the same order of magnitude regardless of the problem and the number of data assimilated.

5.1.4 Summary of EnKF-MCMC procedure

Algorithm 5.1 presents the EnKF-MCMC procedure applied in this chapter. This procedure is repeated for N_f different initial ensembles so that a total of $N_s = N_f \times N_e$ samples are generated. Then, we run the reservoir simulator for all resulting models, compute the normalized objective function for each model and resample using Eq. 5.11. If the final data matches obtained after running the reservoir simulator are still not satisfactory, we can repeat steps 2 and 3 of Algorithm 5.1 until we do not observe significant improvements in the objective functions, i.e., the algorithm can be applied iteratively.

Algorithm 5.1 EnKF-MCMC

1. Perform data assimilation using EnKF with covariance localization.
2. Run the final ensemble from time zero, build the matrix ΔY (Eq. 5.1) and compute the matrices U_r and Λ_r using SVD. These matrices are used to estimate the square root of the state covariance matrix (Eq. 5.6).
3. Build N_e Markov chains to generate N_e samples of the posterior pdf. For each Markov chain,
   - Propose the first state using Eq. 5.2 (or equivalently, Eq. 5.8).
   - Propose the remaining states using local perturbations.
   - Keep the last accepted state as one sample.

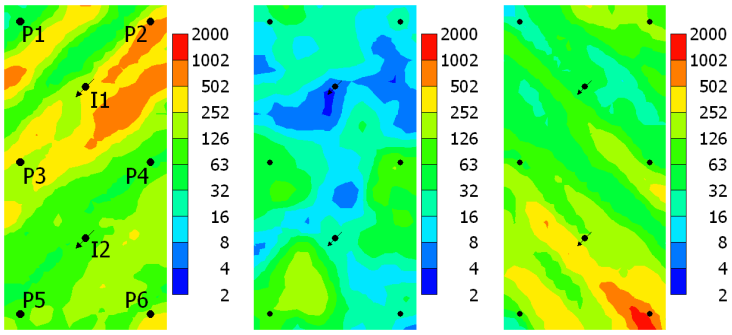
5.2 Example

The example is a synthetic reservoir case on a $15 \times 30 \times 3$ uniform grid. The dimensions of the gridblocks are 200 ft $\times$ 200 ft $\times$ 25 ft. The model parameters are gridblock log-permeabilities. The true model was built using an exponential covariance function with the geostatistical parameters given in Table 5.1. Note that Table 5.1 indicates that the geostatistical parameters of each layer are different. There is no correlation between layers.

Table 5.1: Geostatistical parameters for the reservoir model.

	Layer 1	Layer 2	Layer 3
Prior mean of $\ln(k)$	5.0	3.5	4.5
Prior variance of $\ln(k)$	1.0	1.0	1.0
Major correlation length	12,000 ft	2,000 ft	12,000 ft
Minor correlation length	1,000 ft	2,000 ft	1,000 ft
Azimuth	45^o	0^o	135^o

Fig. 5.1 shows the "true" permeability field used as reference to generate the observed data. For this reservoir, there are six producing wells and two water injectors forming two five spots. All producing wells are controlled by a fixed bottomhole pressure of 2,500 psi. The injectors are controlled by a fixed bottomhole pressure of 3,500 psi. The observed data are 11 years of oil and water production rates and water injection rates. To generate the observed data, we added to the well data predicted by the true model Gaussian random noise with zero mean and standard deviation equal to 5% of the true data. We assumed that the measurement errors are uncorrelated so that the covariance matrix C_{D} is diagonal. We assimilated data every 150 days so for the total historical period, the number of observed data points is $N_d = 378$. The value of the normalized objective function of the true model is $O_N(m_{\mathrm{true}}) = 2.32$.



(a) Layer 1 **(b)** Layer 2 **(c)** Layer 3

Figure 5.1: True permeability field (mD).

5.2.1 Single ensemble

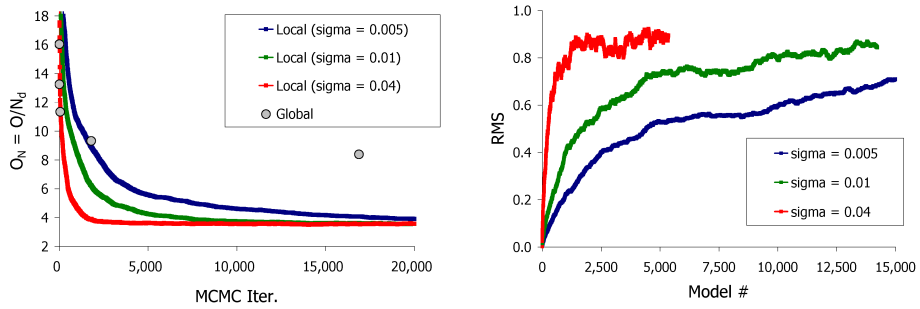
Using the same geostatistical parameters presented in Table 5.1, we generated an ensemble of 100 models and performed data assimilation using EnKF with and without covariance localization. Because localization tends to keep the variability in the final ensemble higher, we used the final ensemble from EnKF with covariance localization to propose states for the Markov chain. For localization, we used the procedure described in Chapter 3 based on the prior correlation and drainage areas.

Before applying the complete EnKF-MCMC algorithm, we performed a sensitivity study to define the parameters of the Markov chains. We ran four chains with a total length of 20,000 proposals each. The first 3 chains were run using local perturbations with different scaling factors, 0.005, 0.01 and 0.04. The fourth chain was run with global perturbations. Fig. 5.2a presents the values of the normalized objective function of the accepted models in the chains. Table 5.2 summarizes the results for each chain. According to the results in Fig. 5.2a and Table 5.2, the chain with local perturbations and a scaling factor of 0.005 seems to be still in the transition period after 20,000 proposals. For scaling factors of 0.01 and 0.04, on the other hand, the chains seem to stabilize after approximately 10,000 proposals for $\sigma = 0.01$ and 4,000 proposals for $\sigma = 0.04$. We assume that this stabilization period represents the burn-in period so that from this point onward, the accepted models

represent samples of the target pdf. The number of accepted proposals in the three chains with local perturbations is very high (see third column of Table 5.2). For $\sigma = 0.04$, we obtained an acceptance rate of 0.268, which is close to the "optimal" value of 0.234. However, as each new proposal is obtained by a small perturbation around the current state, we expect consecutive states in the chain to be highly correlated. It is important to have a high acceptance rate in the Markov chain, but it is more important to have good mixing, i.e., it is necessary to generate samples from all parts of the distribution so that the variability is correct, rather than having a very large number of similar successive states. For global perturbations, all accepted models are independent, but, in this case, only 6 states were accepted and based on the objective function values obtained, the chain does not seem to converge after 20,000 proposals. Although the results are not presented here, we performed some additional experimentation using chains with global perturbations, and we observed that, even for very long chains (more than one million proposals), the number of accepted states after the burn-in period is very small (on the order of 5 to 10 models). In order to evaluate the number of independent states in the chains with local perturbations, we computed the root mean squared (RMS) of the difference between the initial state and all accepted states in the chains. The RMS was computed using

$$\text{RMS}_j = \sqrt{\frac{1}{N_m}\|m_j - m_0\|_2^2}, \text{ for } j = 1, 2, \cdots N_{\text{acc}}, \tag{5.12}$$

with j denoting the jth accepted state and N_{acc} denoting the number of accepted states in the chains. We can interpret RMS_j as a measure of the distance between the first and the jth state in the chain. Fig. 5.2b presents the RMS values obtained for each chain with local perturbations. According to the results in this figure, for the chains with $\sigma = 0.005$ and $\sigma = 0.01$, the values of RMS are still increasing at the end of the chain. This suggests that all states in these chains are correlated with the first state m_0. For $\sigma = 0.04$ after approximately 2,000 accepted states, RMS



(a) Normalized objective function of accepted states

(b) RMS

Figure 5.2: Markov chains for different proposal procedures.

stops increasing and oscillates around 0.87, which suggests that after this point, the states are independent of the first state although any two particular states may be correlated. Because of these long correlation lengths in the chains with local perturbations, only a small percentage of the accepted states are independent.

Table 5.2: Summary of Markov chains.

Case	Burn-in	Average O_N	Accepted states	Acceptance rate
Local ($\sigma = 0.005$)	$>$ 20,000	3.897	15,959	0.798
Local ($\sigma = 0.01$)	$\simeq$ 10,000	3.618	14,602	0.730
Local ($\sigma = 0.04$)	$\simeq$ 4,000	3.556	5,365	0.268
Global pert.	$\gg$ 20,000	8.378	6	0.0003

To apply the complete EnKF-MCMC algorithm, we use chains based on local perturbations ($\sigma = 0.04$) and 10,000 proposals, which, based on the results of Fig. 5.2, is long enough to pass through the burn-in period. Each new sample generated from EnKF-MCMC was obtained by running a new Markov chain starting from an independently proposed initial state. Thus, we ran 100 chains to generate 100 new samples. Fig. 5.3 shows the normalized objective function obtained for each model using EnKF, EnKF with covariance localization and EnKF-MCMC. Note that the values of the normalized objective function presented in Fig. 5.3 correspond to the

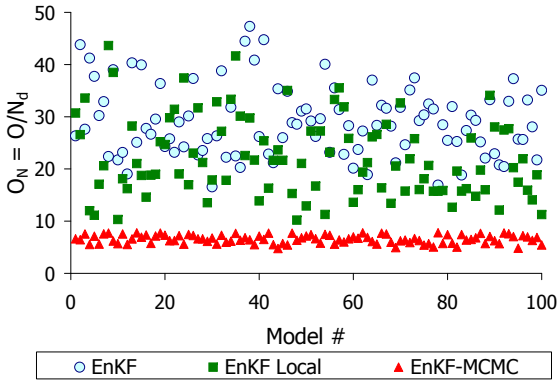


Figure 5.3: Normalized objective function.

actual ones, i.e., they were obtained by running the final set of models from time zero. Table 5.3 presents the average values of the objective functions for each case. In this table, we present model mismatch, $O_m(m)$, the data mismatch, $O_d(m)$, the total, $O(m) = O_m(m) + O_d(m)$, and the normalized objective function, $O_N = O(m)/N_d$. According to the results in Fig. 5.3 and Table 5.3, EnKF gave models with relatively high values of the objective function, especially in the data mismatch part, $O_d(m)$. The final values of the objective function after MCMC are more than three times smaller than the values obtained with the two EnKF procedures. This means that EnKF-MCMC generated samples that have much higher values of posterior pdf than the samples obtained with EnKF.

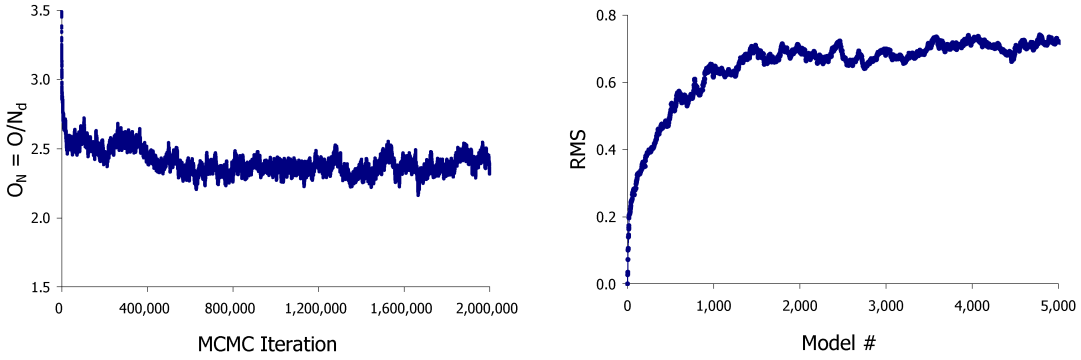
Table 5.3: Average objective function.

Case	$O_m(m)$	$O_d(m)$	$O(m)$	$O_N = \frac{O(m)}{N_d}$
EnKF	2429.0	8550.2	10979.0	29.05
EnKF with localization	1056.5	7412.7	8469.2	22.41
EnKF-MCMC	862.7	1612.0	2474.7	6.55

For comparisons, we ran a Markov chain with two million states proposing models from the square root of the prior covariance matrix, computed using Cholesky decomposition [135, p. 288–289]. In this chain, we used the local perturbation pro-

cedure presented by Oliver et al. [133]. In this procedure, one entry of the random vector corresponding to the current state in the chain is replaced by a sample from $\mathcal{N}(0,1)$. For this chain, each proposal required a reservoir simulation to apply the Metropolis-Hastings acceptance criterion. The total number of accepted states was 206,059 which corresponds to an acceptance rate of 0.103. We initialized the chain with the model obtained by EnKF without covariance localization which gave the lowest value of objective function. Fig. 5.4a presents the normalized objective function of the accepted states in this chain. According to the results in this figure, the transition period lasts until we have accepted about 20,000 proposals; after that, the values of O_N oscillate around 2.4. Here, we assume that this Markov chain is long enough to provide a reasonable approximation of the conditional distributions for predicted cumulative oil and water production. In the rest of this paper, we refer to this case as long MCMC case. Fig. 5.4b presents the root mean square (RMS) of the difference between the initial state and the first 5,000 accepted states in this chain. The results presented in this figure show that after approximately 2,000 accepted states, the RMS stops increasing, which indicates that after this point onward, the models are independent of the initial state. Assuming that this number approximately holds for the whole chain, as we have a total of 206,059 accepted states, we can roughly assume that we have about 103 independent samples. Fig. 5.5 shows the resulting histograms of predicted cumulative oil and water production for the long MCMC case. These histograms are based on a total time of 21 years, 11 years of historical data and 10 years of future predictions. The cumulative production obtained from the true model is also included in these and all other histograms presented in this chapter.

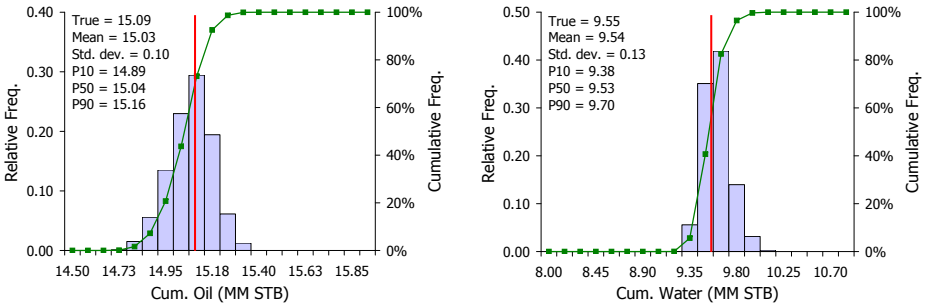
Fig. 5.6 presents the histograms of cumulative oil and water production obtained from EnKF, EnKF with localization and EnKF-MCMC. According to the results shown in Fig. 5.6, EnKF with and without covariance localization resulted



(a) Normalized objective function of accepted states

(b) RMS

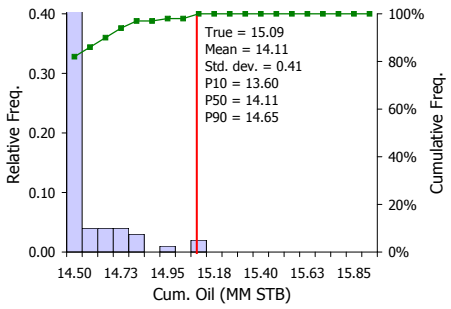
Figure 5.4: Long Markov chain proposing model from prior.



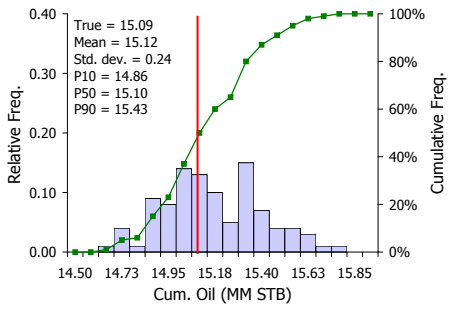
(a) Cumulative oil production

(b) Cumulative water production

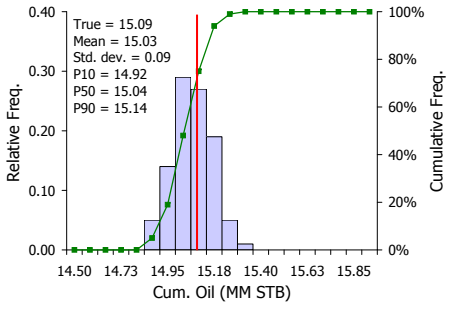
Figure 5.5: Histograms of cumulative oil and water production for the long MCMC case.



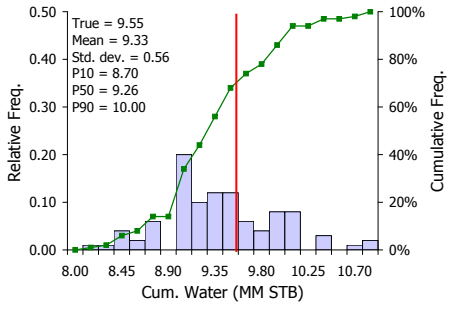
(a) Cumulative oil, EnKF



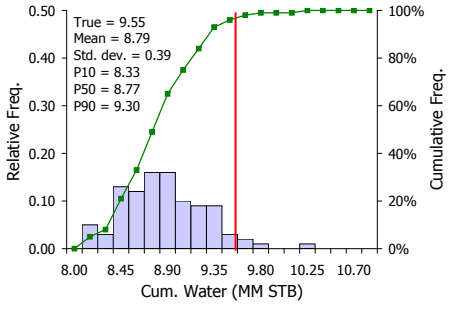
(b) Cumulative oil, EnKF with localization



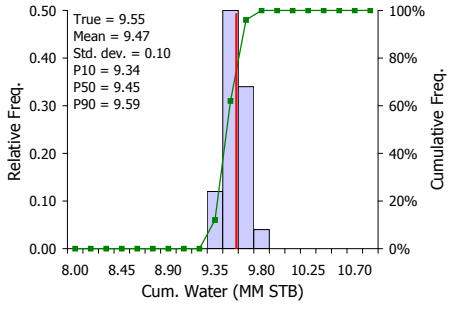
(c) Cumulative oil, EnKF-MCMC



(d) Cumulative water, EnKF



(e) Cumulative water, EnKF with localization

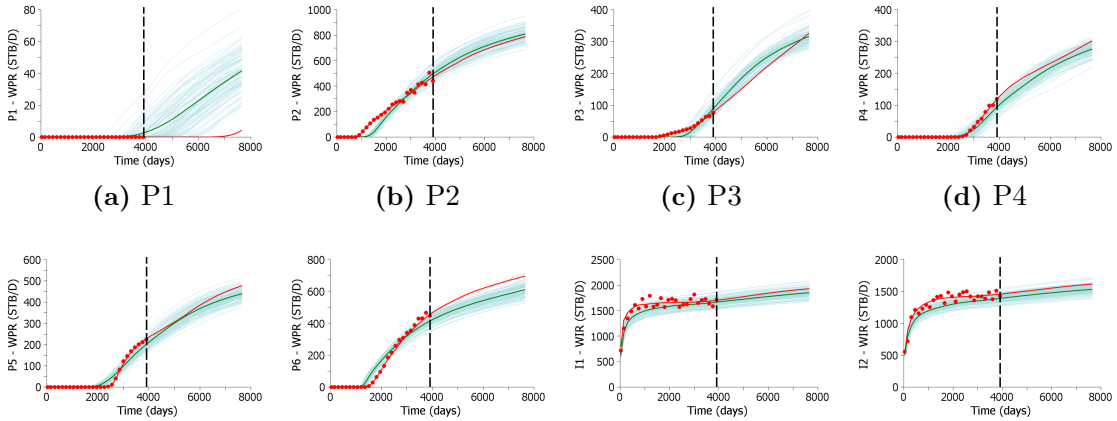


(f) Cumulative water, EnKF-MCMC

Figure 5.6: Histograms of cumulative oil and water production. Figures also include the cumulative oil and water production for the true model (in red).

in histograms with a much larger spread of predicted oil and water production compared to those for the long MCMC case (Fig. 5.5). These large spreads are a consequence of the poor data matches obtained in the EnKF cases. EnKF-MCMC narrowed the spread of predicted oil and water production, resulting in histograms much closer to the ones obtained with the long MCMC case (compare Figs. 5.6c and 5.6f with Figs. 5.5a and 5.5b). Also note that for this example, the histograms from the EnKF cases are biased compared to the long MCMC (see Figs. 5.6a and 5.6e). Figs. 5.7–5.9 present the predicted water rate for EnKF, EnKF with localization and EnKF-MCMC. These figures clearly illustrate the improvement in the data matches obtained with EnKF-MCMC. Unfortunately, it is also clear from these figures that EnKF-MCMC underestimated the uncertainty in the predictions. For example, for wells P1 and P6, none of the methods resulted in predictions that span the water production rate obtained by running the true model.

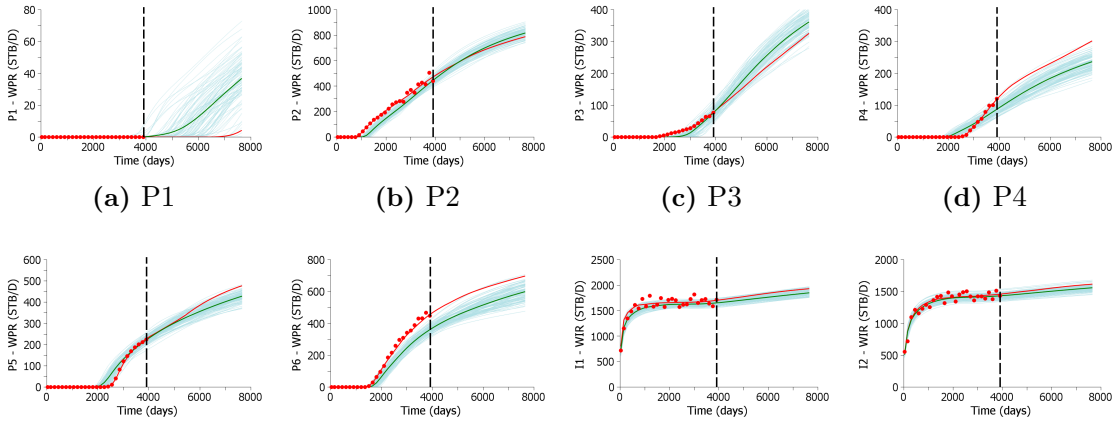
In summary, the application of EnKF-MCMC improved the sampling results obtained from EnKF by generating samples from higher probability regions of the posterior pdf. However, because the proposed models for MCMC are restricted to the subspace spanned by the final ensemble obtained from EnKF, EnKF-MCMC cannot generate samples from all parts of the posterior pdf, which results in underestimation of uncertainty.



(a) P1 **(b)** P2 **(c)** P3 **(d)** P4

(e) P5 **(f)** P6 **(g)** I1 **(h)** I2

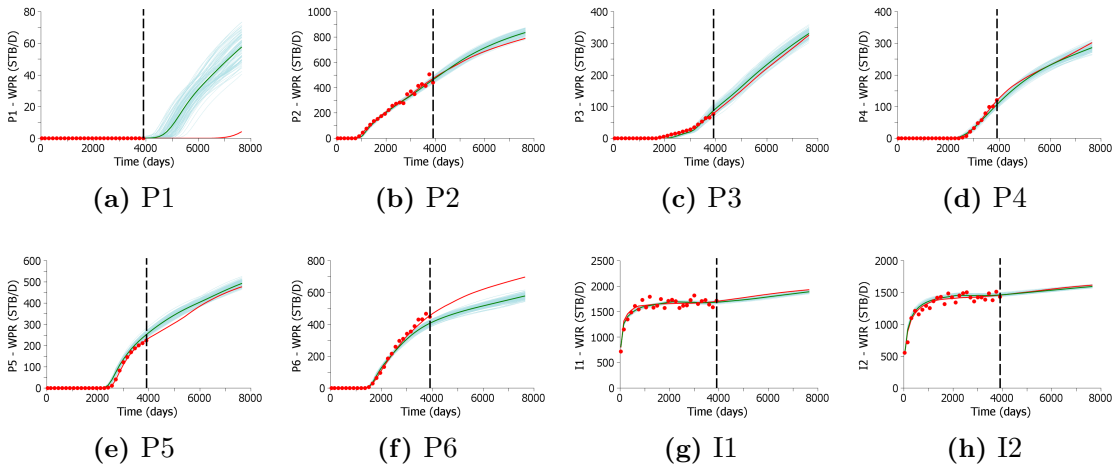
Figure 5.7: EnKF. Predicted water rate by running the final ensemble from time zero (stb/day). Red dots are the history; solid red curve is the true model; solid green curve is the mean prediction computed from the ensemble; and solid light blue curves represent all ensemble predictions. The dashed vertical black line indicates the end of the history and beginning of the forecast period.



(a) P1 **(b)** P2 **(c)** P3 **(d)** P4

(e) P5 **(f)** P6 **(g)** I1 **(h)** I2

Figure 5.8: EnKF with covariance localization. Predicted water rate by running the final ensemble from time zero (stb/day). The colors in this figure have the same meaning as in Fig. 5.7.



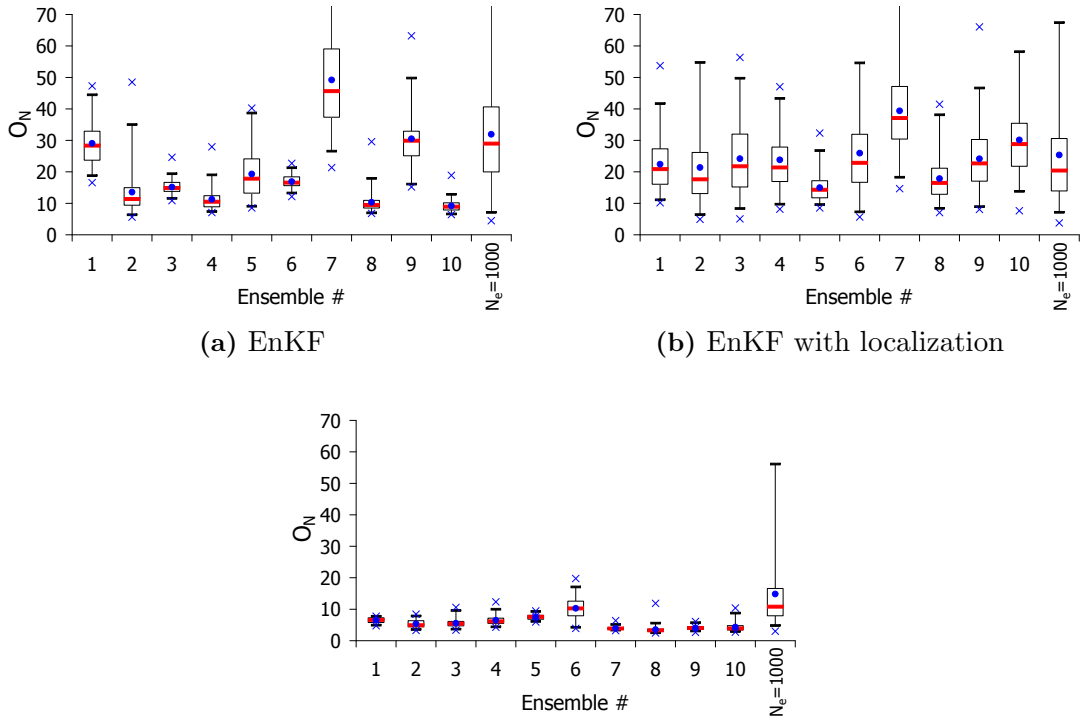
(a) P1 **(b)** P2 **(c)** P3 **(d)** P4

(e) P5 **(f)** P6 **(g)** I1 **(h)** I2

Figure 5.9: EnKF-MCMC. Predicted water rate by running the final ensemble from time zero (stb/day). The colors in this figure have the same meaning as in Fig. 5.7.

5.2.2 Multiple ensembles

From the results of the previous section, we conclude that with one ensemble of 100 models, it is not always possible to obtain a reasonable quantification of the uncertainty in production predictions using EnKF with or without localization, or even with EnKF-MCMC. Here, we extend this example by repeating the process of the previous section with ten different initial ensembles of 100 models each. Then, we combine the results from all ten ensembles and resample based on the normalized objective function. Here, we also consider the case with a single ensemble of 1,000 models. This ensemble was obtained by combining the ten initial ensembles of 100 models. For this single ensemble of 1,000 models, we applied EnKF-MCMC after EnKF with covariance localization by generating 1,000 chains starting with 1,000 different initial states. Fig. 5.10 shows box plots of O_N obtained for each ensemble using EnKF, EnKF with covariance localization and EnKF-MCMC. According to the results in Fig. 5.10, EnKF-MCMC improved significantly the data matches for all ensembles. It is interesting to note that increasing the ensemble size from 100 to 1,000 did not improve the data matches for EnKF and EnKF with localization. We suspect that this happens because, even though a large ensemble improves



(a) EnKF

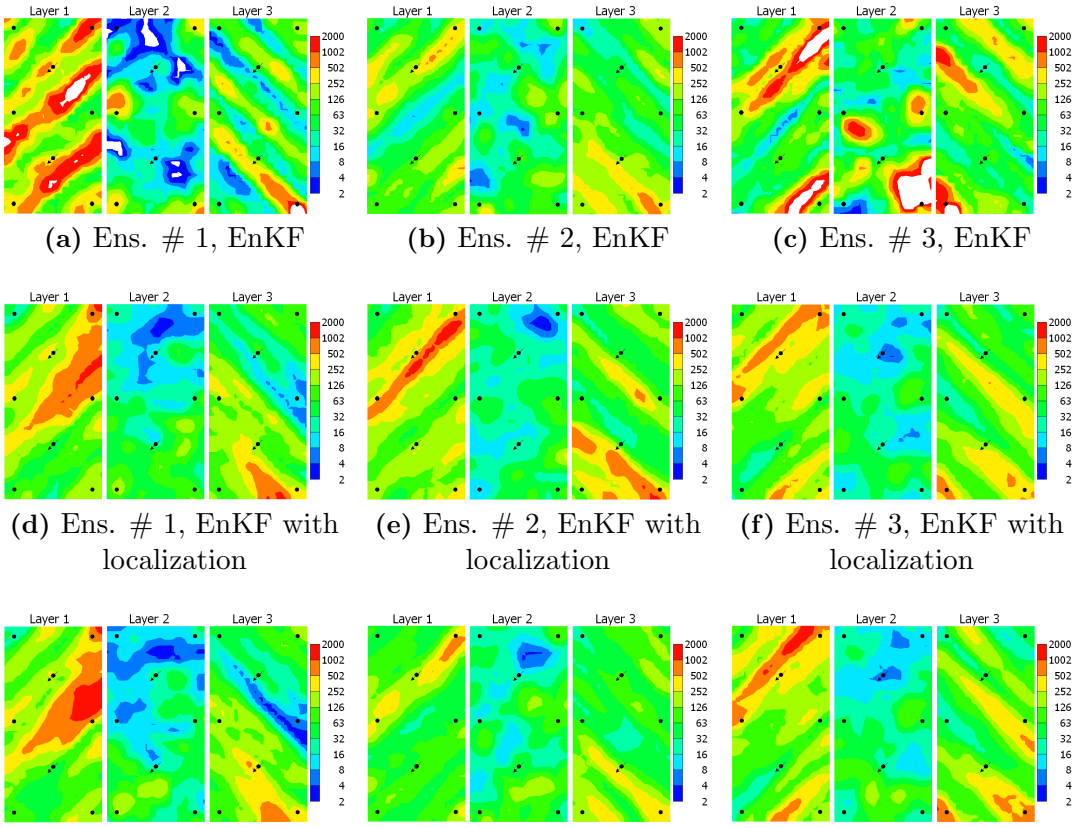
(b) EnKF with localization

(c) EnKF-MCMC

Figure 5.10: Box plots of the normalized objective function for different initial ensembles.

the estimates of cross-covariances, EnKF still requires assumptions of linearity and statistical consistency between model parameters and updated primary simulator variables during data assimilation.

Fig. 5.11 presents the ensemble mean model for each of the first three ensembles obtained with EnKF, EnKF with covariance localization and EnKF-MCMC. According to the results in Fig. 5.11, EnKF resulted in relatively rough ensemble mean models for the first and third ensemble. In this figure, the white areas represent regions of overshooting or undershooting in the permeability fields. Note that it is reasonable to expect that the posterior mean models should be smooth, which is not the case for EnKF. Localization has the effect of regularization on the estimates of cross-covariances during data assimilation, which results in smoother ensemble means. As we started MCMC from the ensembles obtained using EnKF with localization, the final ensemble mean models from EnKF-MCMC exhibit some of the



(a) Ens. # 1, EnKF **(b)** Ens. # 2, EnKF **(c)** Ens. # 3, EnKF

(d) Ens. # 1, EnKF with localization **(e)** Ens. # 2, EnKF with localization **(f)** Ens. # 3, EnKF with localization

(g) Ens. # 1, EnKF-MCMC **(h)** Ens. # 2, EnKF-MCMC **(i)** Ens. # 3, EnKF-MCMC

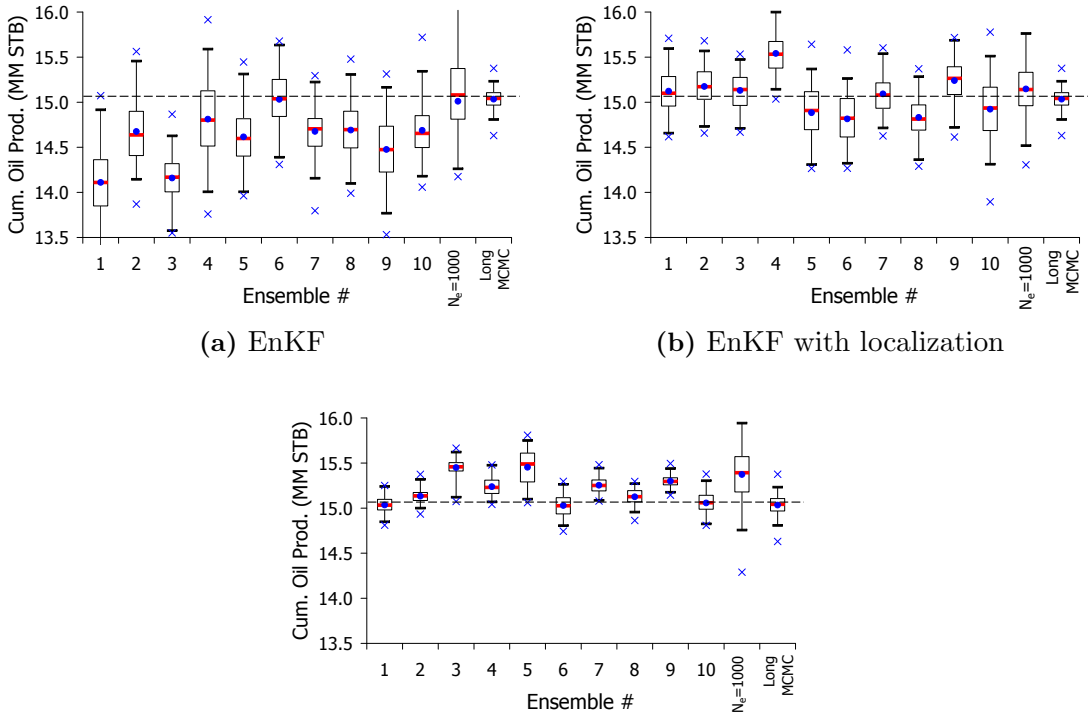
Figure 5.11: Mean permeability fields for the first three ensembles (mD). First row shows models from EnKF; second row shows models from EnKF with covariance localization and third row shows models from EnKF-MCMC.

features of the mean models from EnKF with localization. However, as indicated in Fig. 5.10, these models give data matches approximately three times better than EnKF with localization, and, based on the posterior pdf, they are models with a much higher probability.

Figs. 5.12 and 5.13 present box plots of cumulative oil and water production for all cases, including the long MCMC. These box plots are based on a total time of 21 years, 11 years of historical data and 10 years of future predictions. According to the results in these figures, we observe that different initial ensembles gave very different final distributions of predictions, and, in some cases, the predictions do not span the cumulative oil and water production obtained by the true permeability field,

which is represented by the dashed horizontal black line. In addition, the spreads of the predictions obtained with EnKF and EnKF with localization are too large compared to the spreads in the long MCMC case, which we assume to be correct. Besides that, EnKF with and without localization resulted in biased distributions of cumulative water production for most of the ensembles, i.e., the distributions represented by the box plots underestimate the true water production for most of the ensembles. In summary, we can conclude that, even for such a simple reservoir model, EnKF did not result in a reasonable characterization of uncertainty. Considering the results obtained for an ensemble of 1,000 models, we observe that predictions obtained with EnKF and EnKF with localization span the cumulative oil and water production obtained from the true model. However, again the spreads of theses predictions are too large because most of the models give poor data matches. EnKF-MCMC narrowed the spread of predictions for all ensembles, but we still observe the problem that different initial ensembles resulted in different final distributions.

Figs. 5.14–5.17 present the histograms of cumulative oil and water production obtained by combining all ensembles. These figures include the cases with ten ensembles of 100 models and one single ensemble with 1,000 models. First, we consider the case in which we combine the ensembles by simply taking the predictions of all models without resampling. These results are labeled in each plot as "before resampling." As expected, this procedure gives unrealistic distributions of predicted cumulative oil and water production for EnKF cases. The resulting histograms largely overestimate the variance of the distributions obtained for the long MCMC case. This is also true for the cases with a single ensemble of 1,000 models. Again, this happens because EnKF cases give several models with poor data matches, resulting in an unrealistic large spread of predictions. In fact, we should not consider the predictions from models with poor data matches, or at least, these predictions should be assigned a very low probability, i.e., they should have lower weights when generating the histograms.

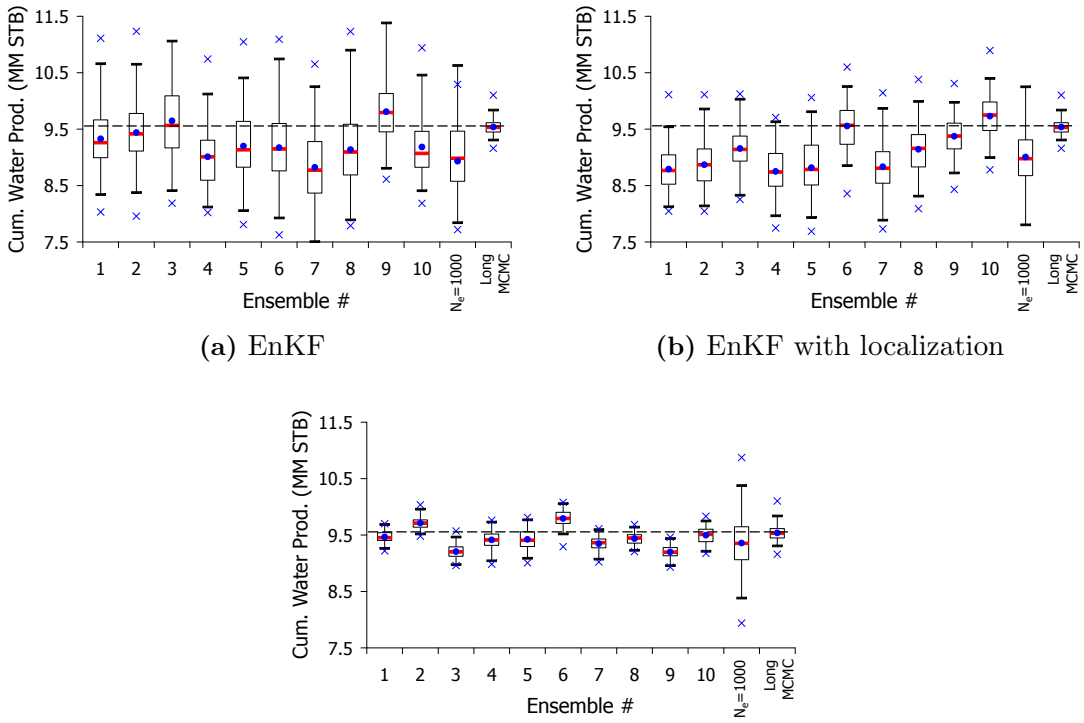


(a) EnKF

(b) EnKF with localization

(c) EnKF-MCMC

Figure 5.12: Box plots of cumulative oil production for different initial ensembles. The dashed horizontal black line indicates the truth prediction.



(a) EnKF

(b) EnKF with localization

(c) EnKF-MCMC

Figure 5.13: Box plots of cumulative water production for different initial ensembles. The dashed horizontal black line indicates the truth prediction.

The resampling procedure based on the values of O_N aims to correct the sampling based on the importance of each model, and the resulting histograms are presented in Figs. 5.14–5.17 labeled as "after resampling." When applied to the case with ten ensembles obtained after EnKF-MCMC, the resampling procedure resulted in histograms of cumulative oil and water production that are close to those obtained from the long MCMC case (compare Figs. 5.14f and 5.16f with Figs. 5.5a and 5.5b). However, the application of the resampling procedure to the ensembles generated from EnKF cases still results in histograms very different from the ones obtained with the long MCMC. The problem is that EnKF cases resulted in several models with high values of O_N. Thus, after resampling, only a few models have non-negligible importance weights. This situation is often referred to as sample impoverishment in the particle filter literature [33, Chap. 1]. A typical measure of the degree of sample impoverishment is the effective sample size, N_{eff} [34], which is given by

$$N_{\text{eff}} = \frac{1}{\sum_{j=1}^{N_s} w_j^2}. \tag{5.13}$$

Table 5.4 presents the values of N_{eff} for all cases. According to the results in this table, because EnKF-MCMC generated more samples with similar values of O_N, it resulted in a larger value of N_{eff}. For the same reason, the N_{eff} is larger for the case with ten ensembles of 100 models than for the case with one ensemble of 1,000 models. As a result, the final histograms from EnKF-MCMC with ten ensembles are much closer to those from the long MCMC case than are the histograms generated with all other procedures. Figs. 5.18–5.20 present the water production rate for wells P3, P4 and P6 for all cases after resampling. In these plots, we show the mean and one and two standard deviations around the mean for each case. The water rate obtained by the long MCMC case is also presented in these figures. These figures illustrate that the best data matches and predictions were obtained by combining the ten ensembles from EnKF-MCMC.

Table 5.4: Effective sample size.

Case	Ten ens. 100 models	One ens. 1,000 models
EnKF	92	15
EnKF with localization	11	13
EnKF-MCMC	386	86

Fig. 5.21 presents the final mean permeability fields obtained with EnKF, EnKF with covariance localization, EnKF-MCMC and long MCMC. For the first three cases, the mean permeability fields presented in this figure were obtained using the ten ensembles of 100 models and resampling. According to the results in Fig. 5.21, all cases resulted in mean permeability fields in reasonable agreement with the long MCMC case. In addition, these models present the main features of the true model (Fig. 5.1), namely, the high permeability channel connecting injector I1 to producers

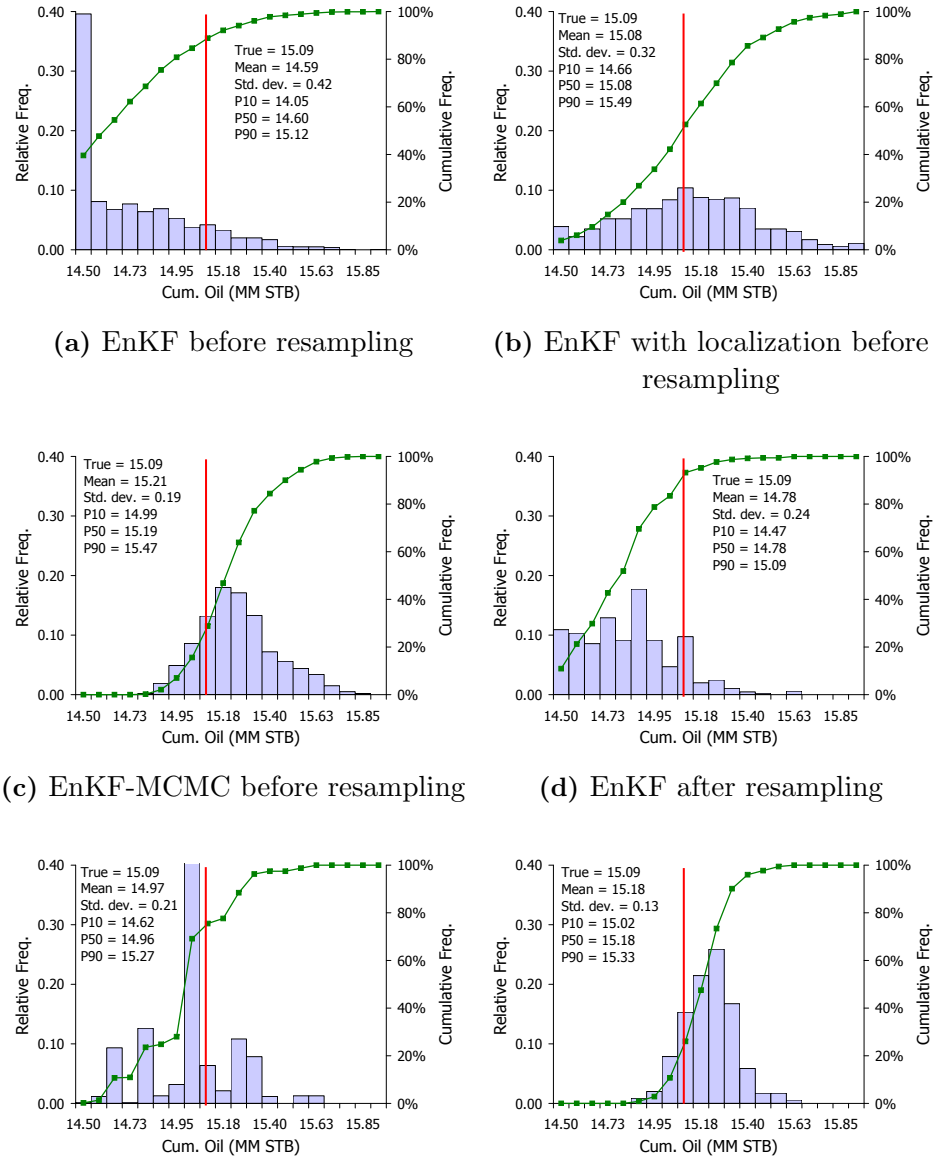
P2 and P3 in the first layer, and the channel connecting injector I2 to producer P6 in the third layer.

Fig. 5.22 presents the final standard deviations of log-permeability, $\ln(k)$, obtained with EnKF, EnKF with covariance localization, EnKF-MCMC and long MCMC. For the first three cases, the standard deviations presented in Fig. 5.22 were obtained using the ten ensembles of 100 models with resampling. According to the results in Fig. 5.22, all cases overestimate the standard deviations of the $\ln(k)$'s compared to those obtained from the long MCMC case. However, this overestimation is clearly much higher for the EnKF case. Note that EnKF resulted in standard deviations larger than one in several gridblocks of the reservoir, which is clearly too high considering that the prior standard deviation of $\ln(k)$ is one. Note that the results presented in Fig. 5.22 were obtained after resampling. Table 5.5 presents the corresponding values of SNV for all cases. According to the results in Table 5.5, combining the final ten ensembles resulting from data assimilation with EnKF resulted in a SNV greater than one. This indicates an unreasonable overestimation of uncertainty in the log-permeability field for this case. The SNV obtained with EnKF-MCMC, on the other hand, is much closer to the SNV obtained with the long MCMC case, even though there is still an overestimation of posterior variances.

Table 5.5: Sum of the normalized variances of log-permeability.

Case	SNV
EnKF	1.244
EnKF with localization	0.666
EnKF-MCMC	0.594
Long MCMC	0.380

It is important to note that the values of SNV presented in Table 5.5 were computed by combining the ten ensembles of 100 models and then resampling based on O_N. If we consider each of the ten ensembles individually, we observe that the



(a) EnKF before resampling

(b) EnKF with localization before resampling

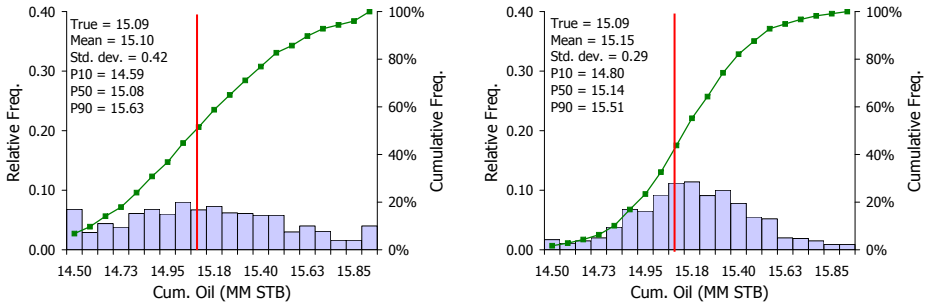
(c) EnKF-MCMC before resampling

(d) EnKF after resampling

(e) EnKF with localization after resampling

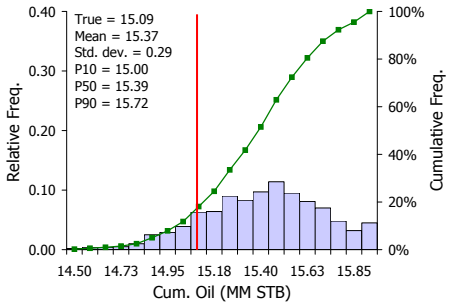
(f) EnKF-MCMC after resampling

Figure 5.14: Histograms of cumulative oil production. Results for ten ensembles of 100 models. Figures also include the cumulative oil production for the true model (in red).

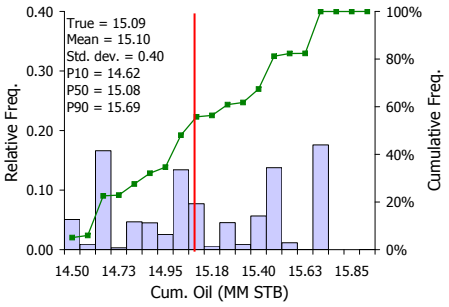


(a) EnKF before resampling

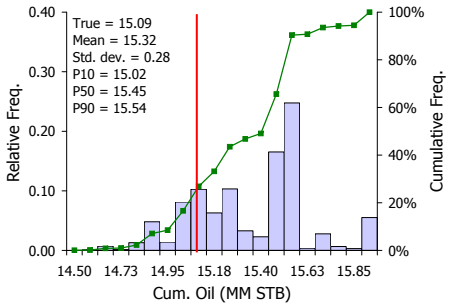
(b) EnKF with localization before resampling



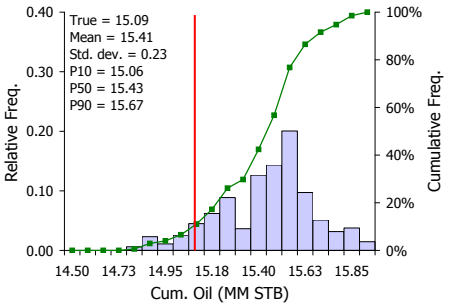
(c) EnKF-MCMC before resampling



(d) EnKF after resampling

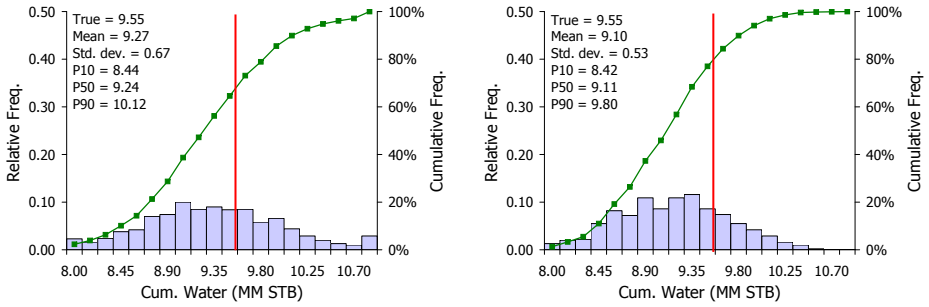


(e) EnKF with localization after resampling



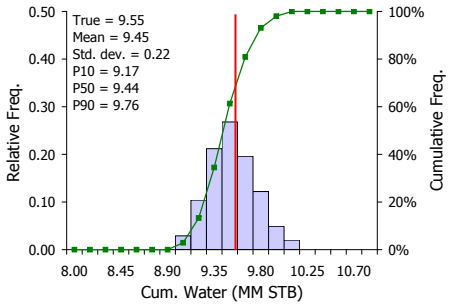
(f) EnKF-MCMC after resampling

Figure 5.15: Histograms of cumulative oil production. Results for one ensemble of 1,000 models. Figures also include the cumulative oil production for the true model (in red).

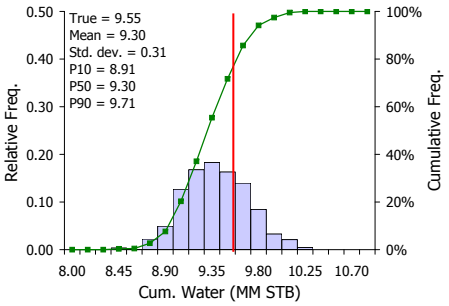


(a) EnKF before resampling

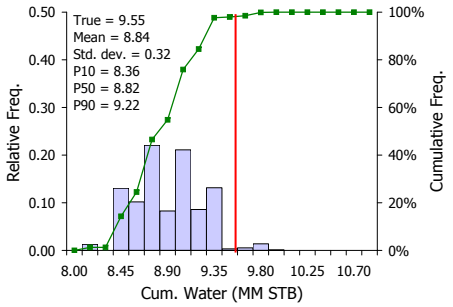
(b) EnKF with localization before resampling



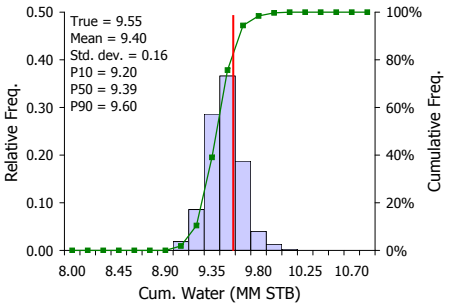
(c) EnKF-MCMC before resampling



(d) EnKF after resampling

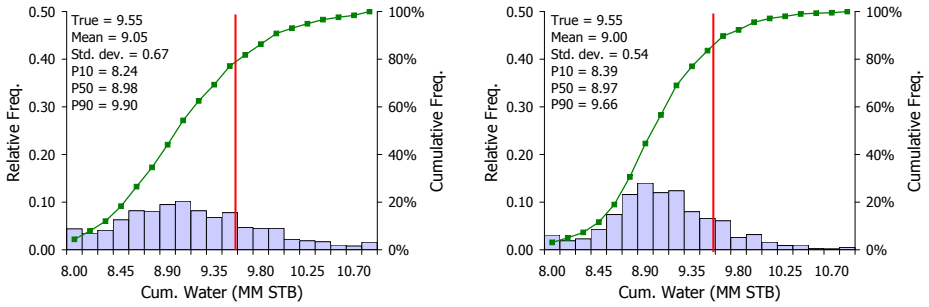


(e) EnKF with localization after resampling



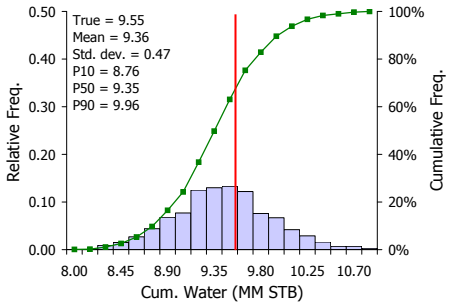
(f) EnKF-MCMC after resampling

Figure 5.16: Histograms of cumulative water production. Results for ten ensembles of 100 models. Figures also include the cumulative water production for the true model (in red).

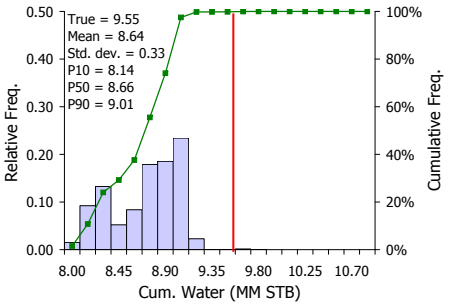


(a) EnKF before resampling

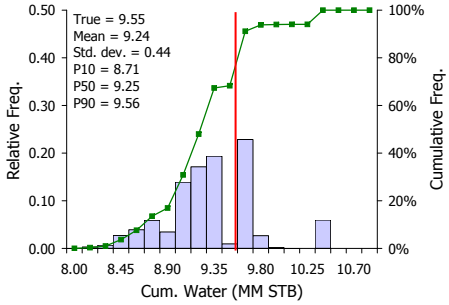
(b) EnKF with localization before resampling



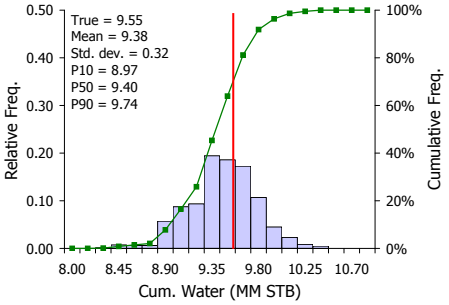
(c) EnKF-MCMC before resampling



(d) EnKF after resampling

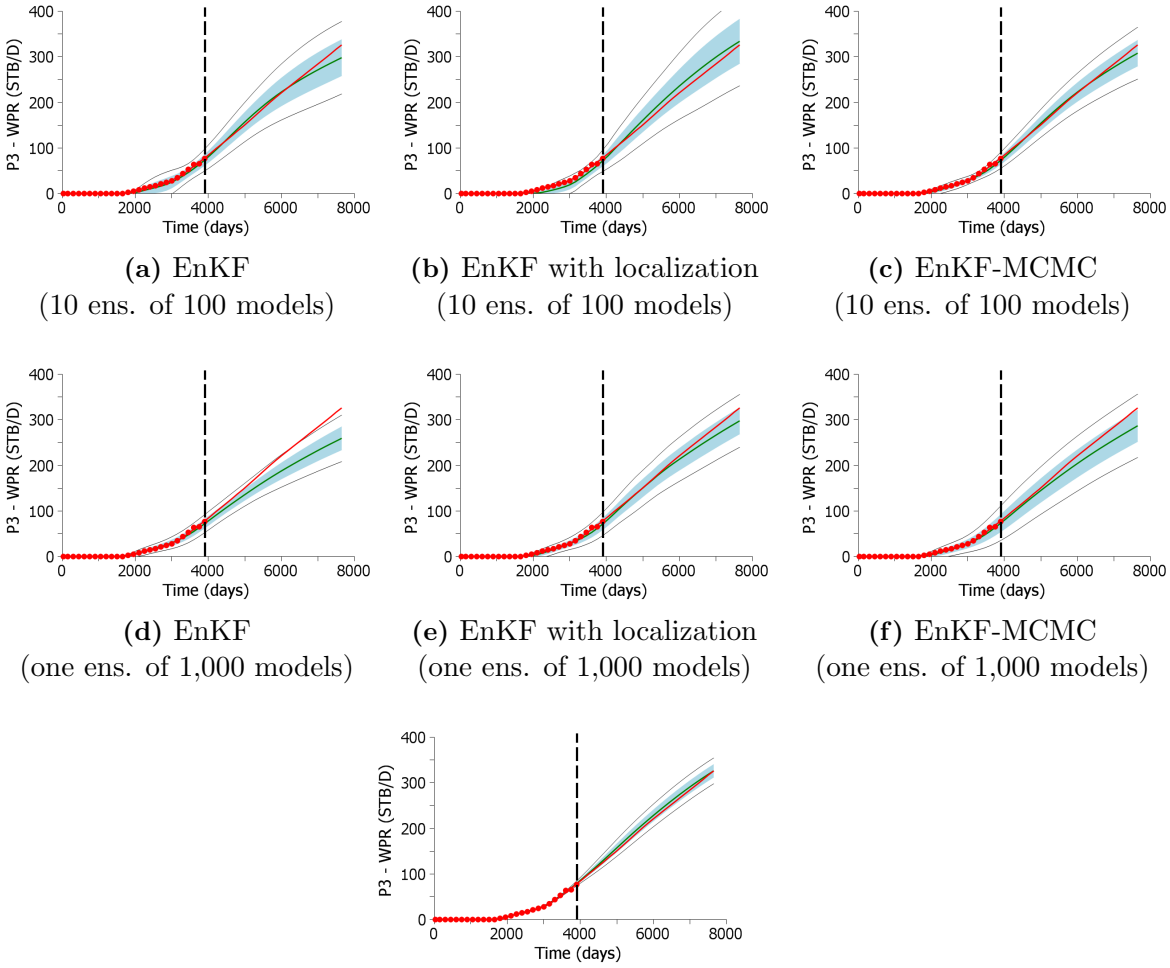


(e) EnKF with localization after resampling



(f) EnKF-MCMC after resampling

Figure 5.17: Histograms of cumulative water production. Results for one ensemble of 1,000 models. Figures also include the cumulative water production for the true model (in red).



(a) EnKF
(10 ens. of 100 models)

(b) EnKF with localization
(10 ens. of 100 models)

(c) EnKF-MCMC
(10 ens. of 100 models)

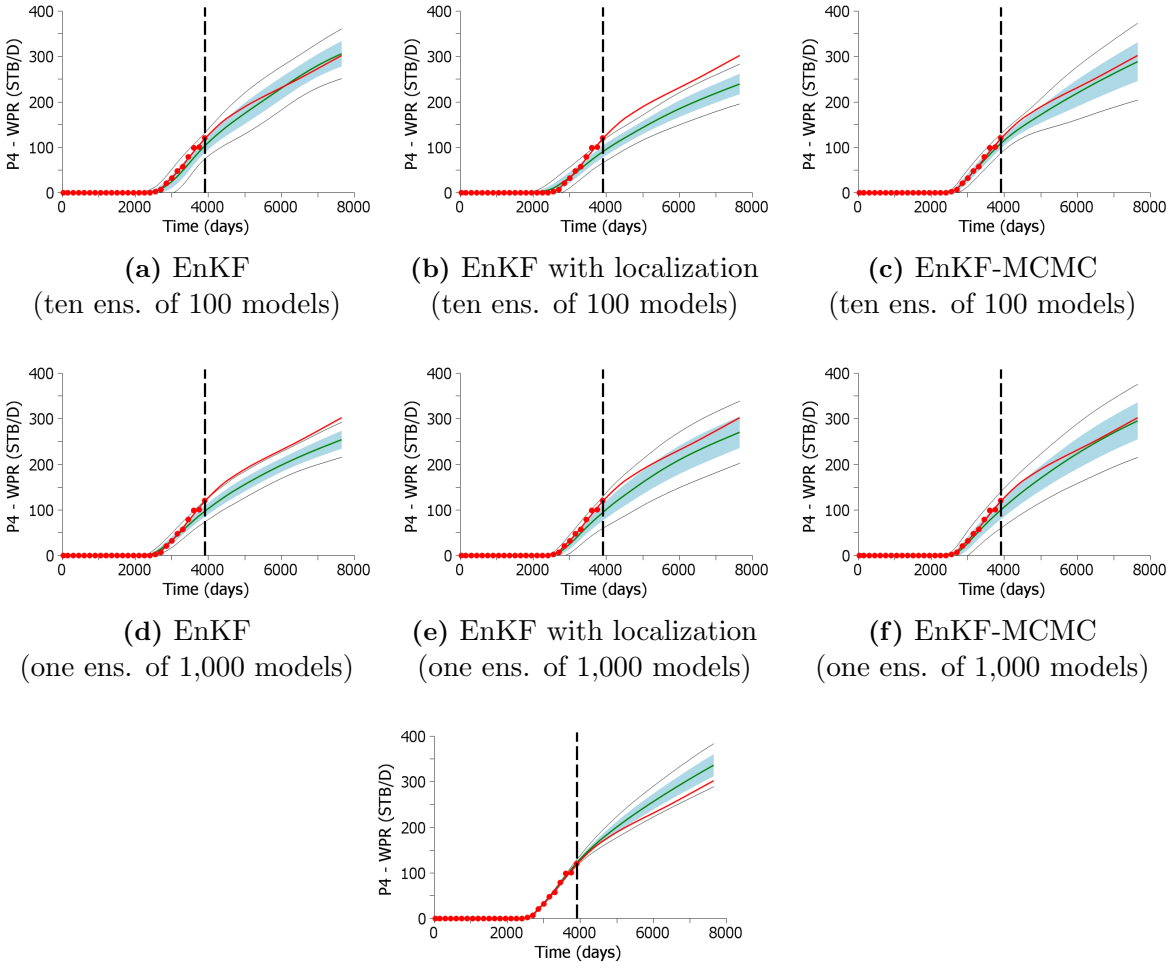
(d) EnKF
(one ens. of 1,000 models)

(e) EnKF with localization
(one ens. of 1,000 models)

(f) EnKF-MCMC
(one ens. of 1,000 models)

(g) Long MCMC

Figure 5.18: Well P3. Predicted water rate by running the final ensemble from time zero (stb/day). Red dots are the history; solid red curve is the true model; solid green curve is the mean prediction computed from the ensemble considering the importance weights. The light blue region corresponds to one standard deviation around the mean. The black lines correspond to two standard deviations around the mean. The dashed vertical black line indicates the end of the history and beginning of the forecast period.



(a) EnKF (ten ens. of 100 models)

(b) EnKF with localization (ten ens. of 100 models)

(c) EnKF-MCMC (ten ens. of 100 models)

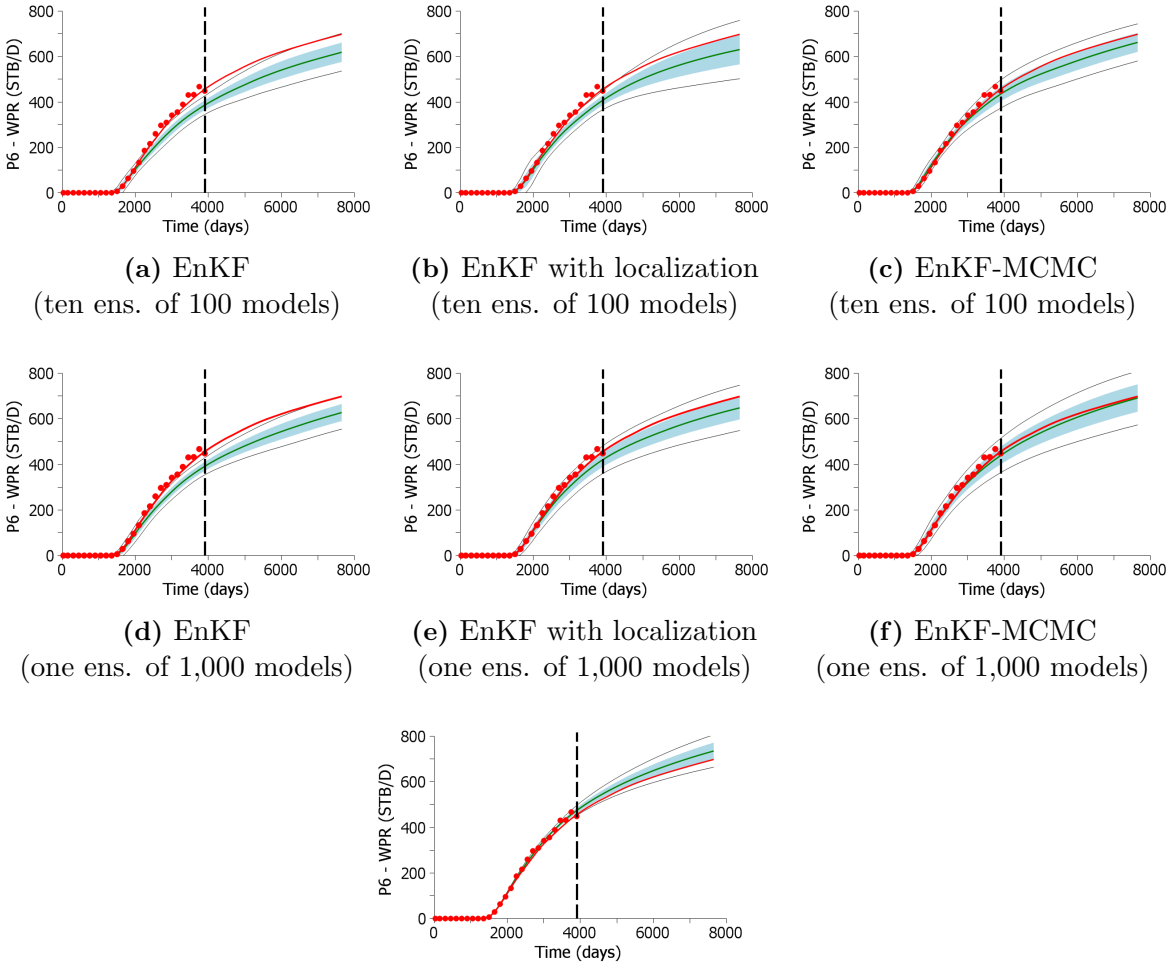
(d) EnKF (one ens. of 1,000 models)

(e) EnKF with localization (one ens. of 1,000 models)

(f) EnKF-MCMC (one ens. of 1,000 models)

(g) Long MCMC

Figure 5.19: Well P4. Predicted water rate by running the final ensemble from time zero (stb/day). The colors in this figure have the same meaning as in Fig. 5.18.



(a) EnKF (ten ens. of 100 models)

(b) EnKF with localization (ten ens. of 100 models)

(c) EnKF-MCMC (ten ens. of 100 models)

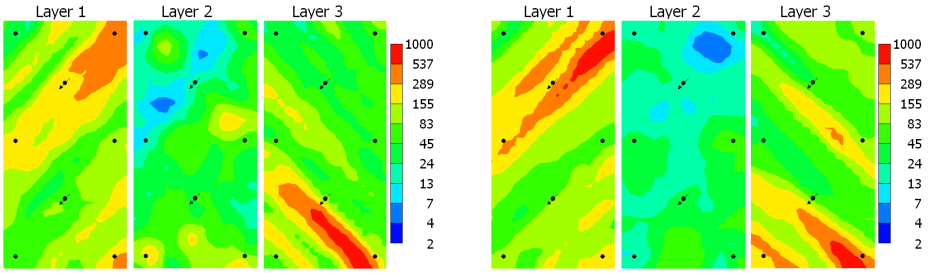
(d) EnKF (one ens. of 1,000 models)

(e) EnKF with localization (one ens. of 1,000 models)

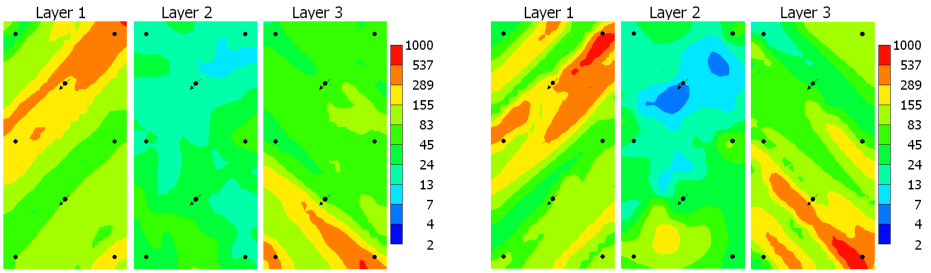
(f) EnKF-MCMC (one ens. of 1,000 models)

(g) Long MCMC

Figure 5.20: Well P6. Predicted water rate by running the final ensemble from time zero (stb/day). The colors in this figure have the same meaning as in Fig. 5.18.

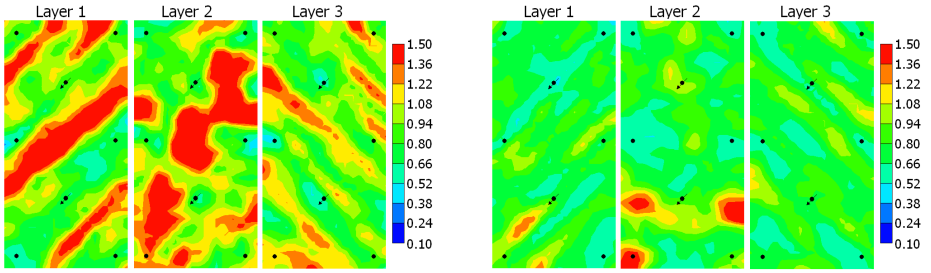


(a) EnKF **(b)** EnKF with localization

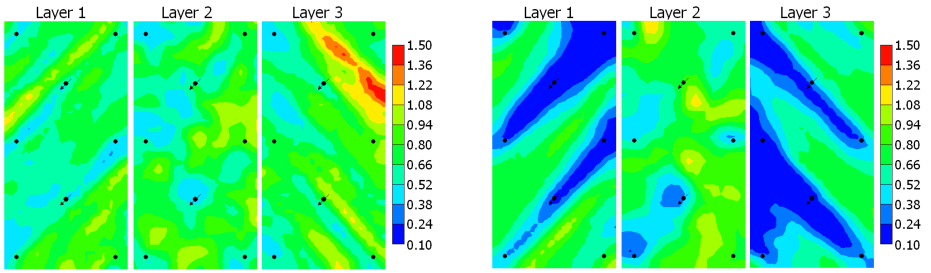


(c) EnKF-MCMC **(d)** Long MCMC

Figure 5.21: Mean permeability fields. For cases (a)–(c) the mean permeability fields were obtained using ten ensembles of 100 models and resampling.



(a) EnKF **(b)** EnKF with localization



(c) EnKF-MCMC **(d)** Long MCMC

Figure 5.22: Standard deviation of log-permeability. For cases (a)–(c) the standard deviations were obtained using ten ensembles of 100 models and resampling.

posterior variances are greatly underestimated. Table 5.6 presents the SNV obtained for each ensemble. The values of SNV presented in Table 5.6 were computed before resampling. According to the results present in this table, for each ensemble, EnKF resulted in SNV's approximately 60% less than the SNV obtained by the long MCMC case. Covariance localization increased significantly the posterior variances resulting in an average SNV approximately 15% less than the SNV obtained by the long MCMC case. After EnKF-MCMC the variance of each ensemble was greatly reduced. The average SNV is 85% less than the one obtained by the long MCMC case. This result shows that we cannot obtain a sampling that reasonably describes uncertainty for this problem using a single ensemble of 100 models. This significant reduction in the ensemble variance occurs because during MCMC, the proposals are restricted to a subspace spanned by the state vectors corresponding to the final ensemble of models obtained by EnKF with covariance localization. Consequently, the variance of the ensemble after MCMC is reduced because there are not many "independent" samples in the subspace spanned by the EnKF ensemble that provide reasonable data matches.

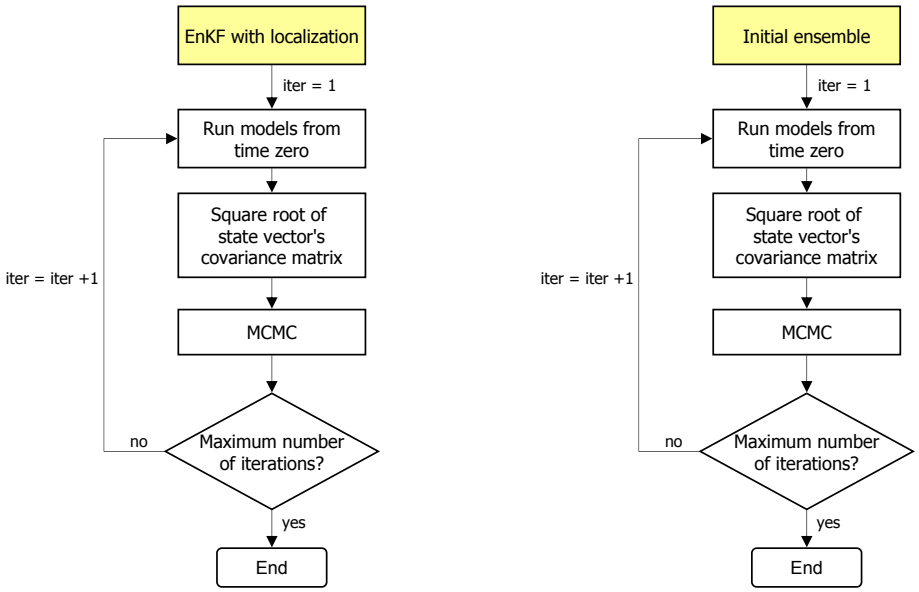
Table 5.6: Sum of the normalized variances of log-permeability for each ensemble.

Ensemble #	EnKF	EnKF with local.	EnKF-MCMC
1	0.152	0.301	0.048
2	0.169	0.334	0.036
3	0.143	0.308	0.053
4	0.141	0.330	0.054
5	0.164	0.311	0.118
6	0.173	0.332	0.074
7	0.146	0.319	0.042
8	0.159	0.325	0.063
9	0.172	0.321	0.036
10	0.161	0.323	0.072
mean	0.158	0.320	0.060

5.2.3 MCMC iterations

In the proposed EnKF-MCMC method, we start with the models resulting from EnKF with covariance localization as an initial sampling. Then, we apply the MCMC procedure where we propose transitions based on the square root of the state vector's covariance matrix to improve sampling. Even though MCMC samples correctly in the limit, the proposed EnKF-MCMC procedure is not guaranteed to sample the correct pdf because we are not computing the correct likelihood. For the example presented in the previous section, this procedure resulted in a significant improvement in the objective functions obtained for all ten ensembles considered. The final average O_N obtained after EnKF-MCMC is 5.77. However, for the long MCMC case, the average O_N is 2.40. Therefore, most of the models obtained with EnKF-MCMC are still samples of regions with lower probability compared to the samples obtained in the long MCMC case, which we assume to be a correct sampling. To improve the accuracy of the sampling procedure, here, we extend the example considered in the previous section by performing additional MCMC procedures. For each of the ten ensembles, we repeated the MCMC procedure staring with the set of models obtained in the previous MCMC procedure. This approach can be viewed as an iterative procedure, which can be repeated until we do not observe significant improvements in the values of the objective function. Fig. 5.23 presents two alternative workflows summarizing this iterative procedure. In the first workflow (EnKF-MCMC, Fig. 5.23a), we start the process using EnKF with covariance localization and then apply multiple iterations of MCMC. The second workflow (MCMC-only, Fig. 5.23b) is similar to the first one, but instead of starting with EnKF with covariance localization, we simply start from the prior ensemble. Here, we applied three iterations of MCMC for EnKF-MCMC and six iterations for MCMC-only. Note that each new iteration requires running the ensemble of models from time zero and then generating the Markov chains, which increases significantly

the computational cost. In Section 5.2.4, we present some comparisons in terms of computational cost for each procedure.

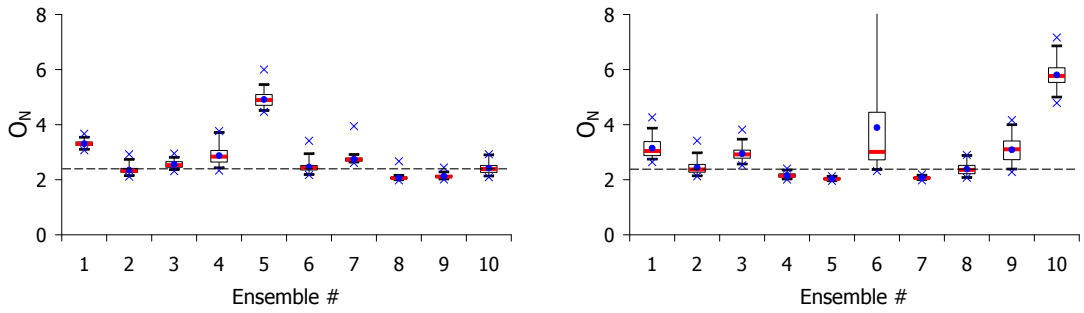


(a) EnKF-MCMC **(b)** MCMC-only

Figure 5.23: Workflows for MCMC iterations.

Fig. 5.24 shows the box plots of the values of O_N obtained for each ensemble after EnKF-MCMC with three iterations and after six iterations of MCMC-only. In this figure, the horizontal dashed line indicates $O_N = 2.4$, which is the average normalized objective function obtained in the long MCMC case. Table 5.7 presents the average and standard deviation of O_N considering all ensembles together. According to the results in Fig. 5.24 and Table 5.7, additional iterations of MCMC resulted in further improvements in the normalized objective functions. For the EnKF-MCMC, after three iterations the average O_N obtained is 2.78. For MCMC-only, after six iterations the average O_N obtained is 3.00.

Figs. 5.25 and 5.26 present the histograms of cumulative oil and water production obtained for EnKF-MCMC and MCMC-only. For comparisons, we also include the histograms obtained with EnKF-MCMC with one iteration and with the long MCMC case in each figure. As indicated in Fig. 5.25b, after three iterations, the



(a) EnKF-MCMC (3rd iteration) **(b)** MCMC-only (6th iteration)

Figure 5.24: Box plots of the normalized objective function for different initial ensembles after MCMC iterations. The dashed black line indicates $O_N = 2.4$, which is the average normalized objective function obtained in the long MCMC case.

Table 5.7: Normalized objective functions for EnKF-MCMC.

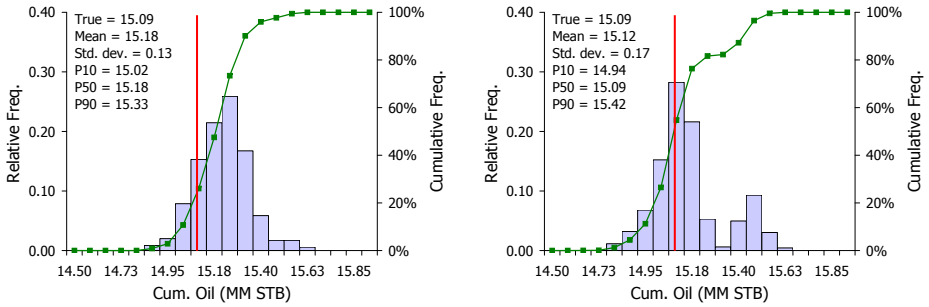
Case	Mean	Std. deviation
EnKF-MCMC (3rd iteration)	2.78	0.81
MCMC-only (6th iteration)	3.00	1.26

histogram of cumulative oil production obtained by EnKF-MCMC shows a second small mode. This mode is caused mainly by the predictions obtained by ensemble number three. After three iterations, the average O_N of the models of this third ensemble decreased from 5.6 to 2.5, thereby increasing the relative weight of those models in the final distributions. Despite the appearance of this second mode, which is not observed in the long MCMC case, the final histograms of cumulative oil and water production obtained after three iterations, are in better agreement with the histograms obtained with the long MCMC case. Note that after the first iteration, the histogram of cumulative oil production (Fig. 5.25a) slightly overestimates the cumulative production compared to the histogram for the long MCMC case (Fig. 5.25d). After three iterations, the main mode of the distribution corresponds to the true cumulative production, which is also the case for the histogram obtained by the long MCMC case. The same comments are valid for the histogram of cumulative water

production (Fig. 5.26b). The histograms obtained after six iterations of MCMC-only slightly overestimate the distributions of the cumulative production of oil and water when compared to the long MCMC case. This may indicate that six iterations were not enough in this case. Fig. 5.27 presents the cumulative distributions of cumulative oil and water production obtained by EnKF, EnKF with localization, EnKF-MCMC (one and three iterations), MCMC-only and long MCMC. This figure shows that the distributions obtained with EnKF-MCMC (three iterations) are the ones that best approximate the distributions obtained for the long MCMC case, although results of roughly the same quality were obtained with six iterations of MCMC-only.

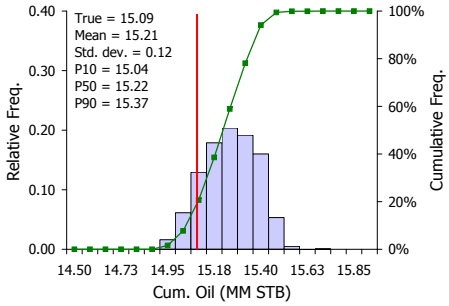
Fig. 5.28, 5.29 and 5.30 present the water production rates for wells P3, P4 and P6 for EnKF-MCMC and MCMC-only after resampling. For comparisons we also include the results after the first iteration of EnKF-MCMC. In these plots, we show the mean and one and two standard deviations around the mean for each case. These figures show the improvements in the data matches obtained for EnKF-MCMC with additional iterations (compare Fig. 5.28a with Fig. 5.28b, Fig. 5.29a with Fig. 5.29b).

Fig. 5.31 presents the final mean permeability fields obtained for EnKF-MCMC and MCMC-only. This figure indicates that there were not significant changes in the final mean permeability fields after additional iterations for the EnKF-MCMC cases (compare Figs. 5.31a and 5.31b). Besides that, all results are in reasonable agreement with the mean permeability field obtained by the long MCMC case. Fig. 5.32 presents the final standard deviation of $\ln(k)$ obtained for EnKF-MCMC and MCMC-only. The final standard deviations still overestimate the ones obtained with the long MCMC case. The same conclusion can be found by comparing the values of the SNV presented in Table 5.8, which indicate an overestimation of the uncertainty in the permeability fields for EnKF-MCMC and MCMC-only.

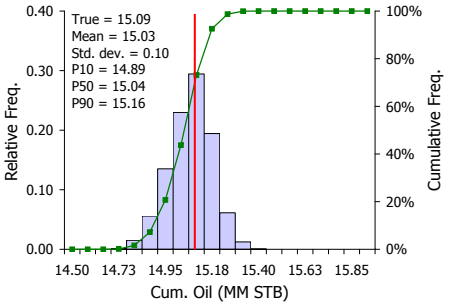


(a) EnKF-MCMC (1st iter.)

(b) EnKF-MCMC (3rd iter.)



(c) MCMC-only (6th iter.)

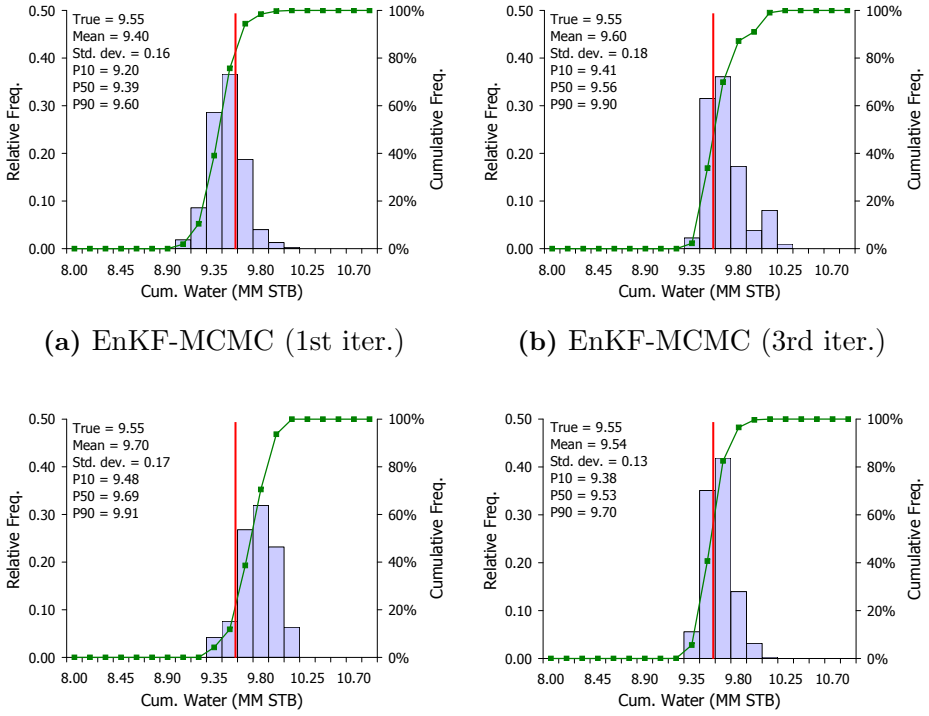


(d) long MCMC

Figure 5.25: Histograms of cumulative oil production. Results for ten ensembles of 100 models using MCMC with iterations and resampling. Figures also include the cumulative oil production for the true model (in red).

Table 5.8: Sum of the normalized variances of log-permeability.

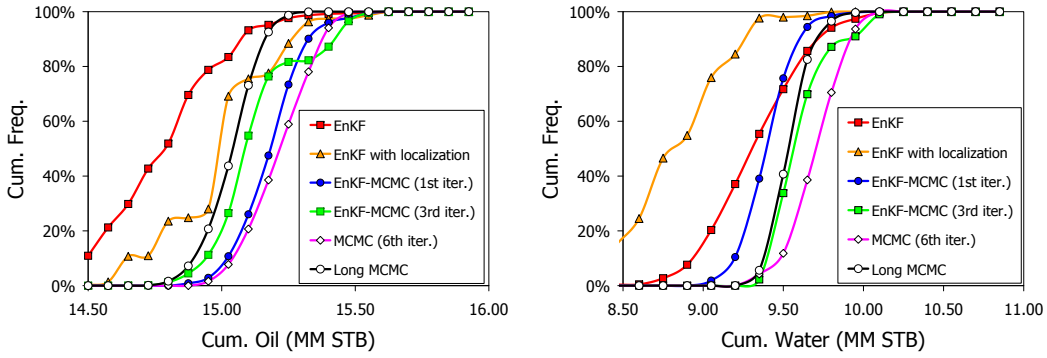
Case	SNV
EnKF-MCMC (1st iter.)	0.594
EnKF-MCMC (3rd iter.)	0.619
MCMC-only (6th iter.)	0.468
Long MCMC	0.380



(a) EnKF-MCMC (1st iter.) **(b)** EnKF-MCMC (3rd iter.)

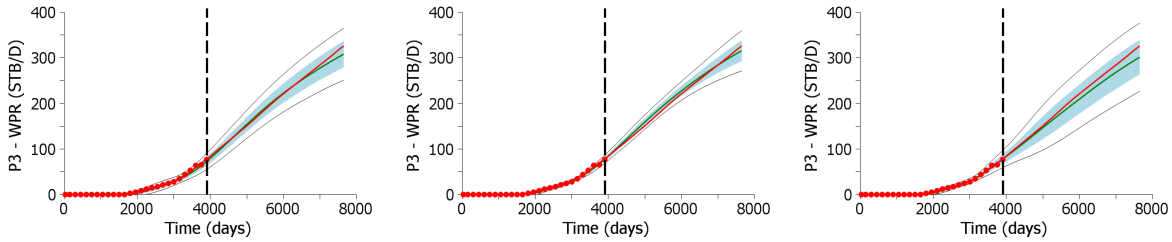
(c) MCMC-only (6th iter.) **(d)** long MCMC

Figure 5.26: Histograms of cumulative water production. Results for ten ensembles of 100 models using MCMC with iterations and resampling. Figures also include the cumulative oil production for the true model (in red).



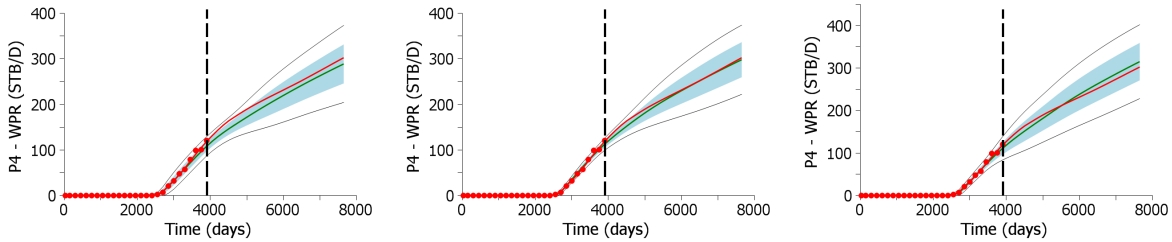
(a) Cumulative oil production **(b)** Cumulative water production

Figure 5.27: Final cumulative distributions. Results for ten ensembles of 100 models and resampling.



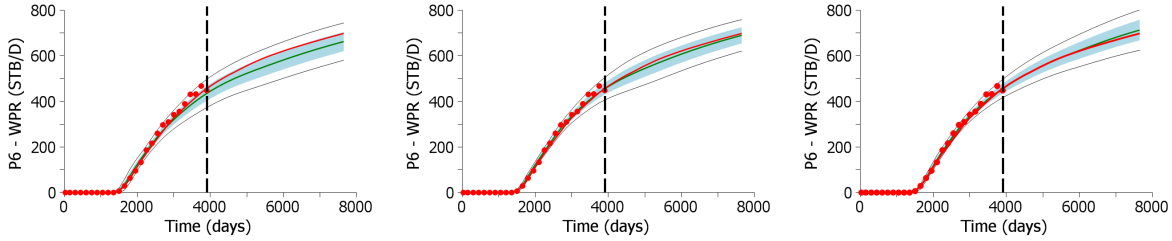
(a) EnKF-MCMC (1st iter.) **(b)** EnKF-MCMC (3rd iter.) **(c)** MCMC-only (6th iter.)

Figure 5.28: Well P3. Predicted water rate by running the final ensemble from time zero (stb/day). Results for the case with ten ensembles of 100 models. Red dots are the history; solid red curve is the true model; solid green curve is the mean prediction computed from the ensemble considering the importance weights. The light blue region corresponds to one standard deviation around the mean. The black lines correspond to two standard deviations around the mean. The dashed vertical black line indicates the end of the history and beginning of the forecast period.



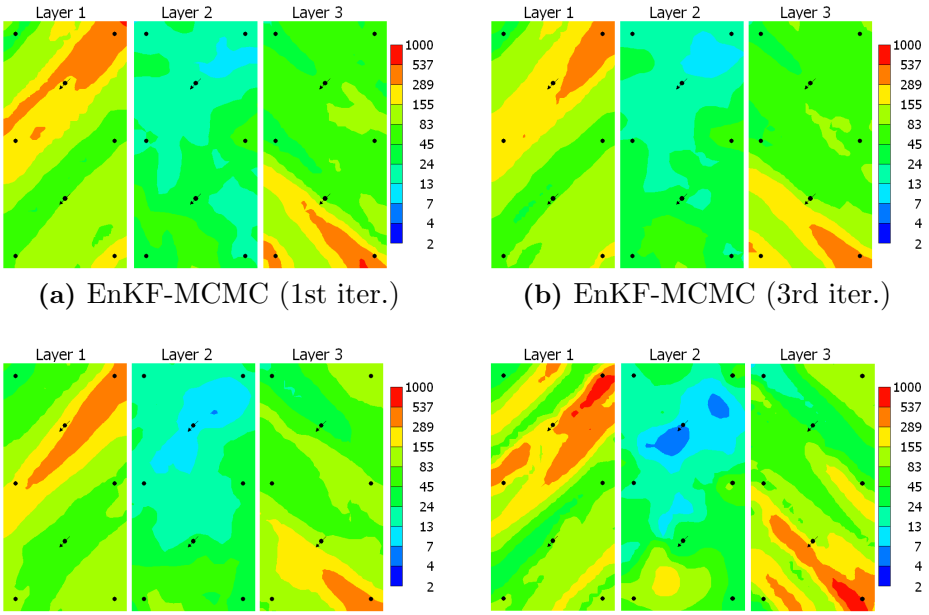
(a) EnKF-MCMC (1st iter.) **(b)** EnKF-MCMC (3rd iter.) **(c)** MCMC-only (6th iter.)

Figure 5.29: Well P4. Predicted water rate by running the final ensemble from time zero (stb/day). Results for the case with ten ensembles of 100 models. The colors in this figure have the same meaning as in Fig. 5.28.



(a) EnKF-MCMC (1st iter.) **(b)** EnKF-MCMC (3rd iter.) **(c)** MCMC-only (6th iter.)

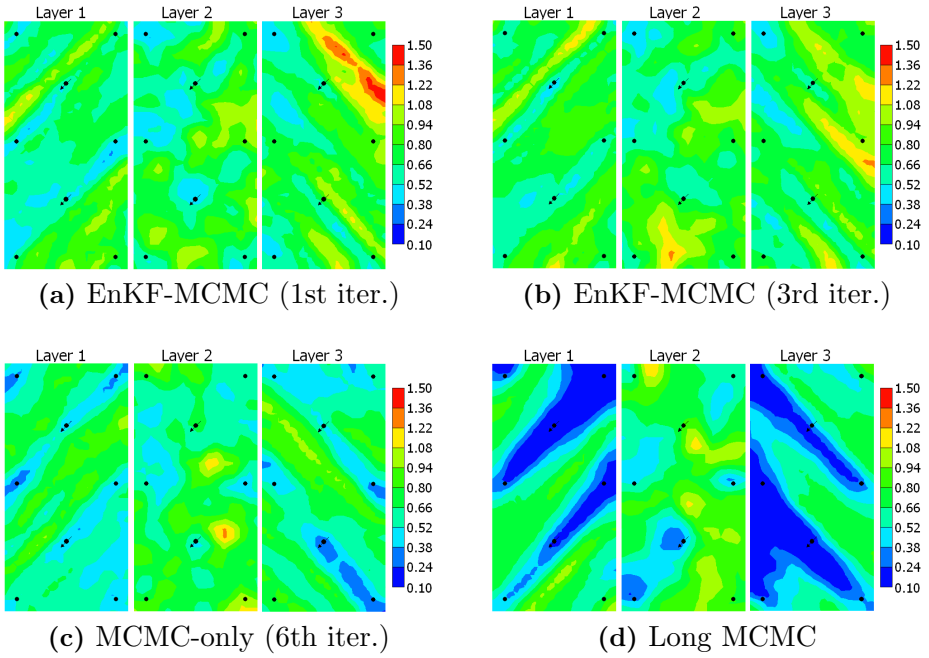
Figure 5.30: Well P6. Predicted water rate by running the final ensemble from time zero (stb/day). Results for the case with ten ensembles of 100 models. The colors in this figure have the same meaning as in Fig. 5.28.



(a) EnKF-MCMC (1st iter.) **(b)** EnKF-MCMC (3rd iter.)

(c) MCMC-only (6th iter.) **(d)** Long MCMC

Figure 5.31: Mean permeability fields. For EnKF-MCMC cases the mean permeability fields were obtained using ten ensembles of 100 models and resampling.



(a) EnKF-MCMC (1st iter.) **(b)** EnKF-MCMC (3rd iter.)

(c) MCMC-only (6th iter.) **(d)** Long MCMC

Figure 5.32: Standard deviation of log-permeability. For EnKF-MCMC cases the mean permeability fields were obtained using ten ensembles of 100 models and re-sampling.

5.2.4 Computational cost

For this example, we estimated the computational cost of each method in terms of equivalent simulation runs. Table 5.9 summarizes the computational cost of generating one ensemble with 100 realizations. The computational costs were estimated by measuring the actual time required for each method divided by the time required to run one ensemble of 100 models from time zero without data assimilation. Note that the number of equivalent simulation runs of EnKF is 385, which is 3.85 times larger than the cost of simply running the ensemble from time zero. This difference is caused mainly by the additional CPU time required to restart the reservoir simulations at every data assimilation step. The computational cost of EnKF with covariance localization is 9% greater than the computational cost of EnKF mainly because, in our implementation, at the end of every data assimilation step, we run the ensemble mean model and compute the drainage areas, which requires calling an external program that reads the simulator results and performs the drainage area

calculations. The computational costs of the EnKF-MCMC cases include the time for the data assimilation using EnKF with covariance localization, the time required to rerun the final ensemble from time zero and the time required during the MCMC procedure. Note that all computational costs presented in Table 5.9 are based on serial simulations; we do not consider the fact that during EnKF and MCMC we can run simulations or propose states in parallel using multiple computers. According to the results in Table 5.9, the computational cost for one iteration of EnKF-MCMC was approximately three times the cost of running one data assimilation with EnKF. For three iterations, the cost of EnKF-MCMC increased to approximately seven times the cost of one data assimilation with EnKF. For MCMC-only the computational cost was approximately 12 times the cost of running one data assimilation with EnKF. These results indicate that MCMC-only is computationally more expensive than EnKF-MCMC. In fact, MCMC-only required six iterations to obtain roughly the same level of data match obtained by EnKF-MCMC with three iterations.

Note that the computational costs presented in Table 5.9 correspond to one ensemble. Thus, the final computational costs for generating the results presented in Section 5.2.2, which are based on combining ten different initial ensembles, are ten times the values presented in the Table 5.9. It is important to emphasize though, that the relative computational costs for the methods presented in Table 5.9 are not generally valid. Because the simulation model was very small, which takes only 10 seconds to run, the computational cost to generate Markov chains, which do not require reservoir simulations, became relatively more important.

Table 5.9: Summary of computational costs in terms of equivalent reservoir simulation runs.

Case	Number of equivalent simulation runs	Number of equivalent EnKF runs
EnKF	385	1.00
EnKF with localization	418	1.09
EnKF-MCMC (1st iteration)	1,183	3.07
EnKF-MCMC (3rd iteration)	2,711	7.05
MCMC-only (6th iteration)	4,585	11.92
Long MCMC	2,000,000	5,194.80

CHAPTER 6

MULTIPLE DATA ASSIMILATION

Even though the EnKF was originally proposed as an alternative to the extended Kalman filter [46, Chap. 4] for applications to nonlinear dynamic systems, the update step in the EnKF is still linear. This linear update may result in a sub-optimal performance for highly nonlinear problems. Although the measurement errors have no relation to the model nonlinearity, it is well known that they influence the level of correction in the states. For example, with EnKF it is typically more difficult to assimilate very reliable data (i.e., with small measurement errors) than data with larger measurement errors.

This chapter presents a procedure in which the same data are assimilated multiple times with the covariance matrix of the measurement errors, C_{D}, increased. This procedure is motivated by the equivalence between single and multiple data assimilation (MDA) for the linear-Gaussian case. In the next section, we establish this equivalence for the case where we assimilate the same data N_a times with C_{D} multiplied by the number of data assimilations, i.e., $\alpha_i = N_a$. Then, we investigate the use of MDA for assimilating time-lapse seismic data with EnKF. After that, we generalize the MDA procedure by deriving the condition that the multiplication coefficients of the data covariance matrix must satisfy in order to guarantee the equivalence between single and multiple data assimilation. Then, we investigate the use of MDA with the ensemble smoother (ES) to assimilate production data.

This work was motivated by the comments made on pages 44, 45, 86 and 109 of Rommelse [149]. Rommelse suggested that when assimilation of accurate data requires a "large jump" between the forecast and analyzed states, the magnitude of

the jump is overestimated by the linear update used in EnKF. He suggested that this overestimation of the magnitude of the jump could be reduced by assimilating the same data multiple times with increased measurement errors. Rommelse derived equations for determining the number of times that the data should be assimilated with increased measurement errors in order to obtain the same posterior variance for the state. His derivation, however, was based on a one-dimensional forecast vector, with a single datum and sensitivity matrix equal to the 1×1 identity matrix. Rommelse applied this idea in a multi-dimensional example, but he did not compare multiple assimilations of data with standard EnKF. He also did not discuss important implementation details. In a sense, the work presented here extends the theoretical equations on pages 44 and 45 of Rommelse [149] to multi-dimensional problems with an arbitrary sensitivity matrix and a general measurement error covariance matrix. However, our derivations assume that we multiply the covariance matrix of the measurement errors by the number of data assimilations, whereas it appears that Rommelse assimilated data one by one using an ensemble square root filter [186]. In this case, each covariance matrix of the measurement errors at each data assimilation is 1×1. Thus, he could assimilate each individual datum a different number of times.

The results presented in this chapter were published in Emerick and Reynolds [41] and Emerick and Reynolds [39].

6.1 Multiple Data Assimilation for the Linear-Gaussian Case

In this section, we show the equivalence between single and multiple data assimilation for the linear case using the Kalman filter (KF), where the full-rank covariance C_Y^f is updated.

Proposition: *For the linear case, with a Gaussian prior and Gaussian noise in the measurements (linear-Gaussian problem), applying the KF to assimilate data N_a times with the measurement error covariance matrix multiplied by N_a is equivalent to assimilating the same data only once with the original measurement error covariance*

matrix.

Proof: First, we note that as shown in Chapter 2 (Section 2.1), for the linear-Gaussian problem the posterior pdf of model parameters conditional to the observations is also Gaussian [135, Chap. 7]. Hence, to prove the proposition, it is only necessary to show that both procedures lead to the same posterior covariance matrix and the posterior mean.

The linear problem refers to the case in which the relation between the vector of predicted data, d^f, and the state vector, y^f, is expressed in the form

$$d^f = Hy^f, \tag{6.1}$$

where H is the $N_n \times N_y$ sensitivity matrix. For the linear-Gaussian problem the posterior covariance, C_{Y}^a, and posterior mean, μ_y^a, are given by [135, Chap. 7]

$$\begin{aligned} C_{\mathrm{Y}}^a &= \left((C_{\mathrm{Y}}^f)^{-1} + H^{\mathrm{T}} C_{\mathrm{D}}^{-1} H\right)^{-1} \\ &= C_{\mathrm{Y}}^f - C_{\mathrm{Y}}^f H^{\mathrm{T}} \left(C_{\mathrm{D}} + H C_{\mathrm{Y}}^f H^{\mathrm{T}}\right)^{-1} H C_{\mathrm{Y}}^f \end{aligned} \tag{6.2}$$

and

$$\begin{aligned} \mu_y^a &= C_{\mathrm{Y}}^a \left((C_{\mathrm{Y}}^f)^{-1}\mu_y^f + H^{\mathrm{T}} C_{\mathrm{D}}^{-1} d_{\mathrm{obs}}\right) \\ &= \mu_y^f + C_{\mathrm{Y}}^f H^{\mathrm{T}} \left(C_{\mathrm{D}} + H C_{\mathrm{Y}}^f H^{\mathrm{T}}\right)^{-1} \left(d_{\mathrm{obs}} - H\mu_y^f\right). \end{aligned} \tag{6.3}$$

To assimilate data N_a times with the measurement error covariance multiplied by N_a, we define

$$\widehat{d}_{\text{obs}} \equiv \begin{bmatrix} d_{\text{obs}} \\ \vdots \\ d_{\text{obs}} \end{bmatrix}, \tag{6.4}$$

$$\widehat{H} \equiv \begin{bmatrix} H \\ \vdots \\ H \end{bmatrix} \tag{6.5}$$

and

$$\widehat{C}_{\text{D}} \equiv \begin{bmatrix} N_a C_{\text{D}} & 0 & \cdots & 0 \\ 0 & N_a C_{\text{D}} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & N_a C_{\text{D}} \end{bmatrix}. \tag{6.6}$$

In the above definitions, we simply repeated the vector d_{obs} and the matrices H and C_{D} N_a times. For assimilating data multiple times, the linear relation (Eq. 6.1) becomes

$$\widehat{d}^f = \widehat{H} y^f. \tag{6.7}$$

Now, we develop an expression for $\widehat{C}^a_{\text{Y}}$, which denotes the posterior covariance for the case with MDA. From Eq. 6.2, we can write $\widehat{C}^a_{\text{Y}}$ as

$$\begin{aligned} \widehat{C}^a_{\text{Y}} &= \left((C^f_{\text{Y}})^{-1} + \widehat{H}^{\text{T}} \widehat{C}^{-1}_{\text{D}} \widehat{H} \right)^{-1} \\ &= C^f_{\text{Y}} - C^f_{\text{Y}} \widehat{H}^{\text{T}} \left(\widehat{C}_{\text{D}} + \widehat{H} C^f_{\text{Y}} \widehat{H}^{\text{T}} \right)^{-1} \widehat{H} C^f_{\text{Y}}. \end{aligned} \tag{6.8}$$

Using definitions (6.4), (6.5) and (6.6) in Eq. 6.8, we can write this posterior covariance matrix as

$$\widehat{C}_{\mathrm{Y}}^{a} = C_{\mathrm{Y}}^{f} - \left[\begin{array}{ccc} C_{\mathrm{Y}}^{f} H^{\mathrm{T}} & \cdots & C_{\mathrm{Y}}^{f} H^{\mathrm{T}} \end{array} \right] C^{-1} \left[\begin{array}{c} H C_{\mathrm{Y}}^{f} \\ \vdots \\ H C_{\mathrm{Y}}^{f} \end{array} \right]. \tag{6.9}$$

In Eq. 6.9, the matrix C is defined by

$$\begin{aligned} C \equiv & \left[\begin{array}{cccc} N_a C_{\mathrm{D}} & 0 & \cdots & 0 \\ 0 & N_a C_{\mathrm{D}} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & N_a C_{\mathrm{D}} \end{array} \right] \\ & + \left[\begin{array}{cccc} H C_{\mathrm{Y}}^{f} H^{\mathrm{T}} & H C_{\mathrm{Y}}^{f} H^{\mathrm{T}} & \cdots & H C_{\mathrm{Y}}^{f} H^{\mathrm{T}} \\ H C_{\mathrm{Y}}^{f} H^{\mathrm{T}} & H C_{\mathrm{Y}}^{f} H^{\mathrm{T}} & \cdots & H C_{\mathrm{Y}}^{f} H^{\mathrm{T}} \\ \vdots & & \ddots & \vdots \\ H C_{\mathrm{Y}}^{f} H^{\mathrm{T}} & \cdots & & H C_{\mathrm{Y}}^{f} H^{\mathrm{T}} \end{array} \right] \\ = & \left[\begin{array}{cccc} N_a C_{\mathrm{D}} + C_{\mathrm{DD}}^{f} & C_{\mathrm{DD}}^{f} & \cdots & C_{\mathrm{DD}}^{f} \\ C_{\mathrm{DD}}^{f} & N_a C_{\mathrm{D}} + C_{\mathrm{DD}}^{f} & \cdots & C_{\mathrm{DD}}^{f} \\ \vdots & & \ddots & \vdots \\ C_{\mathrm{DD}}^{f} & \cdots & & N_a C_{\mathrm{D}} + C_{\mathrm{DD}}^{f}. \end{array} \right]. \end{aligned} \tag{6.10}$$

In Eq. 6.10, we used $C_{\mathrm{DD}}^{f} = H C_{\mathrm{Y}}^{f} H^{\mathrm{T}}$ to simplify notation.

We need to develop an expression for C^{-1} in Eq. 6.9. As C is a real-symmetric positive-definite matrix, its inverse is also real-symmetric positive-definite. Hence, we can write the product $CC^{-1} = I$ as

$$
\begin{aligned}
I &= CC^{-1} \\
&= \begin{bmatrix}
N_a C_{\mathrm{D}} + C^f_{\mathrm{DD}} & C^f_{\mathrm{DD}} & \cdots & C^f_{\mathrm{DD}} \\
C^f_{\mathrm{DD}} & N_a C_{\mathrm{D}} + C^f_{\mathrm{DD}} & \cdots & C^f_{\mathrm{DD}} \\
\vdots & & \ddots & \vdots \\
C^f_{\mathrm{DD}} & \cdots & & N_a C_{\mathrm{D}} + C^f_{\mathrm{DD}}
\end{bmatrix} \\
&\times \begin{bmatrix}
A & B & \cdots & B \\
B & A & \cdots & B \\
\vdots & & \ddots & \vdots \\
B & \cdots & & A
\end{bmatrix} = \begin{bmatrix}
I_{N_n} & 0 & \cdots & 0 \\
0 & I_{N_n} & \cdots & 0 \\
\vdots & & \ddots & \vdots \\
0 & \cdots & & I_{N_n}
\end{bmatrix},
\end{aligned}
\tag{6.11}
$$

where I_{N_n} denotes the $N_n \times N_n$ identity matrix. Thus, in order to compute C^{-1}, it is only necessary to develop expressions for the submatrices A and B. From (6.11), it is straightforward to obtain

$$
\left(N_a C_{\mathrm{D}} + C^f_{\mathrm{DD}}\right) A + \left(N_a - 1\right) C^f_{\mathrm{DD}} B = I_{N_n} \tag{6.12}
$$

and

$$
C^f_{\mathrm{DD}} A + \left(N_a C_{\mathrm{D}} + C^f_{\mathrm{DD}}\right) B + \left(N_a - 2\right) C^f_{\mathrm{DD}} B = 0. \tag{6.13}
$$

Subtracting (6.13) from (6.12), we obtain

$$
A = \frac{1}{N_a} C_{\mathrm{D}}^{-1} + B, \tag{6.14}
$$

and using this result in (6.13), we obtain

$$
B = -\frac{1}{N_a^2} \left(C_{\mathrm{D}} + C^f_{\mathrm{DD}}\right)^{-1} C^f_{\mathrm{DD}} C_{\mathrm{D}}^{-1}. \tag{6.15}
$$

Using Eqs. 6.14 and 6.15 in Eq. 6.9, we obtain

$$
\begin{aligned}
\widehat{C}_{\mathrm{Y}}^{a} &= C_{\mathrm{Y}}^{f} - \left[\begin{array}{ccc} C_{\mathrm{Y}}^{f}H^{\mathrm{T}} & \cdots & C_{\mathrm{Y}}^{f}H^{\mathrm{T}} \end{array}\right] \\
&\times \left[\begin{array}{cccc} \frac{1}{N_a}C_{\mathrm{D}}^{-1}+B & B & \cdots & B \\ B & \frac{1}{N_a}C_{\mathrm{D}}^{-1}+B & \cdots & B \\ \vdots & & \ddots & \vdots \\ B & \cdots & & \frac{1}{N_a}C_{\mathrm{D}}^{-1}+B \end{array}\right] \left[\begin{array}{c} HC_{\mathrm{Y}}^{f} \\ \vdots \\ HC_{\mathrm{Y}}^{f} \end{array}\right] \\
&= C_{\mathrm{Y}}^{f} - \left[\begin{array}{ccc} C_{\mathrm{Y}}^{f}H^{\mathrm{T}} & \cdots & C_{\mathrm{Y}}^{f}H^{\mathrm{T}} \end{array}\right] \left[\begin{array}{c} \left(\frac{1}{N_a}C_{\mathrm{D}}^{-1}+N_aB\right)HC_{\mathrm{Y}}^{f} \\ \vdots \\ \left(\frac{1}{N_a}C_{\mathrm{D}}^{-1}+N_aB\right)HC_{\mathrm{Y}}^{f} \end{array}\right] \\
&= C_{\mathrm{Y}}^{f} - N_aC_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left(\frac{1}{N_a}C_{\mathrm{D}}^{-1}+N_aB\right)HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{f} - C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left[C_{\mathrm{D}}^{-1} - \left(C_{\mathrm{D}}+C_{\mathrm{DD}}^{f}\right)^{-1}C_{\mathrm{DD}}^{f}C_{\mathrm{D}}^{-1}\right]HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{f} - C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left[I_{N_n} - \left(C_{\mathrm{D}}+C_{\mathrm{DD}}^{f}\right)^{-1}C_{\mathrm{DD}}^{f}\right]C_{\mathrm{D}}^{-1}HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{f} - C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left[\left(C_{\mathrm{D}}+C_{\mathrm{DD}}^{f}\right)^{-1}\left(C_{\mathrm{D}}+C_{\mathrm{DD}}^{f}\right) - \left(C_{\mathrm{D}}+C_{\mathrm{DD}}^{f}\right)^{-1}C_{\mathrm{DD}}^{f}\right]C_{\mathrm{D}}^{-1}HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{f} - C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left(C_{\mathrm{D}}+C_{\mathrm{DD}}^{f}\right)^{-1}\left[C_{\mathrm{D}}+C_{\mathrm{DD}}^{f}-C_{\mathrm{DD}}^{f}\right]C_{\mathrm{D}}^{-1}HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{f} - C_{\mathrm{Y}}^{f}H^{\mathrm{T}}\left(C_{\mathrm{D}}+HC_{\mathrm{Y}}^{f}H^{\mathrm{T}}\right)^{-1}HC_{\mathrm{Y}}^{f} \\
&= C_{\mathrm{Y}}^{a}. \qquad (6.16)
\end{aligned}
$$

Eq. 6.16 shows that the posterior covariance matrix obtained by assimilating data N_a times with the measurement error covariance matrix multiplied by N_a is the same as the posterior covariance matrix obtained by assimilating data only once with the original measurement error covariance matrix. Following the same procedure, it is straightforward to show the equivalence of the posterior mean by starting with Eq. 6.3 and using the results in Eqs. 6.14 and 6.15.

EnKF becomes equivalent to the KF for the linear-Gaussian case as the size of

the ensemble goes to infinity [46, Chap. 4]. Thus, the equivalence between single and multiple data assimilation established in this section guarantees that as $N_e \to \infty$, EnKF with MDA is still consistent with the KF.

6.1.1 Sampling the posterior pdf with multiple data assimilation for the linear case

In the previous section, we showed the equivalence between single and multiple data assimilation for the linear-Gaussian case using the KF. In the KF, the mean and covariance are updated sequentially in time. For the EnKF, on the other hand, we sequentially update an ensemble of augmented state vectors generated by sampling the prior distribution in order to obtain a sampling of the posterior distribution. In this section, we demonstrate that for the linear-Gaussian case, MDA samples the posterior pdf correctly if we use the full-rank forecast covariance matrix C_{Y}^f. The derivation presented here explicitly shows that in the multiple data assimilation case, we need to perturb the observations based on the measurement error covariance matrix multiplied by the number of data assimilations. The derivation presented here follows Reynolds et al. [142], where it is shown that RML samples the posterior pdf correctly for the linear-Gaussian case.

For sampling the posterior pdf with MDA, we start with a sample from the prior (forecast) pdf, denoted by y^f, i.e., $y^f \sim \mathcal{N}(\mu_y^f, C_{\mathrm{Y}}^f)$. In addition, we perturb the observations by sampling $d_{\mathrm{uc}} \sim \mathcal{N}(d_{\mathrm{obs}}, N_a C_{\mathrm{D}})$. As before, to assimilate data N_a times, we define

$$\widehat{d}_{\mathrm{uc}} \equiv \begin{bmatrix} d_{\mathrm{uc}}^1 \\ \vdots \\ d_{\mathrm{uc}}^{N_a} \end{bmatrix}, \tag{6.17}$$

where $d_{\mathrm{uc}}^\ell \sim \mathcal{N}(d_{\mathrm{obs}}, N_a C_{\mathrm{D}})$, for $\ell = 1, 2, \cdots N_a$.

The derivation has two parts. In the first part, we define the vector $\widehat{y}^a$ as the minimizer of the RML objective function [142] modified for the multiple data

assimilation case, i.e.,

$$\widehat{y}^a = \arg\min_y \widehat{O}(y), \tag{6.18}$$

where

$$\widehat{O}(y) = \frac{1}{2}\left(y - y^f\right)^{\mathrm{T}} \left(C_{\mathrm{Y}}^f\right)^{-1} \left(y - y^f\right) + \frac{1}{2}\left(\widehat{H}y - \widehat{d}_{\mathrm{uc}}\right)^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-1} \left(\widehat{H}y - \widehat{d}_{\mathrm{uc}}\right). \tag{6.19}$$

$\widehat{H}$ and $\widehat{C}_{\mathrm{D}}$ were defined before (Eqs. 6.5 and 6.6). Minimizing $\widehat{O}(y)$ is equivalent to assimilating $\widehat{d}_{\mathrm{uc}}$ using $\mu_y^f = y^f$ with the KF [191]. In the second part of the derivation, we show that $\widehat{y}^a$ has the correct posterior pdf. Note that $\widehat{y}^a$ is a Gaussian random vector because y^f and $\widehat{d}_{\mathrm{uc}}$ are Gaussian random vectors.

Part 1: Finding $\widehat{y}^a$:

Requiring the gradient of $\widehat{O}(y)$ to vanish, we obtain

$$\begin{aligned} 0 &= \nabla_y \widehat{O}(y) \\ &= \left(C_{\mathrm{Y}}^f\right)^{-1}\left(y - y^f\right) + \widehat{H}^{\mathrm{T}}\widehat{C}_{\mathrm{D}}^{-1}\left(\widehat{H}y - \widehat{d}_{\mathrm{uc}}\right). \end{aligned} \tag{6.20}$$

Solving Eq. 6.20 for y and denoting the result as $\widehat{y}^a$, we obtain

$$\begin{aligned} \widehat{y}^a &= \left(\left(C_{\mathrm{Y}}^f\right)^{-1} + \widehat{H}^{\mathrm{T}}\widehat{C}_{\mathrm{D}}^{-1}\widehat{H}\right)^{-1}\left(\left(C_{\mathrm{Y}}^f\right)^{-1} y^f + \widehat{H}^{\mathrm{T}}\widehat{C}_{\mathrm{D}}^{-1}\widehat{d}_{\mathrm{uc}}\right) \\ &= C_{\mathrm{Y}}^a\left(\left(C_{\mathrm{Y}}^f\right)^{-1} y^f + \widehat{H}^{\mathrm{T}}\widehat{C}_{\mathrm{D}}^{-1}\widehat{d}_{\mathrm{uc}}\right). \end{aligned} \tag{6.21}$$

The last equality of Eq. 6.21 follows from Eqs. 6.2 and 6.16.

The vectors y^f and $\widehat{d}_{\mathrm{uc}}$ are samples of Gaussian distributions and can be

obtained using the square roots of the respective covariance matrices, i.e.,

$$y^f = \mu_y^f + (C_{\mathrm{Y}}^f)^{1/2} z_y \tag{6.22}$$

and

$$\widehat{d}_{\mathrm{uc}} = \widehat{d}_{\mathrm{obs}} + \widehat{C}_{\mathrm{D}}^{1/2} \widehat{z}_d. \tag{6.23}$$

In Eqs. 6.22 and 6.23, z_y and $\widehat{z}_d$ are normally distributed random vectors, i.e., $z_y \sim \mathcal{N}(0, I_{N_y})$ and $\widehat{z}_d \sim \mathcal{N}(0, I_{N_d})$, where I_{N_y} is the $N_y \times N_y$ identity matrix and I_{N_d} is the $N_d \times N_d$ identity matrix, with $N_d = N_n \times N_a$ denoting the number of data points at the nth data assimilation time-step multiplied by the number of times these data are assimilated.

Using Eqs. 6.22 and 6.23 in Eq. 6.21 results in our final expression for $\widehat{y}^a$:

$$\widehat{y}^a = C_{\mathrm{Y}}^a (C_{\mathrm{Y}}^f)^{-1} \left(\mu_y^f + (C_{\mathrm{Y}}^f)^{1/2} z_y \right) + C_{\mathrm{Y}}^a \widehat{H}^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-1} \left(\widehat{d}_{\mathrm{obs}} + \widehat{C}_{\mathrm{D}}^{1/2} \widehat{z}_d \right). \tag{6.24}$$

Part 2: Proving that the pdf for $\widehat{y}^a$ is the correct posterior pdf:

In the second part of the derivation, we show that sampling $\widehat{y}^a$ using Eq. 6.24 is equivalent to sampling the posterior pdf, i.e., $\widehat{y}^a \sim \mathcal{N}(\mu_y^a, C_{\mathrm{Y}}^a)$. Because the posterior pdf is Gaussian, we only need to show that

$$\mathrm{E}\left[\widehat{y}^a\right] = \mu_y^a \tag{6.25}$$

and

$$\mathrm{cov}\left[\widehat{y}^a\right] = \mathrm{E}\left[(\widehat{y}^a - \mu_y^a)(\widehat{y}^a - \mu_y^a)^{\mathrm{T}}\right] = C_{\mathrm{Y}}^a. \tag{6.26}$$

Using $\mathrm{E}[z_y] = 0$ and $\mathrm{E}[\widehat{z}_d] = 0$ and Eq. 6.3, it follows that the expectation of $\widehat{y}^a$ is given by

$$
\begin{aligned}
\mathrm{E}\left[\widehat{y}^a\right] &= C_{\mathrm{Y}}^a (C_{\mathrm{Y}}^f)^{-1} \left(\mu_y^f + (C_{\mathrm{Y}}^f)^{1/2} \mathrm{E}[z_y]\right) + C_{\mathrm{Y}}^a \widehat{H}^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-1} \left(\widehat{d}_{\mathrm{obs}} + \widehat{C}_{\mathrm{D}}^{1/2} \mathrm{E}[\widehat{z}_d]\right) \\
&= C_{\mathrm{Y}}^a (C_{\mathrm{Y}}^f)^{-1} \mu_y^f + C_{\mathrm{Y}}^a \widehat{H}^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-1} \widehat{d}_{\mathrm{obs}} \\
&= C_{\mathrm{Y}}^a (C_{\mathrm{Y}}^f)^{-1} \mu_y^f + C_{\mathrm{Y}}^a \begin{bmatrix} H^{\mathrm{T}} & \cdots & H^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \frac{1}{N_a} C_{\mathrm{D}}^{-1} d_{\mathrm{obs}} \\ \vdots \\ \frac{1}{N_a} C_{\mathrm{D}}^{-1} d_{\mathrm{obs}} \end{bmatrix} \\
&= C_{\mathrm{Y}}^a \left((C_{\mathrm{Y}}^f)^{-1} \mu_y^f + H^{\mathrm{T}} C_{\mathrm{D}}^{-1} d_{\mathrm{obs}}\right) \\
&= \mu_y^a. \qquad (6.27)
\end{aligned}
$$

Note that embedded in Eq. 6.27 is the equality

$$
\mu_y^a = C_{\mathrm{Y}}^a \left((C_{\mathrm{Y}}^f)^{-1} \mu_y^f + \widehat{H}^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-1} \widehat{d}_{\mathrm{obs}}\right). \qquad (6.28)
$$

To obtain the posterior covariance, we first compute $\widehat{y}^a - \mu_y^a$ by subtracting Eq. 6.28 from Eq. 6.24, which leads to

$$
\widehat{y}^a - \mu_y^a = C_{\mathrm{Y}}^a \left[(C_{\mathrm{Y}}^f)^{-\mathrm{T}/2} z_y + \widehat{H}^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-\mathrm{T}/2} \widehat{z}_d\right]. \qquad (6.29)
$$

In Eq. 6.29, we allowed the possibility that the square root of the covariance matrices C_{Y}^f and $\widehat{C}_{\mathrm{D}}$ are based on the Cholesky decomposition, i.e., $C_{\mathrm{Y}}^f = (C_{\mathrm{Y}}^f)^{1/2} (C_{\mathrm{Y}}^f)^{\mathrm{T}/2}$ and $\widehat{C}_{\mathrm{D}} = \widehat{C}_{\mathrm{D}}^{1/2} \widehat{C}_{\mathrm{D}}^{\mathrm{T}/2}$, instead of requiring the symmetric square roots from the spectral decomposition of C_{Y}^f and $\widehat{C}_{\mathrm{D}}$. For this reason, we obtained the transposes in Eq. 6.29.

Using Eq. 6.29 in Eq. 6.26, we obtain

$$
\begin{aligned}
\operatorname{cov}\left[\widehat{y}^a\right] &= \mathrm{E}\left[(\widehat{y}^a - \mu_y^a)(\widehat{y}^a - \mu_y^a)^{\mathrm{T}}\right] \\
&= C_{\mathrm{Y}}^a (C_{\mathrm{Y}}^f)^{-1/2} \mathrm{E}[z_y z_y^{\mathrm{T}}] (C_{\mathrm{Y}}^f)^{-\mathrm{T}/2} C_{\mathrm{Y}}^a + C_{\mathrm{Y}}^a (C_{\mathrm{Y}}^f)^{-1/2} \mathrm{E}[z_y \widehat{z}_d^{\mathrm{T}}] \widehat{C}_{\mathrm{D}}^{-\mathrm{T}/2} \widehat{H} C_{\mathrm{Y}}^a \\
&+ C_{\mathrm{Y}}^a \widehat{H}^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-1/2} \mathrm{E}[\widehat{z}_d z_y^{\mathrm{T}}] (C_{\mathrm{Y}}^f)^{-\mathrm{T}/2} C_{\mathrm{Y}}^a + C_{\mathrm{Y}}^a \widehat{H}^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-1/2} \mathrm{E}[\widehat{z}_d \widehat{z}_d^{\mathrm{T}}] \widehat{C}_{\mathrm{D}}^{-\mathrm{T}/2} \widehat{H} C_{\mathrm{Y}}^a.
\end{aligned}
$$

Noting that

$$
\mathrm{E}\left[z_y z_y^{\mathrm{T}}\right] = I_{N_y}, \tag{6.30}
$$

$$
\mathrm{E}\left[\widehat{z}_d \widehat{z}_d^{\mathrm{T}}\right] = I_{N_d}, \tag{6.31}
$$

$$
\mathrm{E}\left[z_y \widehat{z}_d^{\mathrm{T}}\right] = 0 \tag{6.32}
$$

and

$$
\mathrm{E}\left[\widehat{z}_d z_y^{\mathrm{T}}\right] = 0, \tag{6.33}
$$

Eq. 6.30 reduces to

$$
\operatorname{cov}\left[\widehat{y}^a\right] = C_{\mathrm{Y}}^a \left((C_{\mathrm{Y}}^f)^{-1} + \widehat{H}^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-1} \widehat{H}\right) C_{\mathrm{Y}}^a. \tag{6.34}
$$

In the previous section, we showed that

$$
\widehat{C}_{\mathrm{Y}}^a = \left((C_{\mathrm{Y}}^f)^{-1} + \widehat{H}^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-1} \widehat{H}\right)^{-1} = C_{\mathrm{Y}}^a. \tag{6.35}
$$

Using this result in Eq. 6.34, we obtain

$$
\operatorname{cov}\left[\widehat{y}^a\right] = C_{\mathrm{Y}}^a \left(C_{\mathrm{Y}}^a\right)^{-1} C_{\mathrm{Y}}^a = C_{\mathrm{Y}}^a, \tag{6.36}
$$

which completes the proof.

This result shows that in order to sample the posterior pdf for the linear-Gaussian case with EnKF-MDA, we need to perturb the observations using the inflated covariance matrix, $\widehat{C}_{\mathrm{D}}$. Again, the proof was presented assuming the correct full-rank matrix C_{Y}^{f}, but will apply for EnKF when $N_e \to \infty$.

6.2 Interpretation of Multiple Data Assimilation for the Nonlinear Case

In the previous sections, we established the equivalence between single and multiple data assimilation for the linear-Gaussian case. However, for the nonlinear case, this equivalence does not hold. Intuitively, we can expect some benefit from assimilating data multiple times because we replace one potentially large update in the state vector with multiple smaller updates.

It is well known that when using the Gauss-Newton (GN) method for history matching reservoir models, convergence problems can occur due to overcorrection in the model parameters at early iterations. This overcorrection may result in unreasonably small or large values of some model parameters [188, 105]. This is particularly true when the initial guess for the GN iterative process gives predicted data far from the observations. Reynolds et al. [143] showed that EnKF is similar to applying GN sequentially with a full step and replacing the sensitivity matrix by an average sensitivity matrix obtained from the ensemble. Because of this similarity between the EnKF and GN method, it is reasonable to expect that EnKF analysis may also result in overcorrection in the states, especially if the predicted data are far from the observations at a particular data assimilation time-step.

One way to ameliorate overcorrection when applying gradient-based minimization is to increase the variance of the measurement errors [188, 59]. By using an artificially high value of data measurement errors during early iterations, the objective function that we minimize becomes more nearly quadratic so Newton-type

methods work well. As the algorithm improves the data match, we can eventually use the correct measurement error covariance matrix [59].

Li et al. [105] showed that the Levenberg-Marquardt (LM) algorithm provides a natural way to avoid these convergence difficulties observed in the GN method, but it is clear that any trust-region method [124, 127] could be used to achieve the same result. For the case where the number of measurements is less than the number of model parameters, a convenient way to write the LM update equation for sampling with RML is [192]

$$\begin{aligned} y_j^{\ell+1} &= y_j^\ell + \frac{y_j^f - y_j^\ell}{1+\lambda_\ell} + C_{\mathrm{Y}}^f H_\ell^{\mathrm{T}} \left[(1+\lambda_\ell) C_{\mathrm{D}} + H_\ell C_{\mathrm{Y}}^f H_\ell^{\mathrm{T}}\right]^{-1} \\ &\times \left\{ \frac{H_\ell \left(y_j^\ell - y_j^f\right)}{1+\lambda_\ell} + d_{\mathrm{uc},j} - d_j^\ell \right\}, \end{aligned} \tag{6.37}$$

where $\lambda_\ell \geq 0$ is the LM parameter. Note that by choosing $\lambda_\ell = 0$, Eq. 6.37 becomes the GN update equation with a full step. Because we can interpret EnKF as one GN iteration with an initial guess $y_j^0 = y_j^f$ and average sensitivity matrix, $\widehat{H}$, we can write the EnKF analogous LM equation as

$$y_j^1 = y_j^f + \widetilde{C}_{\mathrm{Y}}^f \widehat{H}^{\mathrm{T}} \left[(1+\lambda_0) C_{\mathrm{D}} + \widehat{H} \widetilde{C}_{\mathrm{Y}}^f \widehat{H}^{\mathrm{T}}\right]^{-1} \left(d_{\mathrm{uc},j} - d_j^f\right). \tag{6.38}$$

The similarity between Eq. 6.38 and the EnKF analysis equation (Eq. 2.49) is evident. Using $y_j^1 = y_j^a$, $\widetilde{C}_{\mathrm{Y}}^f \widehat{H}^{\mathrm{T}} = \widetilde{C}_{\mathrm{YD}}^f$ and $\widehat{H} \widetilde{C}_{\mathrm{Y}}^f \widehat{H}^{\mathrm{T}} = \widetilde{C}_{\mathrm{DD}}^f$, Eq. 6.38 represents the EnKF analysis equation with the covariance of measurement errors increased by the factor $1 + \lambda_0$. Note that we introduced the tilde in the matrices $\widetilde{C}_{\mathrm{Y}}^f$, $\widetilde{C}_{\mathrm{YD}}^f$ and $\widetilde{C}_{\mathrm{DD}}^f$ to emphasize that for EnKF, these matrices are approximated using the ensemble members. From this similarity between Eqs. 6.38 and 2.49, we can interpret the EnKF with MDA as applying the first iteration of the LM method N_a times with

$\lambda_0 = N_a - 1$. Note that because Eq. 6.38 is missing terms involving $y_j^\ell - y_j^f$ (Eq. 6.37), applying Eq. 6.38 is not the same as applying multiple consecutive LM iterations.

Another well-known advantage of the LM algorithm over the GN method is that increasing λ_ℓ tends to decrease the condition number of the matrix inverted in Eq. 6.37. A similar argument can be used for EnKF with MDA. To illustrate this argument, consider the case where the covariance matrix of the measurement errors is a diagonal matrix given by $C_{\mathrm{D}} = \sigma_d^2 I_{N_n}$. It is straightforward to show that in this case the condition number of the matrix $C = \widetilde{C}_{\mathrm{DD}}^f + C_{\mathrm{D}}$ is given by

$$\kappa(C) = \frac{\beta_{\max} + \sigma_d^2}{\beta_{\min} + \sigma_d^2}, \tag{6.39}$$

where $\beta_{\max}$ and $\beta_{\min}$, respectively, are the largest and the smallest eigenvalues of $\widetilde{C}_{\mathrm{DD}}^f$. For MDA, we have a matrix $\widehat{C} = \widetilde{C}_{\mathrm{DD}}^f + N_a C_{\mathrm{D}}$ with its condition number given by

$$\kappa(\widehat{C}) = \frac{\beta_{\max} + N_a\sigma_d^2}{\beta_{\min} + N_a\sigma_d^2}, \tag{6.40}$$

which is a decreasing function of N_a. Thus, $\kappa(\widehat{C}) < \kappa(C)$ for $N_a > 1$. Also note that $\kappa(\widehat{C}) \to 1$ as $N_a \to \infty$.

6.3 Implementation of EnKF with Multiple Data Assimilation

Algorithm 6.1 presents the general procedure of EnKF with multiple data assimilation (EnKF-MDA).

Algorithm 6.1 EnKF-MDA

1. Choose the number of data assimilations, N_a.
2. Set $\ell = 1$, $m_j^{n,1} = m_j^{n,f}$ and $p_j^{n-1,1} = p_j^{n-1,f}$.
3. While $\ell \leq N_a$ do:
   (a) For $j = 1$ to N_e do:
      i. Run the forward model from the last data assimilation time-step, t_{n-1}, until the next data assimilation time-step, t_n, and store the predicted data vector $d_j^{n,\ell}$.
      ii. If $\ell < N_a$, then build the forecast state vector, $y_j^{n,\ell}$, as

$$y_j^{n,\ell} = \begin{bmatrix} m_j^{n,\ell} \\ p_j^{n-1,\ell} \\ d_j^{n,\ell} \end{bmatrix}. \tag{6.41}$$

      else, build $y_j^{n,\ell}$ as

$$y_j^{n,\ell} = \begin{bmatrix} m_j^{n,\ell} \\ p_j^{n,\ell} \\ d_j^{n,\ell} \end{bmatrix}, \tag{6.42}$$

      (See comment 4 of Section 6.3.1.)
      iii. Perturb the observation vector using

$$d_{\text{uc},j}^n = d_{\text{obs}}^n + \sqrt{N_a}(C_{\text{D}}^n)^{1/2} z_n, \tag{6.43}$$

      where $z_n \sim \mathcal{N}(0, I_{N_n})$.
      iv. Update the state vector using EnKF analysis with $N_a C_{\text{D}}^n$ instead of C_{D}^n:

$$y_j^{n,\ell+1} = y_j^{n,\ell} + \widetilde{C}_{\text{Y}}^{n,\ell} H_n^{\text{T}} \left(N_a C_{\text{D}}^n + H_n \widetilde{C}_{\text{Y}}^{n,\ell} H_n^{\text{T}} \right)^{-1} \left(d_{\text{uc},j}^n - d_j^{n,\ell} \right). \tag{6.44}$$

   end (for).
   (b) $\ell = \ell + 1$

   end (while).
4. Set $y_j^{n,a} = y_j^{n,\ell}$ and go to the next data assimilation time-step.

6.3.1 Comments about the EnKF-MDA algorithm

1. The algorithm presented above refers to one data assimilation time-step. The same algorithm is applied for all time-steps.

2. Perhaps the most important thing to clarify about the implementation of the EnKF-MDA algorithm is that $d_j^{n,\ell}$ is included in $y_j^{n,\ell}$ to keep the notation consistent with Eqs. 2.36 and 2.49, where we used the augmented state vector to present EnKF equations. However, $d_j^{n,\ell}$ and $p_j^{n,\ell}$ are computed by running the forward model from time-step t_{n-1} to t_n using the reservoir state $p_j^{n-1,\ell}$ at time-step t_{n-1}. Additional clarification is given below.

3. With EnKF-MDA, at each data assimilation time-step, we do not assimilate the observed data N_a times simultaneously. Instead, we assimilate data N_a times consecutively and, after each of the N_a data assimilations, we rerun the forward model (reservoir simulator) starting from the previous time-step for the ensemble with the updated state vectors. With this procedure, we are effectively updating the "average sensitivity" before the next data assimilation. In a sense, EnKF-MDA can be interpreted as an iterative form of EnKF where the number of iterations is chosen a priori. Note that in the derivations presented for the linear-Gaussian case, we assimilate data N_a times simultaneously. However, for the linear-Gaussian case, consecutive and simultaneous data assimilations are equivalent [46, Chap. 7]. Hence, the derivations are still valid for assimilating data N_a times consecutively for the linear-Gaussian case.

4. In order to use MDA sequentially in time, it is necessary to introduce an additional modification in the data assimilation process to keep the updated ensemble of model parameters statistically consistent with the state of the dynamical system (primary variables of the reservoir simulator). The rigorous way to do this would be to rerun the reservoir simulator from time zero after

each data assimilation, but this is computationally expensive. Instead, in our EnKF-MDA algorithm, we update the primary variables ($p_j^{n-1,\ell}$) at time-step t_{n-1} with Eq. 6.44 for the first $N_a - 1$ data assimilations. These primary variables are used to restart the reservoir simulations during the MDA loop. In the last data assimilation, we update the vector of primary variables at the time-step t_n, i.e., $p_j^{n,\ell}$, which is used to restart simulations in the next time-step.

5. As in standard implementations of EnKF, $d_j^{n,\ell}$ does not need to be updated when using Eq. 2.49.

6. The computational cost of the proposed method is roughly N_a times the computational cost of data assimilation with standard EnKF.

7. For each of the N_a data assimilations, we recompute the perturbed observation vector, i.e., we resample $d_{\mathrm{uc},j}^n \sim \mathcal{N}(d_{\mathrm{obs}}^n, N_a C_{\mathrm{D}}^n)$ instead of using the same $d_{\mathrm{uc},j}^n$ for all N_a data assimilations. Intuitively, we expect that this procedure improves sampling because we reduce bias possibly introduced by matching "outliers" generated when sampling $d_{\mathrm{uc},j}^n$. Recall that the development presented in Section 6.1.1 indicates that we should resample d_{uc} at each of the N_a consecutive data assimilations.

6.4 Multiple Data Assimilation for Seismic Data

Reservoir production data are typically scarce spatially, but dense in time. As shown in Reynolds et al. [143], the application of EnKF is similar to one GN iteration using an average sensitivity matrix and full step in the search direction. Based on this analogy, we conjecture that sequential assimilation is one of the reasons that EnKF gives acceptable results when assimilating production data that are fairly closely-spaced in time. The process works such that at each data assimilation step, one GN correction is done to each realization in the ensemble of states, but because we have production data for several consecutive data assimilation time-steps, we accumulate

several GN corrections, keeping the ensemble conditioned to the production history. In contrast, seismic data are characterized by the large amount of data, spatially dense, but available at only a few times. Thus, we no longer have the beneficial effect of several consecutive data assimilation steps as observed in the production data case. While our conjecture that assimilating production data that are fairly closely spaced in time with EnKF provides an approximation of multiple GN iterations is based on intuition, it is well known that taking a full step at early iterations in the GN method when matching production data can lead to significant overcorrection (overshooting and undershooting) of reservoir parameters and result in unrealistically rough rock property fields [105, 59]. This overcorrection can be avoided either by applying some form of damping or constraining changes in model parameter at early iterations or by inflating the covariance matrix associated with measurement errors. This fact provides a strong motivation for the procedure considered in this paper, in which we apply EnKF to assimilate the same data multiple times with inflated covariance matrices.

6.4.1 Example 1 – Waterflooding

The first example is a two-phase (oil and water) synthetic reservoir model on a 2D uniform grid with 60×60 gridblocks. The dimensions of the gridblocks are 150 ft $\times$ 150 ft $\times$ 25 ft. The model parameters are gridblock log-permeabilities, $\ln(k)$'s. The true model was generated from an anisotropic exponential correlation function with a major correlation length of 3,750 ft (which corresponds to the width of 25 gridblocks) and a minor correlation length of 1,050 ft (i.e, seven gridblocks) oriented at 45°. The prior mean of $\ln(k)$ is 5.0 and the prior variance is 1.0 for all gridblocks. Fig. 6.1 shows the true permeability field used as the reference to generate the observed data. The porosities are constant and equal to 0.25 for all gridblocks. The compressibility of the rock, oil and water are also constant and equal to 5×10^{-6} psi^{-1}, 10^{-5} psi^{-1} and 10^{-6} psi^{-1}, respectively. In this model, there

are five producing wells and two water injectors. All producing wells are controlled by a fixed bottomhole pressure of 1,000 psi. The injectors are controlled by a fixed bottomhole pressure of 3,000 psi.

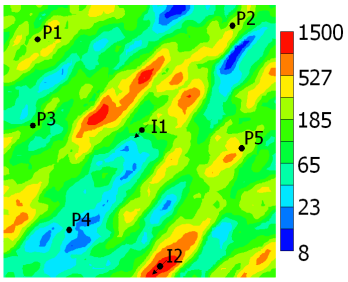
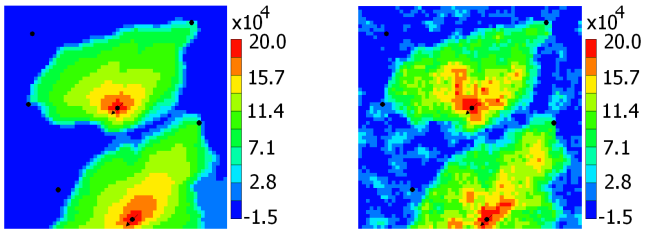


Figure 6.1: True permeability field (mD).

Assimilation of seismic data:

Synthetic time-lapse seismic data were generated based on the true reservoir model. The seismic data correspond to the P-impedance difference (ΔI_P) between a monitor survey after 3,900 days of production and a base survey before the beginning of the production. Fig. 6.2a presents the true ΔI_P data. Correlated random noise was added to the ΔI_P predicted by the true model to generate the synthetic observed data. The noise was generated using an isotropic spherical covariance function with a range equal to 750 ft (five gridblocks) and a standard deviation corresponding to 30% of the average ΔI_P data ($\sigma_{d,s} = 17,320$ lb/ft^{2}s). Fig. 6.2b shows the resulting seismic data with noise added. The seismic data used for history matching correspond to one datum per simulation gridblock.

Data assimilations with standard EnKF and EnKF-MDA assimilating data



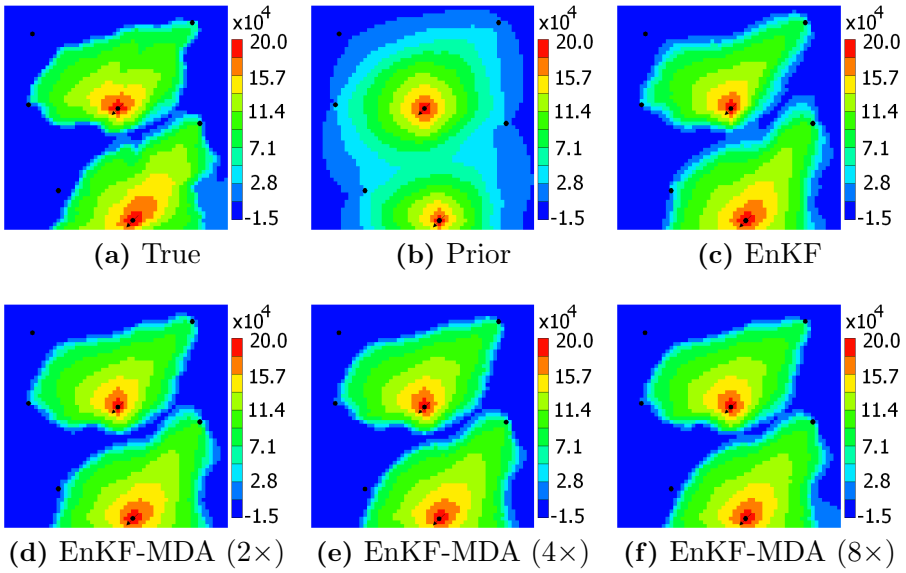
(a) True seismic **(b)** Seismic with noise

Figure 6.2: P-impedance difference (lb/ft^{2}s). True and data with noise.

two, four and eight times were performed to history match the time-lapse seismic with an ensemble size of 100. The models of the initial ensemble were generated using the same prior mean and covariance function used to generate the true model. During data assimilations, we used subspace inversion with rescaling. Truncation was done by retaining the largest singular values corresponding to 99.9% of the sum of the nonzero singular values when applying Eq. 2.75.

Fig. 6.3 presents the mean ensemble predictions of ΔI_P obtained with the prior ensemble and the final ensembles after EnKF and EnKF-MDA. For visual comparison, we also include the true ΔI_P in this figure. For EnKF and EnKF-MDA, the predicted ΔI_P was obtained by running the final ensembles from time zero. This figure shows clearly the improvement in the predicted ΔI_P compared to the prior ensemble. However, it is difficult to see the differences between EnKF and the EnKF-MDA cases. These differences are better visualized in the cross-plots presented in Fig. 6.4. This figure shows that increasing the number of data assimilations improved the match of time-lapse seismic. For four and eight data assimilations, the differences between the true and predicted seismic from the updated permeability fields are less than twice the standard deviation of the measurement errors for almost all reservoir gridblocks. Fig. 6.5 presents the final mean permeability fields obtained for each case. It is interesting to note that all permeability fields present some features of the true permeability, e.g., the low permeability in the region between the wells I1 (at the center of the reservoir) and P4 (at the lower left corner of the reservoir). However, the mean permeability fields obtained with MDA appear to be smoother than the one obtained with single data assimilation. Note, for example, that the case with single data assimilation (Fig. 6.5a) resulted in a relatively high permeability region between wells I1 and P2 (at the upper right corner of the reservoir), whereas after four data assimilations, the permeability in this region is considerably reduced.

In order to further compare the data matches, we repeated the data assim-



(a) True **(b)** Prior **(c)** EnKF

(d) EnKF-MDA (2×) **(e)** EnKF-MDA (4×) **(f)** EnKF-MDA (8×)

Figure 6.3: P-impedance difference (lb/ft^{2}s). True seismic data and mean ensemble predictions.

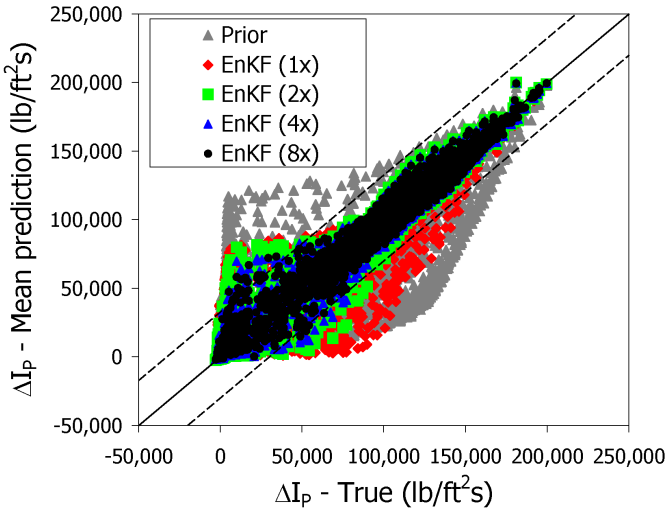
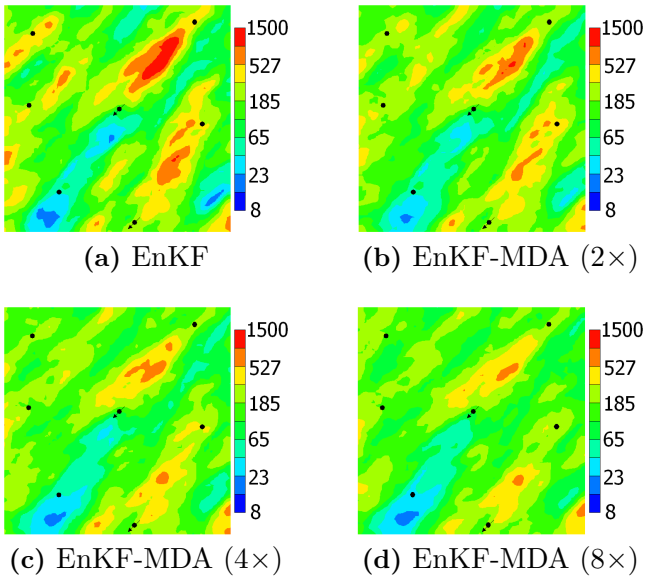


Figure 6.4: Cross-plots between true and mean predicted P-impedance changes (lb/ft^{2}s). The dashed lines correspond to +/− two standard deviations of the measurement errors.



(a) EnKF **(b)** EnKF-MDA (2×)

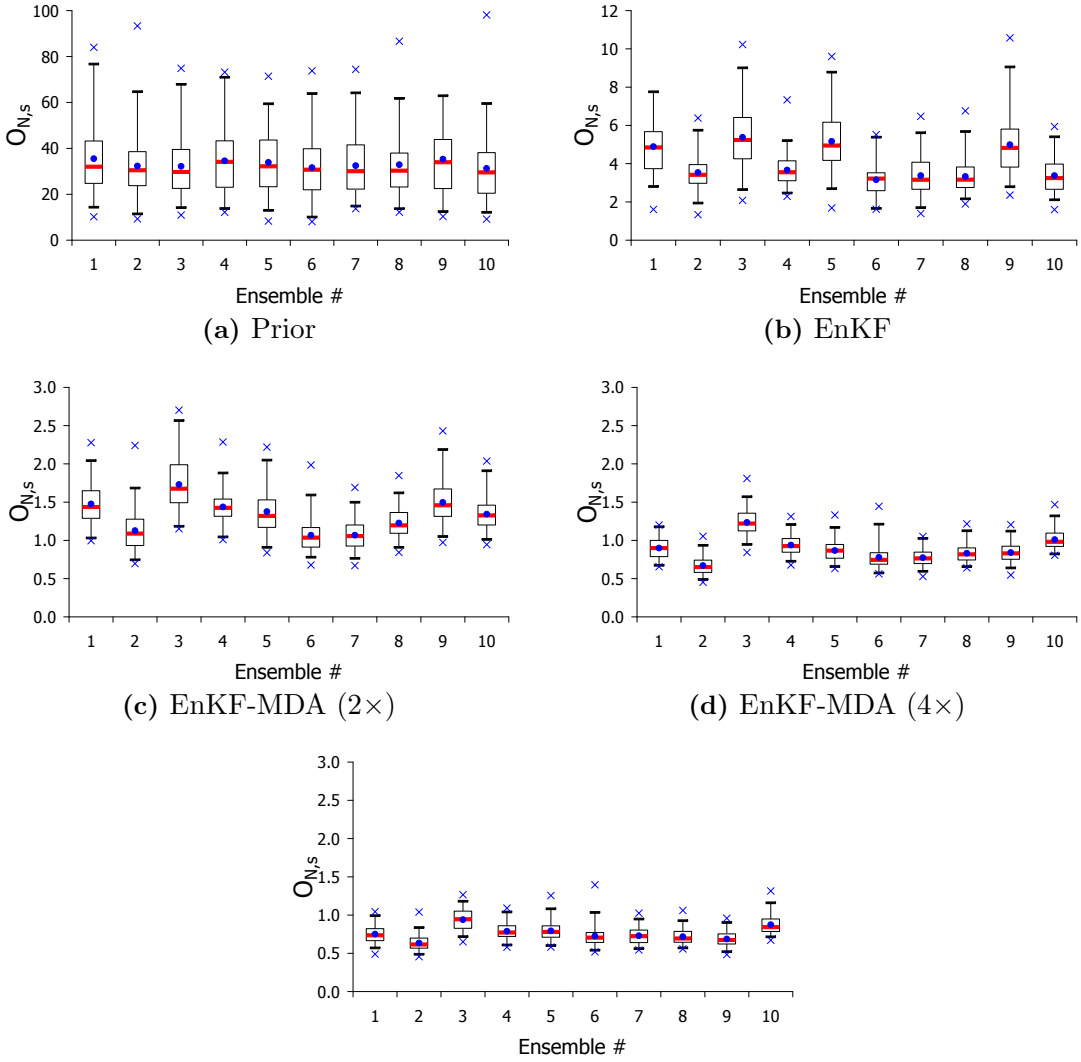
(c) EnKF-MDA (4×) **(d)** EnKF-MDA (8×)

Figure 6.5: Mean permeability fields (mD) after assimilation of seismic data.

ilations with ten different initial ensembles. For each ensemble, we computed the normalized seismic data mismatch objective function, $O_{N,s} = O_d(m)/N_{d,s}$, where $N_{d,s} = 3{,}600$ is the number of data points. Fig. 6.6 presents the box plots of the normalized seismic objective function obtained for each ensemble. Table 6.1 summarizes the results for the ten ensembles. The results presented in Fig. 6.6 and Table 6.1 indicate that increasing the number of data assimilations resulted in a consistently better seismic data match. Again, the main improvement occurred by increasing the number of data assimilations from one (average $O_{N,s} = 4.086$) to two (average $O_{N,s} = 1.335$).

Table 6.1: Mean and standard deviation of $O_{N,s}$ for ten different initial ensembles

Case	Mean	Standard deviation
Prior	33.195	14.261
EnKF	4.086	1.504
EnKF-MDA (2×)	1.335	0.323
EnKF-MDA (4×)	0.885	0.198
EnKF-MDA (8×)	0.764	0.144

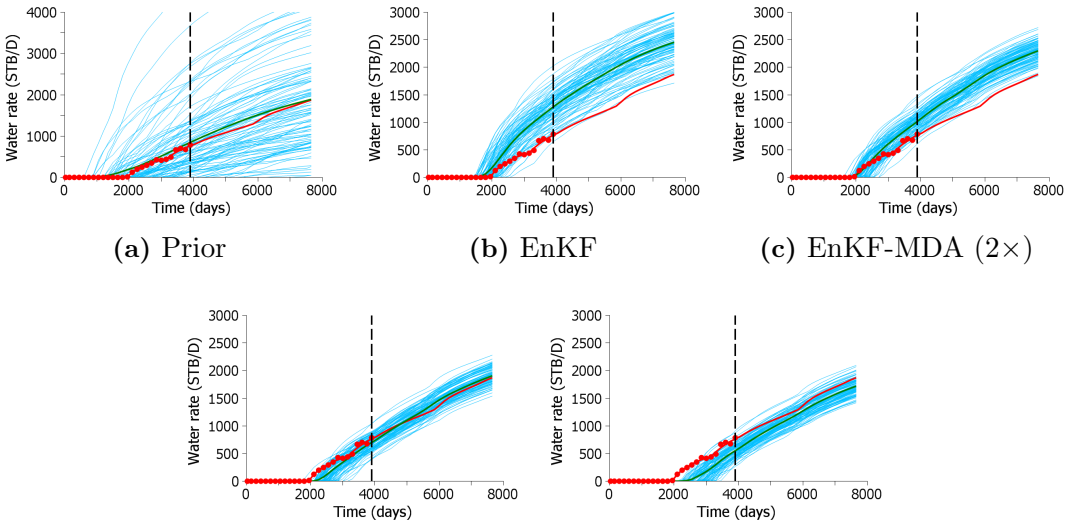


(a) Prior **(b)** EnKF

(c) EnKF-MDA (2×) **(d)** EnKF-MDA (4×)

(e) EnKF-MDA (8×)

Figure 6.6: Box plots of the normalized seismic objective function for ten different initial ensembles. Note that the vertical scales in the plots are different.



(a) Prior **(b)** EnKF **(c)** EnKF-MDA (2×)

(d) EnKF-MDA (4×) **(e)** EnKF-MDA (8×)

Figure 6.7: Field water production rate (stb/day) after assimilation of seismic data. The vertical dashed line indicates the end of the history. The red dots represent the production historical data; the red curve is the prediction from the true model; the green curve is the mean ensemble prediction and the light blue curves are the predictions from the ensemble members.

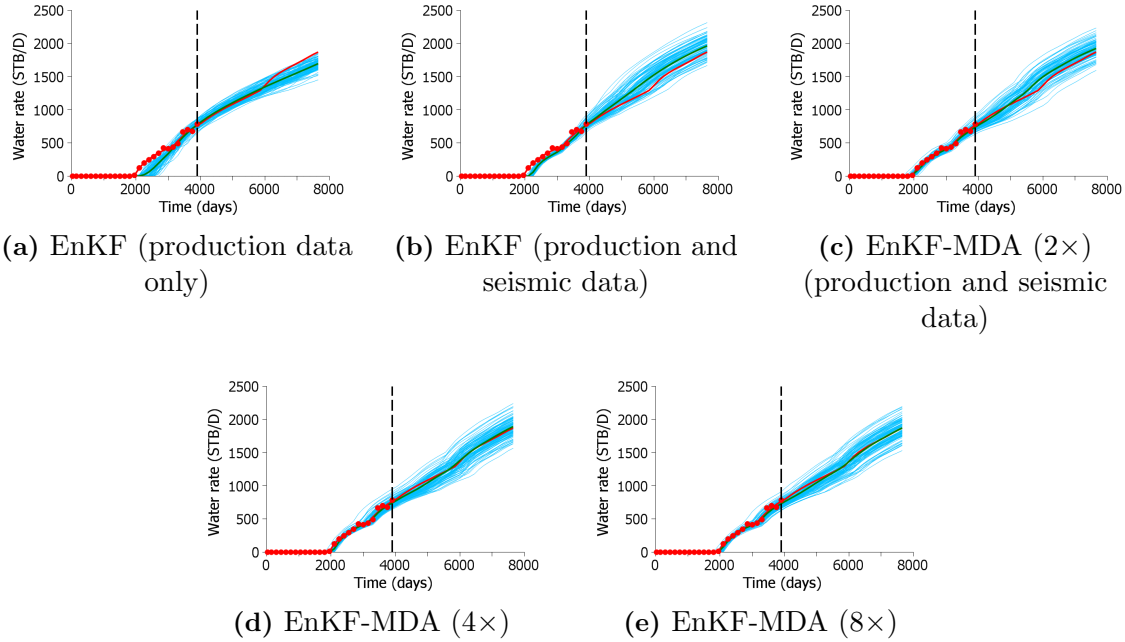
Production data prediction after assimilation of seismic data:

Fig. 6.7 presents the field water production rate obtained by simulation from time zero with the prior ensemble and the final ensembles after assimilation of the seismic data for the first of the ten ensembles considered in the previous section. In this figure, besides the historical period (3,900 days), we also include 3,750 days of forecast. It is important to emphasize that, even though Fig. 6.7 includes the production historical data, only seismic data were assimilated. Note that the assimilation of ΔI_P data with standard EnKF resulted in an ensemble which overestimated the field water production (Fig. 6.7b). Increasing the number of data assimilations of seismic data reduced the variance in the ensemble predictions of field water rate. After eight data assimilations, the ensemble underestimates the field water production. This may be surprising considering that we obtained a better match of seismic data with eight data assimilations. Note, however, that the results presented in Fig. 6.7 are after assimilating seismic data only; no production data were used.

Assimilation of production data:

In order to further investigate the effect of a better seismic data match on the production predictions, we now assimilate production data using the standard EnKF (i.e, single data assimilation) starting from the final ensembles obtained after assimilation of seismic data. For comparison, we also consider the case with assimilation of only production data with standard EnKF. The observed production data correspond to 3,900 days of oil and water production rates and water injection rates. The frequency of data assimilation corresponds to one data assimilation every 150 days. Random normally-distributed noise with zero mean and standard deviation equal to 5% of the true data was added to the true data to define the noisy observations.

Fig. 6.8 presents the field water production rate obtained by running final ensembles after assimilation of the production data for the ensemble corresponding to the results of Figs. 6.3–6.7. The results in this figure indicate that assimilation of seismic data improved the match of water production rate, especially at the breakthrough time (compare Fig. 6.8a with Figs. 6.8b–6.8e). Also, better matches of seismic data resulted in slightly better matches of water rate data (compare Figs. 6.8b and Figs. 6.8e), although the differences are not large. Table 6.2 presents the mean and standard deviation of the normalized production objective function, $O_{N,p} = O_d(m)/N_{d,p}$, obtained for the ten ensembles. $N_{d,p} = 299$ is the number of observed production data points. According to the results presented in Table 6.2, the ensembles obtained with multiple assimilations of seismic data resulted in better production data matches as well. Fig. 6.9 presents the final mean permeability fields for the first of the ten ensembles. Comparing Fig. 6.9 with Fig. 6.5, we observe that for the case with a single assimilation of the seismic data, it was necessary to make larger changes in the permeability field to history match the production data, which resulted in a rougher mean permeability field. Table 6.3 presents the mean and standard deviation of the normalized model mismatch objective function,



(a) EnKF (production data only)

(b) EnKF (production and seismic data)

(c) EnKF-MDA (2×) (production and seismic data)

(d) EnKF-MDA (4×) (production and seismic data)

(e) EnKF-MDA (8×) (production and seismic data)

Figure 6.8: Field water production rate (stb/day) after assimilation of production data. The colors in this figure have the same meaning as in Fig. 6.7. Note that EnKF-MDA refers to multiple assimilations of seismic data only. The production data were assimilated with standard EnKF for all cases.

$O_{N,m} = O_m(m)/N_d$, obtained for the ten ensembles. $N_d = N_{d,s} + N_{d,p}$ is the total number of data points. The results of Table 6.3 indicate that EnKF-MDA resulted in smoother permeability fields.

Table 6.4 presents the mean and standard deviation of the normalized seismic data mismatch objective function obtained after the assimilation of the production data. The results presented in this table indicate only a small deterioration in the seismic data matches. However, the multiple data assimilation cases still present better seismic data matches compared to standard EnKF.

6.4.2 Example 2 – Brugge case

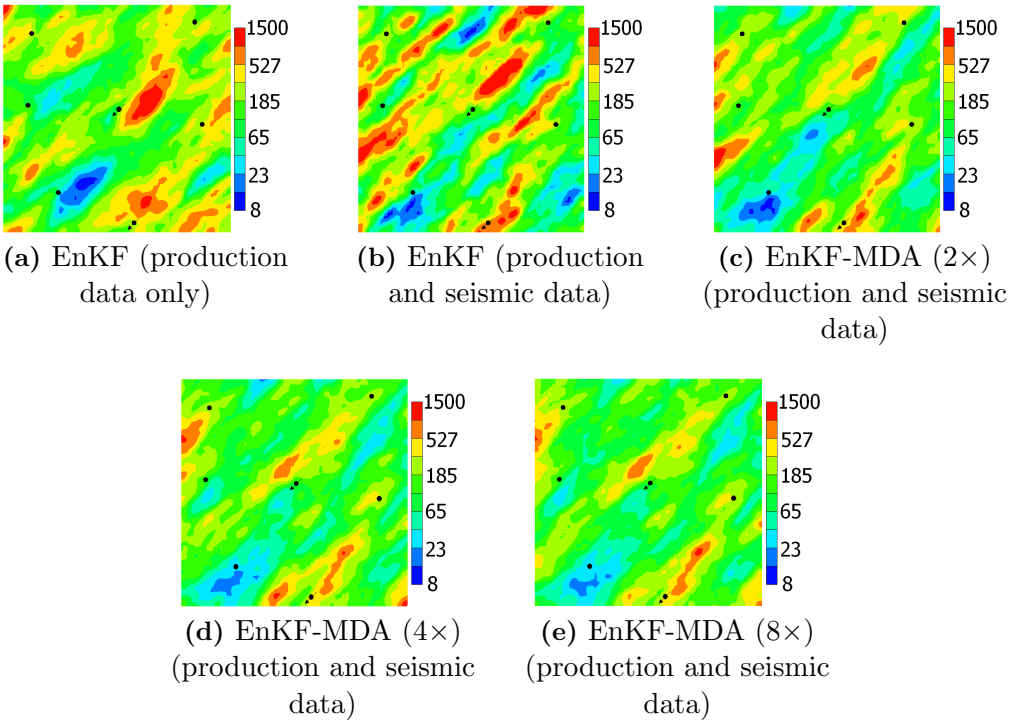
We also tested MDA in the Brugge field (Fig. 6.10) [137]. The Brugge case is a synthetic reservoir designed as a benchmark problem for evaluating methods for waterflooding optimization combined with history matching in a closed-loop work-

Table 6.2: Mean and standard deviation of $O_{N,p}$ for ten different initial ensembles

Case	Mean	Standard deviation
EnKF (production data only)	7.183	1.253
EnKF (prod. and seis.)	2.893	2.946
EnKF-MDA (2×) (prod. and seis.)	2.230	1.806
EnKF-MDA (4×) (prod. and seis.)	1.780	1.460
EnKF-MDA (8×) (prod. and seis.)	1.533	1.135

Table 6.3: Mean and standard deviation of $O_{N,m}$ for ten different initial ensembles

Case	Mean	Standard deviation
EnKF (production data only)	0.402	0.438
EnKF (prod. and seis.)	0.501	0.134
EnKF-MDA (2×) (prod. and seis.)	0.388	0.078
EnKF-MDA (4×) (prod. and seis.)	0.365	0.058
EnKF-MDA (8×) (prod. and seis.)	0.358	0.056



(a) EnKF (production data only)

(b) EnKF (production and seismic data)

(c) EnKF-MDA (2×) (production and seismic data)

(d) EnKF-MDA (4×) (production and seismic data)

(e) EnKF-MDA (8×) (production and seismic data)

Figure 6.9: Mean permeability fields (mD) after assimilation of production data. Note that EnKF-MDA refers to multiple assimilations of seismic data only. The production data were assimilated with standard EnKF for all cases.

Table 6.4: Mean and standard deviation of $O_{N,s}$ for ten different initial ensembles after assimilation of production data

Case	Mean	Standard deviation
EnKF (production data only)	12.643	3.404
EnKF (prod. and seis.)	4.104	0.997
EnKF-MDA (2×) (prod. and seis.)	2.043	0.679
EnKF-MDA (4×) (prod. and seis.)	1.343	0.430
EnKF-MDA (8×) (prod. and seis.)	1.187	0.416

flow. A description of the case can be found in [137]. In the original Brugge dataset, there are 104 realizations of rock properties (porosity, horizontal and vertical permeabilities and net-to-gross ratio), ten years of production history and a synthetic time-lapse seismic. The time-lapse seismic corresponds to pressure and oil saturation changes due to the 10 years of production. According to Peters et al. [137], from the nine research groups that participated in the original benchmark study, six used the seismic data in the history matching. However, no results or discussions about the seismic data matches are presented in [137] nor in other papers published by some of the groups [23, 114, 20].

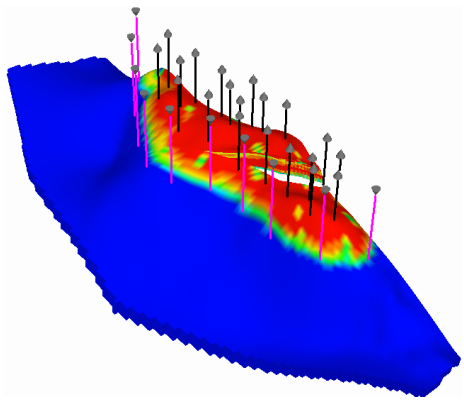


Figure 6.10: Brugge field. In blue the aquifer and in red the oil zone. Water injection wells in pink and oil producing in black.

According to Peters et al. [137], an unintentional bias was introduced in the pressure data because the seismic data were calculated based on an upscaled model.

For this reason, here, we consider only the oil saturation data. The seismic data were provided as vertically averaged data, corresponding to the four geological zones of the reservoir. The provided seismic data were corrupted with an unknown level of noise. Here, we estimated the noise level by smoothing the observed data with a window averaging for each of the reservoir zones and computing the residual between the observed and smoothed data. We tried different sizes of the averaging window and, for each case, we computed empirical variograms. The results indicated that the noise added to the seismic seems to be spatially uncorrelated in the horizontal plane with an average standard deviation of 0.025.

Here, we assimilated only the seismic data (ΔS_o) using standard EnKF and EnKF-MDA assimilating data two and four times (no production data are included). Fig. 6.11 presents a cross-plot between observed and the mean predicted ΔS_o obtained with the prior ensemble and the ensembles after data assimilation. This figure indicates a slight improvement in the predicted ΔS_o using two and four data assimilations. Table 6.5 presents the average values of $O_{N,s}$ for each case. Note that the values of the normalized objective function for the Brugge case are relatively small, even for the prior ensemble. This happens mainly because in the Brugge case, there is a large aquifer in which the oil saturation change is zero, and thus seismic data pertaining to the aquifer is automatically well matched. Nevertheless, the results in Table 6.5 show improvements on the order of 10% (EnKF-MDA 2×) and 15% (EnKF-MDA 4×) in the values of $O_{N,s}$ compared to the standard EnKF. Figs. 6.12–6.16 present the observed ΔS_o and the mean predicted ΔS_o obtained with the prior ensemble and the posterior ensembles obtained with EnKF assimilating data one, two and four times. In these figures, the oil saturation changes are presented for the four reservoir zones (Schelde, Waal, Maas and Schie). In these figures, the negative values (green to blue colors) represent a decrease in oil saturation due to the displacement of oil by injected water.

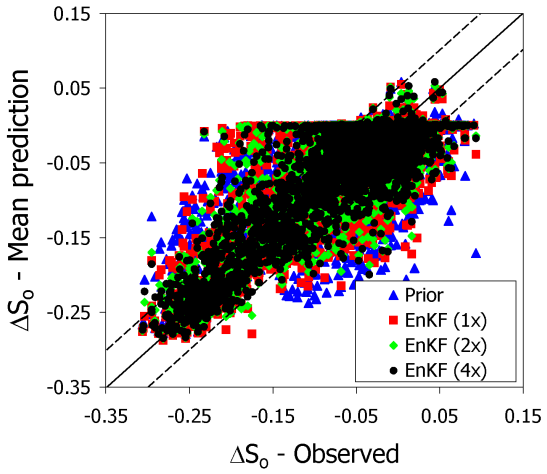


Figure 6.11: Cross-plot between observed and mean predicted oil saturation changes. The dashed lines correspond to +/− two standard deviations of the measurement errors.

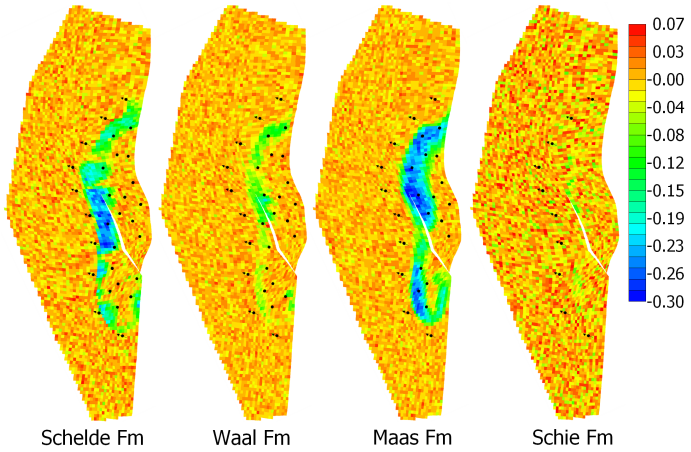


Figure 6.12: Observed seismic (ΔS_o).

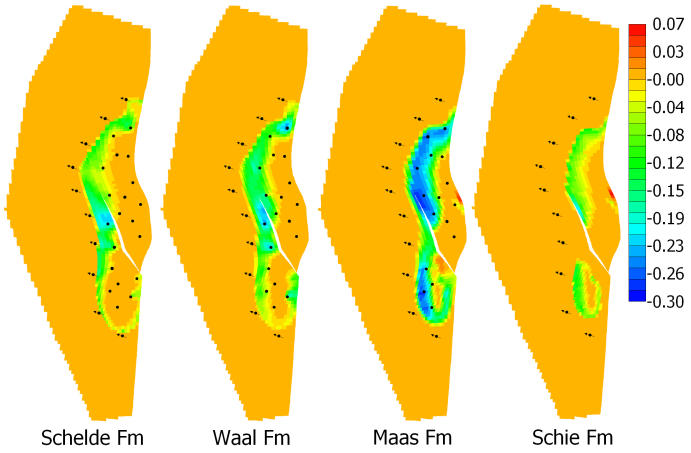


Figure 6.13: Average predicted seismic (ΔS_o) from the prior ensemble.

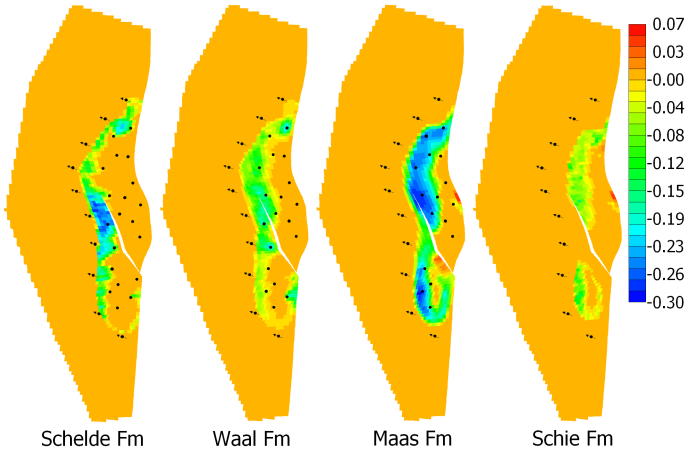


Figure 6.14: Average predicted seismic (ΔS_o) from the final ensemble after EnKF.

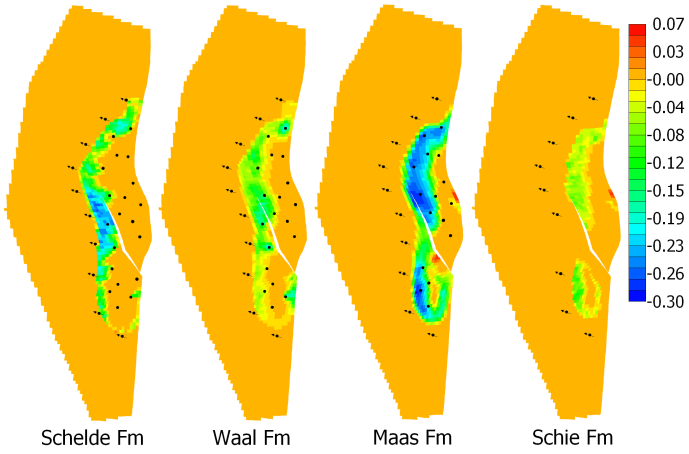


Figure 6.15: Average predicted seismic (ΔS_o) from the final ensemble after EnKF-MDA (2×).

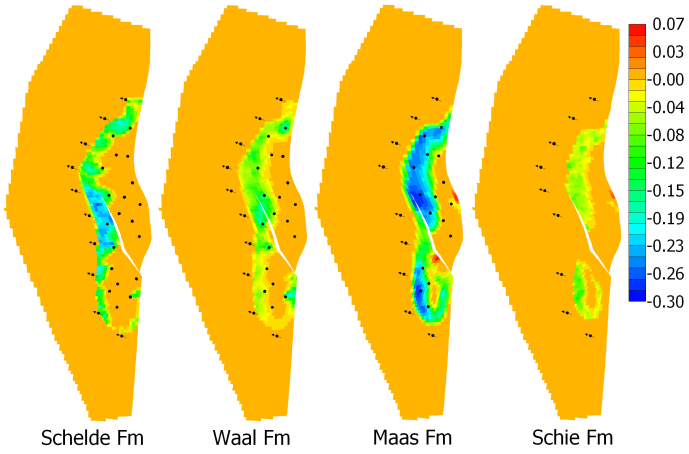


Figure 6.16: Average predicted seismic (ΔS_o) from the final ensemble after EnKF-MDA (4×).

Table 6.5: Mean and standard deviation of $O_{N,s}$ for the Brugge case

Case	Mean	Standard deviation
Prior	0.766	0.053
EnKF	0.606	0.005
EnKF-MDA (2×)	0.540	0.002
EnKF-MDA (4×)	0.510	0.001

6.5 Multiple Data Assimilation as an Iterative Ensemble Smoother

Another potential application of MDA is for the ensemble smoother (ES) [174, 173]. The ES was recently introduced for reservoir history matching by Skjervheim et al. [159]. In the ES, all data are assimilated at once, which means that only a single approximate GN iteration is done to history match all data. One advantage of ES over the standard EnKF is that in ES there is no need to restart the reservoir simulations after every data assimilation time-step. This makes ES much easier to implement and significantly faster than EnKF. Skjervheim et al. [159], for example, reported that ES required approximately 10% of the CPU time required by EnKF. Besides that, ES is an attractive option for data assimilation workflows which integrate different parts of the reservoir modeling process, including seismic, structural and geological modeling with flow simulation, such as the one presented by Zachariassen et al. [190]. These workflows typically require integrating different geomodeling softwares and may include upscaling of the rock properties, which makes the simulation restarts required by EnKF very inconvenient or impossible. However, because in ES all data are assimilated at once, it may result in poor data matches even compared to standard EnKF. In this section, we investigate the use of multiple data assimilation with ES. This procedure can be interpreted as a simple iterative form of ES, which is still reasonably computationally efficient.

6.5.1 Generalization of the MDA procedure

In Section 6.1, we proved for the linear-Gaussian case the equivalence between a single assimilation of data and multiple data assimilation with the covariance matrix of measurement errors, C_{D}, multiplied by the number of data assimilations. Here, we generalize this procedure to the case where C_{D} is increased by a different coefficient, α_i, each time we assimilate data, i.e., we define $\widehat{C}_{\mathrm{D}}$ as

$$\widehat{C}_{\mathrm{D}} \equiv \begin{bmatrix} \alpha_1 C_{\mathrm{D}} & 0 & \cdots & 0 \\ 0 & \alpha_2 C_{\mathrm{D}} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \alpha_{N_a} C_{\mathrm{D}} \end{bmatrix}. \tag{6.45}$$

Using Eq. 6.45 in Eq. 6.24 and computing the expectation of $\widehat{y}^a$, we obtain

$$\begin{aligned}
\mathrm{E}\left[\widehat{y}^a\right] &= C_{\mathrm{Y}}^a \left\{ \left(C_{\mathrm{Y}}^f\right)^{-1} \left(\mu_y^f + \left(C_{\mathrm{Y}}^f\right)^{1/2} \mathrm{E}[z_y]\right) + \widehat{H}^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-1} \left(\widehat{d}_{\mathrm{obs}} + \widehat{C}_{\mathrm{D}}^{1/2} \mathrm{E}[\widehat{z}_d]\right) \right\} \\
&= C_{\mathrm{Y}}^a \left\{ \left(C_{\mathrm{Y}}^f\right)^{-1} \mu_y^f + \widehat{H}^{\mathrm{T}} \widehat{C}_{\mathrm{D}}^{-1} \widehat{d}_{\mathrm{obs}} \right\} \\
&= C_{\mathrm{Y}}^a \left\{ \left(C_{\mathrm{Y}}^f\right)^{-1} \mu_y^f + \begin{bmatrix} H^{\mathrm{T}} & \cdots & H^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \frac{1}{\alpha_1} C_{\mathrm{D}}^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\alpha_{N_a}} C_{\mathrm{D}}^{-1} \end{bmatrix} \begin{bmatrix} d_{\mathrm{obs}} \\ \vdots \\ d_{\mathrm{obs}} \end{bmatrix} \right\} \\
&= C_{\mathrm{Y}}^a \left\{ \left(C_{\mathrm{Y}}^f\right)^{-1} \mu_y^f + \begin{bmatrix} H^{\mathrm{T}} & \cdots & H^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \frac{1}{\alpha_1} C_{\mathrm{D}}^{-1} d_{\mathrm{obs}} \\ \vdots \\ \frac{1}{\alpha_{N_a}} C_{\mathrm{D}}^{-1} d_{\mathrm{obs}} \end{bmatrix} \right\} \\
&= C_{\mathrm{Y}}^a \left\{ \left(C_{\mathrm{Y}}^f\right)^{-1} \mu_y^f + \left(\sum_{i=1}^{N_a} \frac{1}{\alpha_i}\right) H^{\mathrm{T}} C_{\mathrm{D}}^{-1} d_{\mathrm{obs}} \right\}.
\end{aligned} \tag{6.46}$$

The last expression of Eq. 6.46 is equal to the correct posterior mean if, and only if,

$$\sum_{i=1}^{N_a} \frac{1}{\alpha_i} = 1. \tag{6.47}$$

Note that the choice $\alpha_i = N_a$ for $i = 1, 2, \ldots, N_a$ satisfies the condition of Eq. 6.47, but there are infinitely many possible choices for the α_i's. By choosing a set of α_i's that satisfies Eq. 6.47, MDA yields the correct posterior mean for the linear-Gaussian case. i.e.,

$$\mathrm{E}\left[\widehat{y}^a\right] = \mu_y^a. \tag{6.48}$$

The equivalence for the posterior covariance follows from

$$\begin{aligned}
\mathrm{cov}\left[\widehat{y}^a\right] &= C_{\mathrm{Y}}^a \left\{ \left(C_{\mathrm{Y}}^f\right)^{-1} + \begin{bmatrix} H^{\mathrm{T}} & \cdots & H^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \frac{1}{\alpha_1} C_{\mathrm{D}}^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\alpha_{N_a}} C_{\mathrm{D}}^{-1} \end{bmatrix} \begin{bmatrix} H \\ \vdots \\ H \end{bmatrix} \right\} C_{\mathrm{Y}}^a \\
&= C_{\mathrm{Y}}^a \left\{ \left(C_{\mathrm{Y}}^f\right)^{-1} + \left(\sum_{i=1}^{N_a} \frac{1}{\alpha_i} \right) H^{\mathrm{T}} C_{\mathrm{D}}^{-1} H \right\} C_{\mathrm{Y}}^a \\
&= C_{\mathrm{Y}}^a.
\end{aligned} \tag{6.49}$$

6.5.2 ES-MDA procedure

Algorithm 6.2 presents the ES-MDA general procedure. One difficultly with the proposed ES-MDA method is that N_a and the coefficients α_i's must be selected prior to the data assimilation. The simplest choice for α is $\alpha_i = N_a$ for all i. However, intuitively, we expect that choosing α_i in a decreasing order can improve the performance of the method. In this case, we start assimilating data with a large value of α, which corresponds to reducing the magnitude of the initial updates, then, we gradually decrease α.

Algorithm 6.2 ES-MDA

1. Choose the number of data assimilations, N_a, and the coefficients α_i for $i = 1, 2, \ldots, N_a$.
2. For $i = 1$ to N_a:
   (a) Run the ensemble from time zero.
   (b) For each ensemble member, perturb the observation vector using $d_{\text{uc}} = d_{\text{obs}} + \sqrt{\alpha_i} C_{\text{D}}^{1/2} z_d$, where $z_d \sim \mathcal{N}(0, I_{N_d})$.
   (c) Update the ensemble using the ES analysis equation (Eq. 2.129) with C_{D} replaced by $\alpha_i C_{\text{D}}$.

   end (for).

6.5.3 Example 1 – Waterflooding

The first example corresponds to the same reservoir model used in Section 6.4.1. For this case, we performed data assimilations using EnKF, ES and ES-MDA to assimilate the production data. For ES-MDA, we considered the following cases:

- 2×: ES assimilating data twice ($\alpha_1 = \alpha_2 = 2.0$).
- 4×-a: ES assimilating data four times ($\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 4.0$).
- 4×-b: ES assimilating data four times ($\alpha_1 = 9.333$, $\alpha_2 = 7.0$, $\alpha_3 = 4.0$ and $\alpha_4 = 2.0$).

Fig. 6.17 presents the box plot of the normalized objective function (O_N) obtained from each case. These box plots were computed based on combining the results of the ten different initial ensembles. The numbers next to each box plot correspond to the value of the median.

According to the results in Fig. 6.17, ES gave relatively high values of O_N compared to EnKF. This fact supports our conjecture that EnKF performs reasonably well when history matching production data because the sequential assimilation

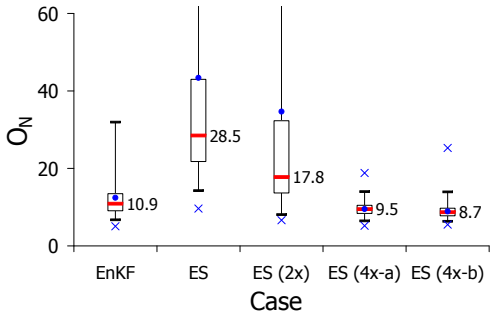
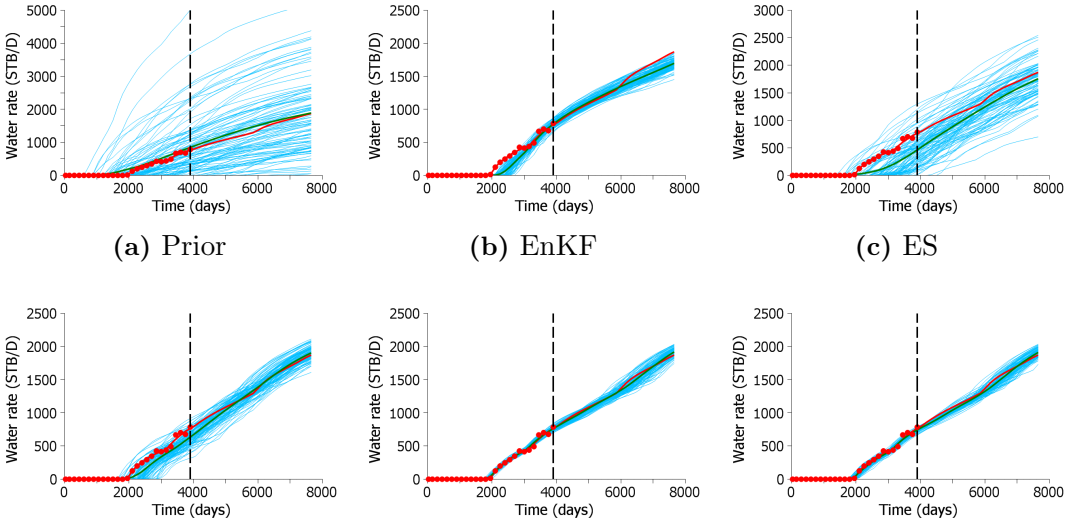


Figure 6.17: Box plots of the normalized objective function. The numbers next to each box plot correspond to the value of the median of O_N.

approximately corresponds to accumulating several GN corrections. Increasing the number of data assimilations improved the data matches and, consequently, reduced the values of O_N obtained by ES. After four data assimilations, the values of O_N are slightly lower than the ones obtained by EnKF. Selecting the coefficients α_i's in a decreasing order resulted in further reduction in the values of O_N although the difference is not large. We tried other combinations of the coefficients α_i's, but no significant differences were obtained.

Fig. 6.18 presents the field water production rate for the first of the ten ensembles. Besides the historical period (3,900 days), we also include 3,750 days of forecast. For comparison, we also show the prediction obtained by the true model in each plot. The results presented in Fig. 6.18 were obtained by running the final ensembles after data assimilation from time zero. Fig. 6.18 shows that ES failed in matching the water breakthrough time for most of the ensemble members. Two data assimilations were not enough to obtain good data matches of the water production rate. Four data assimilations, on the other hand, resulted in very good water rate data matches, which are better than the ones obtained by EnKF. Although we showed the water production curves only for the first of the ten ensembles, these results are representative of what is observed in the other nine ensembles.

We estimated the average computational cost of each procedure in terms of the number of equivalent simulation runs, and the results are presented in Table 6.6. The



(a) Prior **(b)** EnKF **(c)** ES

(d) ES-MDA (2×) **(e)** ES-MDA (4×-a) **(f)** ES-MDA (4×-b)

Figure 6.18: Field water production rate. The vertical dashed line indicates the end of the historical period. Red dots are the history; red curve is the prediction from the true model; green curve is the ensemble mean prediction and blue curves are the ensemble members.

CPU time required by the data assimilation with EnKF is 4.58 times larger than the CPU time used to run an initial ensemble from time zero without data assimilation. This difference occurs mainly because of the additional CPU time required by EnKF to restart reservoir simulations. The CPU time required by ES is practically the same as the CPU time used to run an ensemble without data assimilation. MDA increases the CPU time of ES by the factor equal to the number of data assimilations. For this example, the CPU time of ES-MDA (4×) is slightly lower than the the CPU time used by EnKF. For this example, we used observations at 150 days intervals. Note that increasing the frequency of data would increase the computational cost of EnKF (because we would have more simulation restarts), but would have a small impact on the computational cost of ES-MDA.

6.5.4 Example 2 – Production-logging data

The second example is a single-phase synthetic reservoir model on a 3D uniform grid with $11 \times 11 \times 40$ gridblocks. The dimensions of the gridblocks are 200 ft

Table 6.6: Average computational costs.

Case	Number of equivalent simulation runs
Run prior ensemble	100
EnKF	458
ES	101
ES-MDA (2×)	203
ES-MDA (4×)	405

× 200 ft × 25 ft. In this model, there is a single producing well at the center of the reservoir completed in all 40 layers operating at a constant rate of 1,000 stb/day. The observed data correspond to the oil rate from individual layers at the specific time of 30 days, mimicking the data obtained from a production-logging acquisition. The noise level added to each datum corresponds to 2% of the layer-rate predicted by the true model. We chose a small noise level to make the problem more challenging for data assimilation. The model parameters are the log-permeabilities for each of the 40 layers. Each layer of the reservoir has homogeneous and isotropic permeabilities. The log-permeability of each layer of the true model and the initial ensemble was obtained by sampling $\mathcal{N}(5.0, 1.0)$. Hence, there is no correlation in the permeability between layers. This problem was designed to test our conjecture that sequential assimilation of data with overlapping information content is one of the reasons for the good performance of EnKF when history matching production data. For the problem considered in this section, the production-logging data are available at only one time. In this situation, EnKF becomes equivalent to ES.

We assimilated the production-logging data using ES and ES-MDA with two and four data assimilations. For ES-MDA (2×), we used $\alpha_1 = \alpha_2 = 2.0$. For ES-MDA (4×), we used $\alpha_1 = 9.333$, $\alpha_2 = 7.0$, $\alpha_3 = 4.0$ and $\alpha_4 = 2.0$. The ensemble size is 100. Fig. 6.19 presents the box plot of the normalized data mismatch objective function ($O_{N,d}$) obtained for each case. According to the results in Fig. 6.19, ES with

single data assimilation resulted in an unreasonably high value of $O_{N,d}$. ES-MDA significantly improved the data matches. After four data assimilations, $O_{N,d}$ was reduced to 6.7, which is 219 times smaller than the value obtained with the standard ES. Fig. 6.20 presents the predicted layer-rates for each case. Fig. 6.20b suggests that ES resulted in overcorrection when matching the data from layers 5 and 17. ES-MDA (4×), on the other hand, resulted in excellent data matches. Fig. 6.21 presents the corresponding values of permeability for each reservoir layer showing that ES-MDA (4×) captured the correct permeability field.

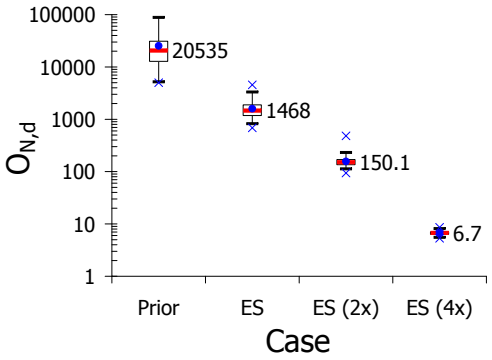
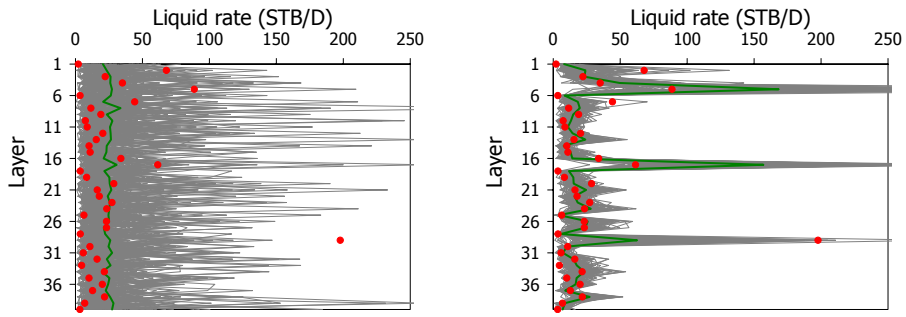
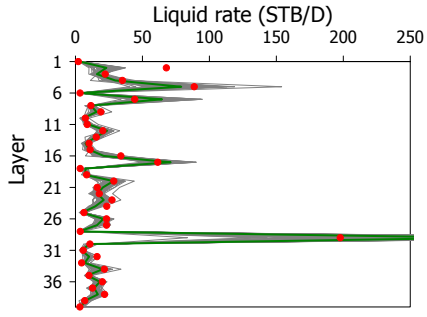


Figure 6.19: Box plots of the normalized data mismatch objective function. The numbers next to each box plot corresponds to the median value of $O_{N,d}$.

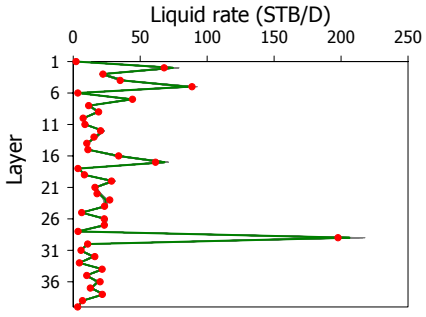


(a) Prior

(b) ES

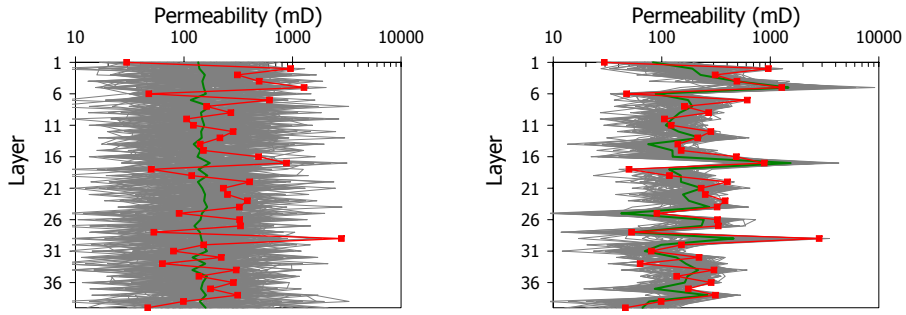


(c) ES-MDA (2×)



(d) ES-MDA (4×)

Figure 6.20: Predicted liquid rate for each reservoir layer. Red dots are the history; green curve is the ensemble mean prediction and gray curves are the prediction from the ensemble members.



(a) Prior **(b)** ES

(c) ES-MDA (2×) **(d)** ES-MDA (4×)

Figure 6.21: Layer-permeabilities. Red curve with squares is the truth; green curve is the ensemble mean and gray curves are the ensemble members.

6.5.5 Example 3 – Brugge case

Here, we consider the history matching of production data for the ten year period provided in the original Brugge case dataset. The production data correspond to "measurements" of the oil and water rates at the producing wells and the bottomhole pressure at producing and water injection wells. We assume that the noise level corresponds to 3% for oil rate data and 5% for water rate data. For bottomhole pressure, we assume a constant measurement error of 0.5 bars (7.25 psi). Even though data are available with a frequency of 30 days, we assimilated data every 120 days using EnKF, ES and ES-MDA (4×). For this last case, the coefficients α_i's are the same as in the case ES-MDA (4×-b) of example 1. Because the number of wells in the Brugge field is large (20 producers and 10 water injectors), we used localization to reduce problems related to sampling errors and limited degrees of freedom. We defined the localization regions using the procedure described in Chapter 3.

Fig. 6.22 presents the box plot of the normalized data mismatch objective function ($O_{N,d}$) obtained for each case. According to the results presented in Fig. 6.22, ES resulted in values of $O_{N,d}$ about four times larger than those based on results from EnKF. ES-MDA (4×) significantly improved the final data matches, and the values of $O_{N,d}$ are about three times lower than EnKF. Figs. 6.23 and 6.24 present the data matches obtained for two producing wells (P5 and P14). These figures illustrate that ES was not able to achieve reasonable data matches. Also, the predictions from the ES-MDA (4×) are in better agreement with the historical data than the ones obtained by EnKF; see, for example, Figs. 6.23a and 6.23b, which show that EnKF results in slightly biased predictions compared to the historical data.

Table 6.7 presents the computational cost in terms of the equivalent number of reservoir simulation runs for EnKF and ES-MDA (4×). The results in this table were obtained by measuring the actual CPU time required during the data assimilations divided by the average CPU time required for one reservoir simulation run. According

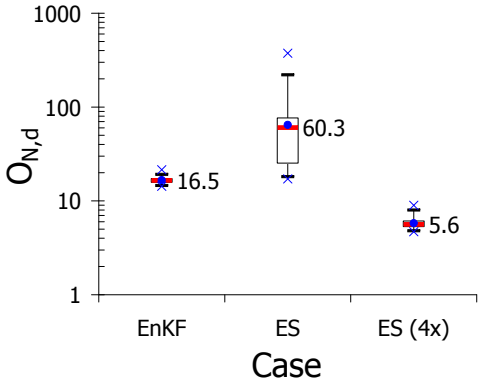


Figure 6.22: Box plots of the normalized data mismatch objective function. The numbers next to each box plot correspond to the value of the median of $O_{N,d}$.

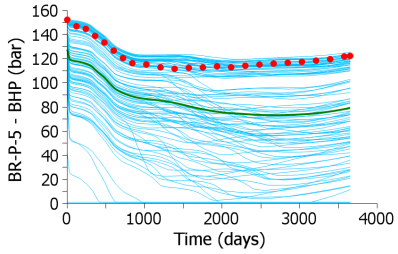
Table 6.7: Computational costs.

Case	Number of equivalent simulation runs
Run prior ensemble	104
EnKF	367
ES-MDA (4×)	430

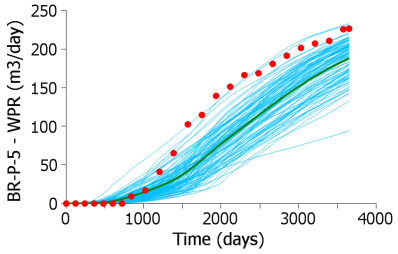
to the results in Table 6.7, the CPU time of ES-MDA (4×) was 17% higher than the CPU time of EnKF, but resulted in far better data matches than were obtained with EnKF.

6.5.6 Comments

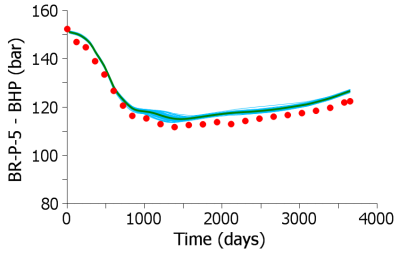
The proposed ES-MDA method was motivated by our conjecture that when sequential data have similar information content, the sequential assimilation of data employed by EnKF provides an approximation of multiple Gauss-Newton iterations. This conjecture yields a second conjecture, namely, EnKF will typically give a better data match than ES because the smoother effectively represents a single Gauss-Newton iteration with an average sensitivity matrix [143]. Although we have not provided a proof of these conjectures, all results we have presented here are consistent with these conjectures.



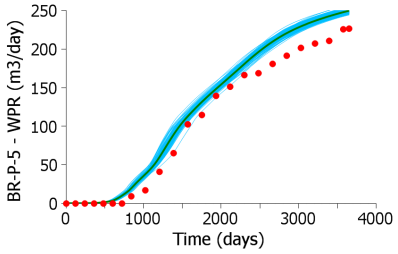
(a) Prior (bottomhole pressure)



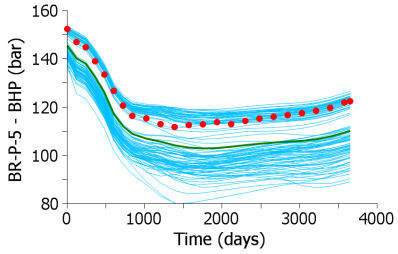
(b) Prior (water rate)



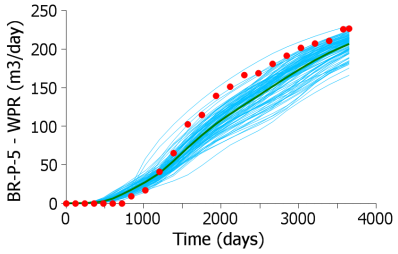
(c) EnKF (bottomhole pressure)



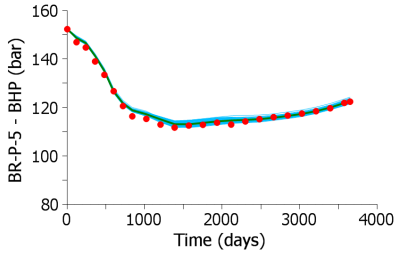
(d) EnKF (water rate)



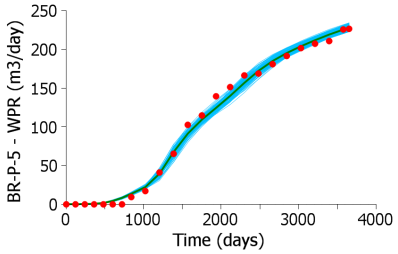
(e) ES (bottomhole pressure)



(f) ES (water rate)

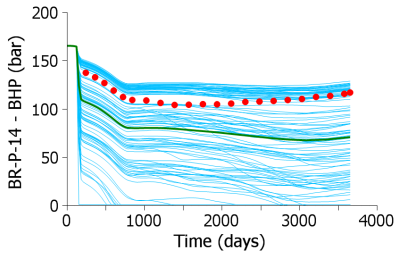


(g) ES-MDA (4×) (bottomhole pressure)

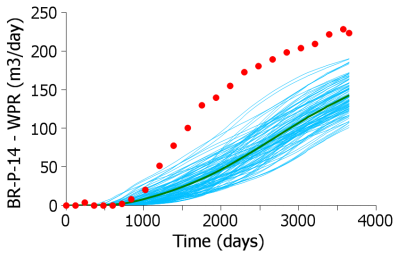


(h) ES-MDA (4×) (water rate)

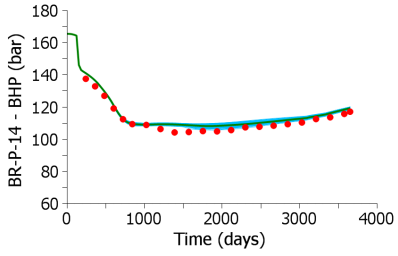
Figure 6.23: Data matches for well P5. Red dots are the history; green curve is the ensemble mean prediction and blue curves are the ensemble members.



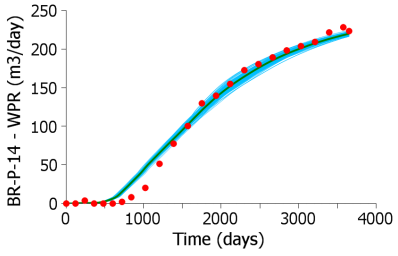
(a) Prior (bottomhole pressure)



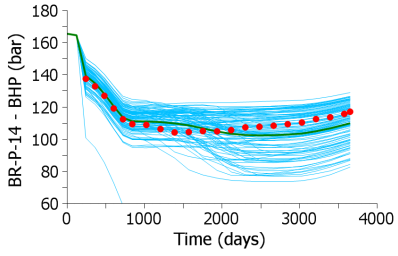
(b) Prior (water rate)



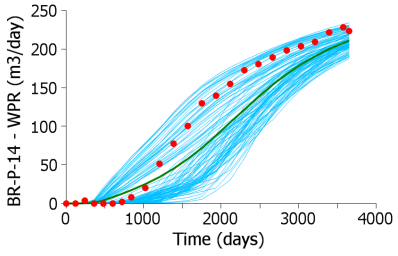
(c) EnKF (bottomhole pressure)



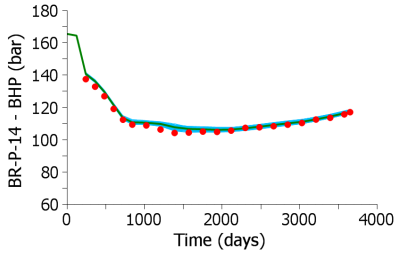
(d) EnKF (water rate)



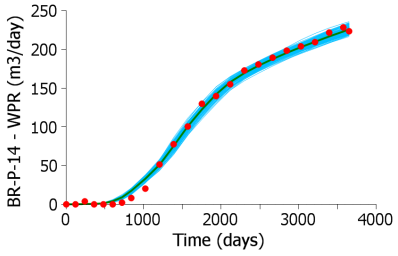
(e) ES (bottomhole pressure)



(f) ES (water rate)



(g) ES-MDA (4×) (bottomhole pressure)



(h) ES-MDA (4×) (water rate)

Figure 6.24: Data matches for well P14. Red dots are the history; green curve is the ensemble mean prediction and blue curves are the ensemble members.

For ES-MDA, we only consider the parameter-estimation problem. Thus, unlike EnKF, the parameters and states are always consistent [167]. This fact helps to explain the better data matches obtained by ES-MDA compared to EnKF.

CHAPTER 7

HISTORY MATCHING OF PRODUCTION AND SEISMIC DATA FOR A REAL FIELD CASE

In this chapter, we present the history matching results of a real field case using EnKF and ES-MDA. The field case corresponds to the upper zone (turbidite system 2) of the same field considered in Chapter 3 (Section 3.3). However, we now consider the assimilation of 3D and 4D seismic data in conjunction with production data. In Chapter 3, we adjusted only the permeability field while history matching the production data. Here, we also include the porosity and net-to-gross ratio (NTG) as model parameters. The focus on Chapter 3 was to compare EnKF with and without localization. Here, we compare the performance of EnKF and ES-MDA. In both cases, we use localization when assimilating the production and the seismic data. In addition, in this chapter, we present a description of the construction of the petroelastic model required to predict seismic data from the results of the reservoir simulation. The results presented in this chapter are summarized in Emerick and Reynolds [40].

7.1 Petroelastic Model

The petroleastic model (PEM) is a set of relationships used to predict the seismic response of a reservoir model. The PEM combines the properties of the rock and fluid in the reservoir to predict elastic properties of the combined media such as pressure-wave impedance (P-impedance or acoustic impedance) and Poisson's ratio. The most widely used model to predict the seismic response of a porous media saturated with fluids is the Gassmann model [62]. In the Gassmann model, the bulk

(K) and shear (G) moduli of the saturated porous media are computed using

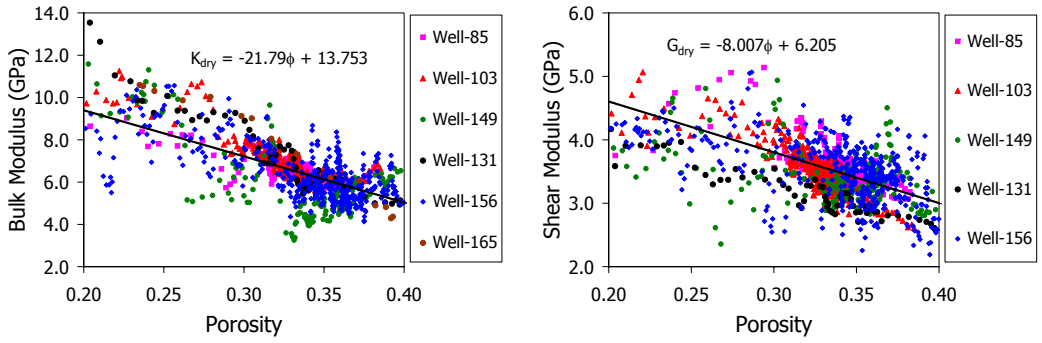
$$K = K_{\text{dry}} + \frac{\left(1 - \frac{K_{\text{dry}}}{K_{\text{min}}}\right)^2}{\frac{\phi_{\text{eff}}}{K_{\text{fluid}}} + \frac{1-\phi_{\text{eff}}}{K_{\text{min}}} - \frac{K_{\text{dry}}}{K_{\text{min}}^2}} \tag{7.1}$$

and

$$G = G_{\text{dry}}. \tag{7.2}$$

In the above equations, K_{dry} is the bulk modulus of the dry rock, i.e., rock frame without fluid in the porous space; K_{min} is the bulk modulus of the mineral that constitutes the rock; K_{fluid} is the bulk modulus of the fluid, i.e., the mixture of water and hydrocarbons; ϕ_{eff} is the effective porosity, i.e., the porosity at the reservoir pressure condition; and G_{dry} is the shear modulus of the dry rock. Eqs. 7.1 and 7.2 assume an isotropic and homogeneous media, fully saturated with fluid. Because we apply the Gassmann model on a gridblock by gridblock basis when computing seismic data from a reservoir model, we are effectively neglecting the anisotropy at the gridblock scale. However, we allow the possibility of anisotropy in the reservoir scale as the rock and fluid properties can vary from gridblock to gridblock [52]. Moreover, the Gassmann model assumes that the seismic frequencies are sufficiently low such that wave-induced pressure gradients are negligible [121] and that the fluid is of sufficiently low viscosity so that it is possible to neglect its effect in the shear modulus.

The dry bulk and shear moduli of the rock are estimated using empirical correlations or based on data from the reservoir (typically, rock cores and sonic logs). Mavko et al. [121] present several empirical correlations between wave-velocities, porosity and clay content that can be used to estimate the values of K_{dry} and G_{dry}. Here, we estimate K_{dry} and G_{dry} using relationships obtained by linear regression with interpreted data from sonic logs obtained in six wells in the field as shown in Fig. 7.1.



(a) K_{dry} **(b)** G_{dry}

Figure 7.1: Relationship between dry bulk and shear moduli with effective porosity.

The bulk and shear moduli of the mineral that constitutes the reservoir rock were estimated using the Hashin-Shtrikaman bounds model [72]. In this model, we compute the upper and lower bounds of K_{min} and G_{min} for a mixture of two minerals as a function of their volume fractions. For the field application considered in this chapter, we computed K_{min} and G_{min} as the mixture of sandstone (quartz) and shale (clay). In this case, the upper and lower bounds are given by

$$K_{\text{min}}^{\text{up}} = K_{\text{qtz}} + \frac{f_{\text{clay}}}{(K_{\text{clay}} - K_{\text{qtz}})^{-1} + f_{\text{qtz}}\left(K_{\text{qtz}} + \frac{4}{3}G_{\text{qtz}}\right)^{-1}}, \tag{7.3}$$

$$K_{\text{min}}^{\text{low}} = K_{\text{clay}} + \frac{f_{\text{qtz}}}{(K_{\text{qtz}} - K_{\text{clay}})^{-1} + f_{\text{clay}}\left(K_{\text{clay}} + \frac{4}{3}G_{\text{clay}}\right)^{-1}}, \tag{7.4}$$

$$G_{\text{min}}^{\text{up}} = G_{\text{qtz}} + \frac{f_{\text{clay}}}{(G_{\text{clay}} - G_{\text{qtz}})^{-1} + \frac{2f_{\text{qtz}}(K_{\text{qtz}}+2G_{\text{qtz}})}{5G_{\text{qtz}}\left(K_{\text{qtz}}+\frac{4}{3}G_{\text{qtz}}\right)}} \tag{7.5}$$

and

$$G_{\text{min}}^{\text{low}} = G_{\text{clay}} + \frac{f_{\text{qtz}}}{(G_{\text{qtz}} - G_{\text{clay}})^{-1} + \frac{2f_{\text{clay}}\left(K_{\text{clay}}+2G_{\text{clay}}\right)}{5G_{\text{clay}}\left(K_{\text{clay}}+\frac{4}{3}G_{\text{clay}}\right)}}. \tag{7.6}$$

In the above equations, $K_{\text{qtz}} = 37.5$ GPa is the bulk modulus of the quartz; $K_{\text{clay}} = 22.78$ GPa is the bulk modulus of the clay; $G_{\text{qtz}} = 45.5$ GPa is the shear modulus of

the quartz; and $G_{\text{clay}} = 12.62$ GPa is the shear modulus of the clay. f_{qtz} and f_{clay} are the volume fractions of quartz and clay, such that $f_{\text{qtz}} + f_{\text{clay}} = 1$. In this field application, for each gridblock of the model, we use f_{qtz} equal to the NTG and f_{clay} equal to $1 - \text{NTG}$. After computing the upper and lower bounds using Eqs. 7.3–7.6, we compute the bulk and shear moduli of the mineral as

$$K_{\text{min}} = \frac{K_{\text{min}}^{\text{up}} + K_{\text{min}}^{\text{low}}}{2} \tag{7.7}$$

and

$$G_{\text{min}} = \frac{G_{\text{min}}^{\text{up}} + G_{\text{min}}^{\text{low}}}{2}. \tag{7.8}$$

The bulk modulus of the fluid, K_{fluid}, is computed using Wood's formula [121], i.e.,

$$\frac{1}{K_{\text{fluid}}} = \frac{S_{\text{o}}}{K_{\text{o}}} + \frac{S_{\text{w}}}{K_{\text{w}}} + \frac{S_{\text{g}}}{K_{\text{g}}}, \tag{7.9}$$

where S_{o}, S_{w} and S_{g} are the saturations of the phases oil, water and gas, respectively. K_{o}, K_{w} and K_{g} are the bulk moduli of each phase, which are computed using the set of empirical correlations proposed by Batzle and Wang [11]. According to Batzle and Wang [11], the bulk modulus of the oil can be estimated using

$$K_{\text{o}} = \rho_{\text{o}} V_{\text{o}}^2, \tag{7.10}$$

and

$$V_{\text{o}} = 2{,}096\sqrt{\frac{\rho_{\text{o}}}{2{,}600 - \rho_{\text{o}}}} - (3.7T) + (4.67p) + 0.0115\left[\left(4.12\sqrt{\frac{1{,}080}{\rho_{\text{o}}} - 1}\right) - 1\right]Tp. \tag{7.11}$$

In the above equations, ρ_{o} is the oil density at reservoir conditions in kg/m^3; V_{o}^2 is the pressure-wave velocity in the oil in m/s; p is the reservoir pressure in MPa and T is the reservoir temperature in $^{\text{o}}$C. Using these units, Eq. 7.10 results in K_{o} given

in Pa.

The bulk modulus of the water is given by

$$K_{\mathrm{w}} = \rho_{\mathrm{w}} V_{\mathrm{w}}^2, \tag{7.12}$$

and

$$V_{\mathrm{w}} = V_{\mathrm{wp}} + C_S S + \left(780 - 10p + 0.16p^2\right) S^{1.5} - 1,820 S^2, \tag{7.13}$$

where ρ_{w} is the water density at reservoir conditions in kg/m^3; V_{w}^2 is the pressure-wave velocity in the water in m/s; p is the reservoir pressure in MPa; S is the water salinity, which must be given as a fraction of unity; C_S is a coefficient computed using

$$C_S = 1{,}170 - 9.6T + 0.055T^2 - 8.5 \times 10^{-5} T^3 + 2.6p - 0.0029Tp - 0.0476p^2, \tag{7.14}$$

where T is the reservoir temperature in $^{\mathrm{o}}$C. V_{wp} is the pressure-wave velocity in the pure-water, which is estimated using

$$V_{\mathrm{wp}} = \sum_{i=0}^{4} \sum_{j=0}^{3} w_{ij} T^i p^j. \tag{7.15}$$

The coefficients w_{ij}'s are given in Table 7.1.

The bulk modulus of the gas is given by

$$K_{\mathrm{g}} = \frac{p\gamma}{1 - \left(\frac{p_r}{Z}\frac{\partial Z}{\partial p_r}\right)_T}, \tag{7.16}$$

where

$$p_r = \frac{p}{4.892 - (0.4048SG)}, \tag{7.17}$$

Table 7.1: Coefficients w_{ij} of Eq. 7.15 [11].

i \ j	0	1	2	3
0	1.403×10^{3}	1.524	3.437×10^{-3}	-1.197×10^{-5}
1	4.871	-1.110×10^{-2}	1.739×10^{-4}	-1.628×10^{-6}
2	-4.783×10^{-2}	2.747×10^{-4}	-2.135×10^{-6}	1.237×10^{-8}
3	1.487×10^{-4}	-6.503×10^{-7}	-1.455×10^{-8}	$1,327 \times 10^{-10}$
4	-2.197×10^{-7}	7.987×10^{-10}	5.230×10^{-11}	-4.614×10^{-13}

$$Z = a \cdot p_r + b + c \cdot d, \tag{7.18}$$

$$\frac{\partial Z}{\partial p_r} = c \cdot d \cdot m + a, \tag{7.19}$$

$$a = 0.03 + 0.00527 \left(3.5 - T_r\right)^3, \tag{7.20}$$

$$b = 0.642 T_r - 0.007 T_r^4 - 0.52, \tag{7.21}$$

$$c = 0.109 \left(3.85 - T_r\right)^2, \tag{7.22}$$

$$d = \exp\left\{-\left[0.45 + 8\left(0.56 - \frac{1}{T_r}\right)^2\right]\frac{p_r^{1.2}}{T_r}\right\}, \tag{7.23}$$

$$m = -1.2\left[0.45 + 8\left(0.56 - \frac{1}{T_r}\right)^2\right]\frac{p_r^{0.2}}{T_r}, \tag{7.24}$$

$$\gamma = 0.85 + \frac{5.6}{p_r + 2} + \frac{27.1}{\left(p_r + 3.5\right)^2} - 8.7 \exp\left\{-0.65\left(p_r + 1\right)\right\}, \tag{7.25}$$

$$T_r = \frac{T + 273.15}{94.72 + (170.75SG)} \tag{7.26}$$

and SG is the specific gravity of the gas at standard conditions. In the above equations, p is given in MPa and T is given in °C.

After computing the bulk and shear moduli with Gassmann model (Eqs. 7.1 and 7.2), we compute the pressure-wave velocity (V_P) using

$$V_P = \sqrt{\frac{K + \frac{4}{3}G}{\rho}}. \tag{7.27}$$

Eq. 7.27 assumes a compatible unit system. We use SI units, which results in V_P given in m/s. In this equation, ρ is density of the saturated rock, which is estimated using

$$\rho = \rho_{\text{fluid}}\phi_{\text{eff}} + \rho_{\text{min}}(1 - \phi_{\text{eff}}), \tag{7.28}$$

where ρ_{fluid} is the fluid density at the pressure and temperature conditions of the reservoir, which is computed using

$$\rho_{\text{fluid}} = S_{\text{o}}\rho_{\text{o}} + S_{\text{w}}\rho_{\text{w}} + S_{\text{g}}\rho_{\text{g}}, \tag{7.29}$$

In Eq. 7.28, ρ_{min} is the density of the mineral that constitutes the reservoir rock. Here, we estimate ρ_{min} based on the NTG and the densities of the quartz ($\rho_{\text{qtz}} = 2{,}650\ \text{kg/m}^3$) and clay ($\rho_{\text{clay}} = 2{,}600\ \text{kg/m}^3$), using

$$\rho_{\text{min}} = \rho_{\text{qtz}}\text{NTG} + \rho_{\text{clay}}\left(1 - \text{NTG}\right). \tag{7.30}$$

The effective porosity, ϕ_{eff}, is computed using

$$\phi_{\text{eff}} = \phi_0\left[1 + c_{\text{rock}}\left(p - p_0\right)\right], \tag{7.31}$$

where ϕ_0 is the gridblock porosity specified at a reference pressure p_0 and c_{rock} is the rock compressibility.

The seismic data used in this field application correspond P-impedance (I_P), which is given by

$$I_P = \rho V_P. \tag{7.32}$$

7.2 Initial Ensemble

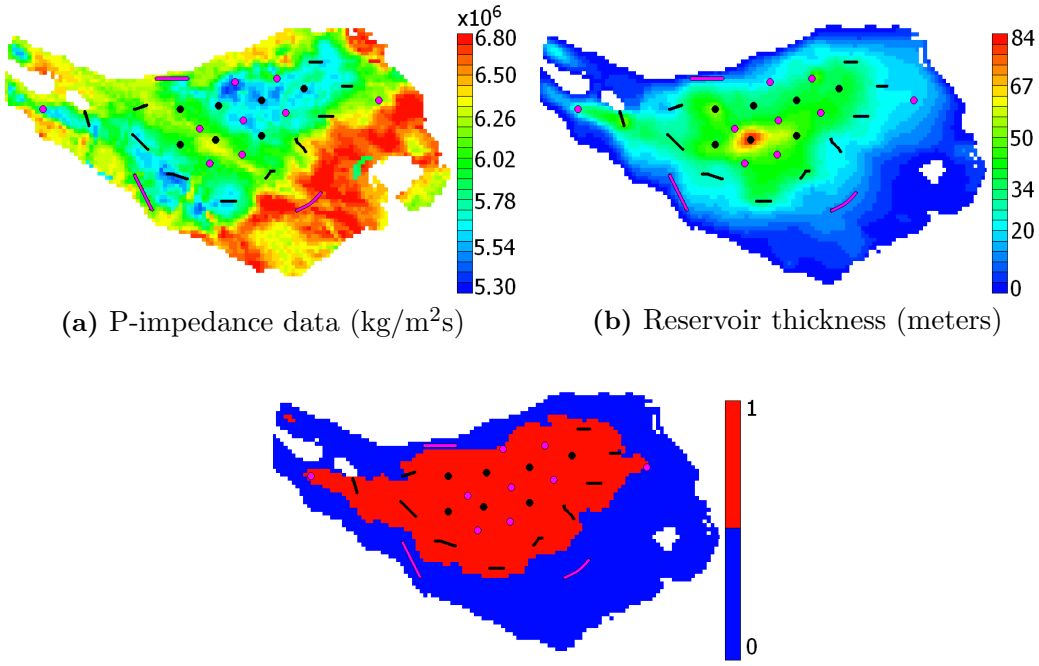
In this history matching study, we consider as model parameters the porosity, NTG and horizontal log-permeability for all reservoir gridblocks. The ensemble size is 200 and the initial ensemble of the gridblock log-permeability field is the same used in the history matching exercise presented in Chapter 3. Recall that this initial ensemble of log-permeabilities was obtained using sequential Gaussian simulation conditioned to interpreted pressure transient data. The initial ensemble of gridblock porosities was also generated with sequential Gaussian simulation. In this case, we used porosity data obtained from sonic logs at seven wells. The initial ensemble of NTG was generated with sequential Gaussian co-simulation [176] with porosity conditioned to NTG data at the same seven wells where we have porosity data. We used the same covariance function for all gridblock properties, which corresponds to an anisotropic exponential correlation function with practical ranges of 2,000 meters (corresponding to the size of 20 gridblocks) in the x-direction and 1,000 meters (10 gridblocks) in the y-direction. In the vertical direction, we assumed that rock properties are correlated over the four reservoir layers by choosing a practical range corresponding to the thickness of four gridblocks. The initial ensemble of log-permeability is not correlated with the initial ensembles of porosity and NTG because we did not find any meaningful correlation in the prior data (well logs and interpreted pressure transient data).

Recall from Chapter 3 that the current model of this reservoir was obtained

by "manually" modifying the permeability field to match the observed history of the water production rate. During this manual-history-matching process, the reservoir simulator was run with each producing well constrained to its historical oil rate and each injector well constrained to its historical water injection rate. Here, we also constrain our simulations to the historical oil and water injection rates. Similarly to Chapter 3, here, we also use the manually history-matched model as reference for comparison.

7.3 Assimilation of 3D Seismic Data

We first assimilated the seismic data from the 3D survey acquired in 1997, before the beginning of production in this reservoir. The seismic data were provided by Petrobras on the same grid used for the reservoir simulation model and inverted to P-impedance. However, because of the low-frequency obtained in the base survey acquisition, the seismic data have low vertical resolution [89]. For this reason, we used for the purpose of data assimilation the average P-impedance over the four reservoir simulation layers. The average P-impedance data were computed weighted by the thickness of the reservoir gridblocks. Fig. 7.2a presents the P-impedance data after the vertical average, while Fig. 7.2b presents a "map" of total thickness of the reservoir. Figs. 7.2a and 7.2b show that regions with low thickness in the reservoir (less than 20 meters) are associated with high values of P-impedance. These high values of P-impedance seem to occur because of the low vertical resolution of the seismic. In fact, for small reservoir thicknesses, the P-impedance data are likely to be corrupted by the rocks above and below the reservoir, making these data unreliable. Therefore, we removed the P-impedance data from regions of the reservoir with a total thickness less than 20 meters in the data assimilation. The red regions in Fig. 7.2c denote the locations corresponding to seismic data that are used for data assimilation. In this figure, P-impedance data pertaining to regions colored in red correspond to reservoir total thickness higher than 20 meters. On the other



(a) P-impedance data (kg/m^{2}s)

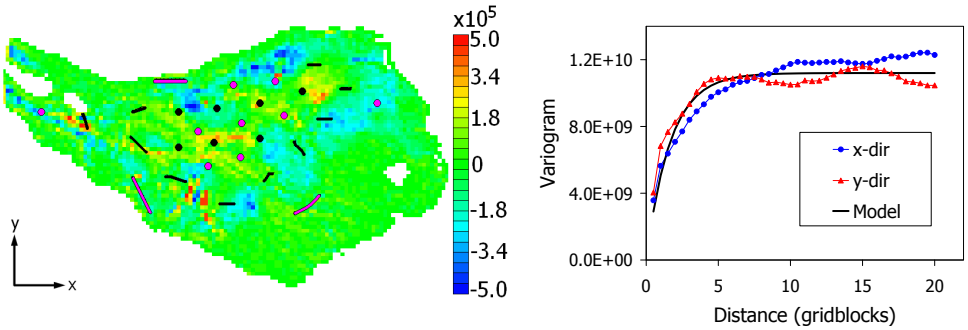
(b) Reservoir thickness (meters)

(c) Active seismic data

Figure 7.2: Observed 3D seismic data, reservoir thickness and active seismic data. In (c), the region in red corresponds to the region with P-impedance data used for data assimilation. The producing wells are shown in black and the water injector wells in pink. Dots represent vertical wells and lines represent horizontal wells.

hand, P-impedance data pertaining to regions colored in blue are not used for data assimilation. We estimated the noise level by smoothing the observed data with window averaging and computing the residual between observed and smoothed data. These residuals represent our estimate of the errors in the measurements. We used a averaging window with size of 3×3 gridblocks. We computed empirical variograms of the residuals and the results indicated that residuals were approximately correlated over a distance corresponding to the width of five gridblocks (500 meters) with a standard deviation on the order of 10^5 kg/m^{2}s. Fig. 7.3 presents the residuals and the corresponding experimental variograms.

We assimilate the 3D seismic data using EnKF and ES-MDA with four data assimilations. For ES-MDA, we used the following multiplication coefficients: $\alpha_1 = 9.333$, $\alpha_2 = 7$, $\alpha_3 = 4$ and $\alpha_4 = 2$, which are the same used in the examples



(a) Residuals between observed and smoothed P-impedance data (kg/m^2s)

(b) Experimental and modeled variograms of the residuals

Figure 7.3: Residuals between observed and smoothed P-impedance data used to estimate the covariance of the measurement errors (3D seismic). In (b), the variogram model is exponential with practical range of 5 gridblocks and sill of $1.1 \times 10^{10}(kg/m^2s)^2$.

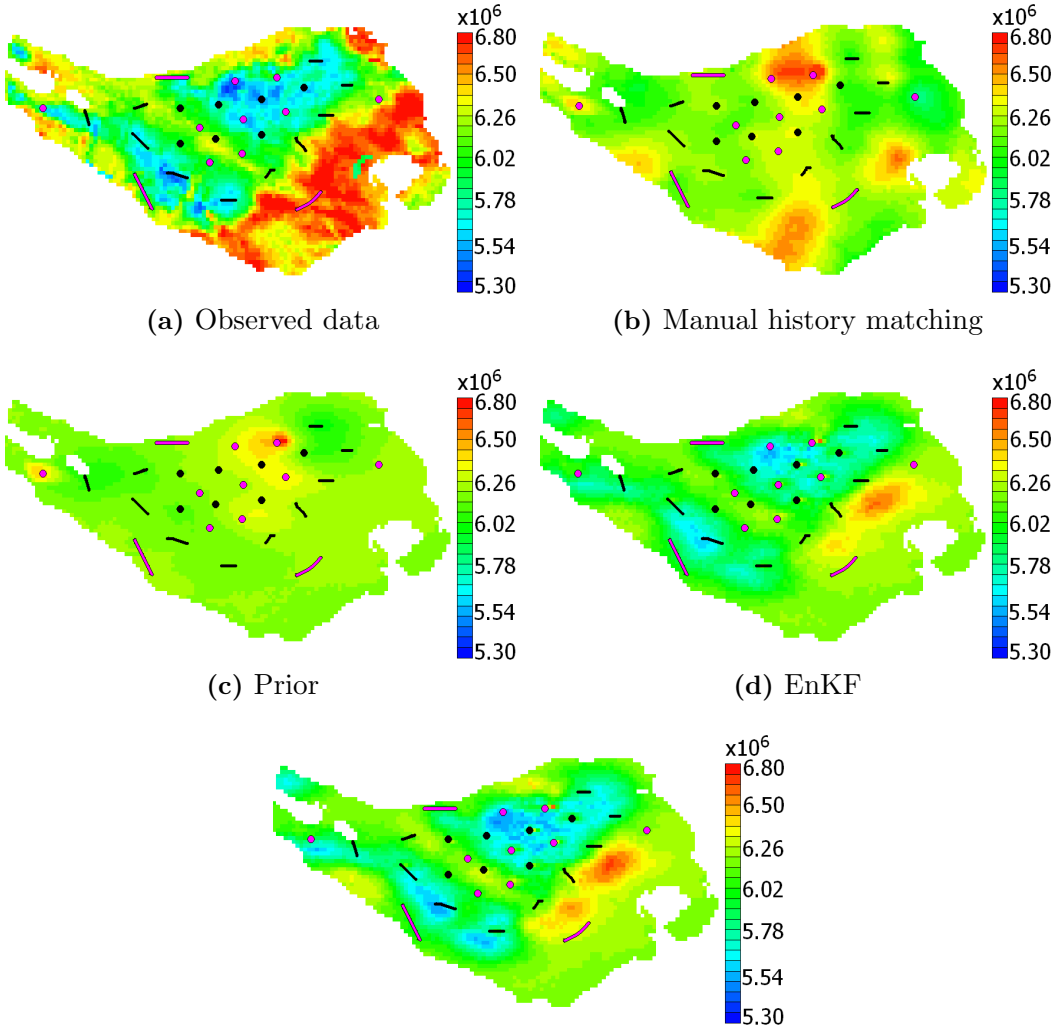
of Sections 6.5.3–6.5.5. We assumed that the initial condition of the reservoir, i.e., initial pressure and saturation distributions, are known (without uncertainty) in this case. Note that because the P-impedance data correspond to a 3D seismic acquisition before the beginning of the reservoir production, there is no need to use sequential data assimilation and EnKF becomes equivalent to ES. During the assimilation of the 3D seismic data, we only update the porosity and NTG fields. There are three main reasons for not including the permeability field in this case: first, because the seismic acquisition was before the beginning of production, the P-impedance data do not contain information about fluid transport. Therefore, the P-impedance data are not sensitive to permeability. Secondly, the prior ensemble of log-permeability is not correlated with the initial ensembles of porosity and NTG. Finally, even though the ensemble of log-permeability is not correlated to the 3D seismic data, including this ensemble in the data assimilation process could still cause some erroneous reduction in the ensemble variance because of spurious correlations possibly introduced by sampling errors. During the data assimilations, we used subspace inversion (Section 2.5.2) because this method is more efficient when the number of data points is large. In this case, we retained the largest singular values corresponding to 99.9%

of the sum of the nonzero singular values when applying Eq. 2.75. We used Kalman gain localization with critical lengths equal to the prior correlation lengths used to build the initial ensemble, i.e., 2,000 meters in the x-direction and 1,000 meters the y-direction. This choice of localization region is justified by the fact that there was no production before the 3D seismic acquisition and the initial pressure and saturation distribution of the reservoir were assumed known and independent of the porosity and NTG fields. In this case, the P-impedance value at a particular gridblock is sensitive only to the porosity and NTG at this gridblock. Therefore, by the argument presented in Chapter 3, in which we defend that localization should be based on both prior model and sensitivity region, we conclude that we only need to include the prior correlations in the localization region to assimilate the 3D seismic data. Fig. 7.4 presents the mean predicted P-impedance data obtained by the prior ensemble and the ensembles after data assimilation with EnKF and ES-MDA (4×). For comparison, we also include the P-impedance data predicted by the manually history-matched model and the observed P-impedance data. Fig. 7.5 presents the cross-plots between the observed and the predicted P-impedance data for all cases. The results in Figs. 7.4 and 7.5 show that the predicted P-impedance data from the manually history-matched model and from the prior ensemble are biased compared to the observed seismic. EnKF and ES-MDA (4×) resulted in significant improvements in the data matches, reducing the bias compared to the observations. After four data assimilations, the differences between the observed and the mean predicted P-impedance data are less than twice the standard deviation of the measurement errors for most of the reservoir gridblocks. Table 7.2 presents the mean and standard deviation of the seismic objective function normalized by the number of data points, $O_{N,s}$. The results in this table show that ES-MDA (4×) resulted in slightly better data matches than standard EnKF. The manually history-matched model obtained a value of $O_{N,s}$ similar to the mean $O_{N,s}$ from the prior ensemble, which may in-

Table 7.2: Mean and standard deviation of $O_{N,s}$ (3D seismic)

Case	Mean	Standard deviation
Manual history matching	9.938	–
Prior	9.238	0.842
EnKF	1.050	0.023
ES-MDA (4×)	0.739	0.025

dicate that the information from the 3D seismic data were not directly used when this model was generated. Figs. 7.6 and 7.7 present the ensemble mean porosity and NTG fields before and after the assimilation of the P-impedance data. We also include the porosity and NTG field from the manually history-matched model in these figures. According to the results in Figs. 7.6 and 7.7, the prior mean porosity and NTG fields were fairly homogeneous and significant changes were necessary to match the P-impedance data. Compared to the manually history-matched model, EnKF and ES-MDA (4×) resulted in some regions with significantly higher values of porosity. Fig. 7.7a shows that the manually history-matched model presents low values of NTG in the east part of the reservoir. However, we do not have hard-data and we assumed that the 3D seismic data are unreliable in this part of the field because of the small thicknesses of the reservoir. For these reasons, our prior model and the resulting models from EnKF and ES-MDA (4×) do not contain these low values of NTG observed in the manually history-matched model.



(a) Observed data

(b) Manual history matching

(c) Prior

(d) EnKF

(e) ES-MDA (4×)

Figure 7.4: Predicted P-impedance data (3D seismic) in kg/m^{2}s. For the prior, EnKF and ES-MDA (4×), the predicted P-impedance data in this figure correspond to the mean prediction from the corresponding ensembles. The producing wells are shown in black and the water injector wells in pink. Dots represent vertical wells and lines represent horizontal wells.

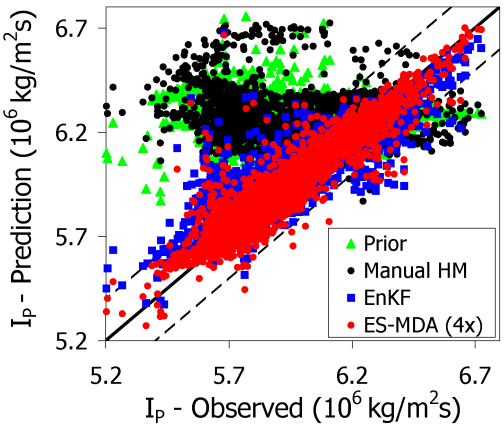
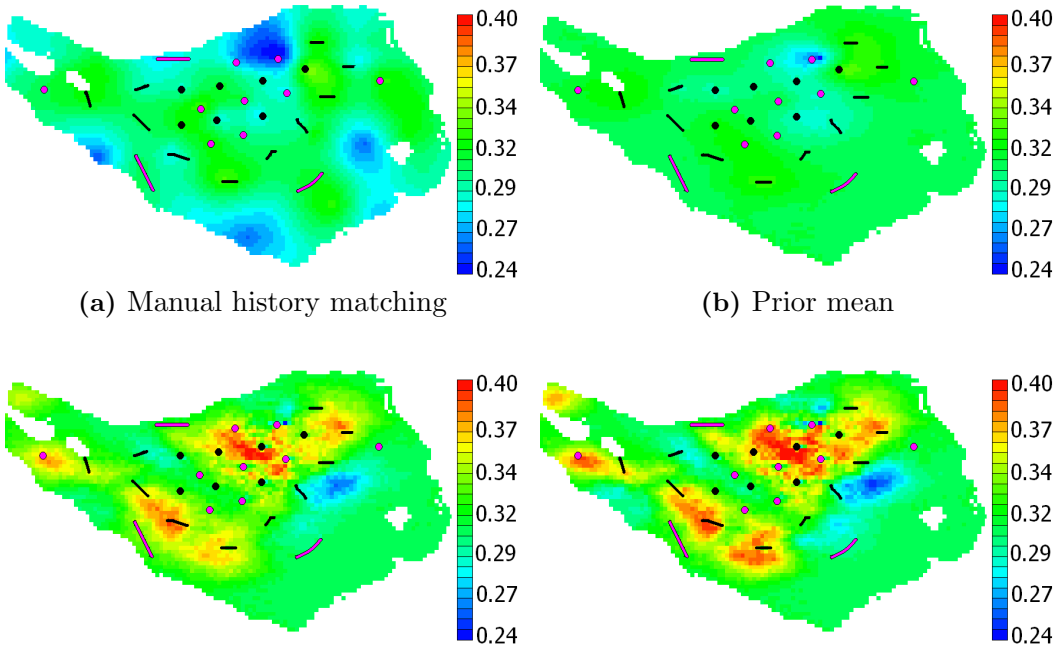


Figure 7.5: Cross-plot between observed and predicted P-impedance data (3D seismic). For the prior, EnKF and ES-MDA (4×), the predicted P-impedance data in this figure correspond to the mean prediction from the corresponding ensembles. The dashed lines correspond to +/− two standard deviations of the measurement errors.



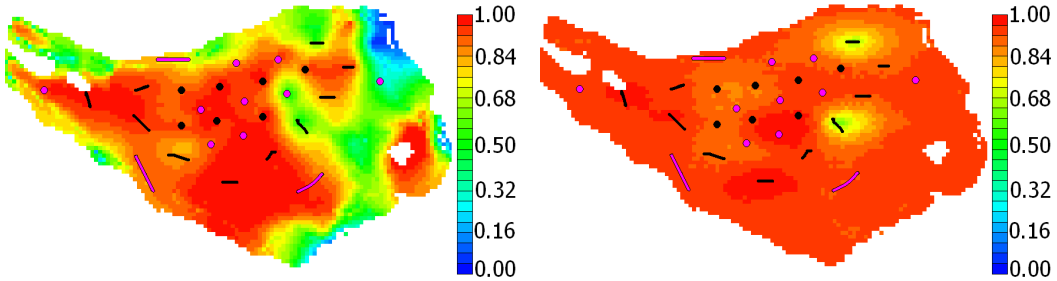
(a) Manual history matching

(b) Prior mean

(c) Posterior mean. EnKF

(d) Posterior mean. ES-MDA (4×)

Figure 7.6: Porosity fields. In this figure, (c) and (d) correspond to the posterior means after assimilation of the 3D seismic data. The producing wells are shown in black and the water injector wells in pink. Dots represent vertical wells and lines represent horizontal wells.



(a) Manual history matching **(b)** Prior

(c) EnKF **(d)** ES-MDA (4×)

Figure 7.7: Net-to-gross fields. In this figure, (c) and (d) correspond to the posterior means after assimilation of the 3D seismic data. The producing wells are shown in black and the water injector wells in pink. Dots represent vertical wells and lines represent horizontal wells.

7.4 Assimilation of Time-lapse Seismic and Production Data

In this field, the time-lapse seismic data correspond to two 3D seismic acquisitions, the base 3D acquisition in 1997 and a monitor survey acquired in May of 2005. Although there is a difference of eight years between these two surveys, the production in the reservoir zone considered in this history matching study began only in June of 2000. Hence, there are approximately 4.9 years of production between the two acquisitions. The two surveys were acquired using streamers with the same azimuth. However, some acquisition parameters were different [89]. Even though seismic processing partially compensated for these differences, the repeatability of the two acquisitions is not ideal for a time-lapse seismic. Moreover, this field produces to large FPSO – Floating Production Storage and Offloading – vessels, which represented obstacles during the monitor acquisition. Specifically, one FPSO vessel over the reservoir zone considered in this history matching study interfered with the monitor seismic acquisition. A considerable effort was made to image these areas with obstacles during the acquisition, however, the ocean currents were very high during the time of the survey, which caused poor repeatability in these areas [89]. Fig. 7.8a presents the time-lapse seismic in terms of the time-difference P-impedance data (ΔI_P). The green to red colors in this figure correspond to increase in P-impedance due to the displacement of oil by injected water. We also show in Fig. 7.8a the region of interference of the FPSO vessel during the monitor seismic acquisition. Note that there are three water injection wells in this region of interference; however, we do not observe a significant increase in the observed P-impedance data inside this region. This indicates that the ΔI_P data from this region are not reliable in this region; hence, we they were not used for data assimilation. Fig. 7.2b shows the regions of the reservoir with time-lapse seismic data used for history matching. Note that besides the region of interference of the FPSO vessel, we also removed data from the areas of the reservoir with thickness less than 20 meters,

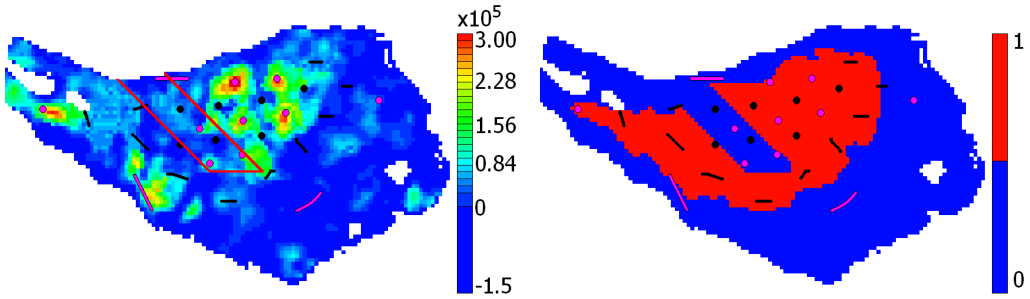
as we did during the assimilation of the 3D seismic data. Similarly to what we did for the 3D seismic data, we use an average ΔI_P data over the four reservoir layer for data assimilation. We estimated the noise level in the ΔI_P data using the same procedure used for the 3D seismic. However, unlike the 3D seismic data, we did not find spatial correlation in the noise of the ΔI_P data. We obtained a standard deviation of the noise of approximately 2.9×10^4 kg/m^{2}s. Fig. 7.9 shows the residuals and the corresponding variograms indicating that the noise is spatially uncorrelated. This result is surprising if we consider that both 3D acquisitions have correlated noise over distances corresponding to approximately the width of 5 gridblocks. In order to analyze this result, let $d_{\text{obs},1}$ and $d_{\text{obs},2}$ denote the base and monitor seismic data. Let us assume that $d_{\text{obs},1} \sim \mathcal{N}(\overline{d}_{\text{obs},1}, C_{\text{D}1})$ and $d_{\text{obs},2} \sim \mathcal{N}(\overline{d}_{\text{obs},2}, C_{\text{D}2})$. Then, let us define the random variable $\Delta d_{\text{obs}} = d_{\text{obs},2} - d_{\text{obs},1}$, which correspond to the observed time-lapse seismic data. The covariance of Δd_{obs} is give by

$$\begin{aligned}
\text{cov}[\Delta d_{\text{obs}}] &= \text{E}[(\Delta d_{\text{obs}} - \overline{\Delta d}_{\text{obs}})(\Delta d_{\text{obs}} - \overline{\Delta d}_{\text{obs}})^{\text{T}}] \\
&= \text{E}[d_{\text{obs},2} d_{\text{obs},2}^{\text{T}}] + \text{E}[d_{\text{obs},1} d_{\text{obs},1}^{\text{T}}] - 2\text{E}[d_{\text{obs},2} d_{\text{obs},1}^{\text{T}}] \\
&- \overline{d}_{\text{obs},2} \overline{d}_{\text{obs},2}^{\text{T}} - \overline{d}_{\text{obs},1} \overline{d}_{\text{obs},1}^{\text{T}} + 2\overline{d}_{\text{obs},2} \overline{d}_{\text{obs},1}^{\text{T}} \\
&= C_{\text{D}1} + C_{\text{D}2} - 2\text{cov}[d_{\text{obs},1}, d_{\text{obs},2}]. \qquad (7.33)
\end{aligned}$$

If $d_{\text{obs},1}$ and $d_{\text{obs},2}$ are independent samples of $\mathcal{N}(\overline{d}_{\text{obs},1}, C_{\text{D}1})$ and $\mathcal{N}(\overline{d}_{\text{obs},2}, C_{\text{D}2})$, respectively, we have $\text{cov}[d_{\text{obs},1}, d_{\text{obs},2}] = 0$ and $\text{cov}[\Delta d_{\text{obs}}]$ becomes

$$\text{cov}[\Delta d_{\text{obs}}] = C_{\text{D}1} + C_{\text{D}2}. \qquad (7.34)$$

Therefore, as $C_{\text{D}1}$ and $C_{\text{D}2}$ have approximately the same correlation length, i.e., 5 gridblocks, we should expect $\text{cov}[\Delta d_{\text{obs}}]$ to be also correlated over a distance of 5 gridblocks. However, the results in Fig. 7.9 indicate that Δd_{obs} is spatially uncorrelated.



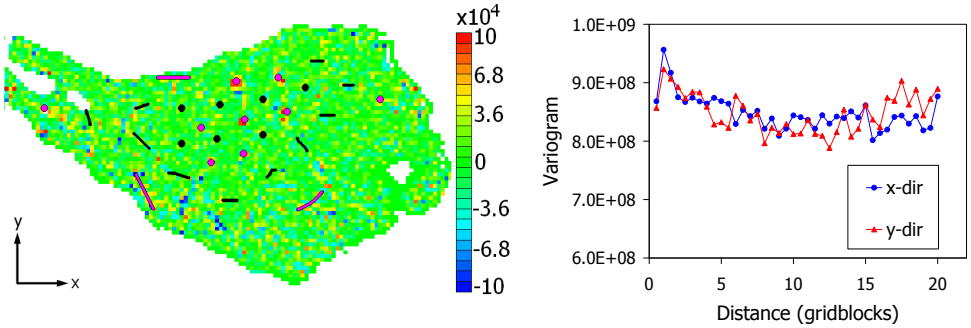
(a) P-impedance difference (kg/m^{2}s) **(b)** Active seismic data

Figure 7.8: Observed time-lapse and active seismic data. In (a) the red lines indicate the region of interference of the FPSO during the seismic acquisition. In (b), the region in red corresponds to the region with ΔI_P data used for data assimilation. The producing wells are shown in black and the water injector wells in pink. Dots represent vertical wells and lines represent horizontal wells.

A possible explanation for the result of Fig. 7.9 is that $d_{\text{obs},1}$ and $d_{\text{obs},2}$ are not independent samples. In fact, we believe that the inversion process from amplitude to P-impedance might have introduced correlation between $d_{\text{obs},1}$ and $d_{\text{obs},2}$. In this case, we have $\text{cov}[d_{\text{obs},1}, d_{\text{obs},2}] \neq 0$, which may explain the fact that we obtained $\text{cov}[\Delta d_{\text{obs}}]$ with approximately zero correlation length. Although we cannot be sure about the conjecture that the inversion process introduced correlation between the base and monitor data, we used the results of Fig. 7.9 to define the covariance of the measurement errors of the time-lapse seismic during data assimilation.

The production data used for history matching are the same used in Chapter 3, i.e., water production rate at 14 wells. As before, from the total production period of 3,621 days, we assimilated data only for the first 2,770 days. The remaining 851 days of history are used for comparing the predictions from the history-matched models.

We assimilated the time-lapse seismic and the production data simultaneously using EnKF and ES-MDA (4×). During data assimilations, we update the porosity, NTG and log-permeability fields. Note that the initial ensembles of porosity and NTG used for assimilation of time-lapse seismic and the production data



(a) Residuals between observed and smoothed ΔI_P data (kg/m^{2}s)

(b) Experimental variograms of the residuals

Figure 7.9: Residuals between observed and smoothed data used to estimate the covariance of measurement errors of the ΔI_P.

correspond to the posterior ensembles after the assimilation of 3D seismic data. We used subspace inversion with rescaling and truncation based on 99.9% of the sum of the nonzero singular values (Section 2.5.2). For localization of the production data, we defined the critical lengths using the same procedure used in Chapter 3, which is based on the prior correlation and drainage areas. For the time-lapse seismic data, we do not have a simple procedure to estimate the sensitivity region, which makes it difficult to design a proper localization strategy. In this case, we chose the critical lengths for localization as twice the correlation lengths of the prior geological model. However, we did not experiment with different choices of critical lengths for localization of the time-lapse seismic data.

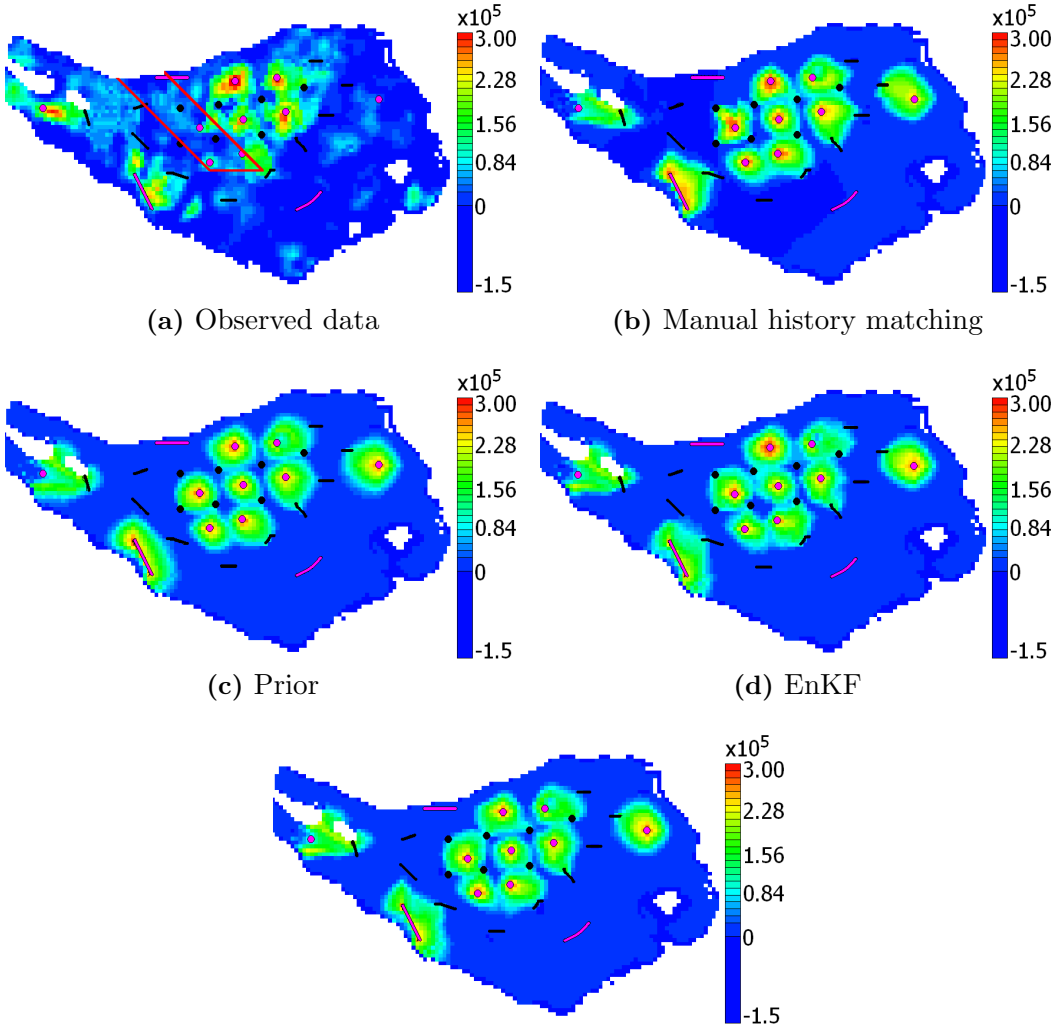
Fig. 7.10 presents the mean predicted ΔI_P data obtained by the prior ensemble and the ensembles after data assimilation with EnKF and ES-MDA (4×). For comparison, we also include the ΔI_P data predicted by the manually history-matched model and the observed ΔI_P data in this figure. Fig. 7.11 presents the cross-plots between the observed and the predicted ΔI_P data and Table 7.2 presents the mean and standard deviation of $O_{N,s}$. The results in Figs. 7.10 and 7.11 and Table 7.2 show that the ΔI_P data predicted by the manually history-matched model and the initial ensemble are in reasonable agreement with the observed time-lapse

Table 7.3: Mean and standard deviation of $O_{N,s}$ (time-lapse seismic)

Case	Mean	Standard deviation
Manual history matching	1.834	–
Prior	1.775	0.236
EnKF	1.274	0.091
ES-MDA (4×)	1.258	0.038

seismic data. Nevertheless, the assimilation of time-lapse seismic data with EnKF and ES-MDA resulted in improvements in the match of ΔI_P data. For example, the average values of $O_{N,s}$ was reduced by approximately 30% after data were assimilated with EnKF and ES-MDA (4×). Compared to each other, EnKF and ES-MDA (4×) resulted in approximately the same level of data match of ΔI_P although the ES-MDA (4×) obtained a slightly lower value of mean $O_{N,s}$.

Fig. 7.12a presents the box plots of normalized production data mismatch objective function, $O_{N,p}$, obtained by the manually history-matched model, the prior ensemble and the ensembles after data assimilation with EnKF and ES-MDA (4×). Note that the results in this figure correspond to the historical period used for data assimilation (2,770 days). Fig. 7.12a shows that ES-MDA (4×) resulted in significantly better production data matches than EnKF. For example, the median value of $O_{N,p}$ obtained by ES-MDA (4×) is almost seven times lower that the median value of $O_{N,p}$ obtained by EnKF and five times lower than the $O_{N,p}$ obtained by the manually history-matched model. The results of Fig. 7.12b, which show the corresponding box plots of $O_{N,p}$ for the forecast period (851 days), indicate that EnKF and ES-MDA (4×) gave values of $O_{N,p}$ which are two times smaller that the $O_{N,p}$ value corresponding to the manually history-matched model. Figs. 7.13–7.20 present the predicted water production rate for eight wells. Fig. 7.21 shows the total predicted water production rate (all wells). In these figures, we include both the historical period (2,770 days) and the forecast period (851 days). These figures illustrate that



(a) Observed data

(b) Manual history matching

(c) Prior

(d) EnKF

(e) ES-MDA (4×)

Figure 7.10: Predicted ΔI_P data (time-lapse seismic) in kg/m^{2}s. For the prior, EnKF and ES-MDA (4×), the predicted P-impedance data in this figure correspond to the mean prediction from the corresponding ensembles. The producing wells are shown in black and the water injector wells in pink. Dots represent vertical wells and lines represent horizontal wells.

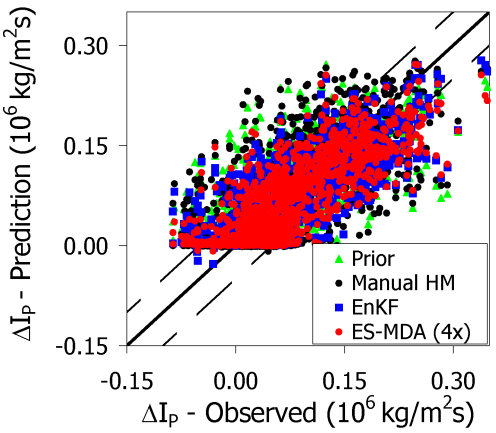
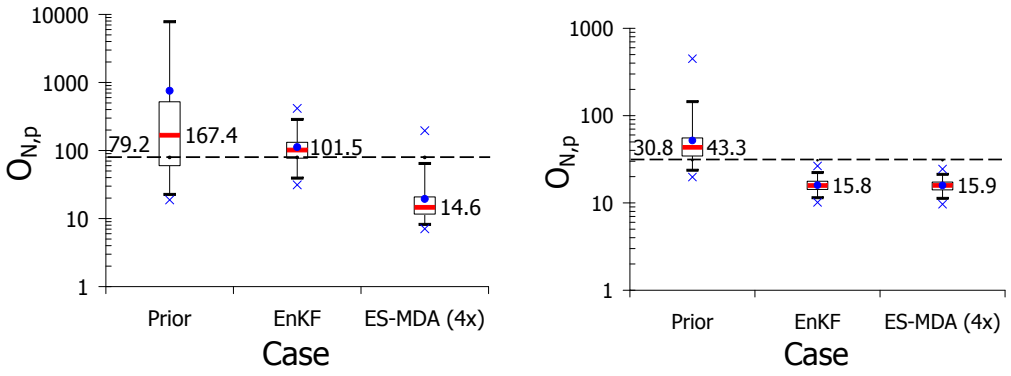


Figure 7.11: Cross-plot between observed and predicted ΔI_P data (time-lapse seismic). For the prior, EnKF and ES-MDA (4×), the predicted ΔI_P data in this figure correspond to the mean prediction from the corresponding ensembles. The dashed lines correspond to +/− two standard deviations of the measurement errors.

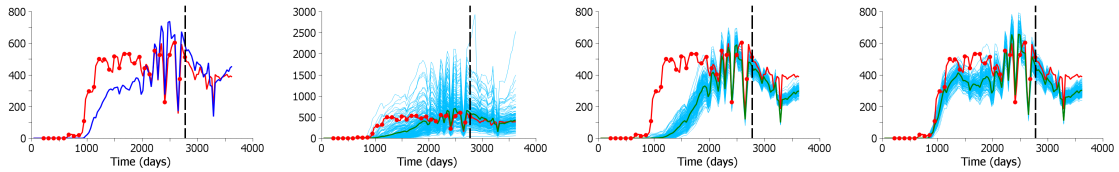
EnKF and ES-MDA (4×) resulted in reasonable data matches and predictions of water production rate for most of the wells, and that, in general, these matches are better than the data matches and predictions obtained by the manually history-matched model. The main differences between the results of EnKF and ES-MDA (4×) occur for wells P-85 and P-86, in which ES-MDA (4×) gave significantly better data matches, especially close to the water breakthrough time; and for well P-141, in which EnKF resulted in better agreement with the observations during the forecast period. Figs. 7.22, 7.23 and 7.24 present the final porosity, NTG and permeability fields, respectively. These figures illustrates that even though significant changes in the prior model were necessary in order to match the production and seismic data, the final rock properties fields obtained by EnKF and ES-MDA (4×) seem plausible. For the porosity fields, the main changes were due to the assimilation of the 3D seismic; compare Fig. 7.6 with Fig. 7.22. For the NTG field, the assimilation of the production and time-lapse seismic data incurred additional significant changes to those observed after the assimilation of the 3D seismic data. The final permeability fields obtained from EnKF and ES-MDA (4×) present some regions with values close 24,000 mD. Although these values are extremely high, they were expected consid-



(a) History (2,770 days) **(b)** Forecast (851 days)

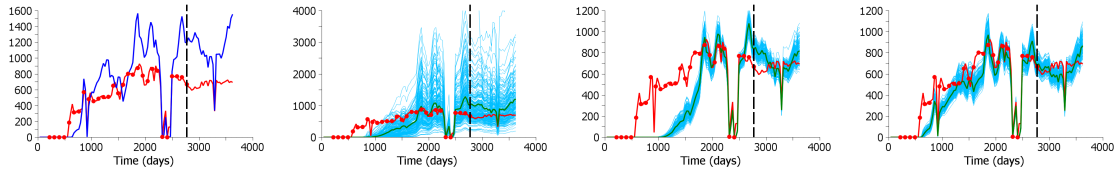
Figure 7.12: Box plots of the normalized production data mismatch objective function. The horizontal dashed lines correspond to the $O_{N,p}$ obtained by the manually history-matched model. The numbers next to each box plot correspond to the values of the median of $O_{N,p}$ and the numbers in the beginning of the horizontal dashed lines correspond to the values of $O_{N,p}$ obtained by the manually history-matched model

ering that the prior information from interpreted pressure transient data indicates values around 20,000 mD. Besides that, these permeability fields are more geologically plausible than the permeability field of the manually history-matched model, which presents several artifacts introduced by the engineers of the field to match the water production rate data.



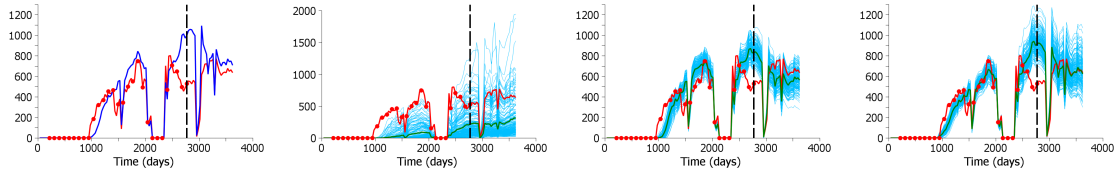
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** ES-MDA (4×)

Figure 7.13: Well P-85. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). Red dots are the data points used for data assimilation; solid red curve is the production history; solid blue curve is the prediction from the manually history-matched model; solid green curves are the mean predictions from the ensembles; solid light blue curves represent predictions from all models of the ensembles. The dashed vertical black line indicates the end of the history and beginning of the forecast period.



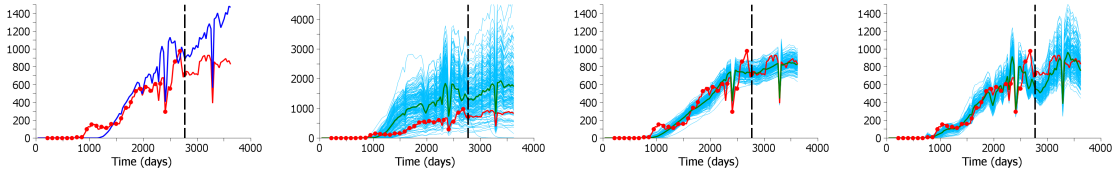
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** ES-MDA (4×)

Figure 7.14: Well P-86. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 7.13



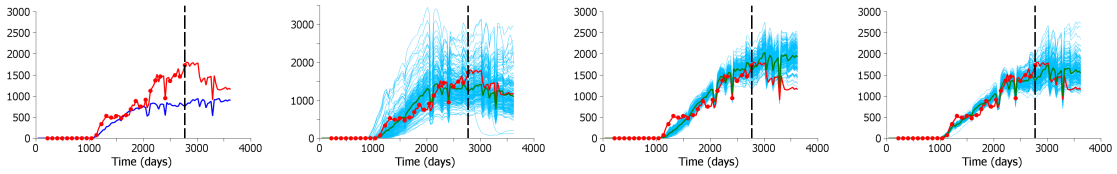
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** ES-MDA (4×)

Figure 7.15: Well P-101. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 7.13



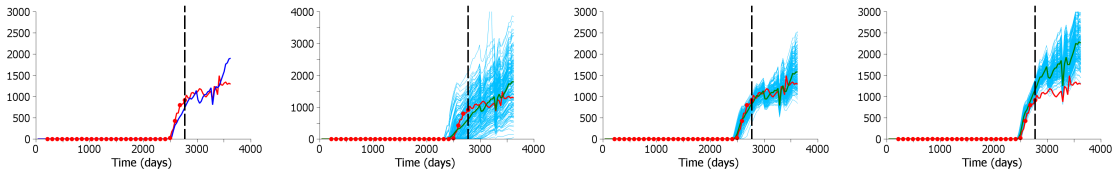
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** ES-MDA (4×)

Figure 7.16: Well P-104. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 7.13



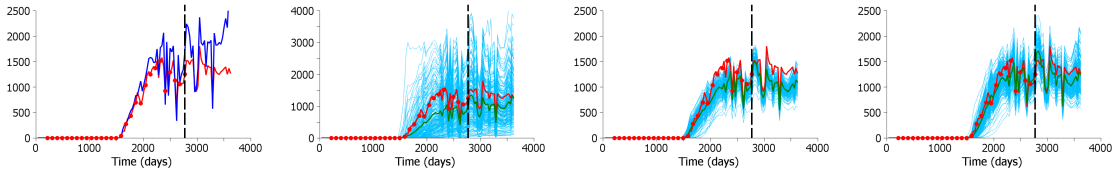
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** ES-MDA (4×)

Figure 7.17: Well P-111. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 7.13



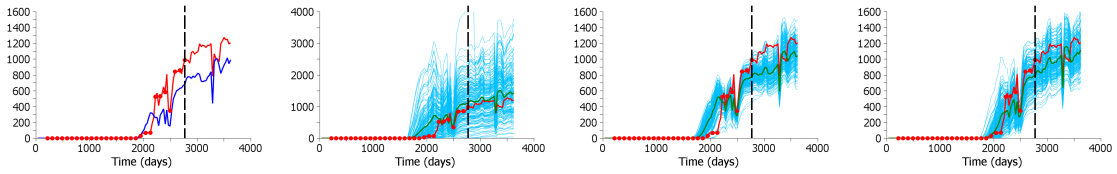
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** ES-MDA (4×)

Figure 7.18: Well P-141. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 7.13



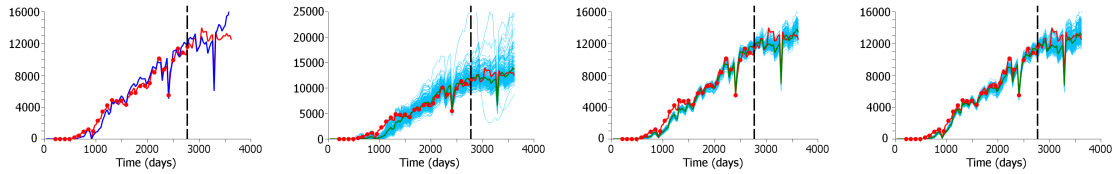
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** ES-MDA (4×)

Figure 7.19: Well P-149. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 7.13



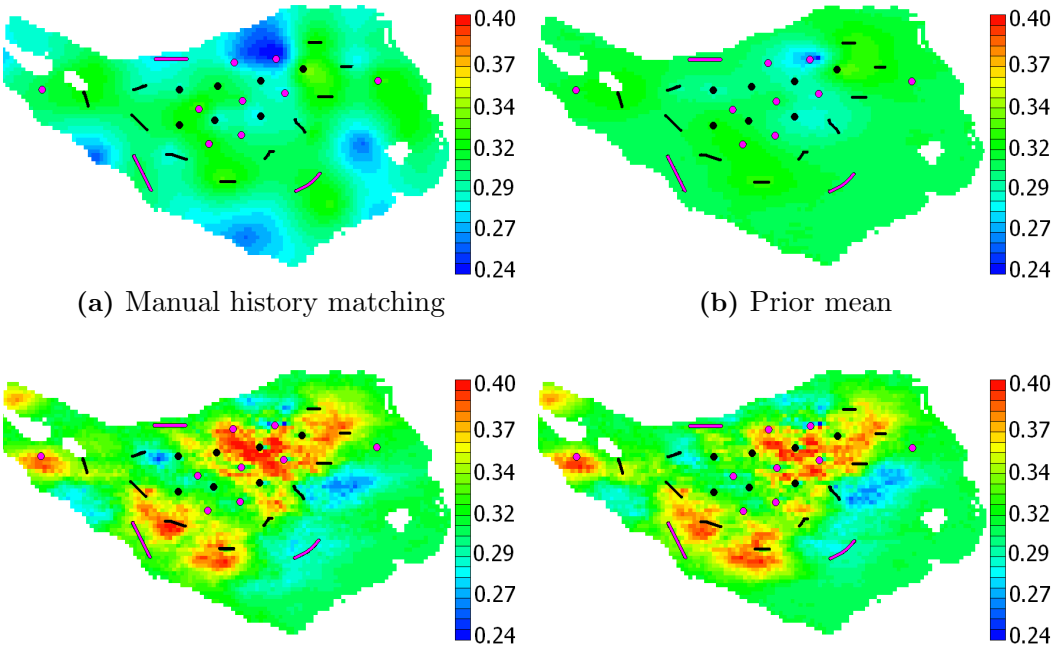
(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** ES-MDA (4×)

Figure 7.20: Well P-152. Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 7.13



(a) Manual history matching **(b)** Prior **(c)** EnKF **(d)** ES-MDA (4×)

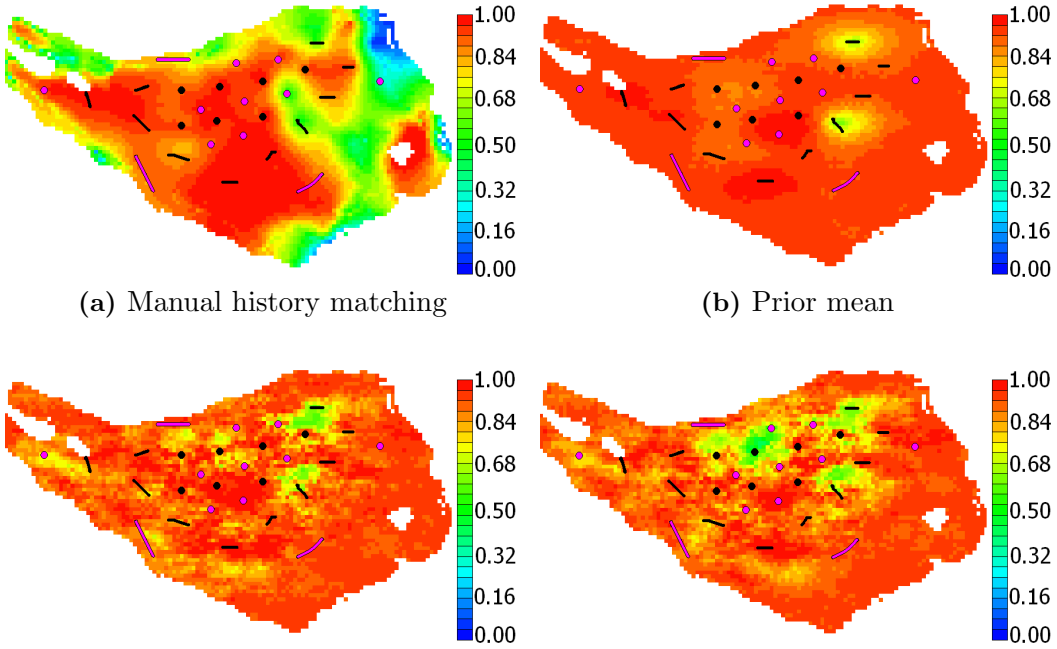
Figure 7.21: Field water production (all wells). Predicted water production rate obtained by running the final ensembles from time zero (std m^3/day). The colors in this figure have the same meaning as in Fig. 7.13



(a) Manual history matching **(b)** Prior mean

(c) Posterior mean. EnKF **(d)** Posterior mean. ES-MDA (4×)

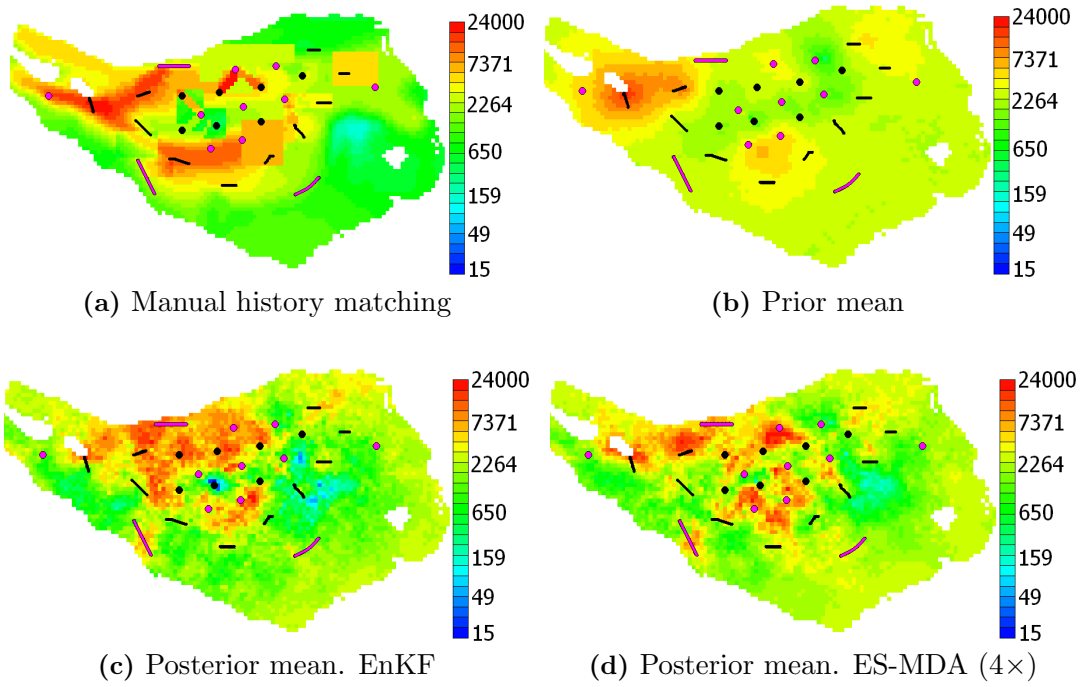
Figure 7.22: Final porosity fields. In this figure, (c) and (d) correspond to the posterior means after assimilation of production and seismic data. The producing wells are shown in black and the water injector wells in pink. Dots represent vertical wells and lines represent horizontal wells.



(a) Manual history matching **(b)** Prior mean

(c) Posterior mean. EnKF **(d)** Posterior mean. ES-MDA (4×)

Figure 7.23: Final net-to-gross fields. In this figure, (c) and (d) correspond to the posterior means after assimilation of production and seismic data. The producing wells are shown in black and the water injector wells in pink. Dots represent vertical wells and lines represent horizontal wells.



(a) Manual history matching **(b)** Prior mean

(c) Posterior mean. EnKF **(d)** Posterior mean. ES-MDA (4×)

Figure 7.24: Final permeability fields. In this figure, (c) and (d) correspond to the posterior means after assimilation of production and seismic data. The producing wells are shown in black and the water injector wells in pink. Dots represent vertical wells and lines represent horizontal wells.

Table 7.4: Computational costs.

Case	Number of equivalent simulation runs
Run prior ensemble	200
EnKF	850
ES-MDA (4×)	885

Table 7.4 presents the computational cost in terms of the equivalent number of reservoir simulation runs for EnKF and ES-MDA (4×). The results in this table were obtained by measuring the actual CPU time required during the data assimilations divided by the average CPU time required to run one reservoir simulation and to compute the corresponding predicted seismic data. According to the results in Table 7.4, the CPU time of ES-MDA (4×) was only 4% higher than the CPU time of EnKF.

CHAPTER 8

ENSEMBLE-BASED PARAMETERIZATION FOR THE RANDOMIZED MAXIMUM LIKELIHOOD METHOD

Even when adjoint code for computing the gradient of the objective function of Eq. 2.21 is available, the randomized maximum likelihood (RML) method can be computationally intensive. Therefore, the development of efficient parameterizations for RML is highly desirable. This chapter introduces an ensemble-based parameterization to reduce the computational cost of RML. This parameterization is inspired by the EnKF, which uses linear combinations of realizations of prior models to history-match data. In addition, this chapter presents a comparison between RML and EnKF for a small reservoir problem.

8.1 Parameterization Based on Ensembles of Prior Realizations

The parameterization is based on a linear combination of realizations generated from the prior model. We write the vector of reservoir model parameters, m, as a linear combination of an ensemble of these realizations from the prior using

$$m = \overline{m} + \frac{1}{\sqrt{N_e - 1}} \Delta M \theta, \tag{8.1}$$

where, as usual, $\overline{m}$ is the prior ensemble mean vector; $\Delta M = M - \overline{M}$ and M is the $N_m \times N_e$ matrix containing the ensemble of vectors of prior model parameters, i.e., each column of M corresponds to a prior realization. $\overline{M}$ is the $N_m \times N_e$ matrix with all columns equal to the ensemble mean, $\overline{m}$; θ is the N_e-dimensional column vector of coefficients of the linear combination.

As discussed in Chapter 2, the MAP estimate is the model that maximizes the posterior pdf defined in the Eq. 2.2, or equivalently minimizes the objective function of Eq. 2.3. Hence, in order to compute the MAP estimate, we look for the vector θ_{MAP} such that

$$\theta_{\text{MAP}} = \arg\min_{\theta} O, \tag{8.2}$$

where O is given by Eq. 2.3. Recall that in Eq. 2.3, the model mismatch part of the objective function, O_m, is given by

$$O_m = \frac{1}{2}(m - m_{\text{pr}})^{\text{T}} C_{\text{M}}^{-1}(m - m_{\text{pr}}). \tag{8.3}$$

First note that $\overline{m} \approx m_{\text{pr}}$. In fact, we correct the ensemble of prior realizations to make $\overline{m} = m_{\text{pr}}$ using

$$m_j^{\text{corr}} = m_j - (\overline{m} - m_{\text{pr}}), \text{ for } j = 1, 2, \cdots N_e, \tag{8.4}$$

where m_j^{corr} is the jth prior realization corrected so that

$$\frac{1}{N_e}\sum_{j=1}^{N_e} m_j^{\text{corr}} = m_{\text{pr}}. \tag{8.5}$$

Then, using Eq. 8.1 in Eq. 8.3, we have

$$O_m \approx \frac{1}{2}\left(\frac{1}{\sqrt{N_e - 1}}\Delta M\theta\right)^{\text{T}} C_{\text{M}}^{-1}\left(\frac{1}{\sqrt{N_e - 1}}\Delta M\theta\right). \tag{8.6}$$

We can write the SVD of ΔM as

$$\Delta M = U\Lambda V^{\text{T}} \tag{8.7}$$

where U is a $N_m \times N_m$ orthogonal matrix; V is a $N_e \times N_e$ orthogonal matrix and Λ is a $N_m \times N_e$ matrix with nonzero elements corresponding to the singular values

of ΔM. Assuming the ensemble approximation for the prior covariance matrix and using Eq. 8.7, we can write the prior covariance matrix as

$$\begin{aligned} C_{\mathrm{M}} \approx \widetilde{C}_{\mathrm{M}} &= \frac{\Delta M \Delta M^{\mathrm{T}}}{N_e - 1} \\ &= \frac{(U\Lambda V^{\mathrm{T}})(U\Lambda V^{\mathrm{T}})^{\mathrm{T}}}{N_e - 1} \\ &= \frac{(U\Lambda^2 U^{\mathrm{T}})}{N_e - 1}. \end{aligned} \tag{8.8}$$

and its inverse becomes

$$C_{\mathrm{M}}^{-1} \approx \widetilde{C}_{\mathrm{M}}^{+} = (N_e - 1)(U\Lambda^{-2}U^{\mathrm{T}}), \tag{8.9}$$

where the superscript "+" means pseudo-inverse.

Using Eq. 8.9 in Eq. 8.6 leads to

$$\begin{aligned} O_m &\approx \frac{1}{2}\left(\frac{1}{\sqrt{N_e - 1}}(U\Lambda V^{\mathrm{T}})\theta\right)^{\mathrm{T}}(N_e - 1)(U\Lambda^{-2}U^{\mathrm{T}})\left(\frac{1}{\sqrt{N_e - 1}}(U\Lambda V^{\mathrm{T}})\theta\right) \\ &= \frac{1}{2}\theta^{\mathrm{T}}(V\Lambda U^{\mathrm{T}})(U\Lambda^{-2}U^{\mathrm{T}})(U\Lambda V^{\mathrm{T}})\theta \\ &= \frac{1}{2}\theta^{\mathrm{T}}\theta. \end{aligned} \tag{8.10}$$

Hence, for computing the MAP estimate the model mismatch part of the objective function can be approximated using the 2-norm of the vector θ, i.e.,

$$O_m \approx \frac{1}{2}\|\theta\|_2^2 = \frac{1}{2}\theta^{\mathrm{T}}\theta. \tag{8.11}$$

Using Eq. 8.11, we avoid solving the matrix problem $x = C_{\mathrm{M}}^{-1}(m - m_{\mathrm{pr}})$, which can be computationally expensive for large reservoir models. For computing the MAP estimate, we simply start the minimization with $\theta = 0$ in Eq. 8.11.

For sampling the posterior pdf with RML, we look for the vector θ_j such that

$$\theta_j = \arg\min_{\theta} O_j, \tag{8.12}$$

where O_j is the RML objective function given by Eq. 2.21. Similarly to the MAP estimate case, the model mismatch part of the objective function can be approximated as

$$O_{m,j} \approx \frac{1}{2} \left(\theta - \theta_{\mathrm{uc},j}\right)^{\mathrm{T}} \left(\theta - \theta_{\mathrm{uc},j}\right). \tag{8.13}$$

In this case, the vector $\theta_{\mathrm{uc},j}$ can be obtained by two different procedures:

1. Sample $\theta_{\mathrm{uc},j}$ from $\mathcal{N}(0, I_{N_e})$, with I_{N_e} denoting the $N_e \times N_e$ identity matrix. For this case, the corresponding vector of model parameters, $m_{\mathrm{uc},j}$, is a sample from $\mathcal{N}(\overline{m}, \widetilde{C}_{\mathrm{M}})$.

2. Choose one of the N_e models from the prior ensemble and set the entry of the vector $\theta_{\mathrm{uc},j}$ corresponding to this model to $\sqrt{N_e - 1}$. The remaining entries of the vector $\theta_{\mathrm{uc},j}$ are set to zero.

Note that sampling $\theta_{\mathrm{uc},j} \sim \mathcal{N}(0, I_{N_e})$ is equivalent to sampling $m_{\mathrm{uc},j} \sim \mathcal{N}(\overline{m}, \widetilde{C}_{\mathrm{M}})$ because

$$\begin{aligned} \mathrm{E}[m_{\mathrm{uc},j}] &= \mathrm{E}\left[\overline{m} + \frac{\Delta M}{\sqrt{N_e - 1}} \theta_{\mathrm{uc},j}\right] \\ &= \overline{m} + \frac{\Delta M}{\sqrt{N_e - 1}} \mathrm{E}[\theta_{\mathrm{uc},j}] \\ &= \overline{m} \end{aligned} \tag{8.14}$$

and

$$
\begin{aligned}
\mathrm{cov}[m_{\mathrm{uc},j}] &= \mathrm{E}\left[\left(m_{\mathrm{uc},j}-\overline{m}\right)\left(m_{\mathrm{uc},j}-\overline{m}\right)^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[\left(\frac{\Delta M}{\sqrt{N_e-1}}\theta_{\mathrm{uc},j}\right)\left(\frac{\Delta M}{\sqrt{N_e-1}}\theta_{\mathrm{uc},j}\right)^{\mathrm{T}}\right] \\
&= \frac{\Delta M}{\sqrt{N_e-1}}\mathrm{E}\left[\theta_{\mathrm{uc},j}\theta_{\mathrm{uc},j}^{\mathrm{T}}\right]\frac{\Delta M^{\mathrm{T}}}{\sqrt{N_e-1}} \\
&= \frac{\Delta M \Delta M^{\mathrm{T}}}{N_e-1} \\
&= \widetilde{C}_{\mathrm{M}}. \qquad (8.15)
\end{aligned}
$$

The main motivation for this parameterization is to reduce the dimensionality of the minimization problem, we have N_e instead of N_m parameters. However, this parameterization also has some other attractive features including the following:

- Using Eq. 8.11 (or Eq. 8.13 for RML case), we avoid the need of solving the matrix problem $x = C_{\mathrm{M}}^{-1}(m - m_{\mathrm{pr}})$, which can be computationally intensive for large reservoir models. Besides that, for real field applications, typically the geologist uses geostatistics to generate realizations conditional to well logs and 3D seismic. For these cases, we do not have a well defined model covariance matrix and the use of the ensemble approximation of the model mismatch objective function is a simple and effective alternative.

- The dimensionality reduction tends to introduce regularization in the minimization problem helping to ameliorate convergence problems related to overcorrection in the model.

- This parameterization has similarities with the parameterization based on the Karhunen-Lòeve expansion of the prior covariance matrix used in [130, 141]. However, this ensemble-based parameterization avoids the eigen-decomposition of the prior covariance matrix, which can be computationally expensive. Be-

sides that, parameterizations based on the main eingenpairs of the prior covariance matrix tend to generate prior realization which are "too smooth" because the high frequency components, which are associated with local reservoir heterogeneities, are discarded. The ensemble-based parameterization avoids this difficulty and the prior realizations generated are visually indistinguishable from realizations generated from the actual prior covariance matrix.

Considering that it is not necessary to run the ensemble of prior realizations and the computational cost of generating an ensemble is low compared to the cost of running simulations, we apply RML using a different initial ensemble for computing each conditional realization. As a result, the final conditional models are linear combinations of different prior models. With this procedure, we expect to obtain independent samples of the posterior pdf, which is typically not the case for EnKF, where the final ensemble is obtained by linear combinations of a single ensemble.

In our implementation, we solve the minimization problem required in the RML method using the LBFGS algorithm [126, 192, 59] with scaling [106, 192] and bracketing cubic line search [135, Chap. 8]. The gradients are computed using the adjoint method [21, 19, 188, 105]. In the adjoint implementation, we compute the gradient of the objective function with respect to the vector of model parameters m. In order to perform the optimization in terms of the vector θ, we need to transform the gradients, which can easily be done applying the chain rule:

$$\nabla_\theta O = \nabla_\theta m^{\mathrm{T}} \nabla_m O. \tag{8.16}$$

In the above equation, $\nabla_\theta O$ denotes the N_e-dimensional gradient vector of the objective function with respect to θ and $\nabla_m O$ denotes the N_m-dimensional gradient vector of the objective function with respect to m. From Eq. 8.1 we can write

$$\nabla_\theta m^{\mathrm{T}} = \frac{1}{\sqrt{N_e - 1}} \Delta M^{\mathrm{T}}. \tag{8.17}$$

Therefore, the transformed gradients become

$$\begin{aligned}\nabla_\theta O &= \frac{1}{\sqrt{N_e-1}}\Delta M^{\mathrm{T}}\nabla_m O \\ &= \frac{1}{\sqrt{N_e-1}}\Delta M^{\mathrm{T}}\left(\nabla_m O_m + \nabla_m O_d\right).\end{aligned} \tag{8.18}$$

where $\nabla_m O_m = C_{\mathrm{M}}^{-1}(m - m_{\mathrm{pr}})$ and $\nabla_m O_d$ is the gradient vector of the likelihood part of the objective function, which is obtained in the adjoint implementation. When using the approximate model mismatch objective function, the expression of the transformed gradients can be simplified to

$$\begin{aligned}\nabla_\theta O &= \nabla_\theta O_m + \nabla_\theta O_d \\ &= \nabla_\theta\left(\frac{1}{2}\theta^{\mathrm{T}}\theta\right) + \frac{1}{\sqrt{N_e-1}}\Delta M^{\mathrm{T}}\nabla_m O_d \\ &= \theta + \frac{1}{\sqrt{N_e-1}}\Delta M^{\mathrm{T}}\nabla_m O_d,\end{aligned} \tag{8.19}$$

which can be calculated without solving the matrix problem $x = C_{\mathrm{M}}^{-1}(m - m_{\mathrm{pr}})$. Eq. 8.19 was derived for the case of the MAP estimate. However, this expression is the same for RML, where we only need to replace $\nabla_m O_d$ by $\nabla_m O_{d,j}$.

For convergence, we require both the change in the objective function and the change in θ over an iteration to be small, i.e.,

$$\frac{|O(\theta^{\ell+1}) - O(\theta^{\ell})|}{O(\theta^{\ell+1})} < 10^{-4} \tag{8.20}$$

and

$$\max_{1\le i\le N_e}\left|\frac{\theta_i^{\ell+1} - \theta_i^{\ell}}{\theta_i^{\ell} + 10^{-8}}\right| < 10^{-3}, \tag{8.21}$$

where the superscript ℓ denotes the iteration index.

8.2 Example

The example is a 2D synthetic reservoir model on a 15×30 uniform grid with two-phase (oil and water) flow. The model parameters are gridblock log-permeabilities, $\ln(k)$. The true model was generated from an anisotropic exponential correlation function with major correlation length of 12,000 ft (which corresponds to the width of 60 gridblocks) and minor correlation length of 1,000 ft oriented at 45°. The prior mean of $\ln(k)$ is 5.0 and the prior variance is 1.0 for all gridblocks. Fig. 8.1 shows the true permeability field used to generate the observed data. In this model, there are six producing wells and two water injectors which form two five spots. The observed data are 11 years of oil and water production rates and water injection rates. Random normally distributed noise with zero mean and standard deviation correspondent to 5% of the true data was added to the true data to define the observed data. We assumed that the measurement errors are uncorrelated, so that the covariance matrix C_D is diagonal. During simulations, all producing wells are controlled by a fixed bottomhole pressure of 2,700 psi. The injectors are controlled by a fixed bottomhole pressure of 3,200 psi.

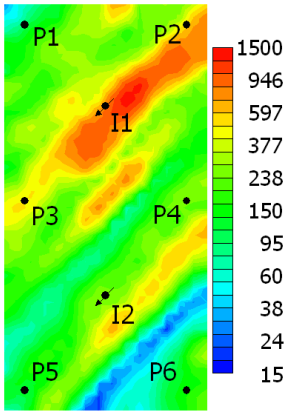


Figure 8.1: True permeability field in mD.

We assimilated data using EnKF and EnKF with covariance localization with an ensemble of 100 models. The prior ensemble was generated with the same covariance matrix and prior mean of the true model. In addition, we applied the standard

RML (with no parameterization) and the RML with the ensemble-based parameterization to generate 100 conditional models. For RML with parameterization, each conditional model is a linear combination of a different set of 100 prior models, and during minimization, the model mismatch part of the objective function was computed using Eq. 8.13. Finally, we also considered a case in which we use each model from the final ensemble of EnKF with covariance localization as an initial guess for RML with parameterization. We refer to this case as EnKF-RML. Again, each RML iteration was performed using a linear combination of a different set of 100 prior models. In order to guarantee that each model from EnKF could be exactly represented as a linear combination 100 models, we simply used the EnKF model as one of the 100 models and set the entry of the initial vector θ_j corresponding to the EnKF model to $\sqrt{N_e - 1}$. The remaining entries of the initial vector θ_j were set to zero. In this case, we expect that conditional models after RML to be close to the models from EnKF, unless the likelihood requires large corrections, which will be the case when EnKF does not provide a reasonable data match.

Table 8.1 presents the mean and standard deviation of the normalized objective function ($O_N = O/N_d$) for each case. For EnKF cases, the results presented in the Table 8.1 were obtained by rerunning the final ensembles from time zero. For this problem, the total number of data points in the history is $N_d = 378$. Hence, according the Eq. 2.132, the value of O_N should be less than 1.364 for samples of the posterior pdf. According to the results in the Table 8.1, the values of O_N obtained by the EnKF cases are, in average, four times larger than the criterion of Eq. 2.132. For the RML cases, on the other hand, the average O_N satisfy this criterion.

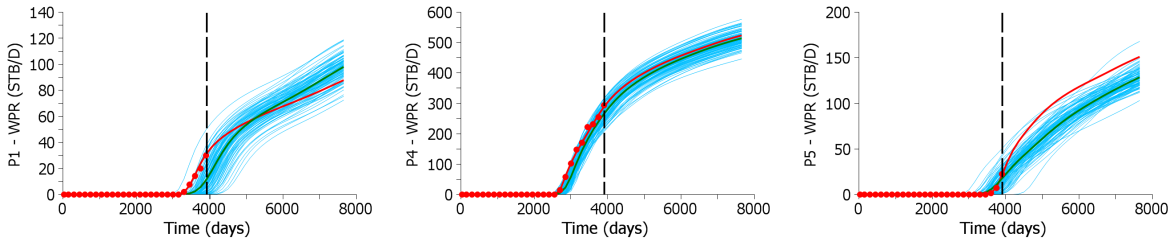
Figs. 8.2–8.6 present the predicted water rate for wells P1, P4 and P5 obtained running the final ensembles of models from time zero. In these figures, besides the 11 years of historical data, there are also ten years of forecast. According to the results presented in the Fig. 8.2, EnKF had problems matching the water breakthrough time.

Table 8.1: Normalized objective function

Case	Mean	Standard deviation
EnKF	5.770	1.287
EnKF with localization	5.256	3.517
RML	1.280	0.089
RML (ens. param.)	1.279	0.119
EnKF-RML	1.344	0.255

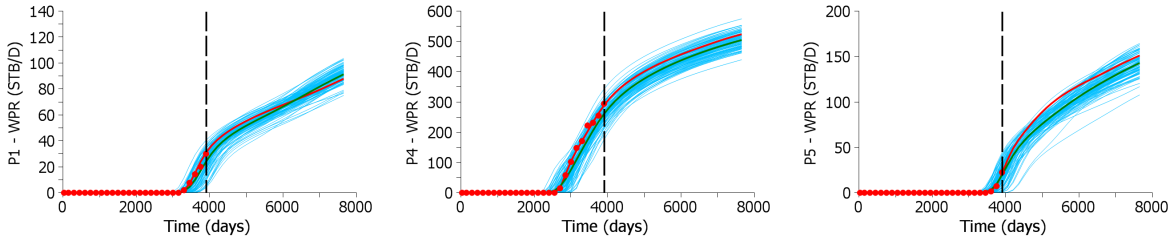
Covariance localization improved the match of water rate, but still gave a significantly large spread (mismatch) of predicted water rate during the historical period. For the RML cases, on the other hand, all models present an excellent match of water data. If we compare the spread of predicted water rate for the forecast period, we note that the EnKF cases result in similar spreads compared to the RML cases. The similarity in the spread can lead to the erroneous conclusion that EnKF and RML result in a similar characterization of uncertainty in the predicted water production rate. However, it is clear from Figs. 8.2 and 8.3 that the spreads of predicted water rate obtained by EnKF and EnKF with localization are mainly a consequence of the fact that not all models of the final ensembles present a reasonable data match. In fact, for both EnKF cases, approximately the same spread observed during the forecast occurs during the history. For the RML cases, on the other hand, all models match data very well and the spread of of predicted water rate increases in the forecast period, which is the typical behavior that one would expect from a set of independently history-matched models.

Figs. 8.7–8.11 present the first three models of the final ensembles for each of the five cases considered. Figs. 8.7 and 8.8 illustrate that EnKF and EnKF with covariance localization result in models that are very similar. This indicates that the final ensembles from EnKF cases do not correspond to independent samples of the posterior pdf. This fact is even more evident if we consider the results presented



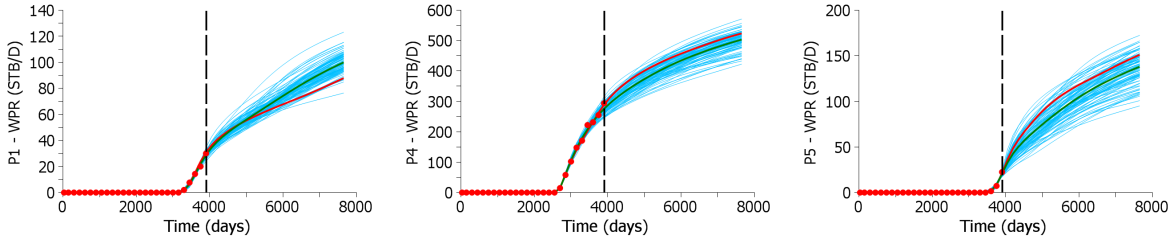
(a) P1 **(b)** P4 **(c)** P5

Figure 8.2: EnKF. Predicted water rate by running the final ensemble from time zero (stb/day). Red dots are the history; red curve is the true model; green curve is the mean prediction computed from the ensemble; and light blue curves represent all ensemble predictions. The dashed vertical black line indicates the end of the history and beginning of the forecast period.



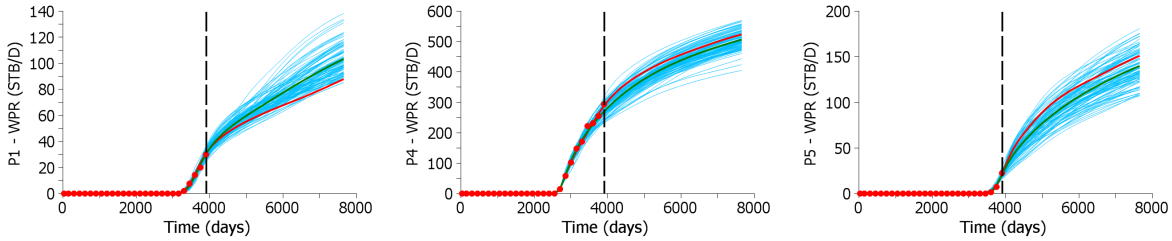
(a) P1 **(b)** P4 **(c)** P5

Figure 8.3: EnKF with localization. Predicted water rate by running the final ensemble from time zero (stb/day). The colors in this figure have the same meaning as in Fig. 8.2.



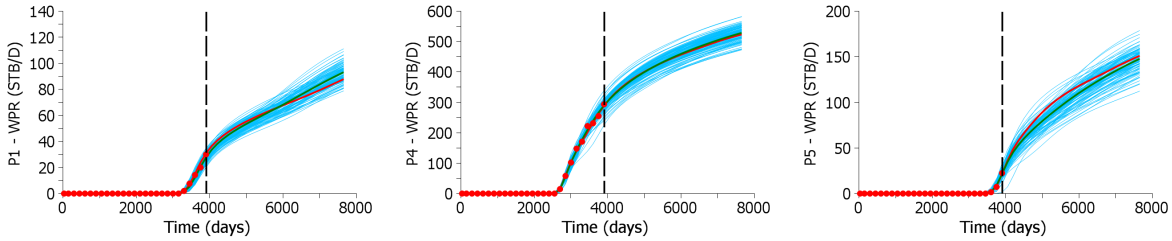
(a) P1 **(b)** P4 **(c)** P5

Figure 8.4: RML (without parameterization). Predicted water rate by running the final ensemble from time zero (stb/day). The colors in this figure have the same meaning as in Fig. 8.2.



(a) P1 **(b)** P4 **(c)** P5

Figure 8.5: RML with ensemble-based parameterization. Predicted water rate by running the final ensemble from time zero (stb/day). The colors in this figure have the same meaning as in Fig. 8.2.



(a) P1 **(b)** P4 **(c)** P5

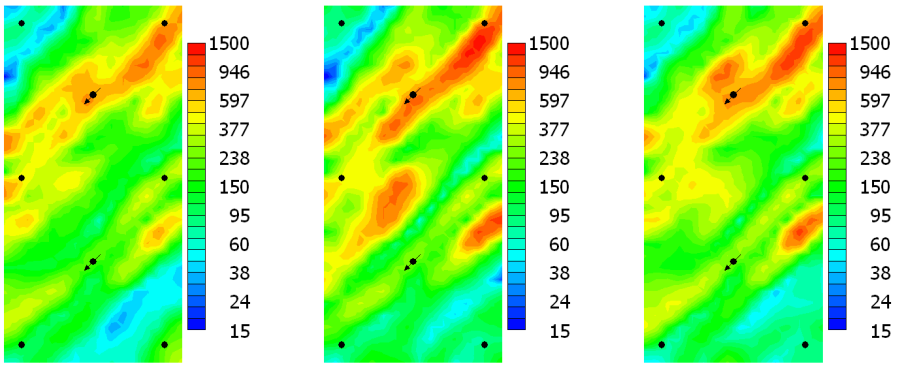
Figure 8.6: EnKF-RML. Predicted water rate by running the final ensemble from time zero (stb/day). The colors in this figure have the same meaning as in Fig. 8.2.

in the Fig. 8.12, which show the ensemble standard deviation of $\ln(k)$ after history matching. Note that the prior standard deviation of $\ln(k)$ is one for all gridblocks. According to the results in Fig. 8.12, EnKF results in very low values of standard deviation, i.e., low variability in the final ensemble, indicating that all models are essentially small variations of the same model. Covariance localization increased the variability in the final ensemble. However, the standard deviations are still much smaller than those computed from the standard RML results (Fig. 8.12c) or RML with parameterization (Fig. 8.12d). These two RML cases result in reasonably close values of standard deviation of $\ln(k)$ (compare Figs. 8.12c and 8.12d). Fig. 8.12e indicates that starting RML with the models resulting from EnKF with localization results in a standard deviation of $\ln(k)$ higher than is obtained using only EnKF with localization, but lower than is obtained using only RML. This probably happens because all initial guesses generated by EnKF are close to each other, making most of the objective function minimizations for RML converge to the same region of the posterior pdf. To further compare the variability in the final ensembles, we computed the SNV for each ensemble and the results are presented in the Table 8.2. According to the results presented in this table, standard EnKF resulted in a significant underestimation of posterior variances of $\ln(k)$ compare to RML. Covariance localization almost doubled the SNV of EnKF, but the variances are still greatly underestimated. RML with parameterization, on the other hand, resulted in a SNV reasonably close to RML without parameterization (7% difference). Starting RML from EnKF models also resulted in underestimation of posterior variances.

These results show that RML with gradient-based optimization provides far better data matching and more reliable characterization of uncertainty than EnKF. However, the computational cost of RML is much higher than EnKF. Table 8.3 presents the total number of reservoir simulations required for the RML cases. Note that each simulation also requires an adjoint solution for gradient computation. Es-

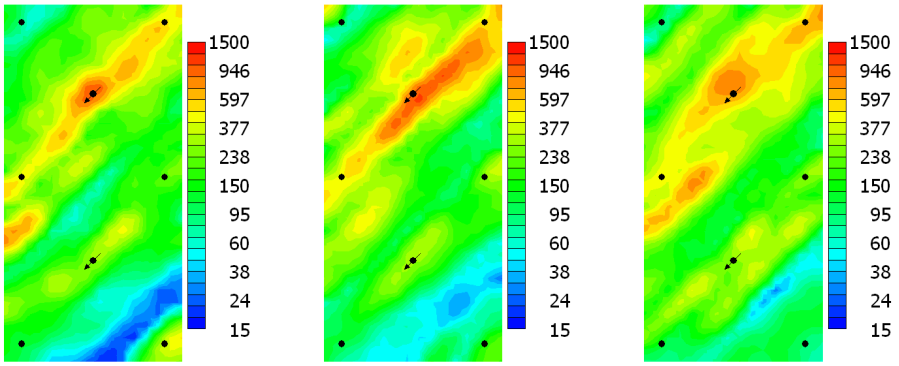
Table 8.2: Sum of the normalized variances of log-permeability.

Case	SNV
EnKF	0.132
EnKF with localization	0.232
RML	0.606
RML (ens. param.)	0.568
EnKF-RML	0.335



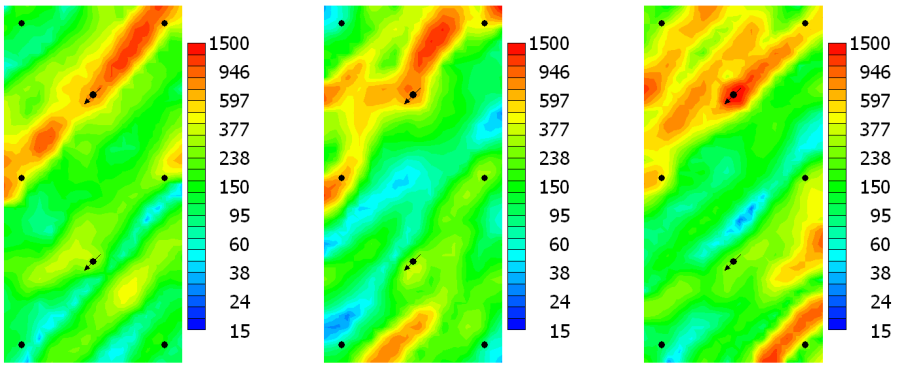
(a) Model # 1 **(b)** Model # 2 **(c)** Model # 3

Figure 8.7: First three models obtained after EnKF. Permeability fields in mD.



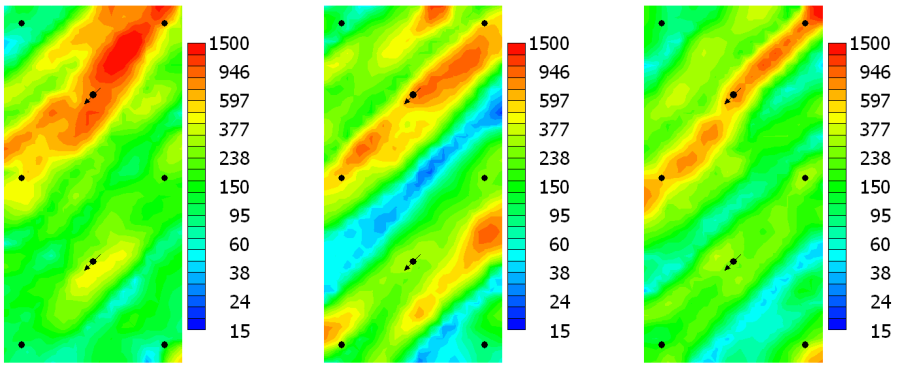
(a) Model # 1 **(b)** Model # 2 **(c)** Model # 3

Figure 8.8: First three models obtained after EnKF with covariance localization. Permeability fields in mD.



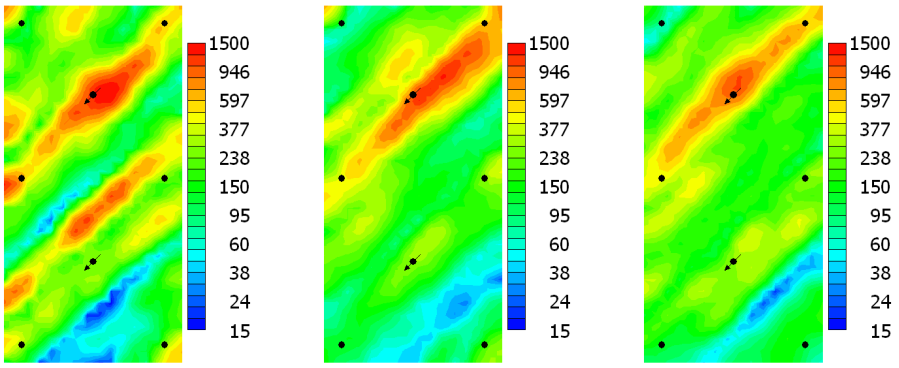
(a) Model # 1 **(b)** Model # 2 **(c)** Model # 3

Figure 8.9: First three models obtained after RML (without parameterization). Permeability fields in mD.



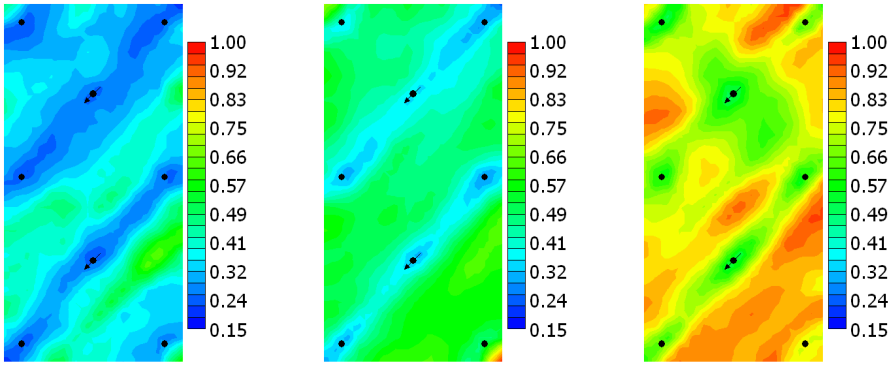
(a) Model # 1 **(b)** Model # 2 **(c)** Model # 3

Figure 8.10: First three models obtained after RML with ensemble-based parameterization. Permeability fields in mD.

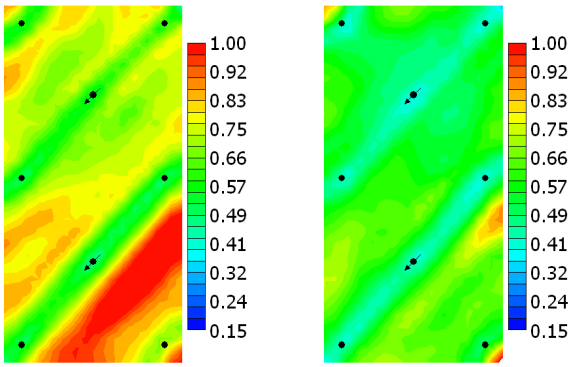


(a) Model # 1 **(b)** Model # 2 **(c)** Model # 3

Figure 8.11: First three models obtained after EnKF-RML. Permeability fields in mD.



(a) EnKF **(b)** EnKF with localization **(c)** RML



(d) RML with ens. parm. **(e)** EnKF-RML

Figure 8.12: Ensemble standard deviation of log-permeability.

sentially, the adjoint solution requires the solution of one linear system for each reservoir simulation time step. In our experience, each adjoint solution requires around 50% of the simulation time required by a forward simulation run for a fully implicit reservoir simulator. According to the results in Table 8.3, RML required on average 205 reservoir simulations (and 205 adjoint solutions) for generating each conditional realization, which correspond to a total of 20,536 reservoir simulations to generate the 100 history-matched models. This computational cost is extremely large compared to the cost of EnKF. Note that for EnKF, besides the computational cost of the reservoir simulations, there is also the additional CPU time required to restart the reservoir simulations every data assimilation step, including the time required for reading and writing files, and the CPU time required by the matrix operations of the EnKF method. In our experience, the computational cost of EnKF is around 3 to 5 times the cost of running the whole ensemble from time zero without data assimilation. This means that the computational cost of EnKF to generate 100 conditional realizations is around 300 to 500 reservoir simulations. Unfortunately, these 100 realizations obtained by EnKF are correlated and do not represent independent samples of the posterior pdf, and do not give a correct characterization of uncertainty.

The ensemble parameterization for RML resulted in a significant reduction in the required number of reservoir simulations compared to the standard RML. The average number of reservoir simulations reduced from 205 (no parameterization) to 106 (with parameterization) for generating each conditional realization. This reduction is mainly explained by the reduction in the dimension of the minimization problem from N_m to N_e. However, because we have a small reservoir model, the total number of model parameters for the case without parameterization is relatively small, $N_m = 450$. Hence, the reduction in the dimensionality of the minimization problem was not large (for this case as we used $N_e = 100$). Nevertheless, the reduction

in the required number of reservoir simulations was significant. We believe that this parameterization also introduces regularization in the objective function, which helps to improve the performance of the minimization algorithm. Note that for field case applications, a reservoir simulation model can easily have tens or hundreds of thousands of gridblocks. In these situations, the computational gain of using this parameterization can be significant. Finally, using the models obtained from EnKF with covariance localization as first guesses for RML resulted in a great reduction in the computational cost. In this case, in average, only 26 reservoir simulations were required to obtain the same level of data matches obtained with only RML. But again, it seems that this procedure does not lead to independent samples of the posterior pdf.

Table 8.3: Number of reservoir simulations for RML cases.

Case	Total	Per model	
	(100 models)	Mean	Std. dev.
RML	20,536	205	79
RML (ens. param.)	10,558	106	27
EnKF-RML	2,610	26	19

8.2.1 Discussion

Compared to RML, the main advantages of EnKF are computational efficiency and easy implementation. EnKF does not require adjoint implementation for computing gradients, which makes the method easy to adapt to different types of reservoir model parameters, different types of data and different reservoir simulators. To some extent, the intense attention that EnKF attracts in the petroleum literature is due to the fact that EnKF can easily be coupled with commercial reservoir simulators. Compared to EnKF, RML requires much more computational time, even with parameterization. However, the requirement of an adjoint implementa-

tion seems to be the main obstacle to the routine application of RML to large-field history-matching problems.

When we compare EnKF and RML in terms of the quality of the results, RML presents a much better performance. First, the values of objective function obtained by RML are much lower than the ones obtained by EnKF. In fact, the reservoir history-matching examples presented in this dissertation indicate that the final values of normalized objective function obtained by EnKF never come close to satisfying the criterion of Eq. 2.132. Large values of the objective function occur because of poor data matches, or rough models, or both. More importantly, a model which gives a high value of the objective function corresponds to a sample of a low probability region of the posterior pdf, which is clearly undesirable, especially if the objective is to generate a good sampling of the posterior pdf, in order to have a reliable characterization of uncertainty. A second important difference between the RML and EnKF results can be seem by comparing the posterior variance of the final set of realizations. Even with localization, EnKF greatly underestimates the posterior variances obtained with RML. As a result, we observe that the final ensemble from EnKF tends to underestimate the uncertainty in the model parameters.

CHAPTER 9

INVESTIGATION ON THE SAMPLING PERFORMANCE OF ENSEMBLE-BASED METHODS

As discussed in the previous chapters of this dissertation, EnKF often fails to provide reasonable data matches for highly nonlinear problems. This fact motivated the development of several iterative ensemble-based methods in the last few years. However, there exists no study comparing the performance of these methods in the literature, especially in terms of their ability to characterize uncertainty correctly. In this chapter, we compare the performance of eight ensemble-based methods in terms of the quality of the data matches, characterization of uncertainty and computational cost. For this purpose, we use a small but highly nonlinear reservoir model so that we can generate the reference posterior distribution of reservoir properties using a very long chain generated by a Markov chain Monte Carlo sampling algorithm. We consider the following ensemble-based methods: EnKF, half-iteration EnKF (HI-EnKF) [182], Lorentzen-Nævdal iterative EnKF (LN-IEnKF) [112], Krymskaya-Hanea-Verlaan iterative EnKF (KHV-IEnKF) [98], EnKF-MCMC (Chapter 5), ES [174], ES-MDA (Chapter 6) and the ensemble randomized maximum likelihood (EnRML) [70]. We also consider one adjoint-based implementation of the randomized maximum likelihood method in the comparisons. The results presented in this chapter are summarized in Emerick and Reynolds [42].

9.1 Test Problem

The test problem is a one-dimensional reservoir model under waterflooding (Fig. 9.1). The number of gridblocks is 31 and the dimensions of all gridblocks

are 50 ft $\times$ 50 ft $\times$ 50 ft. The model parameters are gridblock log-permeabilities, $\ln(k)$. The "true" permeability field (Fig. 9.2a) was generated using an exponential covariance function with a practical range corresponding to the size of 10 gridblocks. The prior mean of $\ln(k)$ is 5.0 and the prior variance equal to 1.0 for all gridblocks. The porosity is constant and equal to 0.25; the oil viscosity is 2 cp and the water viscosity 1 cp. The initial reservoir pressure is 3,500 psi and the compressibility of oil, water and rock are 10^{-5} psi^{-1}, 10^{-6} psi^{-1} and 5×10^{-6} psi^{-1}, respectively. In this synthetic reservoir, there is a water injection well in the first gridblock which is operated at a constant bottomhole pressure of 4,000 psi. In the last gridblock, there is a producing well operated at a constant bottomhole pressure of 3,000 psi. The observations correspond to gridblock pressures at a monitor well located in the center of the reservoir (gridblock number 16). The historical period corresponds to 360 days, with one pressure measurement every 30 days, which results in 12 data points. We added random Gaussian noise with mean equal to zero and standard deviation equal to 1 psi to the data predicted by the true model to define the "measurements." We chose a test case with small measurement errors to make the problem more challenging for data assimilation. The historical period was defined such that we have water breakthrough at the monitor well but not at the producing well. Fig. 9.2b shows the water saturation distribution at the end of the history (360 days) and at the end of a forecast period (750 days).

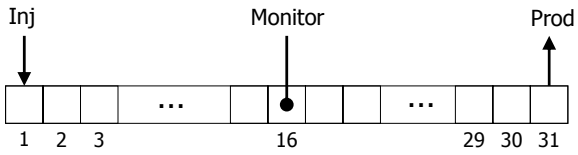
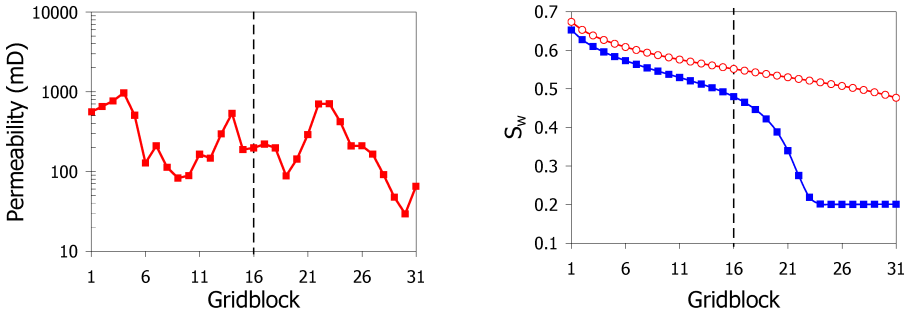


Figure 9.1: Gridblocks and well locations.

Using the same covariance function and the same prior mean used to construct the true model, we generated ten different prior ensembles of $\ln(k)$ with 100 models each. These ensembles were used in all ensemble-based methods presented in the

following sections.



(a) Permeability **(b)** Water saturation

Figure 9.2: Permeability of the true model and water saturation at the end of the historical period (in blue) and at the end of the forecast period (in red). The vertical dashed line indicates the position of the monitor well.

9.2 Markov Chain Monte Carlo

Using the Metropolis-Hastings MCMC algorithm [122, 73], we generated a very long Markov chain with 20 million proposals. We used a proposal mechanism based on local perturbations [64]. Note that each proposal requires one reservoir simulation, but because we have an extremely small problem, where each reservoir simulation takes approximately 0.2 seconds, it is feasible to generate a very long chain. Here, we assume that this chain is long enough to provide a reasonable sampling of the posterior pdf, which serves as the reference for our comparisons. Fig. 9.3 presents the values O_N for the accepted models in the chain after a transitional period (burn-in). The total number of accepted models is 458,648. This gives an acceptance rate of 0.229, which is close to the "optimal" acceptance rate of 0.234 [64, 145, 146]. In MCMC, we used a scaling factor $\sigma = 0.05$. This scaling factor was chosen after some experimentation.

Fig. 9.4a presents the distribution of permeability (percentiles P2, P25, P50, P75 and P98) obtained from MCMC. The results in this figure show a narrow distribution of permeability for the gridblocks to the left of the monitor well. The spread

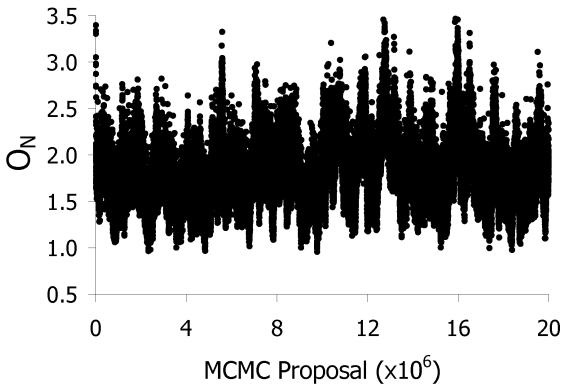
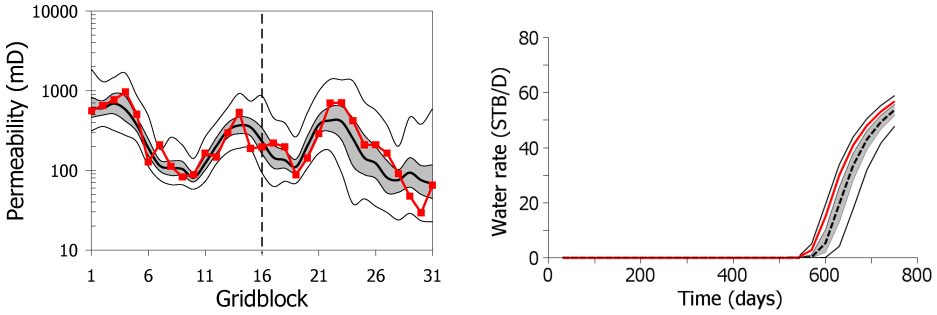


Figure 9.3: Normalized objective function of every 100th accepted model in the Markov chain.

increases for the gridblocks to the right of the monitor well. Fig. 9.4b shows the distributions (percentiles P2, P25, P50, P75 and P98) of predicted water rate at the producing well (q_w). This figures indicates that there is a relatively small uncertainty in q_w. Also, q_w obtained with the true model is above the percentile P75 obtained with MCMC.



(a) Permeability **(b)** Water production rate

Figure 9.4: Distributions of permeability and water production rate obtained from MCMC. In (a), the red curve with solid squares is the true permeability field; the solid black curve is the median; the area shaded gray corresponds to the region between the percentiles P25 and P75; the external bounding curves are the percentiles P2 and P98. In (b), the solid red curve is the prediction obtained with the true model; the dashed black curve is the median; the area shaded gray corresponds to the region between the percentiles P25 and P75. The external bounding curves are the percentiles P2 and P98.

9.3 Randomized Maximum Likelihood

We used RML to sample the posterior pdf by generating ten ensembles of 100 models each. In our implementation, we solve this minimization problem using the LBFGS algorithm [126, 192, 59] with scaling [106, 192] and bracketing cubic line search [135, Chap. 8]. We control possible overcorrection in the model by restricting the step size at early iterations. Specifically, we limit the maximum step size normalized by the 2-norm of the search direction to 0.25 until we obtain $O_N < 1{,}000$. We also apply damping to the objective function [105, 59] by artificially multiplying the standard deviation of the measurement errors by three until the minimization reduces O_N to 5.0. After that, we remove the damping and continue the minimization with the actual standard deviations of the measurement errors. The gradients are computed using the adjoint method [21, 19, 188, 105]. For convergence, we require both the change in the objective function and the change in the vector of model parameters over an iteration to be small, i.e.,

$$\frac{|O(m^{\ell+1}) - O(m^{\ell})|}{O(m^{\ell})} < 10^{-4} \tag{9.1}$$

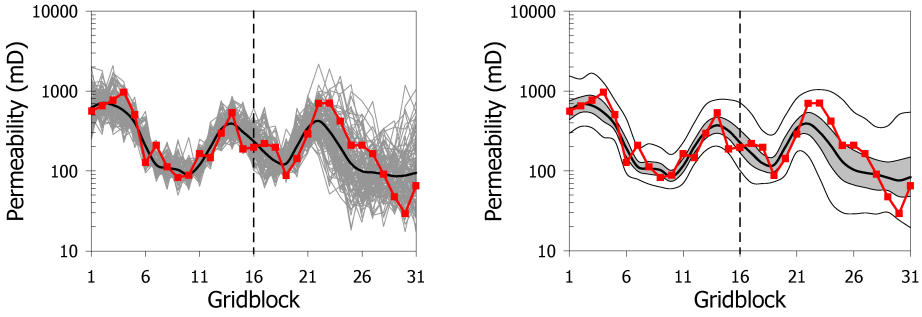
and

$$\max_{1 \le i \le N_m} \left| \frac{m_i^{\ell+1} - m_i^{\ell}}{m_i^{\ell}} \right| < 10^{-3}, \tag{9.2}$$

where the superscript ℓ denotes the iteration index.

Using RML, we sampled the posterior pdf by generating ten ensembles of 100 conditional realizations each. Fig. 9.5a presents the permeabilities of the first ensemble, and Fig. 9.5b shows the distributions of permeability obtained by combining the ten ensembles. The results in Fig. 9.5 indicate that RML obtains a distribution of permeability reasonably close to the distribution obtained from MCMC. Fig. 9.6a presents the box plots of cumulative water production (W_p) after 750 days obtained from each of the ten RML ensembles. In this figure, we also present the box plot

obtained from MCMC. Fig. 9.6a shows that RML results in distributions of W_p very close to the distribution obtained with MCMC. It is important to note that Fig. 9.6a also indicates that distributions obtained by the ten different RML runs are mutually consistent, i.e., there are not significant differences between these ten distributions. Although we did not present the results for all ten ensembles individually, the same conclusion about consistency applies to the resulting permeability fields. This consistency indicates that one ensemble of 100 models suffices for the purpose of characterizing uncertainty with RML for this test problem. Fig. 9.6b presents the distribution of q_w obtained by combining the ten RML ensembles. This figure shows that RML also results in a distribution of q_w that is in good agreement with the distribution obtained from MCMC.

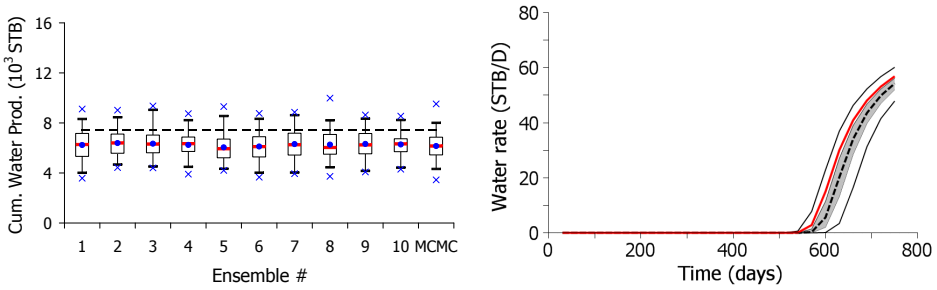


(a) First ensemble **(b)** All ensembles

Figure 9.5: Permeability after RML. In (a), the solid black curve is the ensemble mean and the gray curves are the ensemble members. In (b), the solid black curve is the median; the area shaded gray corresponds to the region between the percentiles P25 and P75; the external bounding curves are the percentiles P2 and P98. The true permeability field is included in both plots (red curve with solid squares).

9.4 Ensemble Kalman filter

Fig. 9.7 presents the permeabilities obtained for the first ensemble and the distributions obtained by combining the ten ensembles after data assimilation with standard EnKF. Compared to MCMC and RML, EnKF resulted in an unreasonably large spread in the permeability distributions. Fig. 9.8 shows the EnKF results in



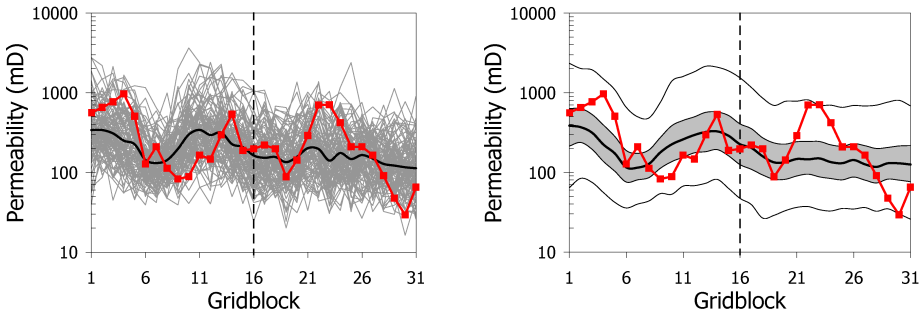
(a) Cumulative water production, W_p **(b)** Water production rate, q_w

Figure 9.6: Water production after RML. In (a), the horizontal dashed line indicates the water production obtained with the true model. In (b), the distributions obtained by combining the ten ensembles. The solid red curve is the prediction obtained with the true model; the dashed curve is the median; the area shaded gray corresponds to the region between the percentiles P25 and P75. The external bounding curves are the percentiles P2 and P98.

terms of the predicted water production obtained by running reservoir simulations from time zero with the final ensembles. According to the results in this figure, EnKF also resulted in an unreasonably large uncertainty in W_p (Fig. 9.8a) and q_w (Fig. 9.8b). Besides that, we observed that the distributions of W_p are not mutually consistent, i.e., each ensemble resulted in significantly different distributions. This problem with EnKF has already been reported in previous studies [113, 168, 38] and is related to the dependence on the prior ensemble. In fact, each EnKF run "searches for solutions" in the specific subspace spanned by the members of the particular initial ensemble for that EnKF run.

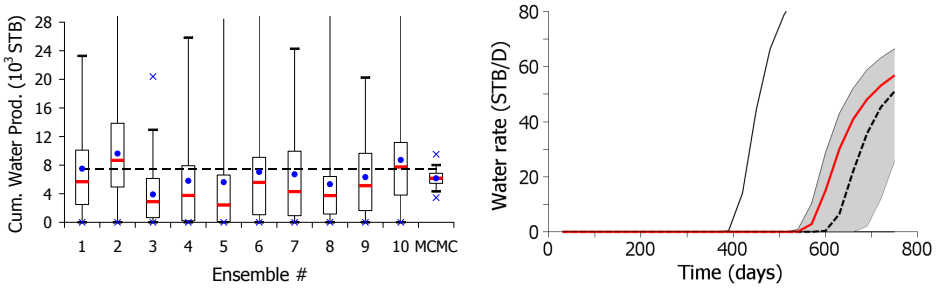
9.5 Half-iteration EnKF

As discussed in Sections 1.1.1 and 3.3.3, one problem that sometimes can occur when EnKF is used for reservoir history matching is inconsistency between the updated vectors of model parameters and states (primary variables of the reservoir simulator). For example, Fig. 9.9 presents the analyzed ensemble of water saturation (S_{w}) at 360 days (last data assimilation time-step) obtained from EnKF for the first of the ten ensembles considered in this paper. These values of S_{w} are used by



(a) First ensemble **(b)** All ensembles

Figure 9.7: Permeability after EnKF. The curves in this figure have the same meaning as in Fig. 9.5.



(a) Cumulative water production, W_p **(b)** Water production rate, q_w

Figure 9.8: Water production after EnKF. The colors have the same meaning as in Fig. 9.6.

EnKF to restart reservoir simulation for the assimilation of subsequent data. This figure shows that even though the mean S_{w} curve is in good agreement with the S_{w} profile obtained with the true model, some ensemble members will be restarted with non-physical values of S_{w}; e.g., S_{w} less than 0.2, which corresponds to the irreducible water saturation, and greater than 0.75, which corresponds to one minus the residual oil saturation. A simple fix for this problem is to use the half-iteration EnKF (HI-EnKF), in which case we simply rerun the reservoir simulator with the latest ensemble of model parameters from time zero after each data assimilation time-step.

We assimilated data using HI-EnKF and the results are presented in Figs. 9.10 and 9.11. These figures show that although HI-EnKF eliminates the parameter-state

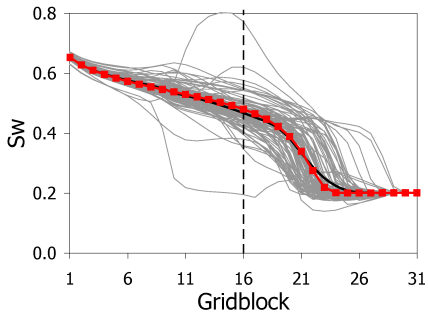


Figure 9.9: Updated ensemble of water saturation at the last data assimilation time-step with EnKF. The red curve with squares is the S_w obtained by the true model. The solid black curve is the ensemble mean and the grey curves are the ensemble members.

consistency issue, the final results are not significantly better than those obtained with the standard EnKF.

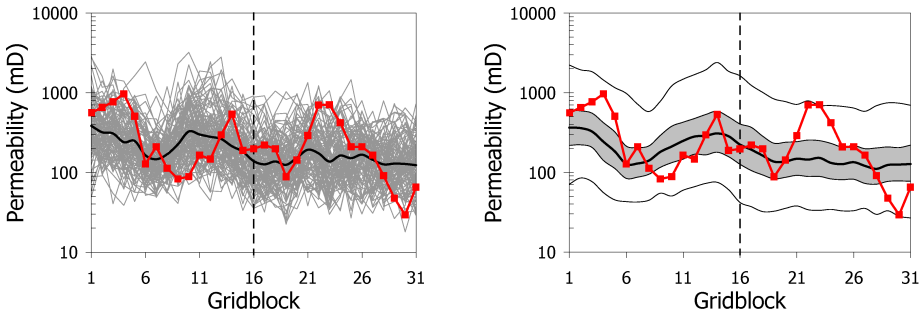
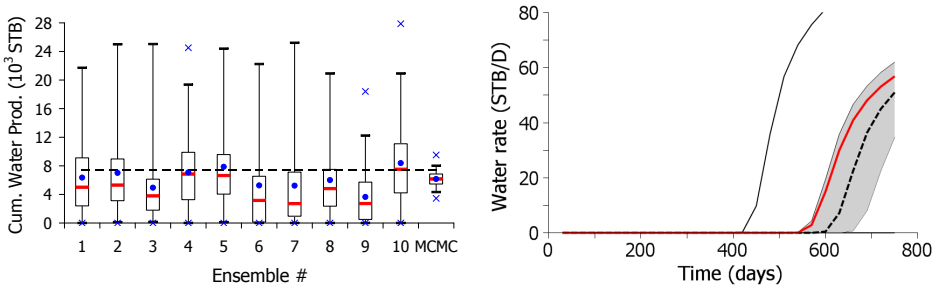


Figure 9.10: Permeability after after HI-EnKF. (a) First ensemble. (b) All ensembles. The curves in this figure have the same meaning as in Fig. 9.5.

9.6 Lorentzen-Nævdal Iterative EnKF

Figs. 9.12 and 9.13 present the results obtained after data assimilation with LN-IEnKF [112]. These figures show that LN-IEnKF did not improve the results obtained with the standard EnKF. The failure of LN-IEnKF to improve results can be explained as follows: (i) the first step of the LN-IEnKF method is the analysis of the forecast state vector using the standard EnKF; (ii) the predicted data correspond to a component of the state vector $y_j^{n,\ell}$; (iii) the measurement errors are small so that



(a) **(b)**

Figure 9.11: Water production after HI-EnKF. (a) Cumulative water production, W_p. (b) Water production rate, q_w. The colors in this figure have the same meaning as in Fig. 9.6.

$\widetilde{C}_{\mathrm{DD}}^{n,f} + C_{\mathrm{D}}^{n} \approx \widetilde{C}_{\mathrm{DD}}^{n,f}$. In this case, it is straightforward to show that after EnKF analysis we have

$$\begin{aligned} d_j^{n,a} &= d_j^{n,f} + \widetilde{C}_{\mathrm{DD}}^{n,f}\left(\widetilde{C}_{\mathrm{DD}}^{n,f} + C_{\mathrm{D}}^{n}\right)^{-1}\left(d_{\mathrm{uc},j}^{n} - d_j^{n,f}\right) \\ &\approx d_j^{n,f} + \widetilde{C}_{\mathrm{DD}}^{n,f}\left(\widetilde{C}_{\mathrm{DD}}^{n,f}\right)^{-1}\left(d_{\mathrm{uc},j}^{n} - d_j^{n,f}\right) \\ &\approx d_{\mathrm{uc},j}^{n}, \end{aligned} \tag{9.3}$$

provided that $\widetilde{C}_{\mathrm{DD}}^{n,f}$ is nonsingular, which is true for the test case because $\widetilde{C}_{\mathrm{DD}}^{n,f}$ is a 1×1 non-zero matrix at each data assimilation time-step. Hence, at the first iteration of LN-IEnKF, we have $d_j^{n,1} = h(y_j^{n,a}) = d_j^{n,a} \approx d_{\mathrm{uc},j}^{n}$. This makes the subsequent updates with Eq. 2.123 very small because $d_{\mathrm{uc},j}^{n} - d_j^{n,\ell} \approx 0$. Consequently, LN-IEnKF gives results close to those obtained with EnKF.

9.7 Krymskaya-Hanea-Verlaan Iterative EnKF

We assimilated data using KHV-IEnKF [98]. Each iteration of KHV-IEnKF corresponds to one data assimilation with EnKF. Hence, we limited the number of iterations to ten. In [98], the authors use the "confirming step" [184] procedure to avoid non-physical values of pressure and fluids saturation when restarting the

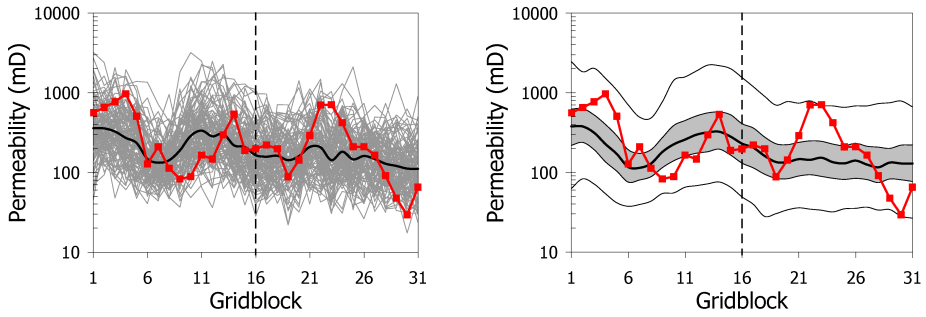


Figure 9.12: Permeability after LN-IEnKF. (a) First ensemble. (b) All ensembles. The curves in this figure have the same meaning as in Fig. 9.5.

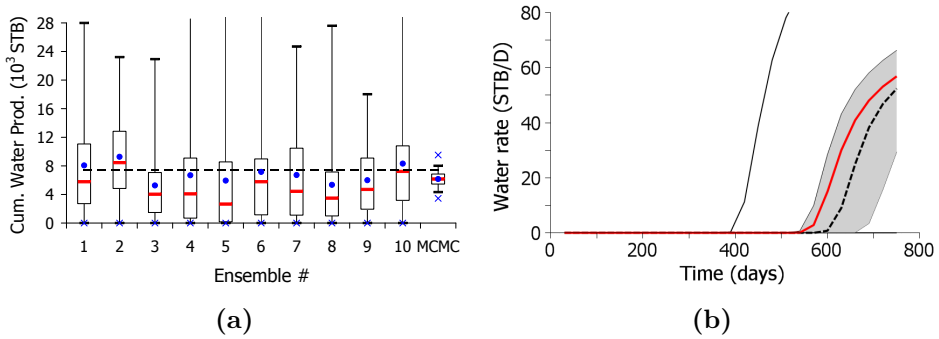


Figure 9.13: Water production after LN-IEnKF. (a) Cumulative water production, W_p. (b) Water production rate, q_w. The colors in this figure have the same meaning as in Fig. 9.6.

reservoir simulations. However, here we do not use the confirming step option because this procedure is inconsistent for the linear-Gaussian case [191]. Fig. 9.14 presents the box plots of O_N after each of the ten iterations of KHV-IEnKF. This figure shows that KHV-IEnKF does not result in a continuously decreasing O_N as a function of the iteration number; however, the iteration does improve the data match. Nevertheless, the largest decrease in O_N occurs at the first iteration, which corresponds to the first data assimilation with EnKF.

Fig. 9.15a presents the permeability obtained with KHV-IEnKF for the first of the ten ensembles. This figure shows that the mean permeability is overly rough, indicating possible overcorrections to the permeability field caused by the iterations.

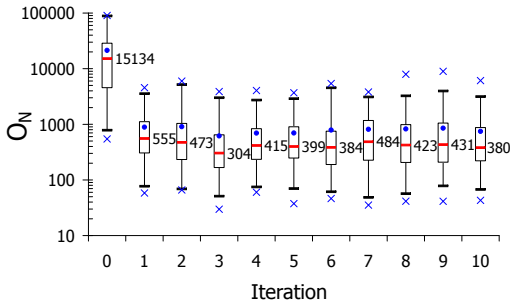


Figure 9.14: Box plots of the normalized objective function after each iteration of KHV-IEnKF (first ensemble). The numbers next to the boxes correspond to the median of O_N.

Fig. 9.15b presents the permeability distribution obtained by combining the ten ensembles. This figures shows that KHV-IEnKF resulted in a distribution very different from the reference distribution obtained with MCMC. Fig. 9.16 presents the distributions of predicted water production showing that KHV-IEnKF also resulted in a large overestimation of uncertainty.

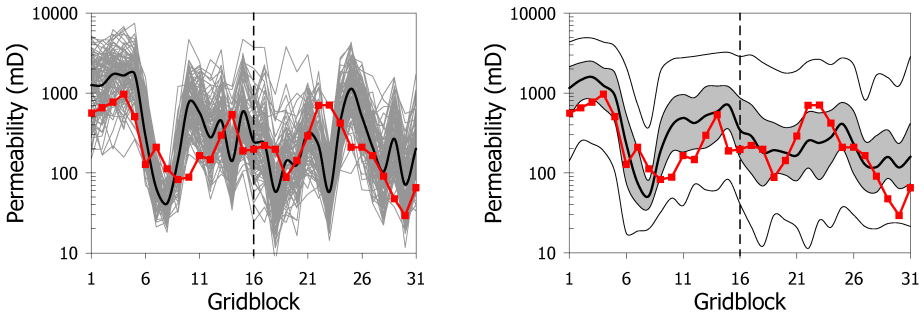
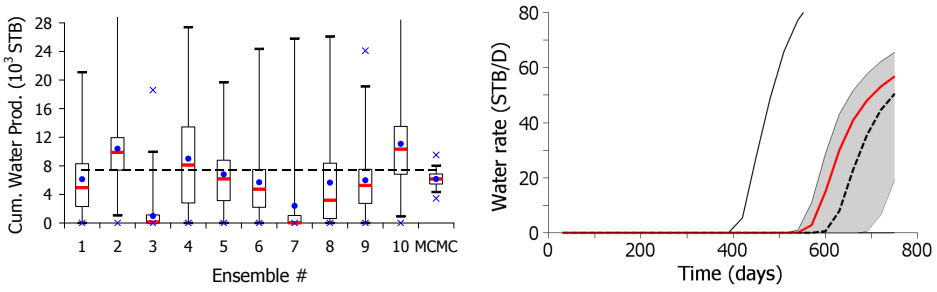


Figure 9.15: Permeability after KHV-IEnKF. (a) First ensemble. (b) All ensembles. The curves in this figure have the same meaning as in Fig. 9.5.

9.8 EnKF-MCMC

Algorithm 9.1 presents EnKF-MCMC procedure used in this test problem. The setup of the Markov chains with 5,000 proposals and $\sigma = 0.05$ was obtained a small number of experiments. The objective is to run each chain long enough to pass through the transitional (burn-in) period and choose a single model as a



(a) **(b)**

Figure 9.16: Water production after KHV-IEnKF. (a) Cumulative water production, W_p. (b) Water production rate, q_w. The colors in this figure have the same meaning as in Fig. 9.6.

sample. Fig. 9.17 presents the box plots of O_N after each of the ten iterations of EnKF-MCMC. This figure indicates that O_N is a continuously decreasing function of the iteration number, but after five iterations, the rate of decreasing is very slow.

Algorithm 9.1 EnKF-MCMC

1. Perform data assimilation using EnKF.
2. For $\ell = 1$ to 10 (where 10 is the total number of iterations):
   (a) Run the ensemble from time zero.
   (b) Build the matrix ΔY (Eq. 5.1) and compute the matrices U_r and Λ_r using SVD. These matrices are used to estimate $\widetilde{C}_{\mathrm{Y}}^{1/2}$ (Eq. 5.6).
   (c) Build $N_e = 100$ Markov chains to generate the new ensemble. For each Markov chain:
      - Start the jth chain with the jth model of the ensemble at the previous iteration.
      - Propose 5,000 states using local perturbations with $\sigma = 0.05$.
      - Keep the last accepted state as one sample.

   end (for).

In Chapter 5, we applied EnKF-MCMC with ten different initial ensembles and resampled the final models based on the value of O_N. Here, we consider the results of EnKF-MCMC before and after resampling.

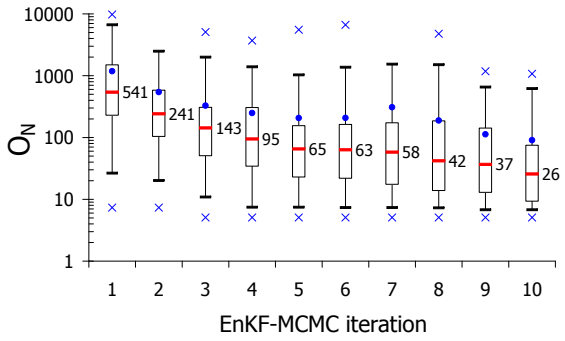
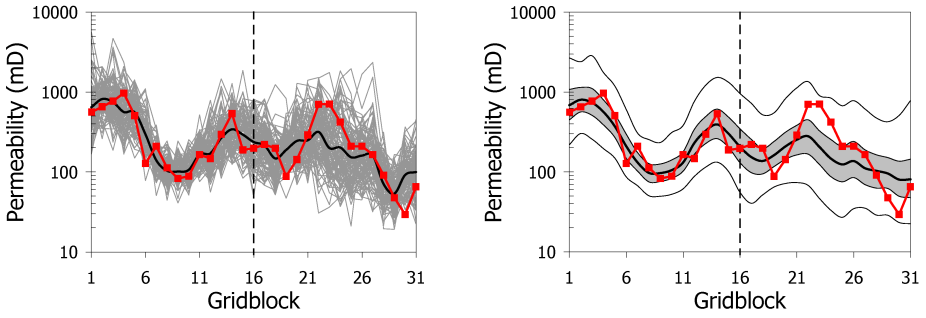
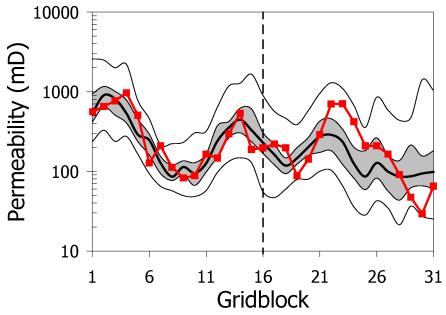


Figure 9.17: Box plots of the normalized objective function for each iteration of EnKF-MCMC (first ensemble). The numbers next to each box plot correspond to the value of the median of O_N

Fig. 9.18 presents the permeabilities obtained for the first of the ten initial ensembles after EnKF-MCMC (Fig. 9.18a) and the permeability distributions obtained by combining the ten ensembles before (Fig. 9.18b) and after (Fig. 9.18c) resampling. The resampling was based on the values of O_N (Eq. 5.11), as described in the Chapter 5. Compared to MCMC, the spreads in the permeability distributions obtained by EnKF-MCMC are slightly overestimated although the results can be considered acceptable. Resampling based on O_N did not improve significantly the permeability distributions. Fig. 9.19 shows the results in terms of water production. Fig. 9.19a shows that each ensemble resulted in significantly different distributions of W_p, and the variance of each distribution is overestimated compared to MCMC. Fig. 9.19b shows that before resampling, EnKF-MCMC greatly overestimates the uncertainty in the predicted water production. After resampling (Fig. 9.19c), the distribution of q_w is in reasonable agreement with MCMC although some overestimation is still observed.

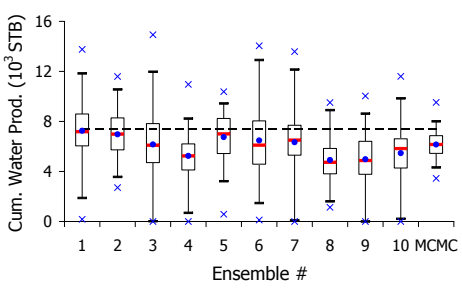


(a) First ensemble (no resampling) **(b)** All ensembles (no resampling)

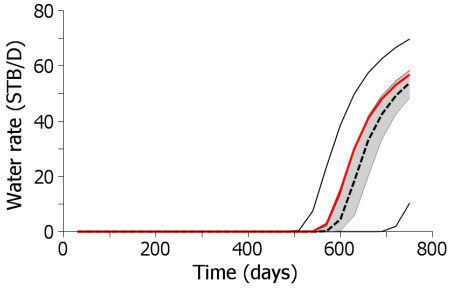


(c) All ensembles (after resampling)

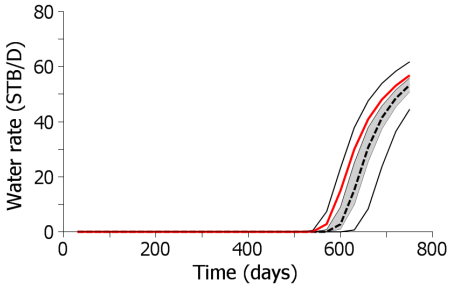
Figure 9.18: Permeability after EnKF-MCMC. The curves in this figure have the same meaning as in Fig. 9.5.



(a) Cumulative water production, W_p (no resampling)



(b) Water production rate, q_w (no resampling)

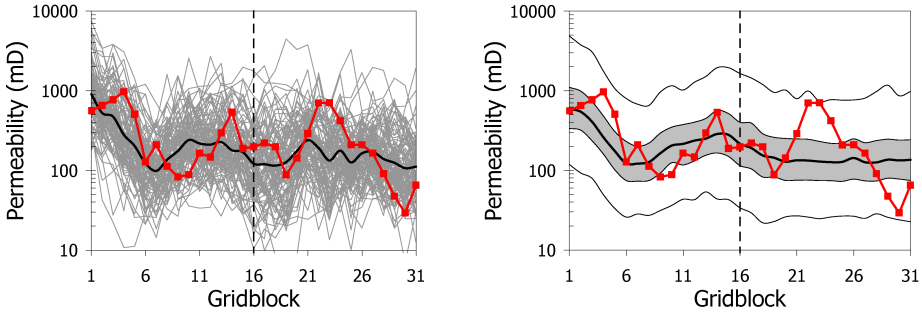


(c) Water production rate, q_w (after resampling)

Figure 9.19: Water production after EnKF-MCMC. The colors have the same meaning as in Fig. 9.6.

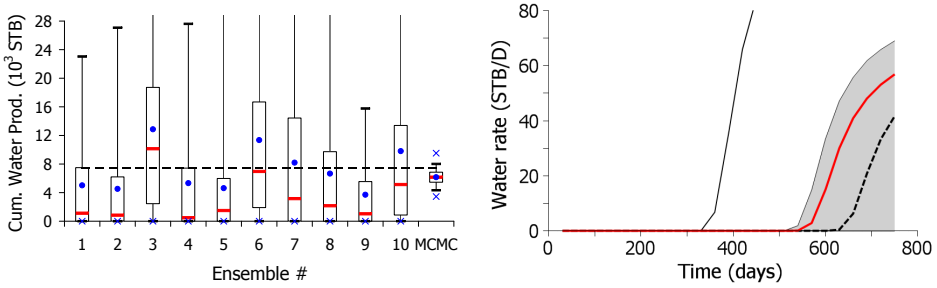
9.9 Ensemble Smoother

Figs. 9.20 and 9.21 present the results obtained after data assimilation with standard ES. For ES, the conclusions are essentially the same as the ones obtained for EnKF, i.e., a large overestimation of uncertainty and inconsistent distributions of W_p. In fact, the overestimation of uncertainty with ES is greater than with EnKF.



(a) First ensemble **(b)** All ensembles

Figure 9.20: Permeability after ES. The curves in this figure have the same meaning as in Fig. 9.5.



(a) Cumulative water production, W_p **(b)** Water production rate, q_w

Figure 9.21: Water production after ES. The colors have the same meaning as in Fig. 9.6.

9.10 Ensemble Smoother with Multiple Data Assimilation

Fig 9.22 presents the box plots of O_N for ES with one, four, eight and ten data assimilations. In these results, the following values of the coefficients α_i's were used:

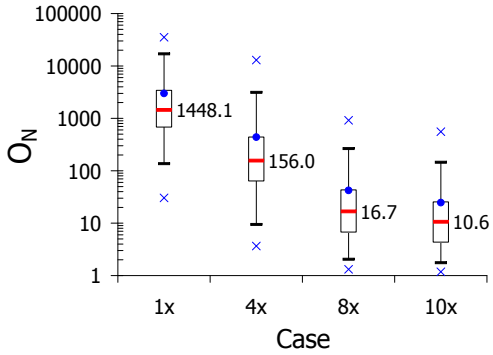
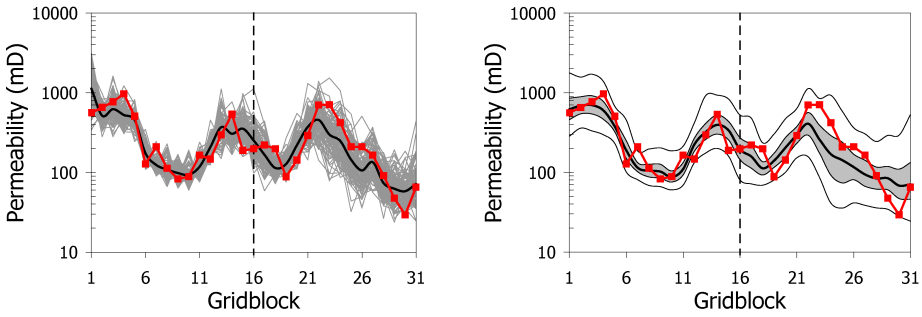


Figure 9.22: Box plots of the normalized objective function for ES-MDA. The numbers next to the boxes correspond to the median of O_N.

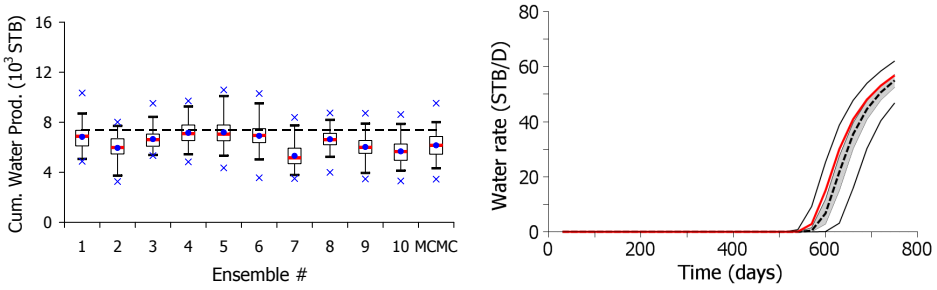
- $4\times$: $\alpha_1 = 9.333$, $\alpha_2 = 7.0$, $\alpha_3 = 4.0$ and $\alpha_4 = 2.0$.
- $8\times$: $\alpha_1 = 20.719$, $\alpha_2 = 19.0$, $\alpha_3 = 17.0$, $\alpha_4 = 16.0$, $\alpha_5 = 15.0$, $\alpha_6 = 9.0$, $\alpha_7 = 5.0$ and $\alpha_8 = 2.5$.
- $10\times$: $\alpha_1 = 57.017$, $\alpha_2 = 35.0$, $\alpha_3 = 25.0$, $\alpha_4 = 20.0$, $\alpha_5 = 18.0$, $\alpha_6 = 15.0$, $\alpha_7 = 12.0$, $\alpha_8 = 8.0$, $\alpha_9 = 5.0$ and $\alpha_{10} = 3.0$.

We have experimented with different coefficients, but we did not obtain any results significantly different from those shown here. According to the results in Fig. 9.22, after ten data assimilations, the median of O_N is reduced from 1,448 to 10.6. Fig. 9.23 presents the permeabilities obtained by ES-MDA ($10\times$). Unlike EnKF and ES, ES-MDA resulted in a fairly low spread of permeabilities, which are in reasonable agreement with the MCMC results. In terms of W_p (Fig. 9.24a), each ES-MDA ($10\times$) ensemble resulted in distributions with roughly the same variance obtained from MCMC. However, there is still an inconsistency problem in the distributions obtained for different ensembles, compare, for example, the box plots for the fourth and seventh ensembles in Fig. 9.24a. This means that a single ensemble is not enough to obtain a reliable characterization of uncertainty. Combining the results of multiple ensembles, on the other hand, resulted in a distribution of q_w very close to the one obtained with MCMC.



(a) First ensemble **(b)** All ensembles

Figure 9.23: Permeability after ES-MDA (10×). The curves in this figure have the same meaning as in Fig. 9.5.



(a) Cumulative water production, W_p **(b)** Water production rate, q_w

Figure 9.24: Water production after ES-MDA. The colors have the same meaning as in Fig. 9.6.

9.11 Ensemble Randomized Maximum Likelihood

The original EnRML method assimilates data sequentially in time. Here, we refer to this method as EnRML-F, where "F" stands for filter. [24] proposed to use EnRML as a smoother, in which case, all data is assimilated simultaneously. We refer to this procedure as EnRML-S. The general EnRML procedure was presented in Chapter 2 (Section 2.7.4). This procedure is essentially the same for EnRML-F and EnRML-S; the only essential difference is that for EnRML-S, the procedure presented in Section 2.7.4 is applied only once for the entire historical period. Here, we used the following termination criteria:

- $\max |m_{i,j}^{\ell+1} - m_{i,j}^{\ell}| < 10^{-5}$ for $i = 1, 2, \ldots, N_m$ and $j = 1, 2, \ldots, N_e$.

- $\left|\frac{\overline{O}_d^{\ell+1}-\overline{O}_d^{\ell}}{\overline{O}_d^{\ell}}\right| < 10^{-4}$.
- Maximum number of iterations = 10.
- Maximum number of step size cuts = 5.

The first two termination criteria were chosen as the same as used in [70].

The estimate $\widetilde{G}$ of the sensitivity matrix is clearly the main approximation introduced in the EnRML method. Fig. 9.25 presents the values of $\widetilde{G}$ calculated for the first of the ten initial ensembles and the actual sensitivity matrix, G, computed using the adjoint method evaluated at the prior mean. According to Fig. 9.25, $\widetilde{G}$ is very noisy. However, if we consider the product $\widetilde{C}_{\mathrm{M}}^{f}\widetilde{G}^{\mathrm{T}}$ (Fig. 9.26), we observe a smoother behavior and a qualitative agreement with the actual product, $C_{\mathrm{M}}G^{\mathrm{T}}$, computed with the adjoint method. Note that, for EnRML, we computed $\widetilde{C}_{\mathrm{M}}^{f}\widetilde{G}^{\mathrm{T}}$ using the ensemble approximation for C_{M}^{f} while for the adjoint case, we used the correct prior covariance matrix C_{M}. Figs. 9.25 and 9.26 illustrate that while the $\widetilde{G}$ estimated from EnRML using Eq. 2.128 is highly inaccurate, the resulting product $\widetilde{C}_{\mathrm{M}}^{f}\widetilde{G}^{\mathrm{T}}$ is reasonably similar to the true $C_{\mathrm{M}}G^{\mathrm{T}}$. In particular, each entry of $\widetilde{C}_{\mathrm{M}}^{f}\widetilde{G}^{\mathrm{T}}$ has the same sign as the corresponding entries of the true $C_{\mathrm{M}}G^{\mathrm{T}}$. In essence, [24] made the same observation, but we provide a simple theoretical explanation of why this occurs. In estimating $\widetilde{C}_{\mathrm{M}}^{f}\widetilde{G}^{\mathrm{T}}$, we use

$$\begin{aligned}\widetilde{C}_{\mathrm{M}}^{f} &= \frac{1}{N_e-1}\sum_{j=1}^{N_e}\left(m_j^f-\overline{m}^f\right)\left(m_j^f-\overline{m}^f\right)^{\mathrm{T}} \\ &= \frac{1}{N_e-1}\Delta M^f\left(\Delta M^f\right)^{\mathrm{T}},\end{aligned} \tag{9.4}$$

and from Eq. 2.128, we have

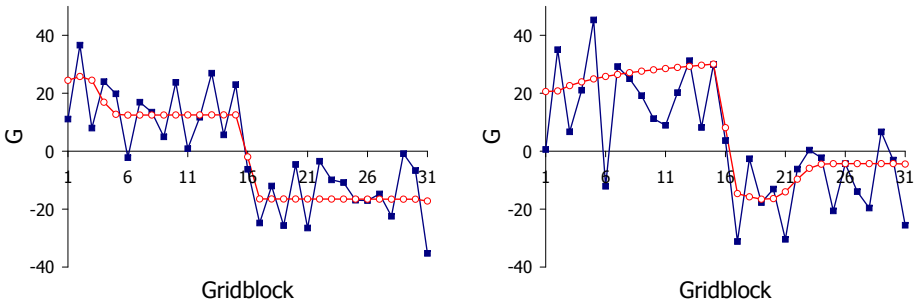
$$\left(\Delta D^{\ell}\right)^{\mathrm{T}} = \left(\Delta M^{\ell}\right)^{\mathrm{T}}\widetilde{G}_{\ell}^{\mathrm{T}}. \tag{9.5}$$

For the first iteration of EnRML, we have $\Delta M^{\ell} = \Delta M^{f}$ and $\Delta D^{\ell} = \Delta D^{f}$. Thus, multiplying $\widetilde{C}_{\mathrm{M}}^{f}$ by $\widetilde{G}_{\ell}^{\mathrm{T}}$, we obtain

$$\begin{aligned} \widetilde{C}_{\mathrm{M}}^{f}\widetilde{G}_{\ell}^{\mathrm{T}} &= \frac{1}{N_e - 1}\Delta M^{f}\left(\Delta M^{f}\right)^{\mathrm{T}}\widetilde{G}_{\ell}^{\mathrm{T}} \\ &= \frac{1}{N_e - 1}\Delta M^{f}\left(\Delta D^{f}\right)^{\mathrm{T}} \\ &= \widetilde{C}_{\mathrm{MD}}^{f}. \end{aligned} \tag{9.6}$$

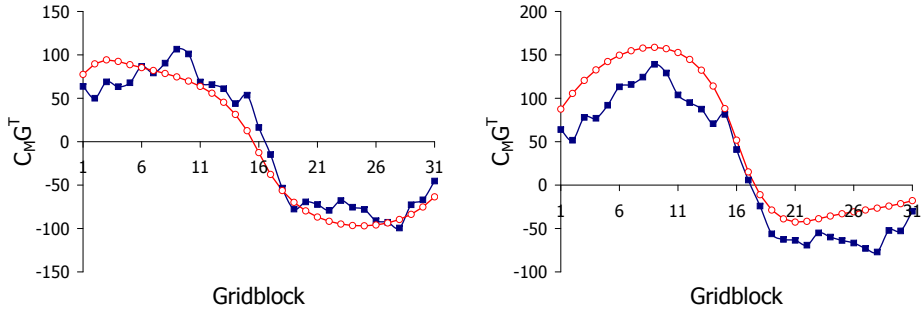
Therefore, the $\widetilde{C}_{\mathrm{M}}^{f}\widetilde{G}_{\ell}^{\mathrm{T}}$ used in the first EnRML iteration is equal to $\widetilde{C}_{\mathrm{MD}}^{f}$ used in EnKF and ES. Because Reynolds et al. [143] showed that $\widetilde{C}_{\mathrm{MD}}^{f}$ can be approximated by $C_{\mathrm{M}}^{f}\overline{G}_{\ell}^{\mathrm{T}}$, where $\overline{G}_{\ell}$ is the sensitivity matrix evaluated at $\overline{m}^{f}$, it follows that $\widetilde{C}_{\mathrm{M}}^{f}\widetilde{G}_{\ell}^{\mathrm{T}} \approx C_{\mathrm{M}}^{f}\overline{G}_{\ell}^{\mathrm{T}}$ so that the approximation $\widetilde{C}_{\mathrm{M}}^{f}\widetilde{G}_{\ell}^{\mathrm{T}}$ used in Eq. 2.126 should be reasonably accurate. Thus, the most unreliable approximation in Eq. 2.126 is the product $\widetilde{G}_{\ell}(m_j^{\ell} - m_j^{f})$. However, the inaccuracy in $\widetilde{G}_{\ell}$ (Fig. 9.25) apparently is not bad enough to destroy the utility of the method.

In EnRML, all ensemble members are updated with the same search direction and step size. Unfortunately, there is no guarantee that the search direction is downhill or the same step size is appropriate for all ensemble members. Fig. 9.27 illustrates this fact for EnRML-S. In Fig. 9.27, we present the values of the normalized data mismatch objective function ($O_{N,d}$) for the first iteration considering different step sizes. Fig. 9.27 shows that a step size $\beta_0 = 1$ corresponds to a decrease in the average $O_{N,d}$. Therefore, this step is accepted and used to update all ensemble members. However, Fig. 9.27 indicates that a step size of unity does not always yield a decrease in $O_{N,d}$ for all models. For example, $O_{N,d}$ increases for the third and 13th ensemble members. In fact, the search direction is not even downhill for the 13th ensemble member. A possible way to ameliorate this problem is to perform a different line search for each ensemble member as suggested by Wang et al. [182].



(a) 30 days **(b)** 360 days

Figure 9.25: Sensitivity obtained from EnRML (in dark blue) and the adjoint method (in red).



(a) 30 days **(b)** 360 days

Figure 9.26: Product of the prior covariance and transpose of sensitivity matrix obtained from EnRML (in dark blue) and the adjoint method (in red).

However, here, we tested only the original EnRML procedure proposed by Gu and Oliver [70].

We first considered the sequential formulation of EnRML (EnRML-F). However, we could not obtain reasonable data matches or estimates of the permeability field with this method. In our tests, EnRML-F failed to update model parameters for some consecutive data assimilation time-steps because five cuts in the step size failed to give a decrease in $\overline{O}_d^{\ell+1}$. In other data assimilation time-steps, EnRML-F resulted in apparent overcorrections of the permeability field. In most cases, these overcorrections occur after the 8th data assimilation time-step, which corresponds to the water breakthrough in the monitor well. We tried different combinations of step sizes and termination criteria. We even tried to use a Levenberg-Marquardt update

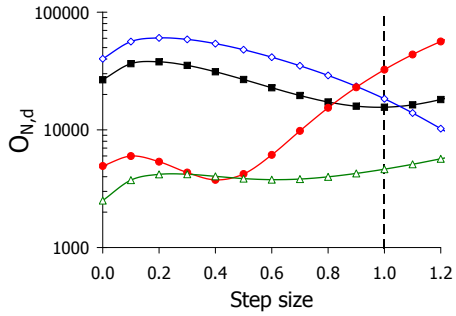


Figure 9.27: Normalized data mismatch objective function versus step size for the first iteration of EnRML-S. This figure presents the values of $O_{N,d}$ for the mean (in black), first (in blue), third (in red) and 13th (in green) ensemble members. The vertical dashed line indicates the full step.

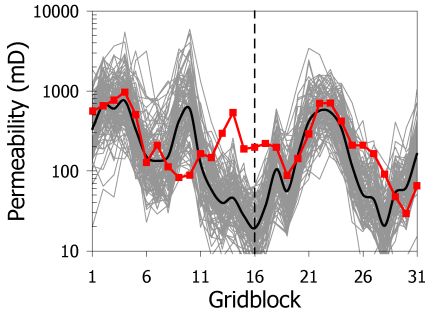
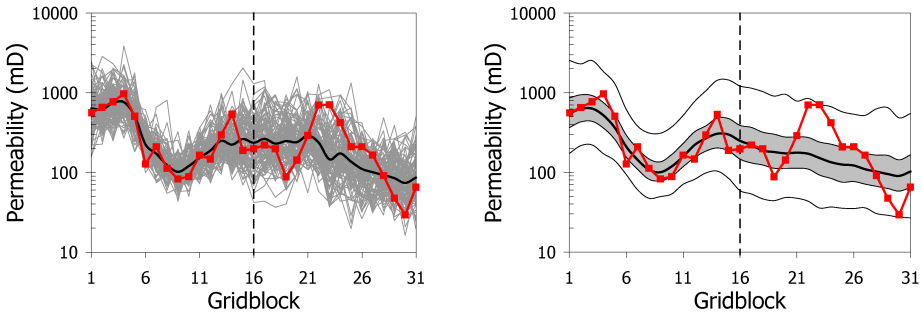


Figure 9.28: Permeability fields after EnRML-F for the first ensemble.

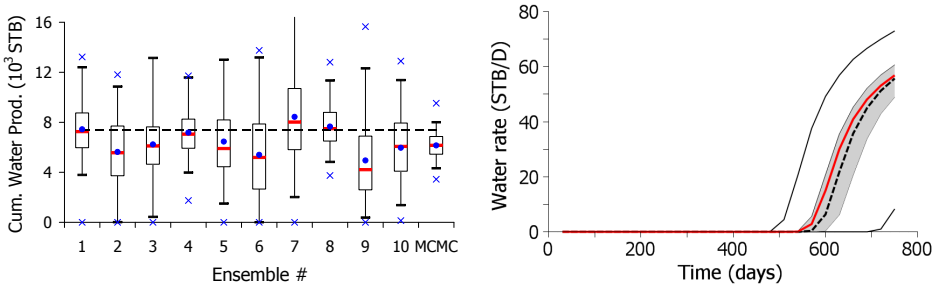
equation, instead of Gauss-Newton, to control the overcorrection issue. Despite these efforts, we could not match data with EnRML-F. Fig. 9.28 shows the final permeability field obtained for the first ensemble with EnRML-F. This figure illustrates the overcorrections obtained in the permeability field close to the monitor well location.

For EnRML-S, on the other hand, we obtained more reasonable estimates of the permeability field. No overcorrection was observed in this case. Fig. 9.29a presents the resulting permeability fields obtained by the first ensemble with EnRML-S while Fig. 9.29b presents the distribution of permeabilities obtained by combining the ten ensembles. Compared to MCMC, EnRML-S resulted in an acceptable permeability distribution although some overestimation in the spread is observed. In terms of predicted water production (Fig. 9.30), EnRML-S resulted in significant overestimation of uncertainty. Similar to the other ensemble-based methods, EnRML-S also



(a) First ensemble **(b)** All ensembles

Figure 9.29: Permeability fields after EnRML-S. The curves in this figure have the same meaning as in Fig. 9.5.



(a) Cumulative water production, W_p **(b)** Water production rate, q_w

Figure 9.30: Water production after EnRML-S. The colors have the same meaning as in Fig. 9.6.

obtained inconsistent distributions of W_p when repeating the data assimilation with ten different initial ensembles. Among the ten EnRML-S runs, eight stopped because of the maximum number of iterations; i.e., after ten iterations. This indicates that perhaps some improvements in the results of EnRML-S can be achieved if we allow more iterations. However, we limited the number of iterations to make the methods comparable in terms of computational cost. We also tried EnRML-S with an initial step-size $\beta_0 = 0.5$, but we did not obtain significantly different results from the ones presented here, which use $\beta_0 = 1$.

9.12 Overall Comparison

Fig. 9.31 presents the values of the variance of log-permeability, var[ln(k)],

after data assimilation for all methods considered in this paper. In order to make this figure clear, we divided the results from the methods into two plots. We assume that MCMC results are correct, so we present the MCMC results in both plots of Fig. 9.31. According to the results in Fig. 9.31a, EnKF, HI-EnKF, LN-IEnKF, KHV-IEnKF and ES gave unreasonably high values of var[ln(k)]. There are almost no differences between the var[ln(k)] obtained with EnKF, HI-EnKF and LN-IEnKF so that the corresponding curves overlap in Fig. 9.31a. Among all methods, KHV-IEnKF resulted in the largest overestimation of var[ln(k)]. According to the results in Fig. 9.31a, KHV-IEnKF obtained var[ln(k)] values larger than one for most of the reservoir gridblocks, i.e., variances higher than the prior variance. Fig. 9.31b shows that, compared to each other, EnKF-MCMC and EnRML-S provided nearly identical values of var[ln(k)] for the gridblocks to the left of the monitor well and fairly similar variance values for the gridblocks to the right of the monitor well. However, EnKF-MCMC and EnRML-S overestimated the presumably correct variances obtained from MCMC although to a much lesser extent than the methods presented in Fig. 9.31a. The best results were obtained by ES-MDA and RML (Fig. 9.31b). These two methods resulted in values of var[ln(k)] very close to the ones obtained from MCMC for all gridblocks. Table 9.1 presents the same results in terms of the sum of normalized variance (SNV) of log-permeability. This table shows that EnKF, HI-EnKF, LN-IEnKF, KHV-IEnKF and ES largely overestimate the value of SNV obtained from MCMC. EnKF-MCMC and EnRML-S also overestimate the SNV obtained from MCMC significantly. ES-MDA, on the other hand, resulted in a value of SNV closer to MCMC, but with an underestimation of approximately 15%. RML resulted in a value of SNV very close to the one obtained from MCMC.

Fig. 9.32 presents the values of the variance of the oil production rate, var[q_o], after data assimilation for all methods considered in this paper. We divided the results into three plots in order to make the figure clear and included the results from

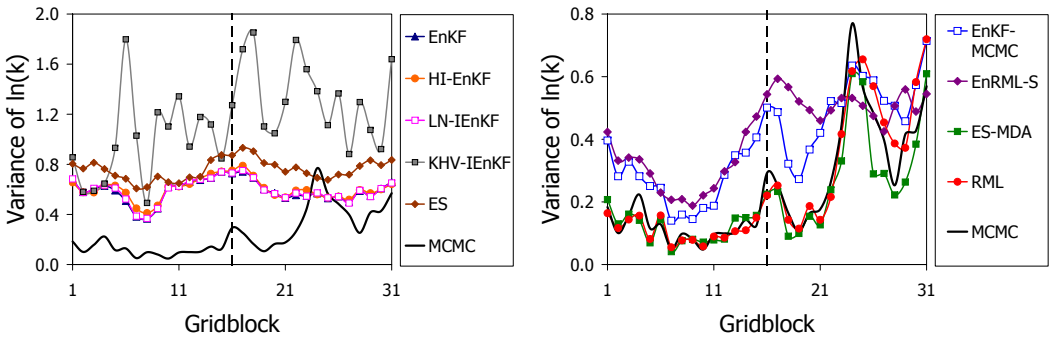


Figure 9.31: Variance of log-permeability. Note that the vertical scale is different in each plot.

Table 9.1: Sum of normalized variances of log-permeability.

Case	SNV
EnKF	0.590
HI-EnKF	0.604
LN-IEnKF	0.592
KHV-IEnKF	1.161
EnKF-MCMC	0.387
ES	0.758
ES-MDA	0.209
EnRML-S	0.412
RML	0.248
MCMC	0.242

MCMC in all plots for comparison. EnKF, HI-EnKF, LN-IEnKF, KHV-IEnKF and ES resulted in unreasonably high values of var$[q_o]$ (Fig. 9.32a). EnKF-MCMC and EnRML-S overestimated the var$[q_o]$ obtained from MCMC (Fig. 9.32b). However, after resampling, EnKF-MCMC resulted in var$[q_o]$ fairly close to the variance obtained from MCMC although some overestimation is still observed for the forecast period. ES-MDA and RML resulted in values of var$[q_o]$ in very good agreement with the results obtained from MCMC (Fig. 9.32c).

Fig. 9.33 presents the box plots of O_N for all methods considered in this paper. In this figure, we sorted the methods in a decreasing order of O_N. In this test problem,

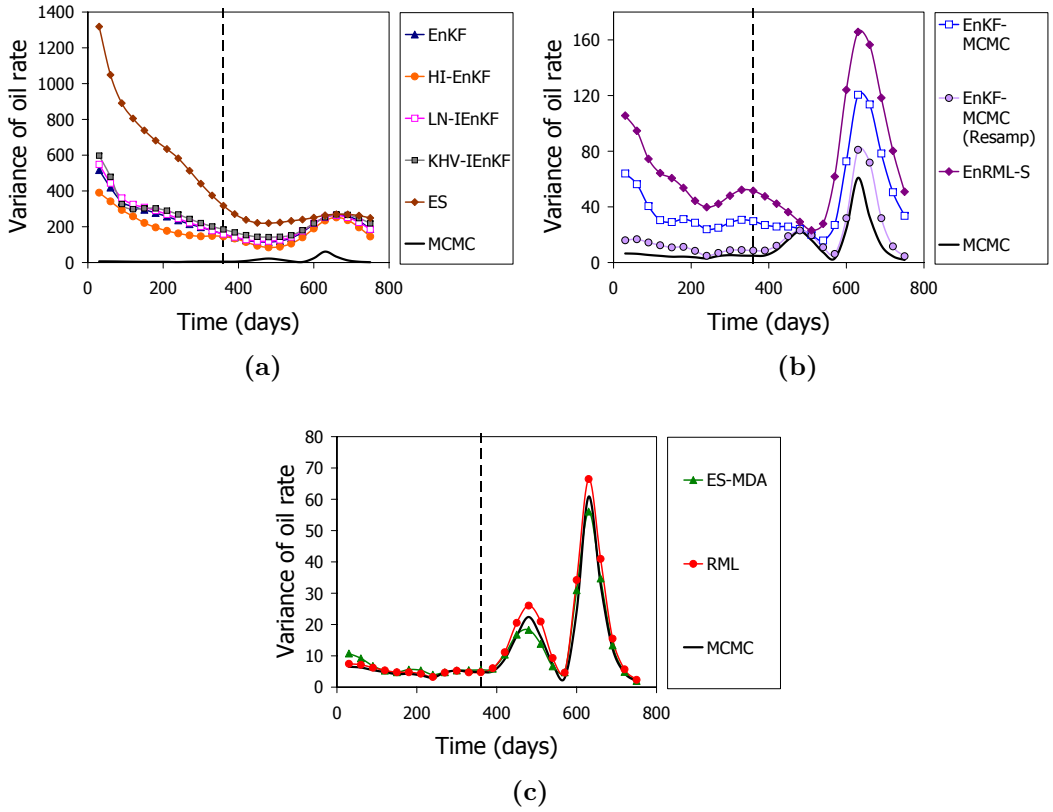


Figure 9.32: Variance of oil production rate in $\text{stb}^2/\text{day}^2$. The vertical dashed line indicates the end of the period with observed data. Note that the vertical scale is different in each plot.

we have 12 measurements and according to the criterion of Eq. 2.132, the values of O_N should be less than 3.06. However, only RML and MCMC were able to satisfy this criterion. All other ensemble-based methods resulted in significantly higher values of the objective function. High values of the objective function are associated with poor data matches. More importantly, a model which results in high value of the objective function gives a small value of the posterior pdf, which suggests that this model is a sample from a low probability region. Among the ensemble-based methods, ES-MDA obtained the lowest values of O_N. ES-MDA is also the ensemble-based method which gives the best characterization of uncertainty. In fact, there is a correlation between the reliability of the uncertainty quantification and the final values of O_N obtained by the ensemble-based methods. Note that EnKF, HI-EnKF, LN-IEnKF,

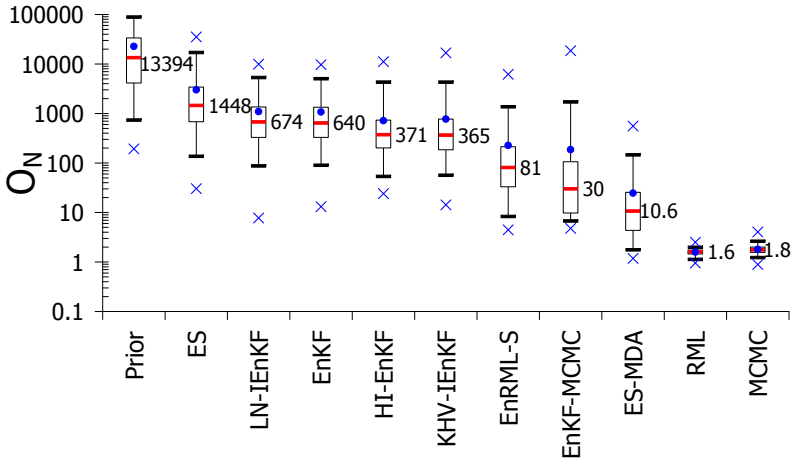


Figure 9.33: Box plots of the normalized objective function. The numbers next to each box plot correspond to the value of the median of O_N.

KHV-InEnKF and ES obtained very high values of O_N and resulted in unreasonably large overestimations of uncertainty. EnKF-MCMC, EnRML-S, ES-MDA and RML obtained better data matches and more reliable uncertainty quantification.

Table 9.2 presents the estimated computational cost to generate an ensemble of 100 realizations with each method. However, because we have a very small simulation model, which requires about 0.2 seconds to run, the relative cost of writing/reading simulation files and matrix operations during the data assimilations becomes relatively important. Note that this is not the typical situation in reservoir history-matching problems where the CPU time required by the reservoir simulation largely dominates the total time of the data assimilation. For this reason, in Table 9.2, we present the computational time in terms of the total number of reservoir simulation runs and in terms of the actual measured CPU time divided by the measured CPU time of the ES method, which is the fastest method. Note that in the results of Table 9.2, we count as one equivalent simulation run the simulation of the total historical period; hence, we count 100 simulation runs for EnKF, even though these simulations require several restarts. According to the results in Table 9.2, ES is the fastest method in terms of both the number of reservoir simulation runs and

the measured CPU time. Although EnKF also requires only 100 reservoir simulations, the total CPU time is 7.5 larger than ES mainly because of the simulation restarts. HI-EnKF requires running each reservoir simulation from time zero after each data assimilation, which makes the equivalent number of reservoir simulations 6.5 times higher than the number required by EnKF. For LN-IEnKF, the iterative process does not require additional reservoir simulations, but the total CPU time is longer than EnKF because of the additional matrix operations required during the iterations. The computational cost of KHV-IEnKF corresponds to ten data assimilations with EnKF and ten reruns of the ensemble. For EnKF-MCMC, the total CPU time includes the time required for one data assimilation with EnKF, ten reruns of the ensemble and the time required to generate the Markov chains. This makes the computational cost of EnKF-MCMC approximately 42 times the cost of data assimilation with ES. ES-MDA with ten data assimilations requires ten times the computational cost of ES. For EnRML-S, the iterative process often requires cutting the step size and rerunning the ensemble, which results in an average of 2,100 reservoir simulations per data assimilation and a total CPU time 37.8 times greater than that used with ES. For RML, each sample of the posterior pdf requires solving one minimization problem. In our implementation, each minimization requires, on average, 245 reservoir simulations. Besides that, during RML, each simulation requires the solution of an adjoint problem to compute gradients. Here, each adjoint solution requires on the order of 50% of the time required for a simulation run. Therefore, the final CPU time of RML is 350 times the CPU time of ES. The computational cost of RML seems extremely high for this test problem. However, the final objective function values obtained with RML are around 700 times lower than those from ES (the fastest method) or four times lower than those from ES-MDA (the ensemble-based method with the best performance). Moreover, RML was the only method with consistent sampling results when we repeat the data assimilation ten times.

This means that for RML, it would be sufficient to generate a single ensemble of 100 realizations. For MCMC, we ran a very long chain with 20 million proposals. We did not measure the actual CPU time; however, it is clear that the direct application of MCMC for any reasonably sized reservoir model is not computationally feasible.

Table 9.2: Estimated computational cost to generate an ensemble of 100 realizations.

Method	Equivalent number of simulation runs	Normalized CPU time
EnKF	100	7.5
HI-EnKF	650	12.9
LN-IEnKF	100	9.9
KHV-IEnKF	2,000	84.9
ES	100	1.0
ES-MDA	1,000	10.0
EnKF-MCMC	1,100	41.9
EnRML-S	2,100	37.8
RML	24,500	350.1
MCMC	2×10^6	–

The data set used in this paper, including the reservoir simulator, the true permeability field, the ten initial ensembles and the results from the long Markov chain are available for download at `http://www.tuprep.utulsa.edu/comparative_study.html`. The objective of this data set is to allow other research groups to reproduce the results in this chapter, test their own implementations and extend the comparative study to other methods.

CHAPTER 10

CONCLUSIONS

The main objective of this dissertation was to investigate the use of ensemble-based methods for history matching and uncertainty characterization in petroleum reservoir models. The research focused on the limitations of these methods and on proposing new improvements. In the following sections, we present the specific conclusions for each chapter of the dissertation.

10.1 Covariance Localization

Chapter 3 introduced a distance-based covariance localization procedure for reservoir history-matching problems. In this procedure, the critical lengths for localization are defined based on the correlation lengths of the underlying geological model and the approximate regions of influence of the data (sensitivity regions). The sensitivity regions are estimated using a pseudo-tracer to calculate drainage regions for each well at each data assimilation time-step. Chapter 3 presented three distinct examples, which demonstrate that the proposed covariance localization procedure reduces long-distance spurious correlations and yields far better estimates of the conditional mean and posterior covariances than are obtained with EnKF without localization. The proposed covariance localization procedure also gives better estimates of the conditional mean and posterior covariance than those that are obtained using localization based only on the prior geology or localization based only on the region of data sensitivity.

Chapter 3 also presented a real field application of the proposed covariance localization procedure. For this field case, our covariance localization procedure

gave a smoother ensemble mean, higher variability in the final ensemble, better data matches and better predictions than were obtained with the standard EnKF. Compared to a manually history-matched model, EnKF with localization resulted in far more geologically realistic models with better data matches and better predictions. We also assimilated data with the HI-EnKF with covariance localization. For this field case, HI-EnKF gave significant improvements in the data matches and predictions. The results also showed that the computational cost of HI-EnKF can be reduced without compromising the results by rerunning the ensemble from time zero only when "large" changes in the state vector occur at a data assimilation time-step.

10.2 Strategies to Reduce Loss of Variance Due to Sampling Errors in the EnKF

Chapter 4 presented a comparison among methods that appeared in the literature for reducing the negative effects of small ensembles by means of two simple problems with a large number of measurements. Based on the results presented in this chapter, we can state the following conclusions:

- For the two test problems considered in Chapter 4, standard EnKF and EnSRF resulted in rough estimates of the posterior mean, poor data matches and severe underestimation of the posterior variances although EnSRF resulted in a smaller underestimation of posterior variances than did EnKF.

- Among the methods considered in Chapter 4, distance-based covariance localization is the most effective. In fact, distance-based covariance localization was the only method which provided significant improvements for the two problems considered in this chapter.

- The two non-distance dependent localization procedures considered in Chapter 4, namely, Furrer and Bengtsson taper and bootstrap hierarchical filter,

failed to eliminate spurious correlations, and gave no appreciable improvements of the results obtained with standard EnKF.

- Even though covariance inflation is often used in the numerical weather prediction literature, it was ineffective for both problems considered in Chapter 4.

- In Chapter 4, we modified the adaptive covariance inflation procedure proposed by Evensen [47] for the case where distance-based covariance localization is also applied. However, this procedure deteriorated the data matches and resulted in apparent overestimation of posterior variances for the reservoir history-matching problem considered in Chapter 4.

10.3 Combining EnKF and MCMC

Chapter 5 introduced a procedure which combines EnKF and MCMC for the purpose of obtaining a more accurate sampling of the posterior pdf for reservoir model parameters and, consequently, a more accurate characterization of uncertainty than can be obtained by EnKF itself. The proposed method is easy to implement, it does not require adjoint implementations and it can be easily coupled with any commercial reservoir simulator. The algorithm was applied to history match production data and to predict future production for a three-dimensional two-phase reservoir problem. This example problem was sufficiently small that the posterior pdf could be sampled by generating a long Markov chain consisting of 2 million states (realizations of the log-permeability field). The results from this long Markov chain are assumed to be a correct sampling. Based on theoretical arguments and results from the example problem, the following conclusions are warranted:

- EnKF gives relatively high values of the objective function. High values of the objective function are associated with poor data matches. More importantly, a model which results in high value of the objective function gives a small value

of the posterior pdf, which suggests that this model is a sample from a low probability region.

- Because we may obtain relatively poor data matches with EnKF, the uncertainty in predictions of future production may be overestimated.

- For the example considered in Chapter 5, the proposed EnKF-MCMC method decreased the final objective functions, resulting in "more probable" models, i.e., higher values of the posterior pdf.

- Repeating the data assimilation with different initial ensembles resulted in distributions of reservoir predictions that are not mutually consistent, i.e., each ensemble resulted in significantly different distributions. Therefore, the uncertainty quantification provided by a single ensemble may be very unreliable. This problem is related to the dependence on the prior ensemble. In fact, each EnKF run "searches for solutions" in the specific subspace defined by the initial ensemble members. The proposed EnKF-MCMC does not ameliorate this problem.

- In Chapter 5, we proposed assimilating data with different initial ensembles and resampling the final set of models based on importance weights computed from the values of the normalized objective function. This procedure did not improve the sampling results obtained by EnKF because EnKF obtained unevenly distributed values of the objective function. Consequently, most of the EnKF samples have weights essentially equal to zero. Because EnKF-MCMC reduced significantly the values of the objective function obtained by EnKF, it resulted in better approximations of the posterior distribution.

- Applying the EnKF-MCMC iteratively resulted in additional improvements in the values of the objective function.

10.4 Multiple Data Assimilation

Chapter 6 introduced a method in which the same data are assimilated multiple times with an increased covariance matrix of measurement errors. In this chapter, we proved that single and multiple data assimilations are equivalent for the linear-Gaussian case and provided computational evidences that MDA improves the data assimilation results for the nonlinear case. The initial motivation for introducing this method was based on the problem of assimilating time-lapse seismic data. Unlike production data, time-lapse seismic data are available only at a few times during the reservoir production history. In this case, we no longer have the beneficial effect of several consecutive data assimilation time-steps as observed in the production data case. The proposed EnKF-MDA procedure was applied to assimilate time-lapse seismic data in two synthetic reservoir cases. The results showed significant improvements in the final data matches compared to standard EnKF.

The results obtained with EnKF-MDA for time-lapse seismic data motivated the use of MDA in conjunction with ES to assimilate production data. The resulting method (ES-MDA) can be interpreted as an iterative form of ES, where the number of "iterations" must be selected a priori. Even though the focus of this dissertation is on reservoir history-matching problems, the ES-MDA method is general and has potential applications in other research areas. Moreover, the method is easy to implement as it requires few modifications to a standard ES implementation. Based on the results from the three synthetic reservoir history-matching problems presented in Chapter 6, the following conclusions are warranted:

- ES performed poorly compared to standard EnKF and ES-MDA.
- ES-MDA resulted in better data matches than were obtained with EnKF with comparable computational cost.
- For the three history-matching problems considered in Chapter 6, four data assimilations were enough for providing good data matches.

- The use of the inflation coefficients, α_i's, in a decreasing order resulted in only small improvements compared to using α_i constant and equal to the number of data assimilations.

10.5 History Matching of Production and Seismic Data for a Real Field Case

Chapter 7 presented a field application of EnKF and ES-MDA for history matching production and seismic data. The results showed that ES-MDA obtained significantly better data matches than EnKF with a computational cost only 4% higher than the computational cost of EnKF.

10.6 Ensemble-based Parameterization for the RML Method

Chapter 8 introduced an ensemble-based parameterization for gradient-based history matching. This chapter described the parameterization for generating the MAP estimate and for sampling the posterior pdf with RML. The parameterization is general and requires very few modifications in a gradient-based implementation. The ensemble-based parameterization was tested in a small reservoir problem and the results suggest a significant improvement in the computational cost compared to a standard RML implementation without parameterization. Chapter 8 also presented a comparison between EnKF and gradient-based RML in terms of computational cost, data matches and uncertainty characterization. The results in this chapter showed that the computational cost of RML is much higher than that of EnKF. However, the results are much more reliable. The main downside of RML is its dependence on the availability on an adjoint implementation for efficient gradient calculation. On the other hand, EnKF can be easily coupled with virtually any commercial simulator. Moreover, EnKF is easy to adapt to different types of data and model variables. These advantages make EnKF a very attractive method for history matching field cases even though it may result in unreliable quantification of uncertainty.

10.7 Investigation on the Sampling Performance of Ensemble-based Methods

Chapter 9 presented a comparison between the adjoint-based RML and eight ensemble-based methods in terms of the data matches, characterization of uncertainty and computational cost for a small, but highly nonlinear, reservoir history-matching problem. Most of the ensemble-based methods, including EnKF, failed to obtain good data matches and resulted in a significant overestimation of uncertainty. Among the ensemble-based methods, ES-MDA obtained the best performance. ES-MDA provided an uncertainty quantification comparable to the adjoint-based RML if we combine the results of the ten data assimilations with different initial ensembles. However, RML was the only method that obtained as good a data-match as the one generated from MCMC.

BIBLIOGRAPHY

[1] Aanonsen, S. I., G. Nævdal, D. S. Oliver, A. C. Reynolds, and B. Vallés, Review of ensemble Kalman filter in petroleum engineering, *SPE Journal*, **14**(3), 393–412, 2009.

[2] Agarwal, B., H. Hermansen, J. Sylte, and L. Thomas, Reservoir characterization of Ekofisk field: A giant, fractured chalk reservoir in the Norwegian North Sea–history match, *SPE Reservoir Evaluation & Engineering*, **3**(6), 534–543, 2000.

[3] Agbalaka, C. and D. S. Oliver, Application of the EnKF and localization to automatic history matching of facies distribution and production data, *Mathematical Geosciences*, **40**(4), 353–374, 2008.

[4] Anderson, J. L., An ensemble adjustment Kalman filter for data assimilation, *Monthly Weather Review*, **129**(12), 2884–2903, 2001.

[5] Anderson, J. L., An adaptive covariance inflation error correction algorithm for ensemble filters, *Tellus A*, **59**(2), 210–224, 2007.

[6] Anderson, J. L., Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, *Physica D: Nonlinear Phenomena*, **230**(1–2), 99–111, 2007.

[7] Anderson, J. L. and S. L. Anderson, A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Monthly Weather Review*, **127**(12), 2741–2758, 1999.

[8] Andrews, A., A square root formulation of the Kalman covariance equations, *AIAA*, **6**(6), 1165–1168, 1968.

[9] Anterion, F., B. Karcher, and R. Eymard, Use of parameter gradients for reservoir history matching, in *Proceedings of the 10th SPE Reservoir Simulation Symposium, Houston, Texas, 6–8 February*, SPE 18433, 1989.

[10] Arroyo-Negrete, E., D. Devegowda, A. Datta-Gupta, and J. Choe, Streamline-assisted ensemble Kalman filter for rapid and continuous reservoir model updating, *SPE Reservoir Evaluation & Engineering*, **11**(6), 1046–1060, 2008.

[11] Batzle, M. and Z. Wang, Seismic properties of pore fluids, *Geophysics*, **57**(11), 1396–1408, 1992.

[12] Bengtsson, T., P. Bickel, and B. Li, Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, in *Probability and Statistics: Essays in Honor of David A. Freedman*, vol. 2, pp. 316–334, IMS Collections, 2008.

[13] Bertino, L., G. Evensen, and H. Wackernagel, Sequential data assimilation techniques in oceanography, *International Statistical Review*, **71**(2), 223–241, 2003.

[14] Bianco, A., A. Cominelli, L. Dovera, G. Nævdal, and B. Vallès, History matching and production forecast uncertainty by means of the ensemble Kalman filter: A real field application, in *Proceedings of the EAGE/EUROPEC Conference and Exhibition, London, U.K., 11–14 June*, SPE 107161, 2007.

[15] Bishop, C. H., B. J. Etherton, and S. J. Majumdar, Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects, *Monthly Weather Review*, **129**, 420–436, 2001.

[16] Bissell, R., Calculating optimal parameters for history matching, in *Proceedings of the 4th European Conference on the Mathematics of Oil Recovery*, 1994.

[17] Bruhn, C. H. L., C. D. Luchesse, J. A. T. Gomes, and P. R. S. Johann, Campos Basin: Reservoir charactertization and management – historical overview and future challenges, in *Proccedings of the Offshore Technology Conference, Houston, Texas*, OTC 15220, 2003.

[18] Burgers, G., P. van Leeuwen, and G. Evensen, Analysis scheme in the ensemble Kalman filter, *Monthly Weather Review*, **126**(6), 1719–1724, 1998.

[19] Chavent, G. M., M. Dupuy, and P. Lemonnier, History matching by use of optimal control theory, *SPE Journal*, **15**(1), 74–86, 1975.

[20] Chen, C., G. Li, and A. C. Reynolds, Closed-loop reservoir management on the Brugge test case, *Computational Geosciences*, **14**(4), 691–703, 2010.

[21] Chen, W. H., G. R. Gavalas, J. H. Seinfeld, and M. L. Wasserman, A new algorithm for automatic history matching, *SPE Journal*, **14**(6), 593–608, 1974.

[22] Chen, Y. and D. Oliver, Cross-covariance and localization for EnKF in multiphase flow data assimilation, *Computational Geosciences*, **14**(4), 579–601, 2009.

[23] Chen, Y. and D. Oliver, Ensemble-based closed-loop optimization applied to Brugge field, *SPE Reservoir Evaluation & Engineering*, **13**(1), 56–71, 2010.

[24] Chen, Y. and D. Oliver, Ensemble randomized maximum likelihood method as an iterative ensemble smoother, *Mathematical Geosciences*, **Online First**, 2011.

[25] Chen, Y. and D. Zhang, Data assimilation for transient flow in geologic forma-

tions via ensemble Kalman filter, *Advances in Water Resources*, **29**(8), 1107–1122, 2006.

[26] Cipra, B. A., The best of the 20th Century: Editors name top 10 algorithms, *SIAM News*, **33**(4), 2000.

[27] Coutinho, E. J., A. A. Emerick, G. Li, and A. C. Reynolds, Conditioning multi-layered gelogic models to well test and production logging data using the ensemble Kalman filter, in *Proccedings of the SPE Annual Technical Conference and Exhibition, Florence, Italy, 19–22 September*, SPE 134542, 2010.

[28] Damiani, M. C., *Determinação de padrões de fluxo em simulações de reservatório de petróleo utilizando traçadores (in Portuguese)*, Master's thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, 2007.

[29] de Marsily, G., G. Lavedan, M. Boucher, and G. Fasanino, Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model, in *Geostatistics for Natural Resources Characterization, Part 2*, (edited by G. Verly, M. David, A. G. Journel, and A. Marechal), pp. 831–849, D. Reidell, Dordrecht, Holland, 1984.

[30] Deutsch, C. V. and A. G. Journel, *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 1992.

[31] Devegowda, D., E. Arroyo-Negrete, A. Datta-Gupta, and S. G. Douma, Efficient and robust reservoir model updating using ensemble Kalman filter with sensitivity-based covariance localization, in *Proceedings of the SPE Reservoir Simulation Symposium, Houston, Texas, 26–28 February*, SPE 106144, 2007.

[32] Dong, Y. and D. S. Oliver, Quantitative use of 4D seismic data for reservoir description, *SPE Journal*, **10**(1), 51–65, 2005.

[33] Doucet, A., N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 2000.

[34] Doucet, A. and A. M. Johansen, *A Tutorial on Particle Filtering and Smoothing: Fifteen years later*, Oxford University Press, 2008.

[35] Emerick, A. A. and A. C. Reynolds, EnKF-MCMC, in *Proccedings of the SPE EUROPEC/EAGE Annual Conference and Exhibition, Barcelona, Spain, 4–17 June*, SPE 131375, 2010.

[36] Emerick, A. A. and A. C. Reynolds, Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications, *Computational Geosciences*, **15**(2), 251–269, 2011.

[37] Emerick, A. A. and A. C. Reynolds, History matching a field case using the ensemble Kalman filter with covariance localization, *SPE Reservoir Evaluation & Engineering*, **14**(4), 423–432, 2011.

[38] Emerick, A. A. and A. C. Reynolds, Combining the ensemble Kalman filter with Markov chain Monte Carlo for improved history matching and uncertainty characterization, *SPE Journal*, **in press**, 2012.

[39] Emerick, A. A. and A. C. Reynolds, Ensemble smoother with multiple data assimilation, *accepted for Computers & Geosciences*, 2012.

[40] Emerick, A. A. and A. C. Reynolds, History matching production and seismic data in a real field case using the ensemble smoother with multiple data assimilation, *To be submitted*, 2012.

[41] Emerick, A. A. and A. C. Reynolds, History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations, *Computational Geosciences*, **Online First**, 2012.

[42] Emerick, A. A. and A. C. Reynolds, Investigation on the sampling performance of ensemble-based methods, *submitted to Computational Geosciences*, 2012.

[43] Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, **99**(C5), 10,143–10,162, 1994.

[44] Evensen, G., The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, **53**, 343–367, 2003.

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

[46] Evensen, G., *Data Assimilation: The Ensemble Kalman Filter*, Springer, Berlin, 2007.

[47] Evensen, G., The ensemble Kalman filter for combined state and parameter estimation, *IEEE Control Systems Magazine*, pp. 83–104, 2009.

[48] Evensen, G., J. Hove, H. C. Meisingset, E. Reiso, K. S. Seim, and O. Espelid, Using the EnKF for assisted history matching of a North Sea reservoir model, in *Proceedings of the SPE Reservoir Simulation Symposium, Houston, Texas, 26–28 February*, SPE 106184, 2007.

[49] Evensen, G. and P. J. van Leeuwen, An ensemble Kalman smoother for nonlinear dynamics, *Monthly Weather Review*, **128**(6), 1852–1867, 2000.

[50] Fahimuddin, A., S. I. Aanonsen, and J.-A. Skjervheim, 4D seismic history matching of a real field case with EnKF: Use of local analysis for model updating, in *Proccedings of the SPE Annual Technical Conference and Exhibition, Florence, Italy, 19–22 September*, SPE 134894, 2010.

[51] Fahimuddin, A., S. I. Aanonsen, and J.-A. Skjervheim, Ensemble based 4D seismic history matching: Integration of different levels and types of seismic data, in *Proccedings of the SPE EUROPEC/EAGE Annual Conference and Exhibition, Barcelona, Spain, 14–17 June*, SPE 131453, 2010.

[52] Falcone, G., O. Gosselin, F. Maire, J. Marrauld, and M. Zhakupov, Petroelastic modelling as key element of 4D history matching: A field example, in *Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, Texas, 26–29 September*, SPE 90466, 2004.

[53] Feller, W., *An Introduction to Probability Theory and Its Applications: Volume I*, third edn., John Wiley & Sons, New York, 1968.

[54] Fletcher, R., *Practical Methods of Optimization*, second edn., John Wiley & Sons, New York, 1987.

[55] Floris, F. J. T., M. D. Bush, M. Cuypers, F. Roggero, and A.-R. Syversveen, Methods for quantifying the uncertainty of production forecasts: A comparative study, *Petroleum Geoscience*, **7**(SUPP), 87–96, 2001.

[56] Foster, D. G., Lessons learnt from over 20 years of 4-D deployment, in *Proceedings of the SPE Indian Oil and Gas Technical Conference and Exhibition, Mumbai, India, 4–6 March*, SPE 113542, 2008.

[57] Furrer, R. and T. Bengtsson, Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants, *Journal of Multivariate Analysis*, **98**(2), 227–255, 2007.

[58] Gao, G., G. Li, and A. C. Reynolds, A stochastic algorithm for automatic history matching, *SPE Journal*, **12**(2), 196–208, 2007.

[59] Gao, G. and A. C. Reynolds, An improved implementation of the LBFGS algorithm for automatic history matching, *SPE Journal*, **11**(1), 5–17, 2006.

[60] Gao, G., M. Zafari, and A. C. Reynolds, Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF, *SPE Journal*, **11**(4), 506–515, 2006.

[61] Gaspari, G. and S. E. Cohn, Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society*, **125**(554), 723–757, 1999.

[62] Gassmann, F., Elastic waves through a packing of spheres, *Geophysics*, **16**, 673–685, 1951.

[63] Gavalas, G. R., P. C. Shah, and J. H. Seinfeld, Reservoir history matching by Bayesian estimation, *SPE Journal*, **16**(6), 337–350, 1976.

[64] Gelman, A., G. O. Roberts, and W. R. Gilks, Efficient Metropolis jumping rules, *Bayesian Statistics. Oxford University Press*, **5**(5), 599–607, 1996.

[65] Golub, G. H. and C. F. van Loan, *Matrix Computations*, 2nd edn., The Johns Hopkins University Press, Baltimore, 1989.

[66] Gosselin, O., S. Aanonsen, I. Aavatsmark, A. Cominelli, R. Gonard, M. Kolasinski, F. Ferdinandi, L. Kovacic, and K. Neylon, History matching using time-lapse seismic (HUTS), in *Proceedings of the SPE Annual Technical Conference and Exhibition, Denver, Colorado, 5–8 October*, SPE 84464, 2003.

[67] Grimstad, A. A., T. Mannseth, S. A. Aanonsen, I. Aavatsmark, A. Cominelli, and S. Mantica, Identification of unknown permeability trends from history matching of production data, *SPE Journal*, **9**(4), 419–428, 2004.

[68] Grimstad, A. A., T. Mannseth, G. Nævdal, and G. Urkedal, Scale splitting approach to reservoir characterization, in *Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, Texas, 11–14 February*, SPE 66394, 2001.

[69] Gruenwalder, M., S. Poellitzer, and T. Clemens, Assisted and manual history matching of a reservoir with 120 wells, 58 years production history and multiple well recompletions, in *Proceedings of the EUROPEC/EAGE Conference and Exhibition, London, U.K., 11–14 June*, SPE 106039, 2007.

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

[71] Hamill, T. M., J. S. Whitaker, and C. Snyder, Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Monthly Weather Review*, **129**(11), 2776–2790, 2001.

[72] Hashin, Z. and S. Shtrikaman, A variational approach to the theory of the elastic behaviour of multiphase materials, *Journal of the Mechanics and Physics of Solids*, **11**(2), 127–140, 1963.

[73] Hastings, W. K., Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**(1), 97–109, 1970.

[74] Haugen, V., G. Nævdal, L.-J. Natvik, G. Evensen, A. M. Berg, and K. M. Flornes, History matching using the ensemble Kalman filter on a North Sea field case, *SPE Journal*, **13**(4), 382–391, 2008.

[75] Haverl, M., M. Aga, and E. Reiso, Integrated workflow for quantitative use of time-lapse seismic data in history matching: A North Sea field case, in *Proceedings of the SPE EUROPEC/EAGE Annual Conference, Madrid, Spain,13–16 June*, SPE 94453, 2005.

[76] Haverl, M. and J.-A. Skjervheim, 4D seismic modeling integrated with the ensemble Kalman filter method for history matching of reservoir simulation model, in *Proceedings of the 11th European Conference on the Mathematics of Oil Recovery, Bergen, Norway*, 2008.

[77] Hegstad, B. K. and H. Omre, Uncertainty in production forecasts based on well observations, seismic data and production history, Tech. Rep. Statistics No. 17/1999, Norwegian University of Science and Technology, 1999.

[78] Hegstad, B. K. and H. Omre, Uncertainty in production forecasts based on well observations, seismic data and production history, *SPE Journal*, **6**(4), 409–424, 2001.

[79] Horn, R. A., The Hadamard product, in *Matrix Theory and Applications*, pp. 87–170, American Mathematical Society, 1990.

[80] Houtekamer, P. L. and H. L. Mitchell, Data assimilation using an ensemble Kalman filter technique, *Monthly Weather Review*, **126**(3), 796–811, 1998.

[81] Houtekamer, P. L. and H. L. Mitchell, A sequential ensemble Kalman filter for atmospheric data assimilation, *Monthly Weather Review*, **129**(1), 123–137, 2001.

[82] Houtekamer, P. L. and H. L. Mitchell, Ensemble Kalman filtering, *Quarterly Journal of the Royal Meteorological Society*, **131**, 3269–3289, 2005.

[83] Hu, L. Y., Gradual deformation and iterative calibration of Gaussian-related stochastic models, *Mathematical Geology*, **32**(1), 87–108, 2000.

[84] Huang, X., R. Will, M. Khan, and L. Stanley, Reservoir characterization by integration of time-lapse seismic and production data, in *Proceedings of the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 5–8 October*, SPE 38695, 1997.

[85] Jacquard, P. and C. Jain, Permeability distribution from field pressure data, *SPE Journal*, **5**(4), 281–294, 1965.

[86] Jahns, H. O., A rapid method for obtaining a two-dimensional reservoir description from well pressure response data, *SPE Journal*, **6**(12), 315–327, 1966.

[87] Jansen, J., D. Brouwer, G. Naevdal, and C. van Kruijsdijk, Closed-loop reservoir management, *First Break*, **23**, 43–48, 2005.

[88] Jansen, J. D., S. D. Douma, D. R. Brouwer, P. M. J. V. den Hof, and A. W. Heemink, Closed-loop reservoir management, in *Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, 2–4 February*, SPE 119098, 2009.

[89] Johann, P., R. Sansonowski, R. Oliveira, and D. Bampi, 4D seismic in heavy-oil, turbidite reservoir offshore Brazil, *The Leading Edge*, **28**(6), 718–729, 2009.

[90] Johann, P., E. Thedy, F. Gomes, and M. Schinelli, 4D seismic in Brazil: Experiences in reservoir monitoring, in *Proceedings of the Offshore Technology Conference, Houston, Texas, 1–4 May*, OTC 18400, 2006.

[91] Johnston, D., J. Shyeh, J. Eastwood, M. Khan, and L. Stanley, Interpretation and modeling of time-lapse seismic data: Lena field, Gulf of Mexico, in *Proceedings of the Offshore Technology Conference, Houston, Texas, 1–4 May*, OTC 12132, 2000.

[92] Kalman, R. E., A new approach to linear filtering and prediction problems, *Transactions of the ASME, Journal of Basic Engineering*, **82**, 35–45, 1960.

[93] Kepert, J. D., On ensemble representation of the observation-error covariance in the ensemble Kalman filter, *Ocean Dynamics*, **54**(6), 561–569, 2004.

[94] Keppenne, C. L. and M. M. Rienecker, Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter, *Journal of Marine Systems*, **40–41**, 363–380, 2003.

[95] Kirkpatrick, S., C. D. G. Jr., and M. P. Vecchi, Optimization by simulated annealing, *Science*, **220**(4598), 671–680, 1983.

[96] Kitanidis, P. K., Quasi-linear geostatistical theory for inversing, *Water Resources Research*, **31**(10), 2411–2419, 1995.

[97] Kjelstadli, R., H. Lane, D. Johnson, O. Barkved, K. Buer, and T. Kristiansen, Quantitative history match of 4D seismic response and production data in the Valhall field, in *Proceedings of the Offshore Europe, Aberdeen, United Kingdom, 6–9 September*, SPE 96317, 2005.

[98] Krymskaya, M. V., R. G. Hanea, and M. Verlaan, An iterative ensemble Kalman filter for reservoir engineering applications, *Computational Geosciences*, **13**(2), 235–244, 2009.

[99] LaVenue, A. M. and J. F. Pickens, Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model, *Water Resources Research*, **28**(6), 1543–1569, 1992.

[100] LaVenue, A. M., B. S. RamaRao, G. de Marsily, and M. G. Marietta, Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, 2. Application, *Water Resources Research*, **31**(3), 495–516, 1995.

[101] Leeuwenburgh, O., G. Evensen, and L. Bertino, The impact of ensemble filter definition on the assimilation of temperature profiles in the tropical pacific, *Quarterly Journal of the Royal Meteorological Society*, **131**, 3291–3300, 2005.

[102] Li, G., M. Han, R. Banerjee, and A. C. Reynolds, Integration of well test pressure data into heterogeneous geological reservoir models, *SPE Reservoir Evaluation & Engineering*, **13**(3), 2010.

[103] Li, G. and A. C. Reynolds, Iterative ensemble Kalman filters for data assimilation, *SPE Journal*, **14**(3), 496–505, 2009.

[104] Li, H., E. Kalnay, and T. Miyoshi, Simultaneous estimation of covariance inflation and observation errors within ensemble Kalman filter, *Quarterly Journal of the Royal Meteorological Society*, **135**(639), 523–533, 2009.

[105] Li, R., A. C. Reynolds, and D. S. Oliver, History matching of three-phase flow production data, *SPE Journal*, **8**(4), 328–340, 2003.

[106] Liu, D. and J. Nocedal, On the limited memory BFGS method for large scale optimization, *Mathematical Programming*, **45**, 503–528, 1989.

[107] Liu, G., Y. Chen, and D. Zhang, Investigation of flow and transport processes at the MADE site using ensemble Kalman filter, *Advances in Water Resources*, **31**(7), 975–986, 2008.

[108] Liu, N., *Assessment of Uncertainty Assessment Methods*, Master's thesis, The University of Tulsa, Tulsa, Oklahoma, 2002.

[109] Liu, N. and D. S. Oliver, Evaluation of Monte Carlo methods for assessing uncertainty, *SPE Journal*, **8**(2), 188–195, 2003.

[110] Lorenc, A. C., The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, **129**(595), 3183–3203, 2003.

[111] Lorentzen, R. J., K. K. Fjelde, J. Frøyen, A. C. Lage, G. Nævdal, and E. H. Vefring, Underbalanced and low-head drilling operations: Real time interpretation of measured data and operational support, in *Proceedings of the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 30 September–3 October*, SPE 71384, 2001.

[112] Lorentzen, R. J. and G. Nævdal, An iterative ensemble Kalman filter, *IEEE Transactions on Automatic Control*, **56**(8), 1990–1995, 2011.

[113] Lorentzen, R. J., G. Nævdal, B. Vallès, A. M. Berg, and A.-A. Grimstad, Analysis of the ensemble Kalman filter for estimation of permeability and porosity in reservoir models, in *Proceedings of the SPE Annual Technical Conference and Exhibition, Dallas, Texas, 9–12 October*, SPE 96375, 2005.

[114] Lorentzen, R. J., A. Shafieirad, and G. Nævdal, Closed loop reservoir management using the ensemble Kalman filter and sequential quadratic programming, in *Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, 2–4 February*, SPE 119101, 2009.

[115] Lorenzatto, R. A., R. Juiniti, J. Gomes, and J. Martins, The Marlim field development: Strategies and challenges, in *Proccedings of the Offshore Technology Conference, Houston, Texas, May, 3–6*, OTC 16574, 2004.

[116] Lumley, D. and R. Behrens, Practical issues of 4D seismic reservoir monitoring: What an engineer needs to know, *SPE Reservoir Evaluation & Engineering*, **1**(6), 528–538, 1998.

[117] Ma, X., M. Al-Harbi, A. Datta-Gupta, and Y. Efendiev, An efficent two-stage sampling method for uncertainty quantification in history matching geological models, *SPE Journal*, **13**(1), 77–87, 2008.

[118] Ma, X., A. Datta-Gupta, and Y. Efendiev, An multistage MCMC with non-parametric error model for efficient uncertainty quantification in history matching, in *Proceedings of SPE Annual Technical Conference and Exhibition, Denver, Colorado, 21–24 September*, SPE 115911, 2008.

[119] Mandel, J., L. Cobb, and J. D. Beezley, On the convergence of the ensemble

Kalman filter, Tech. Rep. CCM Report 278, University of Colorado Denver, 2009, revised May 2011.

[120] Marshall, A. W., The use of multi-stage sampling schemes in Monte Carlo computations, in *Symposium on Monte Carlo Methods*, (edited by M. Meyer), pp. 123–140, Wiley, 1956.

[121] Mavko, G., T. Mukerji, and J. Dvorkin, *The Rock Physics Handbook: Tools for Seismic Analysis of Porous Media*, first edn., Cambridge University Press, 1998.

[122] Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equations of state calculations by fast computing machines, *Journal of Chemical Physics*, **21**, 1087–1092, 1953.

[123] Mezghani, M., A. Fornel, V. Langlais, and N. Lucet, History matching and quantitative use of 4d seismic data for an improved reservoir characterization, in *Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, Texas, 26–29 September*, SPE 90420, 2004.

[124] Moré, J. J. and D. Sorensen, Computing a trust region step, *SIAM Journal on Scientific Computing*, **4**, 553–572, 1983.

[125] Nævdal, G., T. Mannseth, and E. H. Vefring, Near-well reservoir monitoring through ensemble Kalman filter, in *Proceedings of the SPE/DOE Improved Oil Recovery Symposium, 13–17 April*, SPE 75235, 2002.

[126] Nocedal, J., Updating quasi-Newton matrices with limited storage, *Mathematics of Computation*, **35**(151), 773–782, 1980.

[127] Nocedal, J. and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.

[128] Nummelin, E., MC's for MCMCists, *International Statistical Review*, **70**(2), 215–240, 2002.

[129] Oliveira, R. M., D. Bampi, R. C. Sansonowski, N. M. S. R. Junior, P. R. S. Johann, M. S. Santos, and D. M. Ferreira, Marlim field: Incorporating 4D seismic in the geological model and application in reservoir management decisions, in *Proccedings of the SPE Latin American and Caribbean Petroleum Engineering Conference, Buenos Aires, Argentina, April, 15–18*, SPE 108062, 2007.

[130] Oliver, D. S., Multiple realizations of the permeability field from well-test data, *SPE Journal*, **1**(2), 145–154, 1996.

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

[132] Oliver, D. S. and Y. Chen, Recent progress on reservoir history matching: a review, *Computational Geosciences*, **15**(1), 185–221, 2010.

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

[134] Oliver, D. S., N. He, and A. C. Reynolds, Conditioning permeability fields to pressure data, in *Proccedings of the European Conference for the Mathematics of Oil Recovery*, 1996.

[135] Oliver, D. S., A. C. Reynolds, and N. Liu, *Inverse Theory for Petroleum Reservoir Characterization and History Matching*, Cambridge University Press, Cambridge, UK, 2008.

[136] Peaceman, D. W., Interpretation of well-block pressures in numerical reservoir simulation with non-square grid blocks and anisotropic permeability, *SPE Journal*, **23**(6), 531–543, 1983.

[137] Peters, L., R. Arts, G. Brouwer, C. Geel, S. Cullick, R. Lorentzen, Y. Chen, K. Dunlop, F. Vossepoel, R. Xu, P. Sarma, A. Alhuthali, and A. Reynolds, Results of the Brugge benchmark study for flooding optimisation and history matching, *SPE Reservoir Evaluation & Engineering*, **13**(3), 391–405, 2010.

[138] Pham, D. T., Stochastic methods for sequential data assimilation in strongly nonlinear systems, *Monthly Weather Review*, **129**(5), 1194–1207, 2001.

[139] Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd edn., Cambridge University Press, Cambridge, England, 2007.

[140] Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, Hydrologic data assimilation with the ensemble Kalman filter, *Monthly Weather Review*, **130**(1), 103–114, 2002.

[141] Reynolds, A. C., N. He, L. Chu, and D. S. Oliver, Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data, *SPE Journal*, **1**(4), 413–426, 1996.

[142] Reynolds, A. C., N. He, and D. S. Oliver, Reducing uncertainty in geostatistical description with well testing pressure data, in *Reservoir Characterization–Recent Advances*, (edited by R. A. Schatzinger and J. F. Jordan), pp. 149–162, American Association of Petroleum Geologists, 1999.

[143] Reynolds, A. C., M. Zafari, and G. Li, Iterative forms of the ensemble Kalman filter, in *Proceedings of 10th European Conference on the Mathematics of Oil Recovery, Amsterdam, 4–7 September*, 2006.

[144] Ripley, B. D., *Stochastic Simulation*, John Wiley & Sons, New York, 1987.

[145] Roberts, G. O., A. Gelman, and W. R. Gilks, Weak convergence and optimal

scaling of random walk Metropolis algorithm, *The Annals of Applied Probability*, **7**(1), 110–120, 1997.

[146] Roberts, G. O. and J. S. Rosenthal, Optimal scaling of various Metropolis-Hastings algorithms, *Statistical Science*, **16**(4), 351–367, 2001.

[147] Rodrigues, J. R. P., Calculating derivatives for automatic history matching, *Computational Geosciences*, **10**, 119–136, 2006.

[148] Roggero, F. and L. Y. Hu, Gradual deformation of continuous geostatistical models for history matching, in *Proceedings of the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 27–30 September*, SPE 49004, 1998.

[149] Rommelse, J., *Data Assimilation in Reservoir Management*, Ph.D. thesis, Technical University of Delft, Delft, The Netherlands, 2009.

[150] Sakov, P. and L. Bertino, Relation between two common localisation methods for the EnKF, *Computational Geosciences*, **15**(2), 225–237, 2011.

[151] Sakov, P. and P. R. Oke, Implications of the form of the ensemble transformation in the ensemble square root filters, *Monthly Weather Review*, **136**, 1042–1053, 2008.

[152] Sarma, P., L. Durlofsky, K. Aziz, and W. Chen, Efficient real-time reservoir management using adjoint-based optimal control and model updating, *Computational Geosciences*, **10**, 3–36, 2006.

[153] Sarma, P., L. J. Durlofsky, and K. Aziz, Kernel principal component analysis for efficient differentiable parameterization of multipoint geostatistics, *Mathematical Geosciences*, **40**, 3–32, 2008.

[154] Schölkopf, B., A. Smola, and K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Tech. Rep. 44, Max-Planck-Institut für Biologische Kybernetik, 1996.

[155] Seiler, A., S. I. Aanonsen, G. Evensen, and J. C. Reivenæs, Structural uncertainty modeling and updating using the ensemble Kalman filter, *SPE Journal*, **15**(4), 1062–1076, 2010.

[156] Seiler, A., G. Evensen, J.-A. Skjervheim, J. Hove, and J. Vabø, Advanced reservoir management workflow using an EnKF based assisted history matching method, in *Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, 2–4 February*, SPE 118906, 2009.

[157] Shah, P. C., G. R. Gavalas, and J. H. Seinfeld, Error analysis in history matching: The optimum level of parameterization, *SPE Journal*, **18**(6), 219–228, 1978.

[158] Skjervheim, J.-A., G. Evensen, S. I. Aanonsen, B. O. Ruud, and T. A. Johansen, Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter, *SPE Journal*, **12**(3), 282–292, 2007.

[159] Skjervheim, J.-A., G. Evensen, J. Hove, and J. G. Vabø, An ensemble smoother for assisted history matching, in *Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, USA, 21–23 February*, SPE 141929, 2011.

[160] Sonneland, L., C. Signer, H. Veire, R. Johansen, and L. Pedersen, 4D seismic on Gullfaks, in *Proceedings of the Offshore Technology Conference, Houston, Texas, 5–8 May*, OTC 8290, 1997.

[161] Stordal, A. S. and D. S. Oliver, Characterization of permeability and porosity from nanosensor observations, *Advances in Water Resources*, **34**(8), 53–61, 2011.

[162] Strebelle, S., Conditional simulation of complex geological structures using multiple-point statistics, *Mathematical Geology*, **34**(1), 1–22, 2002.

[163] Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. A. Yorke, Assessing a local ensemble Kalman filter: perfect model experiments with the national centers for environmental prediction global model, *Tellus A*, **57**, 528–545, 2005.

[164] Tarantola, A., *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, USA, 2005.

[165] Tavakoli, R. and A. A. Reynolds, Monte Carlo simulation of permeability fields and reservoir performance predictions with SVD parameterization in RML compared with EnKF, *Computational Geosciences*, **15**(1), 99–116, 2011.

[166] Tavakoli, R. and A. C. Reynolds, History matching with parameterization based on the SVD of a dimensionless sensitivity matrix, *SPE Journal*, **15**(12), 495–508, 2010.

[167] Thulin, K., G. Li, S. I. Aanonsen, and A. C. Reynolds, Estimation of initial fluid contacts by assimilation of production data with EnKF, in *Proceedings of the SPE Annual Technical Conference and Exhibition, Anaheim, California, 11–14 November*, SPE 109975, 2007.

[168] Thulin, K., G. Nævdal, H. J. Skaug, and S. I. Aanonsen, Quantifying Monte Carlo uncertainty in the ensemble Kalman field, *SPE Journal*, **16**(1), 172–182, 2011.

[169] Tierney, L., Markov chains for exploring posterior distributions, *Annals of Statistics*, **22**(4), 1701–1728, 1994.

[170] Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, Ensemble square-root filters, *Monthly Weather Review*, **131**, 1485–1490, 2003.

[171] Tjelmeland, H., H. Omre, and B. K. Hegstad, Sampling from Bayesian models in reservoir characterization, Tech. Rep. Statistics No. 2/1994, Norwegian Institute of Technology, Trondheim, Norway, 1994.

[172] van Leeuwen, P. J., Comment on "Data assimilation using an ensemble Kalman filter technique", *Monthly Weather Review*, **127**(6), 1374–1377, 1999.

[173] van Leeuwen, P. J., An ensemble smoother with error estimates, *Monthly Weather Review*, **129**, 709–728, 2001.

[174] van Leeuwen, P. J. and G. Evensen, Data assimilation and inverse methods in terms of a probabilistic formulation, *Monthly Weather Review*, **124**, 2898–2913, 1996.

[175] Černý, V., Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm, *Journal of Optimization Theory and Application*, **45**(1), 41–51, 1985.

[176] Verly, G., *Geostatistics Troia '92*, chap. Sequential Gaussian Cosimulation: A Simulation Method Integrating Several Types of Information, pp. 543–545, Kluwer Academic Publishers, 1993.

[177] Vogel, C. R. and J. G. Wade, Iterative SVD-based methods for ill-posed problems, *SIAM Journal on Scientific Computing*, **15**(3), 736–754, 1994.

[178] Waggoner, J., Lessons learned from 4D projects, *SPE Reservoir Evaluation & Engineering*, **3**(4), 310–318, 2000.

[179] Waggoner, J., A. Cominelli, and R. Seymour, Improved reservoir modeling with time-lapse seismic in a Gulf of Mexico gas condensate reservoir, in *SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 29 September–2 October*, SPE 77514, 2002.

[180] Wang, C., G. Li, and A. C. Reynolds, Production optimization in closed-loop reservoir management, *SPE Journal*, **14**(3), 506–523, 2009.

[181] Wang, X. and C. H. Bishop, A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes, *Journal of the Atmospheric Sciences*, **60**, 1140–1158, 2003.

[182] Wang, Y., G. Li, and A. C. Reynolds, Estimation of depths of fluid contacts by history matching using iterative ensemble-Kalman smoothers, *SPE Journal*, **15**(2), 2010.

[183] Watanabe, S. and A. Datta-Gupta, Use of phase streamlines for covariance localization in ensemble Kalman filter for three-phase history matching, in *Proccedings of the SPE Western North American Region Meeting, Anchorage, Alaska, USA, 7–11 May*, SPE 144579, 2011.

[184] Wen, X.-H. and W. H. Chen, Real-time reservoir model updating using ensemble Kalman filter with confirming option, *SPE Journal*, **11**(4), 431–442, 2006.

[185] Wen, X.-H., C. V. Deutsch, and A. S. Cullick, High-resolution reservoir models integrating multiple-well production data, *SPE Journal*, **3**(4), 344–355, 1998.

[186] Whitaker, J. S. and T. M. Hamill, Ensemble data assimilation without perturbed observations, *Monthly Weather Review*, **130**(7), 1913–1924, 2002.

[187] Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, Ensemble data assimilation with the NCEP global forecast system, *Monthly Weather Review*, **136**, 463–482, 2008.

[188] Wu, Z., A. C. Reynolds, and D. S. Oliver, Conditioning geostatistical models to two-phase production data, *SPE Journal*, **3**(2), 142–155, 1999.

[189] Xue, G. and A. Datta-Gupta, Structure preserving inversion: An efficient approach to conditioning stochastic reservoir models to dynamic data, in *Proceedings of the SPE Annual Technical Conference, San Antonio, Texas, 5–8 October*, SPE 38727, 1997.

[190] Zachariassen, E., J. Skjervheim, J. Vabø, I. Lunt, J. Hove, and G. Evensen, Integrated work flow for model update using geophysical monitoring data, in *Proceedings of the 73rd EAGE Conference & Exhibition, Vienna, Austria, 23–26 May*, 2011.

[191] Zafari, M. and A. C. Reynolds, Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter, *SPE Journal*, **12**(3), 382–391, 2007.

[192] Zhang, F. and A. C. Reynolds, Optimization algorithms for automatic history matching of production data, in *Proceedings of 8th European Conference on the Mathematics of Oil Recovery, Freiberg, Germany, 3–6 September*, 2002.

[193] Zhang, F., A. C. Reynolds, and D. S. Oliver, Evaluation of the reduction in uncertainty obtained by conditioning a 3D stochastic channel to multiwell pressure data, *Mathematical Geology*, **34**(6), 713–740, 2002.

[194] Zhang, Y. and D. Oliver, Improving the ensemble estimate of the Kalman gain by bootstrap sampling, *Mathematical Geosciences*, **42**, 327–345, 2010.

[195] Zhao, Y., A. C. Reynolds, and G. Li, Generating facies maps by assimilating production data and seismic data with the ensemble Kalman filter, in *Proccedings of the SPE Improved Oil Recovery Symposium, Tulsa, Oklahoma, 20–23 April*, SPE 113990, 2008.