Ensemble Methods of Data Assimilation in Porous Media Flow for Non-Gaussian Prior Probability Density

Yanhui Zhang

Dissertation for the degree of philosophiae doctor (PhD)

at the University of Bergen

2015

Dissertation date: 28.08.15

Scientific environment

Abstract

Ensemble-based data-assimilation methods have gained fast growth since the ensemble Kalman filter (EnKF) was introduced into the petroleum engineering. Many techniques have been developed to overcome the inherent disadvantages of the ensemble-based methods including the assumptions of linearity and Gaussianity, to make it more robust and reliable for practical reservoir applications. The current trend in petroleum reservoir history matching is towards taking into account more realistic reservoir models with complex geology. Geologic facies modeling plays an important role in the reservoir characterization as a way to reproduce important patterns of heterogeneity in petroleum reservoirs and to facilitate the modeling of petrophysical properties of reservoir rocks. Because the static data and general geological knowledge are almost always not sufficient to determine the distribution of geologic facies uniquely, it is advantageous to assimilate dynamic data to reduce the uncertainty.

The history-matching problem of geologic facies involves large nonlinearity and non-Gaussianity. When the ensemble Kalman filter (EnKF) or related ensemble-based methods are used to assimilate data in a straightforward way, the updated model variables would lose the geological realism and be contaminated with noise. Therefore, it is necessary to develop some effective measures to adapt the ensemble-based data-assimilation methods to the facies problems. This thesis is motivated by this necessity and explores the ways to conditioning geologic facies to production data using ensemble-based methods. The focus is placed on the post-processing approach.

By modifying the standard randomized maximum likelihood algorithm to accommodate non-Gaussian problems, we introduce a methodology that consists of a straightforward implementation of ensemble-based data assimilation in the first place and a sequential optimization procedure without iteration as a follow-up. In a similar manner, we develop another method for the post-processing of the updated reservoir models after data assimilation using ensemble-based methods. In the post-processing step, the objective function is composed of a weighted quadratic term measuring the distance to the posterior realizations, and a penalty term forcing model variables to take discrete values. A special emphasis is put on the investigation of the importance of the correlation information among the updated model variables introduced by the data which is usually ignored for the probability map approach. All of the proposed methodologies are evaluated via numerical experiments and demonstrated their utilities for improving the assimilation of data to geological facies models.

In this thesis, we also investigate the application of discrete curvelet transform in the denoising problem of updated model variables by the direct use of ensemble-based data-assimilation methods. According to the numerical experiments, the results show that curvelets are useful for denoising in the problem concerned but lose data match unless the covariance is included.

List of papers

Paper A: Yanhui Zhang, Dean S. Oliver, Yan Chen and Hans J. Skaug, *Data assimilation using the EnKF for 2-D Markov chain models*, in Proceedings of the 13th European Conference on the Mathematics of Oil Recovery (ECMOR XIII), Biarritz, France, 10-13 September 2012.

Paper B: Yanhui Zhang, Dean S. Oliver, Yan Chen and Hans J. Skaug, *Data assimilation by use of the iterative ensemble smoother for 2D facies models*, SPE Journal, **20** (01): 169-185, SPE-170248-PA, 2014.

Paper C: Yanhui Zhang, Dean S. Oliver and Yan Chen, *Beyond the probability map: representation of posterior facies probability*, in Proceedings of the 14th European Conference on the Mathematics of Oil Recovery (ECMOR XIV), Catania, Sicily, Italy, 8-11 September 2014.

Paper D: Yanhui Zhang, Dean S. Oliver, Hervé Chauris and Daniela Donno, *Ensemble-based data assimilation with curvelet regularization*, submitted to Journal of Petroleum Science and Engineering, 2015.

Acknowledgements

First and foremost I would like to thank my supervisors: Dean S. Oliver, Yan Chen and Hans J. Skaug for supporting and guiding me throughout my study. I would especially like to offer my sincerest gratitude to my principal supervisor, Dean S. Oliver, for his great knowledge, patience, caring, and excellent guidance.

I would also like to express my special thanks to Hervé Chauris and Daniela Donno (Mines ParisTech), for their warm support and insightful discussions during my academic visit.

In my daily work I have been blessed with a friendly and cheerful group of friends and colleagues at Uni CIPR, who have provided me with an excellent social and research atmosphere during my PhD study. I would especially like to thank Kristian, Svenn, Alina and Dongfang for their accompany and help not only in the study but also in everyday life.

I am also deeply thankful to my parents and elder sister, Yanfen, for supporting me and encouraging me with best wishes. A special thanks to my wife, Weiwei, for always being there cheering me up and standing by me through the good times and bad, and for bringing me the best gift in the world, my little angel, Fanfei.

Contents

"As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality."

Albert Einstein (1879–1955)

Chapter 1

Introduction

1.1   Motivation and research objective

The ensemble Kalman filter (EnKF) has established itself as the most promising method for data assimilation in real reservoir models but, because the method uses only estimates of the first- and second-order moments of the distributions of reservoir properties and states for updating, it is a challenge to apply the method to reservoirs with complex geology.

The complex geology in many real reservoir models are created using geological models that simulate the spatial relationships of one type of geological facies to another. These models often contain regions that are relatively homogeneous compared to the difference with other regions. Geologic facies are commonly described as categorical variables. The petrophysical values like porosity and permeability might be assumed to be discrete. Unfortunately, the ensemble Kalman filter does a poor job at updating discrete variables. In a standard approach, the variables take continuous values after updating. Truncation of continuous values results in models that fail to honor the data and fail to honor spatial transitions. Various solutions, such as the use of truncated pluri-Gaussian models, level sets, and kernel methods have been partially successful but none are general enough to capture the complex transitions between facies. The main objective in this thesis is investigating approaches for assimilation of production data for updating of complex discrete geological reservoir models using modifications to the EnKF-like assimilation methods.

The thesis consists of two chapters. Chapter 1 is focused on the introduction of the research background and basic theory required for the included research papers presented in Chapter 2.

Section 1.2 provides a brief overview of the general procedure in reservoir characterization and puts the emphasis on the description of history-matching problems from a Bayesian point of view.

As the most promising techniques for automatic history matching, ensemble-based data-assimilation methods are introduced in Section 1.3, including the standard formulations of the ensemble Kalman filter and the ensemble smoother (ES), as well as several typical iterative forms of EnKF/ES. In addition, some common issues of ensemble-based methods in history-matching applications are discussed and the available solutions are concisely reviewed.

In Section 1.4, we introduce several typical geostatistical simulation methods that

have been commonly used in geologic facies modeling. In particular, some of these methods have been actively utilized in combination with ensemble-based methods for the problem of conditioning facies distributions to production data.

In Section 1.5, we provide a brief literature review of the techniques that have been proposed in the history matching of geologic facies using ensemble-based data-assimilation methods.

Finally, Chapter 2 is structured into two parts by starting with an introduction of the included papers, followed by the scientific results containing the details of these papers.

1.2   Reservoir characterization in a Bayesian framework

As an indispensable evaluation tool in the oil and gas industry, reservoir simulation has been playing an increasingly important role in the decision-making process involved in the development, planning and management of petroleum reservoirs. The major goal of reservoir simulation is to predict future performance of the reservoir and investigate the ways to improve the oil recovery under various operating strategies. In a simulation study it is of fundamental importance to build reliable and robust reservoir simulation models, which can represent the actual reservoir for particular forecasts with reasonable credibility. For this purpose, all the available data and relevant information about the reservoir should be integrated into the model in a consistent and effective way.

It is generally difficult, if not impossible, to incorporate all types of data simultaneously into a model. The general practice to model a petroleum reservoir consists of two basic steps, which are known as geological modeling and history matching. First, geoscientists build geological models that are conditioned to all of the static data such as geology, geophysics, petrophysics, seismic data and geostatistical information. These models provide a geologically realistic representation of the petroleum reservoir, however, they rarely honor the reservoir's dynamic behavior. Then these models are passed to petroleum engineers who incorporate the dynamic data like well rates, pressures, and 4D seismic survey to improve the reservoir characterization and reduce uncertainty.

History matching is difficult because it poses a nonlinear inverse problem in which the relationship between the reservoir model parameters and the dynamic data is highly nonlinear and there exist multiple solutions. In order to match the dynamic data, often some of the static data (e.g., geological continuity) are disregarded. Although the models after history matching are matched to the dynamic data adequately, they are unrealistic from a geologic point of view. The goal of history matching is not only to obtain history-matched reservoir models, but more importantly to improve the reservoir model's predictive power. Compared with the history-matched models with inconsistent geology, models that match the dynamic data and honor all of the other information are more likely to be predictive. Recently, it has seen growing research in history matching made on generating geologically consistent reservoir models [4, 5, 15, 103].

In this section, we give a brief introduction of the general procedures involved in a reservoir characterization study with the emphasis on the reservoir history matching, which is introduced in a Bayesian flavor. Meanwhile, some key features about geologic facies as the primary history matching target in this dissertation are described.

1.2.1 Geologic modeling

The main goal of geologic modeling is to build a plausible representation of a reservoir that incorporates and quantifies all the geologic characteristics of the reservoir. The geological model is usually constructed from analyses and interpretations of well log, core, seismic and outcrop analogue data based on an interdisciplinary collaboration. The modeling process proceeds sequentially and logically by following a prescribed work flow. In general, the work flow starts with establishing the structural and stratigraphic framework in the fist place, then continues by building a facies model describing the internal architecture and ends up populating petrophysical properties such as porosity and permeability into the established 3D model. As geostatistical methods are extensively used to improve the integration of geologic information, multiple realizations may be generated to quantify uncertainty in the geological model. The resultant models are commonly built at high levels of resolution, based on the recognition that fine-scale features can impact greatly on reservoir performance. Because the high dimensional geological models often exceed the capabilities of standard reservoir simulators, one has to compromise between the detailed geologic description and practical computational capabilities, by upscaling the fine-scale geologic model to obtain an affordable lower-dimensional reservoir flow model.

Although to the geologist, such a comprehensive description generally provide a sufficiently detailed characterization of the overall geologic complexity of the reservoir, the characterization may still prove to be unsatisfactory when the dynamic performance of the field is being considered. The dynamics of fluid flow within a reservoir is closely related to the reservoir heterogeneities, which are small to large scale geological features. These features can comprise variations in lithology, texture and sorting, as well as the presence of fractures, faults and diagenetic effects of different nature, all of which significantly affect the fluid flow at different scales. Due to inadequate and inaccurate data, uncertainty exists at all levels and most history-matching techniques are focused on modifying the smallest scale structures like gridblock permeabilities and porosities. When the dominant feature comes from the larger scales such as facies distributions, the adjustments at the smallest scale may result in a history-matched model but destroy the intended geological realism, leading to an unrealistic model with poor capability in production forecasting and reserves estimation, even though the resulting model has an excellent historical data match. From this point of view, an appropriate characterization of reservoir heterogeneity is one of the key issues of an integrated reservoir study.

1.2.2 Facies

In geology, a facies refers to a body of rock with distinctive characteristics that can be used to identify its depositional environment [23], such as river channel, delta system, submarine fan, and the like. Facies are considered important in reservoir modeling because there often exists a significant correlation between petrophysical properties and facies types. In particular, it is common that the variability and continuity in porosity and permeability distributions is dominated by the facies heterogeneity. As a result the variations in facies may have a dramatic effect on the flow response. Therefore, in a geologic modeling workflow where facies modeling is considered, the facies model

is always established first to constrain the distributions of porosity and permeability so as to narrow down the range of uncertainty. On the other hand, a geologically realistic reservoir model is usually more reliable on the prediction of the reservoir's performance. In this regard, the facies description contributes a significant amount of realism to the model. Since the facies types can only be precisely identified at well locations from well log and core data, the uncertainty on the facies distribution spreads throughout the reservoir. It is therefore advantageous to reduce the uncertainty by including facies into the history-matching process. However, unlike other common targets (e.g., permeability and porosity) in a history-matching study, it is very challenging to condition geologic facies to production data due to its complexity.

In the geological respect, modeling of facies is inherently complicated. There is a wide variety of geostatistical algorithms available for building facies models, such as variogram-based, multiple-point, object-based, and process-mimicking facies modeling techniques [100]. They are also usually categorized as grid-based (e.g., variogram-based and multiple-point-based) and object-based modeling methods. However, there is no single universal method that is able to handle all kinds of facies heterogeneities. The selection of modeling method heavily depends on the complexity of facies features. When the facies have no clear geometrical units like fluvial channels or crevasse splays, variogram-based methods like sequential indicator simulation and truncated Gaussian simulation are commonly the first choice. Multiple-point-based methods may be preferred if more complex and curvilinear features are required. When the facies exhibit clear geometrical patterns or long-range structures, object-based or process-mimicking methods should be used. Of course, geological complexity is not the only deciding factor for modeling method selection. Another important (maybe major) factor constraining the selection is data conditioning. Generally, a modeling method's ability to honor the data is inversely proportional to its ability to characterize the facies complexity. For example, in comparison with object-based methods, grid-based methods are more flexible and more capable in honoring dense data, while these methods lack the ability to match complex geological features like sharp geometries and long-range structures. Other than choosing modeling method, there are still many other aspects that need to be taken into account in practical applications. More importantly, software and modeler's expertise are the prerequisite. The sophisticated facies modeling process imposes great difficulties and leaves less flexibility on the history matching.

In terms of reservoir flow dynamics, conditioning facies to production data usually involves strong nonlinearities. Compared with permeability and porosity, facies is characterized by larger scale heterogeneities. Reservoir properties vary significantly at the locations of facies boundaries where small changes may cause dramatic flow response. Moreover, facies by nature are defined as categorical variables which is intrinsically non-Gaussian. Without proper modifications, most history-matching techniques are unable to be used readily. In summary, history matching of facies is distinguished from other history-matching problems by geological complexity, high nonlinearity and non-Gaussianity.

1.2.3   History matching as an inverse problem

More often than not, there is still considerable uncertainty left in the reservoir model after the geological modeling due to the limited amount of information. Therefore, con-

structing a reservoir simulation model that is consistent with historical dynamic data is necessary to better characterize the reservoir properties and to improve the model's predictive capability. This process is known as history matching during which unknown reservoir properties such as permeabilities and porosities are adjusted so that the obtained model can reproduce the observed historical performance in an optimum way.

History matching is well-known as a type of inverse problem. In its most basic form, if the relationship between the model variables $\mathbf{m} \in \mathbb{R}^{N_m}$ and the observable variables (data) $\mathbf{d} \in \mathbb{R}^{N_d}$ is given and denoted by

$$g(\mathbf{m}) = \mathbf{d}, \tag{1.1}$$

the inverse problem can be expressed as finding $\mathbf{m}$ that is the solution of (1.1) with known $\mathbf{d}$. In order to give a general picture of the distinctive features of history-matching problems in petroleum engineering, it is straightforward to address each term of (1.1) individually. Throughout the dissertation, terms such as "data", "observations" and "measurements" are used interchangeably.

The forward operator $g(\cdot)$ represents an approximation to the true physical relation between reservoir properties and data. The forward model is usually governed by a set of coupled, nonlinear partial differential equations based on Darcy's law, mass conservation and other related conditions. The relationship between $\mathbf{m}$ and $\mathbf{d}$ is typically highly nonlinear. A numerical reservoir simulator is often run to predict outcomes of the forward model. In most cases, the forward prediction requires a very high computational cost and is the most time-consuming part in the history-matching process, especially when the problem involves large-scale and complex reservoirs with a long production history. Because of this, computational efficiency as a key consideration in practice sets a high threshold for the feasibility of methods that one takes to solve history-matching problems, which makes trial-and-error-based approaches mostly impractical.

The model variables $\mathbf{m}$ are customarily organized in the form of a vector containing a variety of uncertain reservoir properties that need to be improved with more data. From the time varying point of view, the model variables are usually divided into model parameters and state variables. The former refers to the variables that are static and not time varying such as porosities and permeabilities, and the latter includes time-dependent variables that define the state of the system like saturations and pressures. Since the state variables can be sufficiently determined by the model parameters, the history-matching problem is traditionally formulated as a parameter estimation problem. Herein we narrow down the definition of $\mathbf{m}$ only referring to model parameters. Under the necessity of reservoir characterization and history matching of the dynamic data, most often the reservoir model contains a large number of model parameters. The inverse problem is often defined in a significant high dimensional model space and becomes severely ill-posed in the sense that the amount of independent data is commonly much smaller than the number of model parameters. On the other hand, there is almost always certain amount of prior information about the reservoir available to assist in restricting the solution space, like positivity, smoothness, and bounds.

The observable data of $\mathbf{d}$ always contain errors. Just like other physical systems, sources of the errors in $\mathbf{d}$, denoted as ε, can come from artificial readings, measuring instrument precisions, data processing or usually a mixture of them all. Taking this into

account, Equation (1.1) is replaced with

$$\mathbf{d}_{obs} = g(\mathbf{m}) + \varepsilon. \tag{1.2}$$

The observations $\mathbf{d}_{obs}$ contain many types of measurements such as production data, time-lapse seismic data, well test data, etc. In this thesis, we consider only the production data which are the most widely used for history matching. Production data include a set of measurements of flow rates, pressures, or ratios of flow rates, made at locations of producing or injecting wells [96]. They are collected in sequence of time along the development of petroleum reservoir. The key features of the production data that make the history-matching problem very special, consist in [92]:

- production data are usually limited in number in the sense that they are only obtained at well locations;

- the information content of production data is relatively low;

- the relationship between production data and model parameters is generally non-local and nonlinear,

Gathering all the information discussed above together, history matching is typically described as an ill-posed large-scale inverse problem, in which uncertain model parameters are estimated from inadequate and inexact data given an approximate relation between the model parameters and the data. Here the ill-posed problem is defined in terms of [52], that is, a problem is ill-posed if it does not satisfy the following conditions simultaneously:

- it has a solution,

- the solution is unique, and

- the solution is a continuous function of the data.

The ultimate goal of history matching is to improve the reliability of reservoir forecasts and to assist the decision making. However, the history-matching process involves various sources of uncertainty. There exist numerous possible combinations of model variables that enable the resulting reservoir models to match the observations equally well. Therefore, it is impossible to make an informed decision on revservoir management and investments without a proper assessment of uncertainty. From a probabilistic point of view, Bayesian statistics provides a straightforward theory for dealing with uncertainty.

1.2.4 Bayesian formulation

In the setting of Bayesian statistics, model parameters are considered as random variables and their distributions are described with probability density functions (PDF). Based on Bayes' theorem, the conditional (posterior) PDF, $p(\mathbf{m}|\mathbf{d}_{obs})$, for model variables $\mathbf{m}$ given observations $\mathbf{d}_{obs}$ is defined as

$$p(\mathbf{m}|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|m)p(\mathbf{m})}{p(\mathbf{d}_{obs})} = \frac{p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m})}{\int_D p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m})d\mathbf{m}} = \mathbf{a}\, L(\mathbf{m}|\mathbf{d}_{obs})p(\mathbf{m}). \tag{1.3}$$

In (1.3), $p(\mathbf{m})$ is the prior PDF for the vector of model parameters and represents the prior information or geostatistical knowledge of the reservoir before incorporating the observations. The prior PDF of $p(\mathbf{m})$ describes the dependency between the model parameters and imposes constraints on the set of possible inverse solutions. The dependency may refer to the spatial structure of $\mathbf{m}$ based on a covariance or a training image. $p(\mathbf{d}_{obs}|\mathbf{m})$ is the conditional PDF of $\mathbf{d}_{obs}$ given $\mathbf{m}$. We also call the function $p(\mathbf{d}_{obs}|\mathbf{m})$ the likelihood function denoted by $L(\mathbf{m}|\mathbf{d}_{obs})$ for $\mathbf{m}$, given $\mathbf{d}_{obs}$. The likelihood function $L(\mathbf{m}|\mathbf{d}_{obs})$ models the stochastic relationship between the observed data d and each particular model $\mathbf{m}$ maintained. $p(\mathbf{d}_{obs})$ is the PDF for the vector of observations and $\mathbf{a}$ is a normalizing constant. In this formulation, the solution to the aforementioned inverse problem is given by the posterior PDF which is proportional to the product of the prior PDF and the likelihood function. A common and instructive assumption is often made that the prior PDF is multi-variate Gaussian in which case the prior distribution is fully defined by the mean and the covariance. In particular, if $g(\mathbf{m})$ is linear and ε is Gaussian, the posterior PDF is also Gaussian and completely characterized by the mean and covariance [48]. The likelihood function measures the uncertainty in measurements and their predictions which normally involves forward simulations of data acquisition processes and corresponding estimations of measurement errors.

Under Gaussian assumptions on the prior PDF and the measurement errors ε in (1.2), we can write $p(\mathbf{m}|\mathbf{d}_{obs})$ as

$$
\begin{aligned}
p(\mathbf{m}|\mathbf{d}_{obs}) &= \mathbf{a}\ \exp\left[-\frac{1}{2}(\mathbf{m}-\mathbf{m}_{pr})^{T}\mathbf{C}_{M}^{-1}(\mathbf{m}-\mathbf{m}_{pr})\right] \\
&\quad \times\ \exp\left[-\frac{1}{2}(g(\mathbf{m})-\mathbf{d}_{obs})^{T}\mathbf{C}_{D}^{-1}(g(\mathbf{m})-\mathbf{d}_{obs})\right], \\
&= \mathbf{a}\ \exp\left[-\frac{1}{2}(\mathbf{m}-\mathbf{m}_{pr})^{T}\mathbf{C}_{M}^{-1}(\mathbf{m}-\mathbf{m}_{pr})\right. \\
&\quad \left. -\frac{1}{2}(g(\mathbf{m})-\mathbf{d}_{obs})^{T}\mathbf{C}_{D}^{-1}(g(\mathbf{m})-\mathbf{d}_{obs})\right], \\
&= \mathbf{a}\ \exp\left[-S(\mathbf{m})\right], \quad (1.4)
\end{aligned}
$$

where $S(\mathbf{m})$ is often termed the objective function [47, 74, 95]

$$
S(\mathbf{m}) = \frac{1}{2}(\mathbf{m}-\mathbf{m}_{pr})^{T}\mathbf{C}_{M}^{-1}(\mathbf{m}-\mathbf{m}_{pr}) + \frac{1}{2}(g(\mathbf{m})-\mathbf{d}_{obs})^{T}\mathbf{C}_{D}^{-1}(g(\mathbf{m})-\mathbf{d}_{obs}). \quad (1.5)
$$

In (1.4), $\mathbf{m}_{pr}$ and $\mathbf{C}_{M} \in \mathbb{R}^{N_m \times N_m}$ are the prior mean and the prior covariance matrix respectively. $\mathbf{C}_{D} \in \mathbb{R}^{N_d \times N_d}$ is the covariance matrix of observations errors and $g(\mathbf{m})$ is the predicted data for a given $\mathbf{m}$. In (1.5), the objective function consists of two terms that the first term is the model mismatch and the second term is the data mismatch. In a sense, the model mismatch term can be seen as a regularization to mitigate the ill-posedness of the problem of the minimization of the data mismatch by penalizing the deviations of model variables from the initial guess. Note that minimizing the objective function in (1.5) is equivalent to maximizing the posterior PDF in (1.4). From the Bayesian viewpoint, the best estimate of $\mathbf{m}$ is the model which maximizes the posterior PDF, or equivalently minimizes the objective function. This estimate is then referred to

as the maximum a posteriori (MAP) estimate of (1.4)

$$\mathbf{m}_{MAP} = \underset{\mathbf{m}}{\operatorname{argmax}}\ p(\mathbf{m}|\mathbf{d}_{obs}),$$
$$= \underset{\mathbf{m}}{\operatorname{argmin}}\ S(\mathbf{m}). \tag{1.6}$$

When the relation between data and model parameters is linear, i.e., $g(\mathbf{m}) = \mathbf{G}\mathbf{m}$ with sensitivity matrix $\mathbf{G} \in \mathbb{R}^{N_d \times N_m}$, $\mathbf{m}_{MAP}$ is also the posterior mean and can be derived in the two following equivalent forms by using the matrix inversion lemmas [96]

$$\mathbf{m}_{MAP} = \mathbf{m}_{pr} + (\mathbf{C}_M^{-1} + \mathbf{G}^T \mathbf{C}_D^{-1} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{C}_D^{-1}(\mathbf{d}_{obs} - \mathbf{G}\mathbf{m}_{pr}) \tag{1.7}$$

and

$$\mathbf{m}_{MAP} = \mathbf{m}_{pr} + \mathbf{C}_M \mathbf{G}^T (\mathbf{C}_D + \mathbf{G}\mathbf{C}_M\mathbf{G}^T)^{-1}(\mathbf{d}_{obs} - \mathbf{G}\mathbf{m}_{pr}). \tag{1.8}$$

The selection of the two equations has a prominent influence on the computational efficiency when the difference between the number of model parameters and the number of data is large. For example, if the number of model parameters is far more than the number of data, i.e., $N_m \gg N_d$, which is the usual case in petroleum reservoir characterization problems, expression (1.8) is more favorable for the computation of $\mathbf{m}_{MAP}$ than (1.7) because the time-consuming part of the matrix inversion involved in (1.8) ($N_d \times N_d$ dimensions) is easier to compute than the one in (1.7) ($N_m \times N_m$ dimensions). In the same way, the covariance matrix $\mathbf{C}_{MAP}$ is given as

$$\mathbf{C}_{MAP} = (\mathbf{C}_M^{-1} + \mathbf{G}^T \mathbf{C}_D^{-1} \mathbf{G})^{-1} \tag{1.9}$$

or equivalently

$$\mathbf{C}_{MAP} = \mathbf{C}_M - \mathbf{C}_M \mathbf{G}^T (\mathbf{C}_D + \mathbf{G}\mathbf{C}_M\mathbf{G}^T)^{-1}\mathbf{G}\mathbf{C}_M. \tag{1.10}$$

In this linear Gaussian case, the posterior PDF is fully characterized by $\mathbf{m}_{MAP}$ and $\mathbf{C}_{MAP}$. However, when $g(\mathbf{m})$ is nonlinear, the posterior PDF may be multimodal which means $S(\mathbf{m})$ may have multiple local minima or even global minima so that the estimate $\mathbf{m}_{MAP}$ is not unique. In this case, the initial guess of $\mathbf{m}$ is critical for the convergence of gradient-based methods.

1.2.5 Sampling the posterior distribution

In a Bayesian framework, uncertainty is measured by probability. In petroleum reservoir characterization, it is generally impractical to evaluate the entire posterior PDF to quantify the uncertainty in reservoir properties and in reservoir predictions. The feasible way to explore the posterior PDF is by sampling. Following this idea, the focal point for uncertainty quantification is shifted to how to generate samples from the posterior PDF correctly and efficiently. If a set of model realizations can be obtained as a sample from the posterior PDF, the uncertainty in reservoir performance predictions are possibly being evaluated by constructing statistics (e.g. histogram, mean and variance) from the predicted outcomes of the sampled model realizations.

Among various sampling methods, Markov chain Monte Carlo (McMC) and the randomized maximum likelihood (RML) methods have been extensively examined to quantify the posterior uncertainty in reservoir characterization.

McMC

McMC is an exact method for sampling which provides an iterative procedure for generating samples from a target PDF, $p(\mathbf{m}|\mathbf{d}_{obs})$ of our interest, by constructing a Markov chain where the probability of introducing a new state in the chain depends only on the current state. Theoretically, under satisfied conditions on the chain, McMC method is guaranteed to sample the desired posterior distribution correctly in the limit as the number of states in the chain goes to infinity regardless of the initial state [49]. The most frequently used scheme is perhaps the Metropolis-Hastings algorithm [55] which is known as a generalization of the Metropolis algorithm [87]. At each iteration, a new state or realization $\mathbf{m}^*$ is sampled from a proposal PDF of $q(\mathbf{m}^*|\mathbf{m})$ and is accepted with the probability of $\alpha(\mathbf{m}, \mathbf{m}^*)$. If the candidate state is rejected, the Markov chain will repeat the current state. A generic implementation procedure of McMC based on Metropolis-Hastings is presented in the Algorithm 1.2.1.

Algorithm 1.2.1 McMC (Metropolis-Hastings)

1. Choose an initial state of the Markov chain $\mathbf{m}_0$ and set the iteration index $\ell = 1$.

2. Generate a sample $\mathbf{m}_\ell^*$ from the proposal PDF, $q(\mathbf{m}_\ell^*|\mathbf{m}_{\ell-1})$

3. Compute the acceptance probability of $\mathbf{m}_\ell^*$ as

$$\alpha(\mathbf{m}_{\ell-1}, \mathbf{m}_\ell^*) = \min\left\{1, \frac{q(\mathbf{m}_{\ell-1}|\mathbf{m}_\ell^*)p(\mathbf{m}_\ell^*|\mathbf{d}_{obs})}{q(\mathbf{m}_\ell^*|\mathbf{m}_{\ell-1})p(\mathbf{m}_{\ell-1}|\mathbf{d}_{obs})}\right\}.$$

4. Take

$$\mathbf{m}_\ell = \begin{cases} \mathbf{m}_\ell^* & \text{with probability } \alpha(\mathbf{m}_{\ell-1}, \mathbf{m}_\ell^*), \\ \mathbf{m}_{\ell-1} & \text{with probability } 1 - \alpha(\mathbf{m}_{\ell-1}, \mathbf{m}_\ell^*). \end{cases}$$

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

In the Metropolis-Hastings algorithm, the chain will asymptotically produce draws from $p(\mathbf{m}|\mathbf{d}_{obs})$ as long as the proposal scheme ensures a positive probability of going from any one state to another in a finite number of transitions. It is worth noting that generating samples from the desired posterior PDF using McMC methods is advantageous when the computation of the normalization constant, i.e. $\mathbf{a}$ in (1.4), is intractable because it cancels out shown at the third step as above. For specific applications, there are some implementation issues that need to be considered before using McMC to explore the target distribution. These may include choosing a sample plan (single long run or multiple relative short runs), determining burn-in, setting the run length and monitoring the performance of the sampler. More detailed discussions on this topic can be found in [49] and [114].

McMC methods have been applied and investigated in many history-matching applications [28, 57, 94, 115]. The major difficulty encountered when implementing McMC particularly in high-dimensional problems is the slow convergence and mixing. A large number of iterations is typically required for McMC to produce reasonable samples from the target distribution. However, in reservoir characterization problems, the computation of acceptance probability of $\alpha(\mathbf{m}, \mathbf{m}^*)$ requires a run of the reservoir simulator to evaluate the likelihood part of $p(\mathbf{m}|\mathbf{d}_{obs})$. This prohibitively restricts the

use of McMC in realistic reservoir problems in terms of the current computing capability. Even though McMC is computationally expensive and unfeasible for uncertainty quantification of reservoir performance, it is often used to validate the performance of approximate methods when the generation of a large number of realizations is manageable.

RML

The randomized maximum likelihood (RML) method was independently proposed by Kitanidis [69] and Oliver et al. [95] as an approximate sampling method. In an effort to improve mixing of McMC algorithm, Oliver et al. [95] presented a two-step proposal mechanism such that the resulting state has a high acceptance probability. It proceeded by firstly generating unconditional realizations of model and data variables from their prior distributions which are known and then calibrating the sampled model variables to the sampled data variables. The Metropolis-Hastings algorithm was used to decide if the calibrated model variables are accepted as a state of the chain. However, because the acceptance rate was very high and the acceptance criterion was difficult to evaluate, the authors suggested that the acceptance test could be ignored and all proposed transitions were accepted.

It can be shown that the RML algorithm samples correctly when the prior PDF is Gaussian and the relationship between the model variables and the data is linear [91, 101]. For nonlinear relationships RML is not guaranteed to sample the posterior PDF correctly and only an approximation, although there is considerable evidence that it is able to sample well for non-Gaussian and nonlinear problems in high dimensions [45, 76, 112, 121]. If we assume the model variables have a Gaussian prior such that $\mathbf{m} \sim \mathcal{N}[\mathbf{m}_{pr}, \mathbf{C}_M]$ and the observation errors are also Gaussian with $\varepsilon \sim \mathcal{N}[0, \mathbf{C}_D]$, samples from the RML algorithm can be generated as follows:

Algorithm 1.2.2 RML

1. Generate an unconditional realization of $\mathbf{m}$ from the prior distribution, $\mathbf{m}^* \sim \mathcal{N}[\mathbf{m}_{pr}, \mathbf{C}_M]$

2. Generate an unconditional realization of the data, $\mathbf{d}^* \sim \mathcal{N}[\mathbf{d}_{obs}, \mathbf{C}_D]$

3. Compute the conditional realization $\mathbf{m}_{rml}$:

$$\mathbf{m}_{rml} = \underset{\mathbf{m}}{\mathrm{argmin}} \left[\frac{1}{2}(\mathbf{m} - \mathbf{m}^*)^T \mathbf{C}_M^{-1}(\mathbf{m} - \mathbf{m}^*) + \frac{1}{2}(g(\mathbf{m}) - \mathbf{d}^*)^T \mathbf{C}_D^{-1}(g(\mathbf{m}) - \mathbf{d}^*) \right].$$

The minimization step is analogous to the computation of the MAP estimate in (1.5), except that the regularization is with respect to the unconditional realization instead of a prior model and the predicted data is matched to the sampled data instead of observed data. The conditional realization, $\mathbf{m}_{rml}$, will stay close to the unconditional realization, $\mathbf{m}^*$, unless the minimization of the data mismatch requires large corrections to the model variables. Additional conditional realizations can be generated by repeating the procedure above with independent unconditional realizations of $\mathbf{m}^*$ and $\mathbf{d}^*$.

The challenging part of RML is to perform the minimization which is required for each sampled realization. There are no specific requirements on the method of minimization in the algorithm, so almost any optimization method could be used to

minimize the objective function. Among the methods used for the minimization in history-matching problems, gradient-based methods are probably the most efficient and computational feasible options for large-scale problems, when the objective function concerned is sufficiently smooth. Newton-like methods including Gauss-Newton and Levenberg-Marquardt algorithms have been used to generate conditional realizations with RML [73, 120]. In these methods, a significant part of the efficiency depends on the algorithm that is used to compute the sensitivity of an objective function or predicted data to model parameters. One of the most efficient methods commonly adopted for accurately computing the sensitivity coefficients, i.e. the derivatives of predicted data with respect to model variables is the adjoint method [74, 96]. In this method,the adjoint state variables are solutions of a linear system of equations and computed only once. However, an implementation of the adjoint system requires specific knowledge of the forward discrete equations, which is not commonly available in commercial reservoir simulators.

1.3 Ensemble-based data-assimilation methods

The use of ensemble-based data-assimilation methods has recently become popular within petroleum engineering and has provided encouraging results in terms of reservoir characterization and uncertainty quantification, since Lorentzen et al. [80] first applied the ensemble Kalman filter (EnKF) to dealing with reservoir problems. Oliver and Chen [92] reviewed the recent developments of history-matching technology with the focus on reparameterization techniques, algorithms for computation of sensitivity coefficients and methods for uncertainty quantification. They demonstrated that great progress has been made over the past decade especially in the ability to history match large-scale reservoir models with substantial amounts of uncertain parameters. They also pointed out that the EnKF method is currently the most promising solution to deal with realistic problems among the various methods proposed for performing history matching.

Ensemble-based data-assimilation methods are based on Monte Carlo techniques among which the most well-known form of ensemble-based assimilation is probably the EnKF algorithm. The EnKF is a sequential data-assimilation method in which the model uncertainties are represented by an ensemble that is a group of realizations for model parameters (e.g., permeability and porosity) and for model states (e.g., saturation and pressure). The ensemble itself provides an empirical estimate of the PDF for the model variables. By continuously assimilating new observations, each realization in the ensemble is updated to be consistent with the data using a Bayesian formulation. The method is able to estimate a large number of model variables and integrate different types of data. It is also readily adapt to any reservoir simulator and easy to code and maintain in the sense that neither a tangent nor an adjoint model is required. Additionally, the sequential updating scheme makes EnKF well-suited for real-time data assimilation, reservoir monitoring and performance prediction that are all required for closed-loop reservoir management. All of these attractive features contribute to the popularity of the EnKF method being explored by a rapidly growing number of researchers as a viable alternative to traditional history matching. Significant research has been done on enhancing the performance of the standard EnKF in dealing with

realistic problems. An thorough review of the development of EnKF in reservoir engineering was provided in [1], and a complementary review of the key developments completed after that paper was supplied in [92].

As an alternative ensemble-based assimilation method, the ensemble smoother (ES) was first applied for petroleum history matching by Skjervheim and Evensen [108], mainly motivated by the computational inefficiency of EnKF caused by the frequent restart of the reservoir simulator at every update step, which leads to a severe disadvantage in some real applications. In the ES methodology, all of the available data are assimilated simultaneously with a single *analysis* step taking the similar form to the EnKF, in which only the model parameters are updated so that simulator restarts are avoided. However, the ES in its original form is not as effective as the EnKF especially when the nonlinearity of the concerned problem is strong and when the changes to the model parameters are dramatic [19]. As a result, several iterative versions of the ES were proposed to improve the performance of the standard ES such as the ensemble randomized maximum likelihood method as an iterative ES (batch-EnRML) [19], the ES with multiple data assimilation (ES-MDA) [32] and the Levenberg-Marquardt form of the ES (LM-EnRML) [21]. The EnKF has been successfully applied to a number of synthetic cases (e.g., PUNQ-S3 and Brugge benchmark cases) for history matching [18, 45, 50, 99]. Moreover, there also exist a number of real field applications of the EnKF and the ES for history matching in the literature [12, 56, 108, 109, 124]. In particular, Chen and Oliver [22] lately applied the LM-EnRML to the history matching of the Norne full-field in which a wide variety of approximately 150,000 model parameters were updated and the resulting data matches were better than those obtained by manual history matching.

In the following section, we introduce some widely used ensemble-based methods in the context of petroleum engineering. We first recall some basic concepts of data assimilation and give a brief description of the classical recursive approaches of the Kalman filter (KF) and the extended KF. Then we introduce the EnKF and ES as a variation of the classical recursive methods and as an alternative to traditional history matching. Several typical iterative forms of EnKF and ES proposed for handling strongly nonlinear problems are addressed. Finally, we discuss some common issues of the ensemble-based methods in the applications of petroleum history matching.

1.3.1 Basic concepts of data assimilation

The general data-assimilation problem can be formulated as a joint parameter and state estimation problem based on Bayesian statistics [37]. Let $\mathbf{y} \in \mathbb{R}^{N_y}$ denote a state vector containing the uncertain model variables

$$\mathbf{y}_k = \begin{bmatrix} \mathbf{m} \\ \mathbf{f}_k \end{bmatrix}, \tag{1.11}$$

where $\mathbf{m}$ represents a N_m-dimensional vector of model parameters which are constant in time and $\mathbf{f}$ represents a N_f-dimensional vector of state variables which describe the dynamic state of the system and hence are time-dependent (so $N_y = N_m + N_f$). The subscript k is introduced to indicate the time dependence on model variables, which are assumed to be discretized in time. Now assume that observations are available at

the same discrete set of times as where the model variables are estimated, and denote the observations up to time step n by $\mathbf{d}_{obs,1:n} = \{\mathbf{d}_{obs,1}, \ldots, \mathbf{d}_{obs,n}\}$. When considering Bayesian assimilation, our aim is to evaluate the posterior distribution $p(\mathbf{y}_t|\mathbf{d}_{obs,1:n})$. The state estimation problem can be classified into three categories depending on the comparison of t and n:

- if $t < n$, $p(\mathbf{y}_t|\mathbf{d}_{obs,1:n})$ is called a *smoother* in which data measured after the time of interest t are used for the estimation of the state vector at time t;

- if $t = n$, $p(\mathbf{y}_t|\mathbf{d}_{obs,1:n})$ is called a *filter* in which the state vector at time t is obtained by using data measured up to and including time t;

- if $t > n$, $p(\mathbf{y}_t|\mathbf{d}_{obs,1:n})$ is called a *predictor* whose aim is to forecast the model state in future (or next observation time) by using data measured up to and including time t.

In its smoother form, the conditional PDF of $p(\mathbf{y}_{0:n}|\mathbf{d}_{obs,1:n})$ can be given by Bayes' theorem

$$p(\mathbf{y}_{0:n}|\mathbf{d}_{obs,1:n}) \propto p(\mathbf{d}_{obs,1:n}|\mathbf{y}_{0:n})p(\mathbf{y}_{0:n}), \tag{1.12}$$
$$\propto p(\mathbf{d}_{obs,n}|\mathbf{y}_{0:n})p(\mathbf{y}_n|\mathbf{y}_{0:n-1})p(\mathbf{y}_{0:n-1}|\mathbf{d}_{obs,1:n-1}), \tag{1.13}$$
$$\propto p(\mathbf{d}_{obs,n}|\mathbf{y}_n)p(\mathbf{y}_n|\mathbf{y}_{n-1})p(\mathbf{y}_{0:n-1}|\mathbf{d}_{obs,1:n-1}), \tag{1.14}$$

where the normalizing constant in the denominator is omitted and hence the equal sign is replaced with the proportional symbol, $\propto$. Expression (1.13) is obtained from (1.12) by assuming that the data collected at different times are independent and the data at a particular time depend only on the states up to and including the this time, i.e.,

$$p(\mathbf{d}_{obs,1:n}|\mathbf{y}_{0:n}) = p(\mathbf{d}_{obs,1}|\mathbf{y}_{0:1})p(\mathbf{d}_{obs,2}|\mathbf{y}_{0:2})\cdots p(\mathbf{d}_{obs,n}|\mathbf{y}_{0:n}),$$
$$= p(\mathbf{d}_{obs,1:n-1}|\mathbf{y}_{0:n-1})p(\mathbf{d}_{obs,n}|\mathbf{y}_{0:n}). \tag{1.15}$$

The expression (1.14) is a simplification of (1.13) by further assuming that the forward model is a first-order Markov process, i.e.

$$p(\mathbf{y}_n|\mathbf{y}_{0:n-1}) = p(\mathbf{y}_n|\mathbf{y}_{n-1}), \tag{1.16}$$

and that $\mathbf{d}_k$ only depends on $\mathbf{y}_k$ for $k = 1, \ldots, n$. In (1.14), $p(\mathbf{y}_{0:n-1}|\mathbf{d}_{obs,1:n-1})$ is the posterior PDF at previous time $n - 1$, and $p(\mathbf{y}_n|\mathbf{y}_{n-1})$ represents the forward model that advances from time $n - 1$ to time n. The product of the these two terms represents the prior distribution at time n. The conditional PDF $p(\mathbf{d}_{obs,n}|\mathbf{y}_n)$ is the likelihood function for the data at time n.

These three equations simply indicate that the smoother solution can be evaluated either by assimilating all data simultaneously with (1.12) or by integrating the data sequentially as they become available with the recursive formulas (1.13) and (1.14). When the assumptions are satisfied, the two approaches are equivalent [40]. Note that the smoother solution $\mathbf{y}_k$ at all times is conditioned to all of the measurements including the ones from subsequent times.

On the other hand, the filter solution may be obtained by integrating (1.14) over $\mathbf{y}_{0:n-1}$ and can be written as

$$
\begin{aligned}
p(\mathbf{y}_n|\mathbf{d}_{obs,1:n}) &\propto \int_{\mathbf{y}_{0:n-1}} p(\mathbf{y}_{1:n}|\mathbf{d}_{obs,1:n}) d\mathbf{y}_{1:n-1}, \\
&\propto p(\mathbf{d}_{obs,n}|\mathbf{y}_n) p(\mathbf{y}_n|\mathbf{d}_{obs,1:n-1}),
\end{aligned}
\tag{1.17}
$$

where $p(\mathbf{y}_n|\mathbf{d}_{obs,1:n-1})$ is the prior PDF given by

$$
p(\mathbf{y}_n|\mathbf{d}_{obs,1:n-1}) = \int_{\mathbf{y}_{0:n-1}} p(\mathbf{y}_n|\mathbf{y}_{n-1}) p(\mathbf{y}_{1:n-1}|\mathbf{d}_{obs,1:n-1}) d\mathbf{y}_{1:n-1}.
\tag{1.18}
$$

In comparison with the smoother solution, the filter solution is calculated based only on the measurements up to and including the current time step. However, the two solutions will be identical at the last data-assimilation time step.

In summary, the Bayesian data-assimilation process can be separated into an *analysis* (or *update*) step and a *forecast* step. At the *analysis* step, an improved estimate of model variables is obtained by applying Bayes' theorem to compute the posterior PDF for the model variables given the observations, for instance, with (1.14) or (1.17). At the *forecast* step, the dynamical model is advanced in time and its result becomes the forecast (or prior) in the next *analysis* cycle. Even though the Bayesian assimilation provides a fundamental framework for updating model variables with observed data, the solution to (1.14) or (1.17) is generally difficult to be found without some simplification. If we put the problem under the assumptions of linear dynamics and Gaussian statistics, the KF method provides a closed-form solution to the filter problem of (1.17).

1.3.2 Kalman filter and extended Kalman filter

The KF is an efficient sequential data-assimilation algorithm which was first introduced by Kalman [67] in an engineering context as a recursive solution to linear filtering and prediction problems. In the original paper [67], the KF was derived as a minimum variance estimator with the orthogonal projection method. However, the KF can also be well interpreted from a Bayesian standpoint [66, 86]. For a linear Gaussian problem, the KF is an optimal sequential assimilation algorithm in the sense that it provides the best linear unbiased estimate (BLUE) of the posterior mean and the posterior covariance. Because the prior and the likelihood are both assumed to be Gaussian, the posterior is also Gaussian which is completely characterized by the mean and the covariance.

Let us consider a linear filtering system for which the state and measurement equations are described as

$$
\mathbf{y}_k = \mathbf{A}_k \mathbf{y}_{k-1} + \omega_{k-1},
\tag{1.19}
$$

$$
\mathbf{d}_{obs,k} = \mathbf{G}_k \mathbf{y}_k + \varepsilon_{\mathbf{k}},
\tag{1.20}
$$

where $\mathbf{A}_k$ represents a linear model operator and $\mathbf{G}_k$ denotes an observation operator which defines a linear relationship between the state vector $\mathbf{y}_k$ and the measurement vector $\mathbf{d}_{obs,k}$. The sequences $\{\omega_k\}_{k=1}^n$ and $\{\varepsilon_k\}_{k=1}^n$ are the model noise and the measurement noise respectively. Now we make the following assumptions on the error statistics:

- Gaussian model and observation noises: $\omega_k \sim \mathcal{N}(0, \mathbf{C}_{\omega_k})$ and $\varepsilon_k \sim \mathcal{N}(0, \mathbf{C}_{\varepsilon_k})$, where $\mathbf{C}_{\omega_k}$ and $\mathbf{C}_{\varepsilon_k}$ are model and observation covariance matrices respectively.

- $\mathbf{y}_k$, ω_k and ε_k are mutually independent: $\mathbb{E}[\mathbf{y}_k \omega_l^T] = 0$ and $\mathbb{E}[\varepsilon_k \omega_l^T] = 0$ for all k, l; $\mathbb{E}[\mathbf{y}_k \varepsilon_l^T] = 0$ for $k \leq l$.

Furthermore, we assume that the initial prior PDF of $\mathbf{y}_0$ is Gaussian $\mathbf{y}_0 \sim \mathcal{N}(\mu_0, \mathbf{C}_{y_0})$. Based on these assumptions, all the PDFs in (1.17) are Gaussian. The mean and the covariance matrix of $p(\mathbf{y}_k|\mathbf{d}_{obs,k})$ are calculated by

$$\mathbb{E}[\mathbf{d}_{obs,k}|\mathbf{y}_k] = \mathbb{E}[\mathbf{G}_k \mathbf{y}_k + \varepsilon_k] = \mathbf{G}_k \mathbf{y}_k, \tag{1.21}$$

$$\mathrm{Cov}[\mathbf{d}_{obs,k}|\mathbf{y}_k] = \mathrm{Cov}[\varepsilon_k|\mathbf{y}_k] = \mathbf{C}_{\varepsilon_k}, \tag{1.22}$$

and hence the likelihood $p(\mathbf{d}_{obs,k}|\mathbf{y}_k)$ can be written as

$$p(\mathbf{d}_{obs,k}|\mathbf{y}_k) \propto \exp\left[-\frac{1}{2}(\mathbf{d}_{obs,k} - \mathbf{G}_k \mathbf{y}_k)^T \mathbf{C}_{\varepsilon_k}^{-1}(\mathbf{d}_{obs,k} - \mathbf{G}_k \mathbf{y}_k)\right]. \tag{1.23}$$

Given the posterior distribution at time $k-1$, $p(\mathbf{y}_{k-1}|\mathbf{d}_{obs,1:k-1})$,

$$\mathbf{y}_{k-1} \sim \mathcal{N}(\mathbf{y}_{k-1}^a, \mathbf{C}_{\omega_{k-1}}^a), \tag{1.24}$$

the mean $\mathbf{y}_k^f$ and the covariance $\mathbf{C}_{y_{k-1}}^f$ of the prior $p(\mathbf{y}_k|\mathbf{d}_{obs,1:k-1})$ are calculated by

$$\mathbf{y}_k^f = \mathbb{E}[\mathbf{y}_k|\mathbf{d}_{obs,1:k-1}] = \mathbf{A}_k \mathbf{y}_{k-1}^a, \tag{1.25}$$

$$\mathbf{C}_{y_k^f} = \mathbb{E}[(\mathbf{y}_k - \mathbf{y}_k^f)(\mathbf{y}_k - \mathbf{y}_k^f)^T|\mathbf{d}_{obs,1:k-1}] = \mathbf{A}_k \mathbf{C}_{y_{k-1}}^a \mathbf{A}_k^T + \mathbf{C}_{\omega_{k-1}}, \tag{1.26}$$

and then

$$p(\mathbf{y}_k|\mathbf{d}_{obs,1:k-1}) \propto \exp\left[-\frac{1}{2}(\mathbf{y}_k - \mathbf{y}_k^f)^T \mathbf{C}_{y_k^f}^{-1}(\mathbf{y}_k - \mathbf{y}_k^f)\right]. \tag{1.27}$$

By substituting equations (1.23) and (1.27) into (1.17), the posterior can be written as

$$p(\mathbf{y}_k|\mathbf{d}_{obs,1:k}) \propto \exp\left[-\frac{1}{2}(\mathbf{d}_{obs,k} - \mathbf{G}_k \mathbf{y}_k)^T \mathbf{C}_{\varepsilon_k}^{-1}(\mathbf{d}_{obs,k} - \mathbf{G}_k \mathbf{y}_k)\right.$$
$$\left. -\frac{1}{2}(\mathbf{y}_k - \mathbf{y}_k^f)^T \mathbf{C}_{y_k^f}^{-1}(\mathbf{y}_k - \mathbf{y}_k^f)\right]. \tag{1.28}$$

The best estimate, $\mathbf{y}_k$, is the state vector that maximizes the posterior PDF in (1.28), which is often referred to as the MAP estimate mentioned in Section 1.2.4. In particular, the MAP estimate is also the posterior mean for the Gaussian case. Taking the derivative of $p(\mathbf{y}_k|\mathbf{d}_{obs,1:k})$ or equivalently $\ln p(\mathbf{y}_k|\mathbf{d}_{obs,1:k})$ with respect to $\mathbf{y}_k$, and setting it to zero, i.e.

$$\left.\frac{\partial \ln p(\mathbf{y}_k|\mathbf{d}_{obs,1:k})}{\partial \mathbf{y}_k}\right|_{\mathbf{y}_k = \mathbf{y}_k^a} = 0, \tag{1.29}$$

the MAP estimate of $\mathbf{y}_k^a$ and the corresponding posterior covariance, $\mathbf{C}_{y_k^a}$, are obtained [96]

$$\mathbf{y}_k^a = \mathbf{y}_k^f + \mathbf{K}_k(\mathbf{d}_{obs,k} - \mathbf{G}_k \mathbf{y}_k^f), \tag{1.30}$$

$$\mathbf{C}_{y_k^a} = \mathbf{C}_{y_k^f} - \mathbf{K}_k \mathbf{G}_k \mathbf{C}_{y_k^f}, \tag{1.31}$$

where $\mathbf{K}_k \in \mathbb{R}^{N_y \times N_d}$ is known as the Kalman gain that is given by

$$\mathbf{K}_k = \mathbf{C}_{\mathbf{y}_k^f} \mathbf{G}_k^T (\mathbf{G}_k \mathbf{C}_{\mathbf{y}_k^f} \mathbf{G}_k^T + \mathbf{C}_{\varepsilon_k})^{-1}. \tag{1.32}$$

As the derivation shows, the KF scheme consists of two steps. At the *forecast* step, the mean vector and the covariance matrix characterizing the posterior PDF at previous time (or the initial prior PDF) are updated with (1.25) and (1.26) to predict the prior PDF of model variables before conditioning to new data. At the following *analysis* step, the prior estimates of the mean and the covariance matrix are improved by assimilating the new data using equations (1.30-1.32). Because the posterior distribution, $p(\mathbf{y}_k|\mathbf{d}_{obs,1:k})$, is Gaussian, it is fully determined by the mean $\mathbf{y}_k^a$ and the covariance $\mathbf{C}_{y_k^a}$.

There are two major difficulties in applying the KF to geophysical data-assimilation problems: handling large amounts of model variables and data, and dealing with nonlinear dynamics. As we discussed above, the KF requires repeated computation of covariance matrices at each update step. When the number of model variables and/or observations is very large (it is typical for reservoir history matching), the computational cost and the storage on the covariance matrices at the *forecast* and *analysis* steps will become unaffordable. In addition, the KF is based on the assumption of linear model dynamics, which will not hold for the reservoir flow problems.

The extended KF is an adaptation of the KF method for nonlinear problems [66, 85]. Let us consider the nonlinear state-space model

$$\mathbf{y}_k = h_k(\mathbf{y}_{k-1}) + \omega_{k-1}, \tag{1.33}$$
$$\mathbf{d}_{obs,k} = g_k(\mathbf{y}_k) + \varepsilon_k, \tag{1.34}$$

where h_k and g_k represent the nonlinear model operator and the nonlinear measurement operator respectively. The extended KF uses linearization techniques to approximate (1.33) and (1.34) by defining

$$\mathbf{A}_k = \left. \frac{\partial h_k(\mathbf{y}_{k-1})}{\partial \mathbf{y}_{k-1}} \right|_{\mathbf{y}_{k-1}=\mathbf{y}_{k-1}^a}, \tag{1.35}$$

$$\mathbf{G}_k = \left. \frac{\partial g_k(\mathbf{y}_k)}{\partial \mathbf{y}_k} \right|_{\mathbf{y}_k=\mathbf{y}_k^f}. \tag{1.36}$$

Based on the linearized dynamics, the conventional KF is further employed. Unlike the KF, the extended KF is not an optimal filter in any sense. For weakly nonlinear problems, the extended KF usually reasonable good performance when used carefully. However, in cases of highly nonlinear models the extended KF may become problematic and even fail. Moreover, the extended KF is not suitable for large-scale systems either, due to the problems with maintaining the covariance matrix.

1.3.3 Ensemble Kalman filter

The EnKF method was originally introduced by Evensen [34] as an improvement to the extended KF to handle data assimilation in high-dimensional nonlinear systems. The EnKF is a modification of the KF based on Monte Carlo method in which an ensemble of model states is employed to represent the probability density. The basic idea

of the EnKF is to use the ensemble of states to approximate mean vectors and covariance matrices, based on which each member of the ensemble is estimated using the KF update equation and integrated forward in time independently using a forecast model. The predicted ensemble provides an approximation of the prior PDF before assimilation of new data for the subsequent *analysis* cycle. Because the mean and covariance are approximated with the sample mean and covariance computed from the ensemble, the EnKF avoids the disadvantages of storing and evolving the covariance matrix explicitly required by the KF and the extended KF. For typical EnKF applications, the number of ensemble members is often much smaller the number of uncertain model variables, so the computational savings can be significant compared with the use of KF or the extended KF (if they are still applicable). Moreover, the EnKF does not require any sensitivity calculations which enable the method to be easily adaptable to different physical systems and model variables. Therefore, the EnKF methodology is very suitable and practical for solving large-scale problems in many science fields.

Instead of introducing the EnKF in a general sense, here we describe the method specially in the context of petroleum reservoir history matching. It is useful to define an augmented state vector, by adding the predicted data $\mathbf{d}_k$ into the vector of model parameters and state variables, when applying the EnKF to nonlinear problems, i.e.,

$$\widetilde{\mathbf{y}}_k = \begin{bmatrix} \mathbf{y}_k \\ \mathbf{d}_k \end{bmatrix} = \begin{bmatrix} \mathbf{m} \\ \mathbf{f}_k \\ \mathbf{d}_k \end{bmatrix}. \tag{1.37}$$

where

$$\mathbf{f}_k = h_k(\mathbf{m}), \qquad \mathbf{d}_k = g_k(\mathbf{m}). \tag{1.38}$$

For a typical reservoir history-matching problem,

- $\mathbf{m}$ consists of the time-invariant variables for rock properties and flow system such as the gridblock permeabilities, porosities, etc.,

- $\mathbf{f}_k$ includes the time-varying variables describing the dynamic states of the reservoir such as gridblock pressures, saturations, etc.,

- $\mathbf{d}_k$ contains the predictions of observations or theoretical data like water and oil production rates, bottom-hole pressures, and so on.

Note that the sate variables and data are both expressed as functions of the model parameters in (1.38), because the state variables can be determined by the model parameters when the initial sate of the reservoir is given (which is a common assumption).

With this definition of $\widetilde{\mathbf{y}}_k$, we can rewrite the relation between the state vector and the data in (1.34) as a linear equation [36]

$$\mathbf{d}_k = \mathbf{H}\widetilde{\mathbf{y}}_k, \tag{1.39}$$

where

$$\mathbf{H} = [\mathbf{0} \ \mathbf{I}_{N_d}], \tag{1.40}$$

and $\mathbf{0} \in \mathbb{R}^{N_d \times (N_m + N_f)}$ is the null matrix with only zero elements and $\mathbf{I}_{N_d} \in \mathbb{R}^{N_d \times N_d}$ is the identity matrix. Although expanding the state vector with the predicted data allows

an easy derivation of the EnKF within the framework of the standard KF methodology, Li and Reynolds [72] show that the EnKF can sample correctly from the marginal PDF of $\mathbf{y}_k(\cdot)$ only when the relation between $\mathbf{y}_k(\cdot)$ and $\mathbf{d}_k(\cdot)$ is linear at each updating step. So using the augmented state vector will not remove the influence of the nonlinearity on the performance of the EnKF.

From the implementation point of view, the EnKF algorithm can be divided into two stages: generation of the initial ensemble and sequential data assimilation. Let us denote the ensemble of augmented state vectors at time step k by $\mathbf{Y}_k$:

$$\mathbf{Y}_k = [\widetilde{\mathbf{y}}_{k,1} \ \widetilde{\mathbf{y}}_{k,2} \ \cdots \ \widetilde{\mathbf{y}}_{k,N_e}] \in \mathbb{R}^{N_{\widetilde{\mathbf{y}}} \times N_e}, \tag{1.41}$$

where N_e is the number of ensemble members and $N_{\widetilde{\mathbf{y}}} = N_m + N_f + N_d$. At the first stage, an initial ensemble of N_e independent samples $\mathbf{m}_{0,j}$ for $j = 1, \ldots, N_e$ is generated consistent with prior knowledge of the initial state and its probability distribution. In reservoir history-matching problems, the initial state of the reservoir is often assumed known and the joint probability of reservoir parameters before assimilation of data can be characterized. Therefore, the estimation of the prior PDF of $p(\mathbf{y}_0)$ is available. For the vector of model parameters m, the posterior solution after the *analysis* step of EnKF at all times is a linear combination of the initial ensemble, so it is important that the initial ensemble of model parameters characterizes the prior uncertainty properly.

By advancing each member of the initial ensemble usually in a parallel manner forward to the first observation time, we have already entered the *forecast* step of the second stage which is a recursive two-step process. Let $\widetilde{\mathbf{Y}}_k^f$ denote the forecast ensemble of augmented state vectors at time step k, for $k = 1, 2, \ldots, n$, i.e.,

$$\widetilde{\mathbf{Y}}_k^f = [\widetilde{\mathbf{y}}_{k,1}^f \ \widetilde{\mathbf{y}}_{k,2}^f \ \cdots \ \widetilde{\mathbf{y}}_{k,N_e}^f], \tag{1.42}$$

with

$$\widetilde{\mathbf{y}}_{k,j}^f = \Gamma(\widetilde{\mathbf{y}}_{k-1,j}^a), \tag{1.43}$$

where $\Gamma(\cdot)$ represents the reservoir simulator and the the superscript a denotes the posterior (analyzed) solution from the previous time step. It is easy to compute the mean $\overline{\widetilde{\mathbf{Y}}}_k^f$ and the covariance matrix $\mathbf{C}_{\widetilde{\mathbf{y}}_k^f}$ from the forecast ensemble:

$$\overline{\widetilde{\mathbf{Y}}}_k^f = [\overline{\widetilde{\mathbf{y}}}_k^f \ \overline{\widetilde{\mathbf{y}}}_k^f \ \cdots \ \overline{\widetilde{\mathbf{y}}}_k^f], \qquad \text{with} \quad \overline{\widetilde{\mathbf{y}}}_k^f = \frac{1}{N_e} \sum_{j=1}^{N_e} \widetilde{\mathbf{y}}_{k,j}^f, \tag{1.44}$$

$$\mathbf{C}_{\widetilde{\mathbf{y}}_k^f} = \frac{(\widetilde{\mathbf{Y}}_k^f - \overline{\widetilde{\mathbf{Y}}}_k^f)(\widetilde{\mathbf{Y}}_k^f - \overline{\widetilde{\mathbf{Y}}}_k^f)^T}{N_e - 1} = \frac{\Delta\widetilde{\mathbf{Y}}_k^f (\Delta\widetilde{\mathbf{Y}}_k^f)^T}{N_e - 1}. \tag{1.45}$$

At the following *analysis* step, each ensemble member is updated using the KF analysis equation based on the sample approximation to the mean and covariance,

$$\widetilde{\mathbf{y}}_{k,j}^a = \widetilde{\mathbf{y}}_{k,j}^f + \mathbf{K}_k(\mathbf{d}_{obs,k}^j - \mathbf{H}\widetilde{\mathbf{y}}_{k,j}^f), \qquad \text{for } j = 1, 2, \ldots, N_e, \tag{1.46}$$

and

$$\mathbf{K}_k = \mathbf{C}_{\widetilde{\mathbf{y}}_k^f} \mathbf{H}^T (\mathbf{H}\mathbf{C}_{\widetilde{\mathbf{y}}_k^f} \mathbf{H}^T + \mathbf{C}_{d_{obs}})^{-1}, \tag{1.47}$$

where $\mathbf{d}_{obs,k}^j \sim \mathcal{N}(\mathbf{d}_{obs,k}, \mathbf{C}_{d_{obs}})$ is a sample from the measurement distribution obtained by adding stochastic perturbations to the observed data vector $\mathbf{d}_{obs,k}$ [13, 59]. Like the KF, the *analysis* step of EnKF is only optimal for linear model with Gaussian statistics and it can shown that the analyzed ensemble mean and covariance from EnKF asymptotically converges to the corresponding estimates given by the KF as the ensemble size goes to infinity [14].

In fact, the covariance matrix $\mathbf{C}_{\widetilde{\mathbf{y}}_k^f}$ is not necessary to be explicitly formed during the computation. In order to emphasize the computational efficiency of EnKF, we rewrite the update equation (1.46) in the following matrix form:

$$\begin{bmatrix} \mathbf{M}_k^a \\ \mathbf{F}_k^a \end{bmatrix} = \begin{bmatrix} \mathbf{M}_k^f \\ \mathbf{F}_k^f \end{bmatrix} + \begin{bmatrix} \mathbf{C}_{md}\mathbf{X}_k \\ \mathbf{C}_{fd}\mathbf{X}_k \end{bmatrix}, \qquad (1.48)$$

and

$$\mathbf{X}_k = (\mathbf{C}_{dd} + \mathbf{C}_{d_{obs}})^{-1}(\mathbf{D}_k - \mathbf{H}\widetilde{\mathbf{Y}}_k^f), \qquad (1.49)$$

where the matrices $\mathbf{M}_k$ and $\mathbf{F}_k$ are the ensembles of model parameters and state variables respectively. $\mathbf{C}_{md}$ is the covariance between model parameters and data, $\mathbf{C}_{fd}$ is the covariance between state variables and data, and $\mathbf{C}_{dd}$ is the covariance of simulated data, where the time step k is not attached for notational simplicity. Because $\mathbf{C}_{md}$ and $\mathbf{C}_{fd}$ are estimated from the ensemble

$$\mathbf{C}_{md} = \frac{\Delta\mathbf{M}_k\Delta\mathbf{D}_k}{N_e - 1}, \qquad \mathbf{C}_{fd} = \frac{\Delta\mathbf{F}_k\Delta\mathbf{D}_k}{N_e - 1}, \qquad \mathbf{C}_{dd} = \frac{\Delta\mathbf{D}_k\Delta\mathbf{D}_k}{N_e - 1}, \qquad (1.50)$$

and do not need to be formed in implementation (only $\Delta\mathbf{M}_k \in \mathbb{R}^{N_m \times N_e}$, $\Delta\mathbf{F}_k \in \mathbb{R}^{N_f \times N_e}$, $\Delta\mathbf{D}_k \in \mathbb{R}^{N_d \times N_e}$ used and stored), the computational cost of EnKF largely depends on the ensemble size (typically 10 to 100) which is much smaller than the number of model variables in most cases. Thus, the EnKF algorithm is very efficient, easy to code and robust on storage for high-dimensional models. However, the ranks of $\mathbf{C}_{md}$ and $\mathbf{C}_{fd}$ cannot be larger than $N_e - 1$ since they are calculated from the ensemble and the quality of the estimates highly depends on the size of the ensemble.

One advantage of the EnKF is that the ensemble forecast of reservoir performance after each *analysis* step can be made by running the forward model from the most recent data-assimilation time without the necessity of rerunning the whole past history. When the relationship between model parameters and state variables is linear, it can be shown that the updated state variables is consistent with those obtained by rerunning the forward model from the initial time [51, 113]. Otherwise, the consistency may not hold. To overcome this problem in nonlinear problems, Wang et al. [118] proposed a half iterative EnKF (HIEnKF) in which only model parameters are updated using the EnKF analysis equation (first row of (1.48) above) and the predicted data after each data-assimilation time are obtained by rerunning the simulator from time zero with the updated ensemble of model parameters. The HIEnKF retains the sequential feature of EnKF, however, at the expense of repeatedly restarting simulation runs. As shown in [119], the HIEnKF is equivalent to the first iteration of the ensemble randomized maximum likelihood (EnRML) [51] with a full step length.

1.3.4 Ensemble smoother

The ES was first introduced by van Leeuwen and Evensen [117] in the context of oceanographic and atmospheric applications. Similar to the EnKF, the ES approximate the same Bayesian formulation using Monte Carlo methods via a variance-minimizing update scheme. However, ES differs from EnKF by assimilating data at all times simultaneously instead of integrating data sequentially in time as in EnKF. So the ES requires only a single update step. If we ignore the errors in the forward model which is common in reservoir history-matching problems, we only need to consider the parameter estimation problem while applying the ES. In this case, the state vector only includes model parameters and the analysis equation is similar to the one for EnKF, written as

$$\mathbf{m}_j^a = \mathbf{m}_j^f + \mathbf{C}_{md}(\mathbf{C}_{dd} + \mathbf{C}_{d_{obs}})^{-1}(\mathbf{d}_{obs,j} - \mathbf{d}_j), \qquad \text{for } j = 1, 2, \ldots, N_e. \tag{1.51}$$

Since the same notations are used here, it is necessary to clarify some definitions of the terms in (1.51) which are a little different from the ones for the EnKF. Other than including only the measurements at a certain assimilation time as for EnKF, $\mathbf{d}_{obs,j} \in \mathbb{R}^{N_n}$ in (1.51) contains all of the measurements made at all times, where $N_n = \sum_{i=1}^{n} N_{d_i}$ and N_{d_i} denotes the number of measurements at ith time step. The cross-covariance matrix $\mathbf{C}_{md} \in \mathbb{R}^{N_m \times N_n}$ is computed based on the initial ensemble of model parameters and its corresponding ensemble of predicated data. The matrices $\mathbf{C}_{dd} \in \mathbb{R}^{N_n \times N_n}$ and $\mathbf{C}_{d_{obs}} \in \mathbb{R}^{N_n \times N_n}$ are the covariances of predicted data and observed data, respectively.

For linear problems, the ES provides identical results with the EnKF [38]. For nonlinear problems, however, the EnKF has been shown to perform better than the ES. van Leeuwen and Evensen [117] first tested the ES on an ocean circulation model where they found that the ES was outperformed by the EnKF. The inferior performance of the ES was further proved on the following tests on a Lorentz model [35, 40]. The best explanation for the better performance of EnKF was attributed to its sequential updating scheme which appears to mitigate the chaotic dynamics of the problem and keep the model closer to the true solution. Skjervheim and Evensen [108] introduced the use of ES for petroleum history matching, motivated by the fact that reservoir-simulation models are different from the chaotic dynamic systems that were previously used to test ES. They compared ES and EnKF on both a synthetic case and a real North-Sea field application where both methods gave comparable results. However, this is not always the case as shown in [19, 32] that ES performed poorly compared to EnKF. Nevertheless, in many real petroleum reservoir applications there is a strong practical motivation to use ES instead of EnKF [108]. This is due to the fact that the cost of frequent restarts of reservoir simulator and convergence issues related to the updates to state variables can be significant while performing EnKF in some applications. In contrast, the ES avoids these disadvantages because in its implementation all data are assimilated at a single step and only the model parameters are updated. The two examples presented in [108] showed that ES required only 10% of the simulation time used by EnKF.

1.3.5 Iterative forms of EnKF / ES

The standard EnKF has been shown to give acceptable history matches and uncertainty estimations in many reservoir applications [39, 50, 90]. Even though the EnKF was

originally proposed as an alternative to the extended KF for applications in nonlinear dynamical systems, the update step in the EnKF is still linear. For highly nonlinear problems, however, the EnKF may yield unphysical updates of model variables and poor data matches. For example, porosity and saturation may be lower than zero or greater than one. In such cases, iterative EnKF methods are usually used in order to overcome the high nonlinearity and improve the quality of data match. On the other hand, the ES often requires iteration to achieve satisfactory matches to data especially when the problem is strongly nonlinear [19]. Many iterative methods based on EnKF and ES have been proposed, most of which are achieved through the way of minimizing a stochastic objective function or equivalently maximizing the posterior probability of each realization [1]. Here, we introduce several typical iterative forms of EnKF and ES based on RML and multiple data assimilation (MDA).

EnRML

The ensemble randomized maximum likelihood (EnRML) was first introduced by Gu and Oliver [51] as an iterative EnKF. The method was then generalized by Chen and Oliver [19] who propose to use EnRML as an iterative ES. Following the notations used in [19], we refer to [51] as seq-EnRML, and [19] as batch-EnRML. In a recent paper, Chen and Oliver [21] introduced an improved version of batch-EnRML by using a modified Levenberg-Marquardt (LM) method instead of Gauss-Newton (GN) method, which is called LM-EnRML. Here, we give a brief introduction of these three closely related methods.

As shown in Section 1.2.5, under assumptions of Gaussian PDFs for model parameters and data, one can generate samples of model parameters from the posterior PDF with the RML method by minimizing the objective function (1.5) . RML method by minimizing the objective function (1.5) in which the prior mean has been replaced by a sample from the prior and the observed data has been replaced by a perturbed observation as in Algorithm 1.2.2. Using the GN method, the estimate of a sample of model parameters at the $\ell + 1$th iteration can be written as [111]

$$\mathbf{m}^{\ell+1} = \mathbf{m}^\ell + \beta_\ell \delta\mathbf{m}^{\ell+1}, \tag{1.52}$$

with

$$
\begin{aligned}
\delta\mathbf{m}^{\ell+1} = &-(\mathbf{G}_\ell^T \mathbf{C}_D^{-1} \mathbf{G}_\ell^T + \mathbf{C}_M^{-1})^{-1} \\
&\times [\mathbf{C}_M^{-1}(\mathbf{m}^\ell - \mathbf{m}_{pr}) + \mathbf{G}_\ell^T \mathbf{C}_D^{-1}(g(\mathbf{m}^\ell) - \mathbf{d}_{obs})], \tag{1.53}
\end{aligned}
$$

$$
\begin{aligned}
= &\; \mathbf{m}_{pr} - \mathbf{m}^\ell - \mathbf{C}_M \mathbf{G}_\ell^T (\mathbf{G}_\ell \mathbf{C}_M \mathbf{G}_\ell^T + \mathbf{C}_D)^{-1} \\
&\times [g(\mathbf{m}^\ell) - \mathbf{d}_{obs} - \mathbf{G}_\ell(\mathbf{m}^\ell - \mathbf{m}_{pr})], \tag{1.54}
\end{aligned}
$$

where $g(\cdot)$ denotes the nonlinear relationship between data $\mathbf{d}$ and model parameters $\mathbf{m}$. $\mathbf{G}$ is the sensitivity of data to model variables and β is the step length parameter whose optimal value can be determined using standard methods [24]. In (1.53), the matrix

$$\mathcal{H} = \mathbf{G}_\ell^T \mathbf{C}_D^{-1} \mathbf{G}_\ell^T + \mathbf{C}_M^{-1}, \tag{1.55}$$

is the Hessian matrix. The two forms (1.53) and (1.54) are equivalent and obtained by using the matrix inversion lemmas. As discussed in Section 1.2.4, the computational

efficiency of these two equations depends on the distinction between the dimensions of model parameters and data. In many cases, the dimension of model space is much larger than the dimension of the data space, so the formula (1.54) is often used.

Taking advantage of ensemble approaches like the EnKF, the prior model covariance $\mathbf{C}_M$ and the sensitivity $\mathbf{G}$ can be approximated from a finite number of ensemble members, which are written as

$$\mathbf{C}_M^e = \frac{\Delta \mathbf{M}_{pr} \Delta \mathbf{M}_{pr}^T}{N_e - 1}, \tag{1.56}$$

$$\mathbf{G}_\ell^e = \Delta \mathbf{D}_\ell \Delta \mathbf{M}_\ell^\dagger, \tag{1.57}$$

where $\Delta \mathbf{M}_{pr} \in \mathbb{R}^{N_m \times N_e}$ represent the matrix of deviation realizations of model parameters from the prior mean. The columns of $\Delta \mathbf{M}_\ell \in \mathbb{R}^{N_m \times N_e}$ and $\Delta \mathbf{D}_\ell \in \mathbb{R}^{N_d \times N_e}$ are deviation realizations of model parameters and deviation realizations of predicated data from the mean at ℓth iteration, respectively. Because $\Delta \mathbf{M}_\ell$ is generally not invertible, pseudo-inverse of $\Delta \mathbf{M}_\ell$ is used which is denoted by the superscript $\dagger$ and can be computed with singular value decomposition (SVD) [51]. Typically, $N_e \ll N_m$, so the cost to compute the SVD is reasonable. Utilizing the ensemble approxmiations $\mathbf{C}_M^e$ and $\mathbf{G}_\ell^e$, (1.52) becomes

$$\mathbf{m}^{\ell+1} = \beta_\ell \mathbf{m}_{pr} + (1 - \beta_\ell) \mathbf{m}^\ell - \beta_\ell \mathbf{C}_M^e \mathbf{G}_\ell^{eT} (\mathbf{G}_\ell \mathbf{C}_M^e \mathbf{G}_\ell^{eT} + \mathbf{C}_D)^{-1} \tag{1.58}$$
$$\times [g(\mathbf{m}^\ell) - \mathbf{d}_{obs} - \mathbf{G}_\ell^e (\mathbf{m}^\ell - \mathbf{m}_{pr})],$$

where the search direction $\delta \mathbf{m}^{\ell+1}$ is computed by (1.54). If we set $\ell = 0$, $\beta_0 = 1$ and $\mathbf{m}^0 = \mathbf{m}_{pr}$, the estimate at the first iteration of (1.58) is

$$\mathbf{m}^1 = \mathbf{m}_{pr} + \mathbf{C}_M^e \mathbf{G}_0^{eT} (\mathbf{G}_0 \mathbf{C}_M^e \mathbf{G}_0^{eT} + \mathbf{C}_D)^{-1} [\mathbf{d}_{obs} - g(\mathbf{m}_{pr})]. \tag{1.59}$$

This is the same as the EnKF upate equation (1.46-1.50) except that both state variables and model parameters are updated in EnKF. When the data $\mathbf{d}_{obs}$ contains all the measurements, (1.59) is the ES update identical with (1.51). One advantage of EnKF and ES is that neither $\mathbf{C}_M^e$ nor $\mathbf{G}_\ell^e$ is computed explicitly, which makes EnKF and ES very computationally efficient. However, in the EnRML, $\mathbf{C}_M^e$ and $\mathbf{G}_\ell^e$ need to be computed separately in (1.58). As shown in (1.56-1.56), $\mathbf{C}_M^e$ is only computed once based on the prior ensemble of model parameters before assimilation of data, while $\mathbf{G}_\ell^e$ varies with iterations.

seq-EnRML and batch-EnRML

Like the EnKF, seq-EnRML integrates the data sequentially in time. The iterative estimate of each realization of model parameters at data-assimilation time k is obtained using (1.58) as

$$\mathbf{m}_j^{\ell+1} = \beta_\ell \mathbf{m}_{k-1,j} + (1 - \beta_\ell) \mathbf{m}_j^\ell - \beta_\ell \mathbf{C}_{M,k-1}^e \mathbf{G}_{k,\ell}^{eT} (\mathbf{G}_{k,\ell} \mathbf{C}_{M,k-1}^e \mathbf{G}_{k,\ell}^{eT} + \mathbf{C}_{D,k})^{-1} \tag{1.60}$$
$$\times [g_k(\mathbf{m}_j^\ell) - \mathbf{d}_{obs,k,j} - \mathbf{G}_{k,\ell}^e (\mathbf{m}_j^\ell - \mathbf{m}_{k-1,j})], \qquad \text{for } j = 1, 2, \ldots, N_e,$$

where $g_k(\cdot)$ represents the nonlinear relationship between the the data at time k and model variables. $\mathbf{C}_{M,k-1}^e$ is the prior covariance before the assimilation of data at time k

and $\mathbf{G}^e_{k,\ell}$ is the sensitivity matrix which is a linearization of $g_k(\cdot)$ at the $\ell + 1$th iteration. $\mathbf{d}_{obs,k,j}$ are sampled from normal distribution with mean $\mathbf{d}_{obs,k}$ and covariance $\mathbf{C}_{D,k}$. seq-EnRML requires to rerun the reservoir simulator with the updated ensemble of model parameters from time zero to predict the data and state variables at each iteration.

Alternatively, batch-EnRML implements the EnRML in a similar manner to the ES, in which all data collected at different times are assimilated together. The update equation is similar to (1.60) given as

$$\mathbf{m}^{\ell+1}_j = \beta_\ell \mathbf{m}_{pr,j} + (1 - \beta_\ell)\mathbf{m}^\ell_j - \beta_\ell \mathbf{C}^e_M \mathbf{G}^{eT}_\ell (\mathbf{G}_\ell \mathbf{C}^e_M \mathbf{G}^{eT}_\ell + \mathbf{C}_D)^{-1} \qquad (1.61)$$
$$\times [g(\mathbf{m}^\ell_j) - \mathbf{d}_{obs,j} - \mathbf{G}^e_\ell(\mathbf{m}^\ell_j - \mathbf{m}_{pr,j})], \qquad \text{for } j = 1, 2, \ldots, N_e,$$

where $\mathbf{d}_{obs,j}$ contains all of the available data that is a sample of the normal distribution with mean $\mathbf{d}_{obs}$ and covariance $\mathbf{C}_D$. Compared with (1.60), $\mathbf{G}^e_\ell$ is expanded to include sensitivity of data at different times. Analogous to the ES, batch-EnRML has a practical motivation in applications when the expense of repeatedly updating variables and restarting simulation runs is substantial for seq-EnRML.

In both seq-EnRML and batch-EnRML, the same ensemble-based sensitivity matrix $\mathbf{G}^e$ is used for all realizations. Thus, seq-EnRML and batch-EnRML may not sample multimodal distribution well. Li and Reynolds [72] presented two iterative EnKF methods based on adjoint methods. In these methods, the ensemble is primarily used to approximate the Hessian at each iteration. Because each ensemble member is updated using a different gradient, these two methods are able to sample a PDF with multiple peaks. However, for problems with multiple local minima the ensemble-based sensitivity might be better than the adjoint-based methods, because it is less likely to get stuck at local minima [7, 19, 77].

LM-EnRML

Chen and Oliver [19] showed that the ensemble approximation of the sensitivity is usually poor and quite noisy for large-scale problems. But they also found that the sensitivity matrix $\mathbf{G}^e$ in (1.58) generally occurs as the product of $\mathbf{C}_M \mathbf{G}^e$, leading to the instability alleviated somewhat by the model covariance. So the influence of computing the sensitivity from a small ensemble did not draw much attention at the beginning. In a later paper [21], however, they found that the poor approximation of sensitivity can cause numerical instability and affect the rate of convergence especially for large-scale problems, which partially explains the slow convergence rate of batch-EnRML. Chen and Oliver [21] introduced an efficient iterative ES algorithm called LM-EnRML that avoids the explicit computation of the sensitivity matrix and shows significant improvements compared to batch-EnRML.

Similar to the gradient-based history matching using adjoint methods [44, 73], it is important to restrict the roughness arising in the early iterations for iterative ensemble-based assimilation methods, when the initial data mismatch is large or large amounts of data are integrated at the same time. In this respect, the LM algorithm has been shown to give good performance in many history-matching applications [8, 73]. Similar to the GN equations (1.53-1.54), the LM implementation can also be expressed in the

following two equivalent forms:

$$\delta\mathbf{m}^{\ell+1} = -(\mathbf{G}_\ell^T \mathbf{C}_D^{-1} \mathbf{G}_\ell^T + (1+\lambda_\ell)\mathbf{C}_M^{-1})^{-1}$$
$$\times [\mathbf{C}_M^{-1}(\mathbf{m}^\ell - \mathbf{m}_{pr}) + \mathbf{G}_\ell^T \mathbf{C}_D^{-1}(g(\mathbf{m}^\ell) - \mathbf{d}_{obs})], \tag{1.62}$$

$$= \frac{\mathbf{m}_{pr} - \mathbf{m}^\ell}{1+\lambda_\ell} - \mathbf{C}_M \mathbf{G}_\ell^T [\mathbf{G}_\ell \mathbf{C}_M \mathbf{G}_\ell^T + (1+\lambda_\ell)\mathbf{C}_D]^{-1}$$
$$\times [g(\mathbf{m}^\ell) - \mathbf{d}_{obs} - \frac{\mathbf{G}_\ell(\mathbf{m}^\ell - \mathbf{m}_{pr})}{1+\lambda_\ell}], \tag{1.63}$$

where λ_ℓ is the damping parameter. When $\ell = 0$, $\mathbf{m}^0 = \mathbf{m}_{pr}$. As before, the computational efficiency of the two formulas depends on the dimensions of model space and data space. Following [21], method based on this LM scheme is called LM-EnRML (orig). Note that when $\lambda_\ell = 0$, LM-EnRML (orig) is identical with batch-EnRML with full step size.

Chen and Oliver [21] developed a modified LM regularization scheme in which the Hessian term (1.59) is changed to the following form:

$$\mathcal{H} = \mathbf{G}_\ell^T \mathbf{C}_D^{-1} \mathbf{G}_\ell^T + (1+\lambda_\ell)\mathbf{P}_\ell^{-1}, \tag{1.64}$$

with

$$\mathbf{P}_\ell^e = \frac{\Delta\mathbf{M}_\ell \Delta\mathbf{M}_\ell^T}{N_e - 1}, \tag{1.65}$$

where $\mathbf{P}_\ell$ is the covariance of model parameters at the ℓth iteration, which is estimated from the ensemble denoted by $\mathbf{P}_\ell^e$. When $\ell = 0$, $\mathbf{P}^0 = \mathbf{C}_{pr}$. Note that the only difference between LM-EnRML and LM-EnRML (orig) is that $\mathbf{C}_{pr}^{-1}$ is replaced $\mathbf{P}_\ell^{-1}$ in LM-EnRML.

The benefits of the modifications made in LM-EnRML are at least twofold. First, similar to the standard LM algorithm, adjusting λ influences not only the step size but also the search direction. By increasing the value of λ, the search direction gets closer to the steepest descent direction with a decreasing step size, so the updates are small and the convergence rate is slow. Conversely, reducing the value of λ makes the search direction less likely to be descent and imposes large corrections on the model variables, but it may accelerate the convergence rate if the current iterate is not far from the solution. So an intuitive tuning strategy is that λ starts with a relatively large value and "smartly" decreases as the minimization gets more stable with iterations. Second, replacing $\mathbf{C}_{pr}^{-1}$ with $\mathbf{P}_\ell^{-1}$ avoids the explicit computation of the sensitivity $\mathbf{G}_e$, which turns out to be a significant unstable factor in the minimization. Chen and Oliver [21] compared LM-EnRML with LM-EnRML (orig), in a one-dimensional multiphase flow problem. The experiment shows that LM-EnRML (orig) stopped before reducing the data mismatch to an acceptable level, while LM-EnRML resulted in a rapid and steady decrease in the data mismatch and achieved a much better result.

In view of the fact that the implementation of LM-EnRML method has been well described in the included papers, we will not repeat the description here and refer the reader to these papers or the original paper [21] for more information.

EnKF-MDA and ES-MDA

Multiple data assimilation (MDA) method was introduced by Emerick and Reynolds [31, 32] to improve the performance of EnKF and ES for nonlinear problems. The MDA procedure seeks the improvement simply by assimilating the same data multiple times with an inflated covariance matrix of measurement errors. It was proved that MDA is equivalent to the single data assimilation for the linear-Gaussian case, under the condition that the covariance matrix of measurement errors in MDA is multiplied by the number of data assimilations [31]. Emerick and Reynolds [32] generalized the procedure and presented the following condition that the inflation coefficients should satisfy in order to keep the equivalence for the linear-Gaussian case,

$$\sum_{\ell=1}^{N_\tau} \frac{1}{\tau_\ell} = 1, \tag{1.66}$$

where N_τ is the total number of data assimilations that need to be predefined. τ_ℓ is the inflation coefficient at the ℓth data assimilation by which the measurement error covariance is multiplied. Under this condition, the previous work in [31] becomes a special case of (1.66) when $\tau_\ell = N_\tau$, for $\ell = 1, 2, \ldots, N_\tau$. MDA is very simple for implementation and only requires minor modifications on the update equation for both EnKF and ES. Take the ES-MDA for example, in which the update equation becomes

$$\mathbf{m}_j^a = \mathbf{m}_j^f + \mathbf{C}_{md}(\mathbf{C}_{dd} + \tau_\ell \mathbf{C}_{d_{obs}})^{-1}(\mathbf{d}_{obs,j} - \mathbf{d}_j), \qquad \text{for } j = 1, 2, \ldots, N_e. \tag{1.67}$$

with

$$\mathbf{d}_{obs,j} = \mathbf{d}_{obs} + \sqrt{\tau_\ell}\mathbf{C}_{d_{obs}}^{1/2}z_d, \qquad z_d \sim \mathcal{N}(0, \mathbf{I}_{N_n}). \tag{1.68}$$

In comparison with the ES update equation (1.51), we see that the only differences are that $\mathbf{C}_{d_{obs}}$ replaced with $\tau_\ell \mathbf{C}_{d_{obs}}$, and that the observation vector is perturbed at each iteration using (1.68).

For the nonlinear case, however, the equivalence does not hold. The use of MDA for nonlinear problems is motivated by the connection between EnKF (ES) and GN method as we discussed before shown by (1.59). The performance of EnKF (ES) in reservoir history-matching problems can be explained in a similar way as GN method. Analogous to the overcorrection problem of GN method at early iterations, EnKF (ES) may render excessive modifications to reservoir models, especially when the nonlinearity is strong and the initial data mismatch is large. In this sense, MDA is similar to those improving strategies used for gradient-based methods. For example, Wu et al. [120] artificially inflated the data covariance matrix during early iterations to relieve the overcorrection of model variables with the GN method, which resulted in either slow convergence rate or unacceptable final data mismatch.

Another way is to switch GN to LM [73] as we see in LM-EnRML. For the purpose of illustration, we take the ES-MDA for example. Recall the update equation for LM-EnRML (orig) shown in (1.63), if we set $\ell = 0$ (i.e., first iteration), it becomes

$$\delta\mathbf{m}^1 = -\mathbf{C}_M^e\mathbf{G}_0^{eT}[\mathbf{G}_0^e\mathbf{C}_M^e\mathbf{G}_0^{eT} + (1 + \lambda_0)\mathbf{C}_D]^{-1} \times [g(\mathbf{m}_{pr}) - \mathbf{d}_{obs}],$$
$$= \mathbf{C}_{md}[\mathbf{C}_{dd} + (1 + \lambda_0)\mathbf{C}_D]^{-1} \times [\mathbf{d}_{obs} - g(\mathbf{m}_{pr})]. \tag{1.69}$$

The notations used here are probably a little confusing. We need to remind that equations (1.67-1.69) are all computed from ensemble no matter with or without the superscript e. Comparing (1.67) and (1.69), it is clear that ES-MDA and LM-EnRML (orig) share similar forms at the first iteration. Moreover, ES-MDA is very similar to the LM-EnRML (approx) in [21] where the $\mathbf{C}_M^e$ is replaced with $\mathbf{P}_\ell^e$.

Although EnKF-MDA and ES-MDA generally work well [33], there are some remaining issues that worth further investigation. For example, it is necessary to specify the number of iterations in MDA prior to performing the data assimilation. If the results are not satisfactory, it may be necessary to discard the results and start over with a larger number. Conceptually, there are numerous possible choices for the inflation coefficients, however, it is unclear how to choose them optimally. Le et al. [70] shed some light in this direction.

1.3.6 Sampling errors and rank deficiency

One fundamental issue for ensemble-based data-assimilation methods relates to the fact that the number of ensemble members is usually quite small compared to the dimension of the model space. Theoretically, increasing the ensemble size could solve much of the difficulty, however, it would sacrifice the efficiency of EnKF and its variants, which is valuable in practical applications. So the goal is always to use as small as an ensemble as possible. It has been well recognized that two problems become prominent in a standard implementation of EnKF with a limited ensemble size [1].

- Sampling errors produced by using a small size of the ensemble could result in spurious correlations arising in the ensemble approximation to the covariance matrix. These spurious correlations, when carrying out the *analysis* step using the Kalman gain, give rise to changes to model or state variables in regions that should not be updated, and collapse of ensemble variability. For example, Lorenc [79] investigated the effect of noise in sampled covariances on the estimation of EnKF, in which an ensemble of 100 realizations is updated with a single perfect observation (i.e., zero error). He found that the global variance estimate increased after assimilation of data, although the analysis error decreased in the local area around the observation. He concluded that the harmful effect of spurious covariances on distant grid points overtakes the local benefit.

- The number of degrees of freedom is only as large as the size of the ensemble, and the updates in EnKF are confined in the subspace spanned by the realizations of forecast ensemble, so the ability to assimilate large amounts of accurate, independent data for a standard EnKF is severely constrained. The problem with the rank-deficiency becomes acute when the number of measurements is large as in the 4D seismic or in fields with many wells and a long history.

Attempts to reduce effectively the sampling error due to small ensembles and meanwhile maintain an ensemble spread that realistically describes error statistics have triggered the development of various techniques [1, 29, 92].

Distance-based covariance localization

The most common method of eliminating the influence of spurious correlations on the update is to apply some type of distance-based covariance localization. This kind of method relies on the assumption that any covariance beyond a certain distance is purely a result of sampling errors. In its most basic form, the localization is performed by limiting the data that will be used for the update of a variable to those data that lie within a particular distance of the variable to be updated, or conversely by limiting the variables that will be updated to those that are within a particular distance of the observation location. The first application of localization in the EnKF [59] took a simple approach in which a cutoff radius was applied to the Kalman gain, so that only model variables within a critical distance of the observation were updated. According to their results, the optimal radius of the cutoff increased as the ensemble size increased. In most current applications of localization, it has been carried out by means of multiplying the ensemble estimate of the covariance matrix by a localization function through the Schur/Hadamard product [46]. Houtekamer and Mitchell [60] have used the fifth-order compactly supported correlation function of Gaspari and Cohn [46] as the localization function, and pointed out that with this procedure, the updated models were relatively smooth compared to those obtained using the distance cutoff or using no localization at all.

In this type of localization, the localized covariance $\mathbf{C}_y^{loc}$ can be expressed as

$$\mathbf{C}_y^{loc} = \rho \circ \mathbf{C}_y. \tag{1.70}$$

where $\circ$ denotes the Schur product with which $\mathbf{C}_y$ is multiplied element-wise by the localization matrix ρ. One advantage of using the element-wise multiplication is the Schur product theorem [58] which states that, *if matrix $\mathbf{A}$ is positive definite and matrix $\mathbf{B}$ is positive semidefinite with all of its main diagonal entries positive, then the product $\mathbf{A} \circ \mathbf{B}$ is positive definite*. Once $\mathbf{C}_y$ is positive semidefinite with positive values on the diagonal in the EnKF, the Schur product theorem guarantees that $\mathbf{C}_y^{loc}$ is positive definite if we use a localization matrix ρ which is positive definite. As a consequence, localization allows the model update to be obtained from a much larger space than the one spanned by the ensemble. The localized Kalman gain in the EnKF is then given by

$$\mathbf{K}^{loc} = (\rho \circ \mathbf{C}_y)\hat{\mathbf{H}}^T [\hat{\mathbf{H}}(\rho \circ \mathbf{C}_y)\hat{\mathbf{H}}^T + \mathbf{C}_{d_{obs}}]^{-1}, \tag{1.71}$$

$$= (\rho_{yd} \circ \mathbf{C}_{yd})(\rho_{dd} \circ \mathbf{C}_{dd} + \mathbf{C}_{d_{obs}})^{-1}, \tag{1.72}$$

$$= \rho_K \circ [\mathbf{C}_{yd}(\mathbf{C}_{dd} + \mathbf{C}_{d_{obs}})^{-1}]. \tag{1.73}$$

To keep the equivalence of these three equations, the localization functions, ρ, ρ_{yd} and ρ_K, are generally different. Note that, because the state vector $\mathbf{y}$ only contains the model and state variables, the observation operator $\hat{\mathbf{H}}$ is introduced for symbolic use to retain the consistency with the original Kalman gain equation (1.47), where the augmented state vector $\widetilde{\mathbf{y}}$ is used.

Since the covariance matrix $\mathbf{C}_y$ is never formed in EnKF methods, localization with (1.71) is rarely used. The implementation of localization methods with the Schur product is hence mainly focused on (1.72) and (1.73), but far from standardized. Hamill et al. [53] localized $\mathbf{C}_{yd}$ but not $\mathbf{C}_{dd}$. They justified the partial use of localization in

view of the fact that the observation operator $\hat{\mathbf{H}}$ in their study is a simple linear operator so that only minor approximation is induced. Houtekamer and Mitchell [60] (also e.g., [61], [41]) localized both $\mathbf{C}_{yd}$ and $\mathbf{C}_{dd}$ but they assume

$$(\rho \circ \mathbf{C}_y)\hat{\mathbf{H}} \approx \rho \circ \mathbf{C}_{yd}, \tag{1.74}$$

$$\hat{\mathbf{H}}(\rho \circ \mathbf{C}_y)\hat{\mathbf{H}}^T \approx \rho \circ \mathbf{C}_{dd}. \tag{1.75}$$

Chen and Oliver [18] (also e.g., [2], [124]) applied localization directly to the Kalman gain instead of covariance matrices under the assumption of

$$\rho_K \approx \rho_{yd}. \tag{1.76}$$

The use of the direct localization to the Kalman gain guarantees zero updates from data to regions that are not highly sensitive to the data. However, as shown in [17], the direct localization of the Kalman gain tends to introduce artifacts in the updates and causes an increase in variability when the size of the localization area is smaller than a single pattern or only slightly bigger than a single pattern. They anticipate the influence of using (1.76) would decrease if the size of the localization area is large, based on the fact that the artifacts observed did not occur in [18]. In particular, Agbalaka and Oliver [2] also distinguished between localization of updates to the model parameters and localization of updates to the state variables, while applying EnKF to the problem of adjusting facies boundaries in a 3D model. Facies constraints were iteratively enforced using the Gaspari-Cohn distance-dependent localization of the Kalman gain to update the Gaussian random fields for facies. But updates are not applied to the state variables during iterations.

On the other hand, the selection of localization function is mainly based on an ad hoc basis. The Gaspari-Cohn function is a common choice. When this kind of function is used for localization, the only free parameter is the range parameter for the correlation which determines the radius of localization. The range parameter is usually chosen on the basis of experience. Lorenc [79] showed that the optimal choice for the range parameter increases as the ensemble size increases because there is less need for localization when the ensemble size is large. For an ensemble of 100 members that is typical in many EnKF applications, the optimal range is almost twice the (practical) range of the prior covariance. Other types of correlation functions are also used such as the exponential covariance function [16], the unit-step function [20] and the like. Instead of using one particular correlation function for localization, Furrer and Bengtsson [42] derived a relationship between a nearly optimal localization function and the true prior covariance. The optimal localization functions determined by the relationship also involves the ensemble size explicitly so that the experimentation with ensemble size is avoided. However, the need to know the true covariance is a strong constraint in some applications. Chen and Oliver [22] (also e.g., [18]) used the method with a spherical covariance function representing the cross-covariance between the data and state variables, to prevent ensemble collapse and to reduce the effect of spurious correlations on the updates in history matching a field case using the LM-EnRML. Some localization functions are based on the regions of sensitivity. Arroyo-Negrete et al. [10] and Devegowda et al. [26] proposed a covariance localization method for petroleum reservoir flow that uses sensitivities from the streamline simulation method to quantify the region of influence of model parameters on the observed data. The cross-covariance

between production observations and model parameters is assumed to be nonzero only in regions where the magnitude of the sensitivity (in any of the realizations) exceeds a small threshold value. By localizing the updates to regions of nonzero sensitivity, the method was shown to be able to reduce the effect of spurious correlations. However, it has been shown that the localization based only on these regions of influence embraces the possibility of keeping model updates too localized around wells, especially when the prior model covariance has a long correlation range [17, 30].

Chen and Oliver [17] gave a detailed study of distance-based location for applications of EnKF in porous media flow. They considered both ρ_{yd} and ρ_{dd} in (1.72), and investigated the true sensitivity and cross-covariance between data and model and state variables for various problems. The pointed out that, since localization is applied to the cross-covariance, it cannot be determined by the region of sensitivity alone, but the contribution of the prior covariance must be considered. The choice of localization function needs to reconcile the effect of the prior data assimilation on the prior covariance and the ensemble approximation to the sensitivity. This means that different localization functions might be needed for different types of data, different model and state variables and different data assimilation times.

Non-distance-based covariance localization

When the localization function is appropriately chosen, distance-based localization usually works well. However, the selection of the localization function can be complicated so that it is difficult to balance the harm done by spurious covariance resulting from the small ensembles against the harm done by excluding real correlations. In addition, there exist model or state variables for which a distance interpretation is not sensible, such as the variables define the shape of the relative permeability curve, or the depth of the oil-water contact. In these situations, non-distance-based covariance localization method are more general and proper to use.

Furrer and Bengtsson [42] proposed two types of localization, one of which is not distance-dependent. The approach they took involves the minimization of the difference between the true covariance matrix and the localized ensemble estimate of the covariance:

$$\min_{\rho} \|\mathbf{C}_{\bar{y}} - \rho \circ \mathbf{C}_y\|_E^2, \tag{1.77}$$

where

$$\|\mathbf{C}_{\bar{y}} - \rho \circ \mathbf{C}_y\|_E^2 = \mathbb{E}[\text{tr}((\mathbf{C}_y - \rho \circ \mathbf{C}_y)^2)], \tag{1.78}$$

and tr represents the trace of the matrix which is the sum of the diagonal elements, and $\mathbf{C}_{\bar{y}}$ denotes the true covariance. Furrer and Bengtsson [42] showed that it was possible to minimize (1.77) term by term if ignoring the positive-definite constraint, and they obtained an expression for each element of the optimal localization. Although the method may not provide a positive definite taper matrix, it is not restricted to distance-dependent localization so it can be applied to nonspatial parameters.

Anderson [6] developed a hierarchical ensemble filter method in which a small number of ensembles are used to compute a regression confidence factor (RCF) to estimate the reliability of the regression coefficients. The set of RCFs for each observation forms a "regression confidence envelope", which can be thought of as a type of localization. By using the RCF, the effect of sampling errors on the updates is reduced. This method

provides a mechanism for discriminating between real and the spurious correlations, but the extra computational cost of running a group of ensembles limits its practical applicability. Based on the work by [6], Vàlles and Nævdal [116] evaluated the use of the hierarchical EnKF (HEnKF) for updating real size synthetic reservoir models. In HEnKF, a damping factor (i.e., RCF in [6]) is computed by minimizing the variance of the estimate of Kalman gain from different ensembles. The damping factor $\alpha_{min,n}$ at time n is chosen to be positive, by truncating the negative one to zero, for not changing the sign of the Kalman gain and defined as

$$\alpha_{min,n} = \max\left(\left\{\left[\left(\sum_{i=1}^{N_g} k_i\right)^2 \bigg/ \sum_{i=1}^{N_g} k_i^2\right] - 1\right\} \bigg/ (N_g - 1), 0\right), \qquad (1.79)$$

where k_i is an element of the Kalman gian matrix for the ith ensemble at time n, and N_g is the number of sub-ensembles. The *analysis* step of HEnKF is then given by

$$\mathbf{y}_n^{ak,j} = \mathbf{y}_n^{fk,j} + \mathbf{A}_n \circ \mathbf{K}_{k,n}(\mathbf{d}_n^j - \hat{\mathbf{H}}_n \mathbf{y}_n^{fk,j}), \qquad (1.80)$$

where $\mathbf{A}_n$ has the same dimensions as the Kalman gain and is composed of the damping factors. k and j indicate the indices of sub-ensemble and ensemble member in the sub-ensemble, respectively. Vàlles and Nævdal [116] compared the results with those obtained using the standard EnKF, and found that spurious correlations were avoided when using HEnKF and it was simpler and more straightforward to use than distance-based localization approaches.

Zhang and Oliver [122] used the similar idea for damping the spurious correlations but used bootstrap sampling to generate multiple estimate of the Kalman gain matrix with almost no computational cost. They also derived an alternative version of objective function which ensures the the damping factor is positive, so that the truncation using in [6, 116] is avoided. Moreover, Zhang and Oliver [123] evaluated and compared the estimation errors when the distance-based localization and the bootstrap-based screening are applied on the covariance and the Kalman gain. They concluded that, when the localization (ρ_{yd}, ρ_{dd}) is applied to the covariance matrices, a consistency condition must be satisfied. However, for nonlocal observations, it is generally difficult to satisfy the consistency condition for distance-dependent covariance localization. They also found that the estimate of Kalman gain based on the covariance localization is subject to greater errors than the estimate of the Kalman gain based on direct regularization. The results showed that the performance of screening Kalman gain is comparable as the performance of localization mehods, but screening Kalman gain is more advantageous in terms of generality for application and no assumption required for the prior covariance.

1.4 Geostatistical methods for facies modeling

Facies are often important in reservoir modeling because the petrophysical properties of interest are highly correlated with facies type. Knowledge of facies constrains the range of variability in porosity and permeability. As we discussed before, however, the facies modeling process is inherently complicated which is often the comprehensive result of

elements like available data, software, geoscientist's knowledge and experience, and project goals. Many geostatistical tools have been developed to simulate the facies distribution and facilitate the modeling process. In this section, we introduce several typical geostatiscal facies modeling methods, some of which have been actively utilized to adapt EnKF-like methods to history matching of facies with production data.

Conditional simulation was initially introduced to correct for the smoothing effect produced by the kriging algorithm [25]. Unlike kriging that yields the "best" local estimate of an attribute with minimum error variance, conditional simulation algorithms reproduce global statistics and spatial features of the attribute such as a variogram, training image and the like. A set of conditionally simulated realizations provides a measure of uncertainty about the spatial distribution of the attribute. The simulation process is described as conditional if it produces realizations that honor sample data at their locations. Otherwise, the simulated results are called unconditional realizations.

1.4.1 Sequential Indicator simulation

Sequential indicator simulation (SIS) is a widely used technique for categorical variable models. Let us consider a facies model which contains N_F mutually exclusive facies categories $z_k, k = 1, \ldots, N_f$. The binary indicator variable $i(\mathbf{u}_j; z_k)$ at a particular location $\mathbf{u}_j$ for a particular facies z_k is defined as the probability of facies z_k prevailing at that location:

$$i(\mathbf{u}_j; z_k) = \text{Prob}\{\text{facies } z_k \text{ being present}\},$$
$$= \begin{cases} 1 & \text{if facies } z_k \text{ is present at } \mathbf{u}_j \\ 0 & \text{otherwise} \end{cases} . \tag{1.81}$$

The expected value $\mathbb{E}[i(\mathbf{u}_j; z_k)] \in [0\ 1]$ is the marginal prior probability of facies z_k, denoted as p_k. If p_k is given, direct kriging of the indicator variable $i(\mathbf{u}_j; z_k)$ provides an estimate for the probability of facies z_k at location $\mathbf{u}_j$. For example, if the simple indicator kriging is used, a model of uncertainty at unsampled location $\mathbf{u}$ is given by:

$$p^{IK}[i(\mathbf{u}_j; z_k)] = \sum_{\alpha=1}^{n} \lambda_\alpha(k)[i(\mathbf{u}_\alpha; z_k) - p_k] + p_k, \qquad \text{for } k = 1, \ldots, N_f, \tag{1.82}$$

where n is the number of all available data including the previously simulated values. The computation of the weights λ_α requires a variogram measure of correlation for each facies. Note that the estimated probabilities $p^{IK}(z_k), k = 1, \ldots, N_f$ using indicator kriging are not guaranteed to be non-negative and sum to one. So the negative value needs to be truncated to zero and then the estimated probabilities are reset according to

$$p^*(z_k) = \frac{p^{IK}(z_k)}{\sum_{k=1}^{N_F} p^{IK}(z_k)}, \tag{1.83}$$

which ensures they sum to one. SIS algorithm is obtained by extending the principle of sequential simulation to the indicator-based model of uncertainty. Algorithm 1.4.1 shows a general procedure of SIS for generating realizations of facies. Multiple realizations can be generated by repeating Algorithm 1.4.1 with a different random number seed each time.

Algorithm 1.4.1 SIS

---

1. Generate a random path and start the simulation from the first random location.

2. Search for nearby data and previously simulated grid nodes

3. Calculate the probability of each facies being present at the current location by building the conditional distribution with kriging, e.g., Equation 1.82.

4. Draw a simulated facies from the set of probabilities.

5. Move to next grid node following the random path and repeat steps $2 \sim 4$ till all grid nodes are visited

---

SIS is simple, flexible and able to constrain spatial continuity for each facies. However, it cannot account for ordering relationships between different facies types because the indicator variogram is only a measure of probability of transition from the current facies type to any other facies types. To reproduce facies ordering relationships, truncated Gaussian simulation or multiple-point simulation (MPS) is favorable.

1.4.2 Truncated Gaussian simulation

The key idea with the truncated Gaussian simulation [84] is to generate realizations of a continuous Gaussian variable and then truncate them at a set of thresholds to create categorical facies realizations. Here, we use a 1D schematic example shown in Figure 1.1, to illustrate some basic concepts behind the truncated Gaussian simulation.
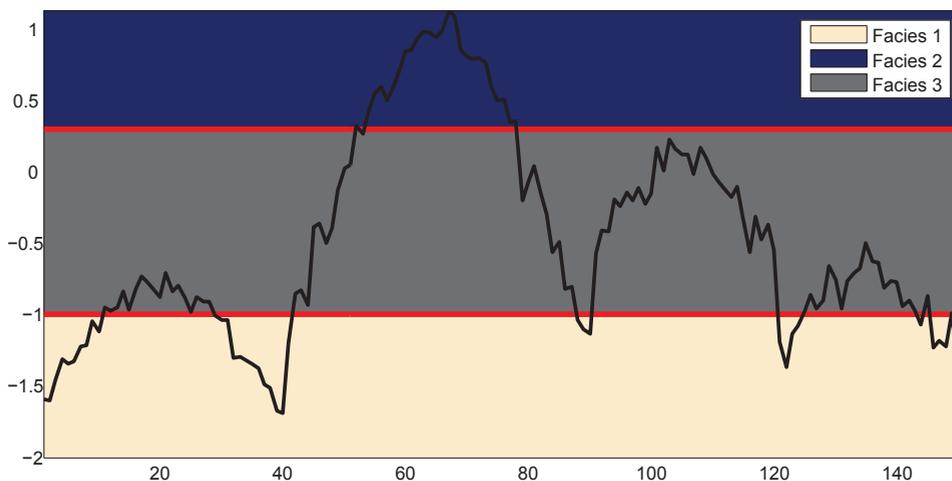
Figure 1.1: Schematic illustration of one Gaussian function using two thresholds.

In this 1D reservoir model of 150 gridblocks, each gridblock is assigned a random Gaussian variable. Thresholds are set at -1 and 0.3 (i.e., red horizontal line in Fig. 1.1). The plane is divided into three regions, each of which corresponds a type of facies. Then gridblocks with Gaussian variables falling into each region are assigned the corresponding facies type. It is clear that a slight perturbation of the thresholds will first

change the facies type of the grids at the boundary of facies regions. This figure also reflects an important feature of truncated Gaussian simulation: the ordering of the resultant facies models. It means that the facies represented by nonadjacent scales of Gaussian values can never be in direct contact with each other in the facies field map. This feature makes the truncated Gaussian simulation very suitable for reproducing facies where the ordering is evident and predictable. Meanwhile, however, the method (single Gaussian random function is truncated) proves to be too restrictive, for example, if there is no natural sequence in the facies or if certain facies can be in contact with more than two facies.

In the truncated Gaussian simulation, there is a one-to-one relationship between the facies proportions and the thresholds, once the order has been established. Assume that the proportion of each ordered facies is known at each location $\mathbf{u}$ in the layer L, denoted by $s_{z_k}(\mathbf{u})$, $\mathbf{u} \in L$, $k = 1, \ldots, N_f$. The cumulative proportions is hence

$$cs_{z_k}(\mathbf{u}) = \sum_{i=1}^{k} s_{z_k}(\mathbf{u}), \tag{1.84}$$

where $cs_{z_0} = 0$ and $cs_{z_{N_f}} = 1$. The $N_F - 1$ thresholds for transforming the continuous Gaussian variables to facies are given by

$$\iota_{z_k}(\mathbf{u}) = \begin{cases} -\infty & k = 0 \\ \Phi^{-1}(cs_{z_k}(\mathbf{u})) & 0 < k < N_f \\ \infty & k = N_f \end{cases} . \tag{1.85}$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function for the standard normal distribution. Given a Gaussian variable $w(\mathbf{u})$ at $\mathbf{u}$, these thresholds can be used to assign a facies code:

$$\text{facies at } \mathbf{u} \leftarrow z_k, \qquad \text{if } \iota_{z_{k-1}}(\mathbf{u}) < w(\mathbf{u}) \leq \iota_{z_k}(\mathbf{u}). \tag{1.86}$$

If the facies are stationary, the thresholds are constant in space; otherwise they vary. In practice, the facies proportions can be obtained experimentally and then be used to deduce the thresholds. The general procedure for implementation of the method is summarized as below [100]:

Algorithm 1.4.2 Truncated Gaussian simulation
---
1. Determine the thresholds to honor facies proportions and ordering relations .

2. Infer a single continuous variogram to represent the spatial continuity of all the facies.

3. Compute a continuous simulation conditioned to the facies data which is transformed into continuous variables, such as sequential Gaussian simulation.

4. Truncate the continuous Gaussian realization with the thresholds and obtain the categorical facies realization.

Pluri-Gaussian simulation

Pluri-Gaussian simulation (PGS) [43, 71] extends the truncated Gaussian simulation for multiple, non-nested, facies utilizing multiple Gaussian random functions. In most

pluri-Gaussian applications, two gaussian random functions are used. The two random functions can be either independent or correlated. The strength of the correlation between these two functions controls the degree of ordering of the resultant facies realizations. It is possible to use more than two Gaussian random fields but would add much more difficulties to compute the variograms, to estimate the parameters and to carry out the simulations [9].

Unlike the truncated Gaussian simulation in which the truncation is effectively defined by the values of thresholds, the situation becomes more complex in PGS where the truncation is represented graphically via the rock type rule [9] defining the relations between the facies in a diagram. The choice of a rock type rule is therefore a major step of the methodology. In practice, this is based on both the analysis of well logs and on a geological conceptual model. It becomes more and more difficult as the number of facies increases.

The thresholds and the variogram model of the underlying gaussian variables are two key factors that control PGS. To compute the value of the thresholds precisely, one needs to determine the relative proportions of the different facies that will be simulated. If the facies field in nonstationary, these proportions are not constant and may vary vertically and laterally because of the existence of trends in the geological processes. The rock type rule is then locally updated by adjusting the values of the thresholds to match the target proportions while preserving the respective positions of the facies. Estimating the parameters of the variogram of underlying Gaussian variables involves an inverse procedure which can be cumbersome.

PGS has demonstrated the flexibility and utility in being able to honor higher-order connectivity relations in many applications. In particular, PGS has been utilized to assist EnKF-like methods in history matching of geological facies. Because PGS provides an underlying gaussian framework for updating non-Gaussian categorical facies variables, on which EnKF-like methods can be applied. More detail in this respect is presented in Sections 1.5.1.

1.4.3   Multiple-point simulation

Among the stochastic methods for facie modeling, multiple-point simulation (MPS) [110] stands in a special position. One the one hand, compared with previously mentioned stochastic methods which are all based on two-point statistics, MPS utilizes multiple-statistics so that it can offer improved ability to honor more complex geological concepts (e.g., curvilinear features). On the other hand, because MPS is still a cell-based simulation method, it preserves more flexibility to honor the conditioning of data than object-based methods, although they are able to reproduce higher geologic complexity and heterogeneity.

In MPS, reservoir heterogeneities are represented by a training image which provides a direct and intuitive way to estimate the spatial statistics of the reservoir. The training image need not be locally accurate or conditional. It need only reflect a prior geological concept and can come from different sources, for example, interpreted outcrop data or a geologist's judges. In addition, training images should be large enough so that sufficient replicates of patterns are available to compute the local conditional probability distributions by sampling the training image with a multiple-point template. In practice, training images are usually generated using unconditional object-based or

process-based simulation programs [100].

One commonly used implementation of MPS is the SNESIM algorithm [110]. The introduction of the search tree is the technical breakthrough that makes SNESIM the first practical implementation of MPS. The algorithm scans the training image using a predefined data template to extract training image events. For every data event, the algorithm searches for replicates of that event, then retrieve the corresponding conditional probability, which is computed directly from the frequency of data events associated with a specific occurrence at the central location of the template, divided by the total number of data events. Once data events and their associated central values are retrieved from the training image, SNESIM stores them in the search tree (a dynamic data structure). The algorithm, then, follows a typical sequential simulation scheme, visiting unsampled nodes using a random path and simulating these nodes conditional on available original data and previously simulated values. The search tree is applied to look up required conditional probabilities for all simulated nodes during simulation.

The utility of MPS for reservoir modeling has been demonstrated in various applications, and it is still experiencing ongoing development. As the core of the method, how to select or generate training images is a major challenge.

1.4.4 Object-based simulation

Object-based simulation (see e.g., [100]) is a totally different framework for facies modeling, in comparison with the cell-based simulation including all aforementioned stochastic methods. It provides a direct control on large-scale geometry. So the facies models obtained with object-based methods are visually attractive because the resulting facie realizations mimic idealized geometries interpreted in outcrops and high-resolution seismic analogs. This is not possible with cell-based methods.

A fundamental assumption with object-based simulation methods is that the shapes of the specific depositional unit are clear and can be characterized by parametric geometries. For example, a fluvial channel system may be composed of multiple channel fills, levees, or a massive sand with shale lenses. All these architectures are converted to parametric geometries, defined by length, width, depth and the like. In general, these parameters are considered uncertain and may be related to each other. For example, channel geometries may be characterized by a uniform distribution of length ranging between an interval, while the width and depth of the channel are related to length by certain ratios. After parameterizing the geometric objects, the next step would be how to simulate with them. This is a major distinction with the cell-based methods, which assign data to the model cells and proceed along a random path simulating at other cells. However, object-based modeling initializes the model space with a background facies and place geometries randomly until some criteria like a global proportion of objects is met. The placement usually involves complicated statistical constraints.

There are many practical situations that are well-suited for object-based simulation including stochastic shales, fluvial channels, and deep-water channels and lobes. The main challenge in object-based simulation methods is the data conditioning which is not automatic and guaranteed as is the case with cell-based methods.

1.5   General approaches for handling non-Gaussian priors

Since ensemble-based data-assimilation methods using the traditional Kalman filter update suffer from the constraints of Gaussian and linear assumptions, it is difficult to condition geological facies to production data for history matching. There are many methods that have been proposed to overcome this challenge and it is still experiencing ongoing developments. Generally, the existing methods for dealing with non-Gaussianity for ensemble-based data-assimilation methods can be categorized into two main approaches: parameterization and post-processing. In this section, we give a brief review of previous applications of the EnKF to the problem of assimilating data into facies models.

1.5.1   Parameterization

The main idea behind the parameterization approach is to apply a transform to the non-Gaussian facies variables to obtain a different set of latent variables with appropriate distribution that can be more easily updated using EnKF. Here, we introduce some techniques that have been combined with EnKF for handling the problem of history matching of geologic facies.

Truncated Pluri-Gaussian

As introduced in Section 1.4.2, truncated pluri-Gaussian method (TPG) for modeling geologic facies is useful not only in generating a variety of textures and shapes, but also in providing an underlying Gaussian framework that maps from a continuous state space to a discrete state space (indicators) and honors the underlying covariances, cross-covariances and facies proportions.

Liu and Oliver [77, 78] first introduced the use of TPG in combination with EnKF for updating reservoir facies models and demonstrated the superiority of EnKF in the history-matching problem of estimating facies boundaries over the gradient-based minimization method. In the TPG model, two Gaussian random fields (GRF) are used and the truncation thresholds are defined by three randomly generated intersecting lines, which divide the plane into seven regions indicating that up to seven different kinds of facies can be modelled with appropriate relative percentage. Facies are hence assigned to each gridblock depending on the region of the truncation threshold map occupied by the two GRFs. Liu and Oliver [78] suggested using the GRFs in the state vector instead of the petrophysical properties (permeabilities and/or porosities) because the GRFs are more consistent with the Gaussian assumptions required by EnKF and the geological realism is preserved as well. In their study, the geological model consists of three facies and is fully specified by the threshold truncation map and the covariance models for the two GRFs. All the parameters of the stochastic model are assumed to be known, and the petrophysical properties are distributed homogeneously within each facies. Both hard data (i.e., facies observations at the wells) and production data are considered in the experiments. Because the facies data are nonnumerical, they proposed a solution by replacing the simulated facies with the facies mismatch f, which is equal to zero if the simulated and real observed facies types are identical and is equal to one otherwise. This strategy also ensures that all facies mismatch terms are equally weighted in

the *analysis* step. The results proved the usefulness of the method in solving reservoir history matching problems with geologic facies

Based on the work by Liu and Oliver [77, 78], Agbalaka and Oliver [2] extended the application to a 3D reservoir with greater nonlinearity in the production data, and investigated some key issues related to the difficulty in applying the history matching to facies. First, to better handle the assimilation of facies observations and production, they propose a sequential global assimilation scheme in which production data are integrated first and facies data afterward in every *analysis* step. An iterative enforcement is only applied to the facies data to achieve a perfect match of facies observations at well locations, where the state vector is reformulated excluding the dynamic state variables (i.e., saturation and pressure). In doing so, they seem to solve the problems of improper weighting of the production data and inappropriate adjustments to the state variables. Second, they apply the localization to the Kalman gain matrix through a Schur product using a Gaspari-Cohn damper function, which turned out to be effective in reducing the need for very large ensemble size and the risk of excessive reduction of variance in the ensemble, as previously observed in [77, 78].

Zhao et al. [125] proposed a different approach to ensure that facies observations at wells are honored at each data assimilation time. The procedure consists of redoing the EnKF *analysis* step with pseudodata whenever the normal EnKF update violates the observed facies. These pseudodata are the pairs of GRF values which are derived to be as close as possible to the currently updated values but also are consistent with the observations of facies. In addition, they used two different truncation maps in the TPG model to generate facies realizations, one of which is the same as the one introduced by Liu and Oliver [77, 78] and the other is defined by two intersecting ellipses shown to be able to model the connectivity of a channelized reservoir. An iterative EnKF is applied to assimilate production data together with seismic data into these facies models, which shows better data match and reliable predictions compared to the standard EnKF or only production data used. Zhao et al. [125] also, for the first time, shows the possibility of jointly updating the facies and the petrophysical properties for each facies, although they assumed that each facies is homogeneous and that permeability is isotropic.

Following this direction, Agbalaka and Oliver [3, 4] described a method for updating both the facies variables and the petrophysical properties with the EnKF, where both heterogeneous petrophysical properties and non-stationary facies proportions are considered. The approach is based on the use of pseudoproperties of the facies which allows consistent updates to the multimodal petrophysical properties and the GRFs that parameterize the facies. To address the problem of getting stuck in a local minimum during the iterative loop for constraining the mismatch of facies observations, they propose to add random perturbations to the facies proxies (i.e., facies mismatch f) to compensate the loss of variability so that the data match can be obtained. Compared to two alternative approaches based on adjusting only GRFs or petrophysical properties directly, the results show that the pseudoproperties approach is better in obtaining geologically plausible models with acceptable matches to production data. Moreover, they observed that the assumption of stationarity or nonsationarity for facies estimation during history matching appears to be trivial, when the correlation lengths and direction of anisotropy are specified correctly. The facies model in this paper is generated from the truncated Gaussian simulation, but the proposed method can be used with other simulation algorithms like the TPG.

Astrakova and Oliver [11] used the pseudoproperties approach for joint updating of permeability and facies variables where they considered the problem of history matching of 3D TPG models with strongly correlated hard data. The synthetic model consists of 5 different facies and 22 layers. Rectangular truncation maps [9] are used and vary vertically to honor the facies proportions. In order to preserve the match of a large amount of hard data at well locations, different from [2, 78, 125], they adopted the interior-point formulation on the objective function by expressing the hard data conditioning as a set of inequality constraints on the model parameters associated with well locations. The derived update equation is similar to the LM-EnRML. The results showed that the data match of facies observations at well locations are maintained during the iteration by adjusting the weight of the interior-point penalty term.

Sebacher et al. [107] constructed a truncation map for three facies using two GRFs from a probabilistic approach. The probabilities field for the occurrence of each facies is introduced into the TPG model by transforming the GRFs through a projection function. They derived two equivalent truncation maps in the two fields by honoring the maximization criterion, that is assigning each gridblock a facies type with the highest occurrence probability in that gridblock. The truncation map depends only on two parameters which, if changed, do not perturb the geometry of the map but can alter the facies proportions. So the truncation parameters can be introduced in the state vector to improve the estimation process and the description of uncertainty. In addition, the observation operator of the facies types at the well locations measures the occurrence probability of each facies type which appears to show more flexibility than the proxy function. However, this method assumes that any two of the three facies have a contact and seems difficult to be extended to model more complex geology.

It is clear that great progress has been made in the application of TPG model in combination with EnKF for the history matching of geologic facies. Especially, the research proceed taking more and more realistic aspects of the problem into account. However, there are some disadvantages of the TPG method that limit its utility in some situations. For example, it is difficult to determine truncation maps and correlation structures of the GRFs that are suitable for very complex reservoirs such as meandering channels, or carbonate systems.

Level set

The level set method was originally introduced by Osher and Sethian [98] for studying surface motion problem. Moreno and Aanonsen [88] first combined the level set method with EnKF to condition facies models to production data.

For a region $\Omega(\tau)$, the level-set function $\phi(x, \tau)$ is commonly defined as a signed distance function

$$\begin{cases} \phi > 0 & \phi \in \Omega \\ \phi = 0 & \phi = \Theta(\tau) \\ \phi < 0 & \phi \notin \Omega \end{cases}, \tag{1.87}$$

where τ denotes the time variable and $\Theta(\tau)$ represents the boundary of the domain $\Omega(\tau)$, which corresponds to the zero level set of the function ϕ. The evolution of the level set function can be described by a partial differential equation,

$$\frac{\partial \phi(x, \tau)}{\partial \tau} + V|\nabla \phi(x, \tau)| = 0. \tag{1.88}$$

The velocity term *V* contains the available information on texture and structure within the image which informs the algorithm on how, in what direction, and at what speed the interface $\Theta(\tau)$ should be relocated. Therefore, a proper choice of *V* is crucial for the level set method which would reflect the important information for edges and connectivity [102]. In the context of facies modeling, $\Theta(\tau)$ represents the boundary between two different facies types. Since the resulting function is not necessarily a signed distance function after evolving the level-set function with (1.88), the reinitialization equation [97] is usually used

$$\frac{\partial \phi(x,\tau)}{\partial \tau} + \text{sgn}[\phi(x,0)](|\nabla \phi(x,\tau)| - 1) = 0. \tag{1.89}$$

Moreno and Aanonsen [89] assumed that there exists a template model based on which level-set functions are derived. The template model may either be an initial best guess model or a more generic model representing prior knowledge of facies distributions. The boundary velocity is modeled as a GRF and different realizations of velocity give different reservoir facies models. Hence, the initial ensemble of facies realizations is generated by Gaussian perturbations of the original level-set function of the template model. The GRFs, instead of porosities and/or permeabilities, are updated by EnKF while matching the production data. This ensures better agreement with the Gaussian assumptions of EnKF. At each *analysis* step, the new updated velocity fields (i.e., GRFs) are applied to the same original level-set function to get new realizations. The authors have tested their approach on two 2D synthetic models composed of two and four facies respectively. The petrophysical properties such as permeabilities and porosities within each facies are assumed known and constant. The results showed reasonably good match to measured data and strong similarities with the reference model in some experiments. However, because only perturbations of a fixed based model are updated instead of the signed distances directly, the results of this method somehow depend on the quality of the template model. Chang et al. [16] proposed a different application of the level-set parameterization in the EnKF for history matching of facies. They predefined a representing node system, on which the values of level-set function are treated as Gaussian random variables with known distribution, and the sign of the level-set function determines the facies type at the nodes. Then, using a linear interpolation procedure, the values of level-set functions at the other grid nodes are obtained. The values of level-set functions at the representing nodes are updated using the EnKF. This approach is applied to 2D reservoir with two or three facies. For the case with three facies, two level set functions are used and the facies models are constructed based on the combination of the signs of those function values. It is demonstrated that this method is able to capture the main features of the reference facies distributions.

In Lorentzen et al. [81], a signed distance function is defined as: (1) the value of the function is the shortest euclidian distance from each grid node to the boundary separating facies types; (2) the sign of the function differentiates two different facies types. By directly updating the distances with EnKF, the proposed method allows the initial ensemble to be generated using any suitable geostatistical tool, which provides more flexibility on quantifying the prior uncertainty of facies distributions compared with [89]. In addition, the authors proposed several techniques for improving the results including boundary treatment and conditioning on specific statistical measurement of prior distribution. The conditioning of facies observations at well locations is also pre-

served. The methodology is evaluated in two synthetic models with channel structure where the initial facies realizations are generated by SNESIM, and gives promising results. Lorentzen et al. [82] extended the method to handle multiple facies by assigning each facies type a level-set function. The facies type in a given grid node is determined by the maximum of the distances in this grid node.

One limitation on the multiple facies parameterization with two or more level-set functions in [16, 89] is that it is difficult to estimate the facies models in which all types of transitions between facies types occur. Mannseth [83] proposed a hierarchical level-set representation which avoids the aforementioned topological constraints and Lien and Mannseth [75] first applied the method in conjunction with the HIEnKF [118] to the problem of facies estimation in 2D and 3D synthetic models. More importantly, Mannseth [83] compared the level set and TPG for facies representation and revealed that there exist strong similarities between these two methodologies.

Other approaches

Jafarpour and McLaughlin [65] introduced the use of the discrete cosine transform (DCT) for geological reservoir parametrization to reduce the number of parameters that must be updated, and to improve the representation of facies geometry. The DCT originates from images processing and is widely used for image compression. It is a Fourier related transform that uses orthonormal cosine basis functions for representing an image, in the context of history matching, which can correspond to the spatial distribution of model parameters or state variables. Jafarpour and McLaughlin [65] showed the powerful compression property of DCT using several examples, in which the DCT allows for retaining only a few basis functions to give a reasonably good approximation. The authors combined the DCT with EnKF for history matching and evaluated the approach in a 2D channelized reservoir with two facies. The coefficients of the retained cosine basis functions representing the permeability field are updated instead of permeability. The updated gridblock permeabilities are then constructed from the updated coefficients. The results suggest that the DCT approach improves the ill-posedness of history-matching problems and reproduction of the spatial continuity of geological facies. Based on similar idea, Jafarpour [63] investigated the wavelet parameterization for the inverse problem of reconstructing geologic facies from the dynamic multiphase flow observations using EnKF.

Since the distribution of petrophysical properties within distinct facies in a reservoir are typically multimodal, Dovera and Della Rossa [27] extended the EnKF to sample from the Gaussian mixture models (GMM). The GMMs are models in which the PDFs of model parameters are parametrically described as weighted sums of Gaussian PDFs [54]. For linear inverse problems, with GMM prior and Gaussian likelihood, the posterior PDF is also a GMM. To combine the GMM with EnKF, Dovera and Della Rossa [27] reformulated the EnKF update equations assuming Gaussian mixture priors and derived an analogous expressions for the GMM update. They proved that the updated ensemble is a correct sample of the posterior Gaussian mixture distribution, under an assumption of a linear inverse problem. The method was tested in the Lorenz model and a 2D synthetic reservoir example with two facies types, both of which showed improved performance in estimation of the posterior distribution and the uncertainty. Hu et al. [62] parameterized geostatistical models generated by MPS using the underlying

uniform random field that is used to draw the local conditional probability in the MPS model. After transformation into Gaussian random fields, the values are updated in the EnKF. The method was evaluated on a 2D synthetic channelized reservoir problem using a small number of parameters and the results provided plausible facies realizations of channels with improved match to production data.

Other approaches to the problem of updating reservoir models with facies have been focused on inclusion of higher moments in the updating equation. In Sarma and Chen [104], the kernel principle component analysis (kernel PCA) is used to derive a generalized EnKF (KEnKF) that is capable of representing non-Gaussian random fields characterized by multi-point geostatistics. The basic idea is to map the data in the input space (state space of the EnKF) to a so-called feature space through a nonlinear map, and then apply a linear algorithm in the feature space, which is chosen such that the linear algorithm is more appropriate to be applied than the input space. Sarma and Chen [104] developed computationally efficient expressions for the nonlinear update equations and nonlinear Kalman gain, where the polynomial kernel [106] is used although it is not limited to this kernel function only. The approach was applied to two synthetic examples based on MPS, and the results indicate that as the order of the kernel is increased, important geological features are better reproduced by the updated ensemble, however, there exists a balance between geological realism and quality of data match. The KEnKF was then further extended by Sarma and Chen [105] to efficiently handle constraints on the state and other variables of the reservoir model that arise from physical or other reasons.

1.5.2 Post-processing

Contrary to the parameterization approach, ensemble-based data-assimilation methods can be used to adjust the model variables directly, ignoring their non-Gaussian distribution. After updating, it would then be necessary to apply a post-processing step to the variables to ensure that the values are consistent with the prior probability distribution. This type of approaches has some practical advantages. In particular, it is sometimes useful to be able to update geological reservoir realizations that have been created from a very general stochastic algorithm. Post-processing of EnKF results is particularly useful when the prior model is uncertain, or when it contains hyperparameters that have uncertainty.

Oliver et al. [93] proposed a modified RML algorithm to sample conditional realizations from models with non-Gaussian priors on the model variables. The objective function in the algorithm consists of two components in a product form. The first component is a ratio of the prior PDF to its Gaussian approximation, which captures the non-Gaussian relationships in the prior probability density that are otherwise ignored in the EnKF. The second component is equivalent to the standard RML in which the prior distribution for the model variables is Gaussian. So it can be obtained from the EnKF or other ensemble-based data-assimilation methods. When the prior PDF of model variables is Gaussian (the first component equals 1), the modified RML is the same as the standard RML. The specific implementation of the methodology is then divided into two steps: (1) data assimilation using EnKF without regard for the discrete nature of the variables, or the probability of transition from one state to another; (2) projection of the updated model variables back to the discrete state space based on the

Viterbi algorithm. Oliver et al. [93] demonstrated the application of the method in a 1D nonlinear fluid flow problem for Markov-chain models and obtained promising results. However, the extension of the methodology to 2D and to multiple facies levels requires a more general modeling method and an efficient optimization algorithm, where the Viterbi algorithm is not directly applicable in higher dimensions. Our work presented in this thesis can be seen as the 2D/3D extension of this methodology.

Jafarpour and Khodabakhshi [64] presented a data assimilation method using the EnKF to condition MPS realizations of geologic facies. The initial ensemble of unconditional facies realizations is first drawn from the training image conditioned on a uniform probability map and equal weights, in which the permeability is assumed to be constant within each facies. The gridblock log-permeabilities are then updated with EnKF analysis equations. They propose, at each data assimilation time step, to update the ensemble of log-permeability fields whose mean field is used to derive a facies probability field, then to use the MPS algorithm to generate a new ensemble of log-permeability fields conditioned to the facies probability field for the next data assimilation time step. With this method, the dynamic data are actually used as a soft constraint for generating MPS realizations. The final ensemble has a realistic geometry structure with a reduced data mismatch. However, the variability and the randomness introduced by the MPS model does not ensure a good data match for the regenerated ensemble members. As pointed out in [64] that a reliable training image is critical for the history matching problems using the proposed method, Khodabakhshi and Jafarpour [68] generalized the method to the use of multiple uncertain training images based on a Bayesian mixture model.

Chapter 2

Introduction to the papers

Several research papers have been produced as a part of this study. In this chapter, we provide a brief summary of these papers. The specific detail about each paper is presented in the later sections (2.4-2.7).

2.1 Summary of Papers A and B

Paper A
Title: *Data assimilation using the EnKF for 2-D Markov chain models*
Authors: Yanhui Zhang, Dean S. Oliver, Yan Chen and Hans J. Skaug

Paper B
Title: *Data assimilation by use of the iterative ensemble smoother for 2D facies models*
Authors: Yanhui Zhang, Dean S. Oliver, Yan Chen and Hans J. Skaug

In Paper A, we develop an ensemble-based data-assimilation method for updating reservoir models containing facies variables using production data. The method takes advantage of the efficiency of the ensemble-based methods for data assimilation, while removing some limitations on multi-Gaussianity of the distributions. In the proposed method, model parameters are first updated using the standard EnKF without regard to the categorical nature of the facies variables, or the discrete distribution of permeability. The realizations in the updated ensemble from the EnKF usually match the production data very well, but do not honor prior constraints on the distribution of properties. To correct for this, an optimization step is used to project the updated ensemble members back to discrete variables by maximizing the product of the prior probability, which only allows discrete variables, and an ensemble approximation of the likelihood function, which penalizes models that do not match data. Because the optimization step is non-iterative and is applied to each ensemble member independently (easily parallelized), the proposed method is quite efficient. Paper A was presented at 13th European Conference on the Mathematics of Oil Recovery in Biarritz, France.

The methodology is based on a modification of the RML algorithm and consists of a series of steps as follows:

1. *Generation of initial ensemble.* One advantage of the proposed methodology is its flexibility on the generation of facies models from which arbitrary method

can be used. In this paper, we generate the initial ensemble using the SNESIM algorithm.

2. *Modeling of prior probability.* Ideally, the methodology would work best when the stochastic mode to represent the prior probability is consistent with the model used to generate the initial ensemble. However, it is of practical importance if the results after updating are good even when the two models are different. This point is better illustrated in Paper B where both conditions are evaluated in the experiments. In both of these two papers, a maximum entropy model (bivariate transition probability function) is used to represent the joint probability of the prior facies model.

3. *Continuous updating of petrophysical properties.* In Paper A, the standard EnKF is used to update reservoir variables for consistency with production data. Because the facies estimation problem often involves large nonlinearity, the standard EnKF would fail to work and hence iteration is necessary. In Paper B, we adopt an iterative ensemble smoother (LM-EnRML) to improve the model update and data conditioning.

4. *Optimization for projection onto prior pdf.* We develop an sequential optimization algorithm without iteration, in which the facie type within each gridblock is determined sequentially to maximize the posterior probability given the approximate prior (from step 2) and likelihood (from step 3).

In Paper A, we demonstrated the approach with conditioning two synthetic channel models with two facies types to both linear and nonlinear observations. The results showed that the distribution of facies after data assimilation honors data much better than before assimilation, and the transitions among facies are consistent with the prior model.

Paper B is an extension of Paper A and includes some improvements on the procedure and deeper evaluation of the methodology. Apart from the aforementioned changes in Paper B, we also introduced the use of graphical lasso algorithm to estimate the posterior covariance which is required in the optimization step.

The methodology was tested with four 2D history matching problems. Two different types of geological models are considered. The first three test cases are typical binary channelized facies models, which are not geologically complex, but are highly nonlinear so that the data are highly sensitive to facies in a few cells in addition to being sensitive to averages over large regions. The emphasis of the first two examples is on the influence of the data on the performance of the proposed method: one example has sufficient data such that the location of facies boundaries is relatively well determined, the second example includes a region at which data provide little information. The third example investigates a case in which channel width is reduced so that the approximation of the prior probability based on two points statistics is not particularly good. In the three channel examples, the initial realizations are generated using a different stochastic model than the one used in the optimization step. The final example, a three-facies model of sand dunes, illustrates the properties of the methodology when the model used for generation of the initial ensemble and for optimization/projection are consistent. In all cases, the only unknown variables are the facies type in each grid cell and the only data are production data.

The methodology worked well when applied to four test cases. We investigated the use of both truncated singular value decomposition and the graphical lasso. In general, better results were obtained from the graphical lasso method, which encourages sparsity of the estimate and eliminates correlations between variables.

The results also showed that the maximum entropy model worked quite well for channels when the dimensions of the channels were large compared to well spacing (examples 1 and 2), but resulted in loss of continuity when channels were meandering and thin (example 3). Best results are obtained if the same probabilistic model can be used for generating initial realizations and for optimization (example 4, in which the maximum entropy model is used for both).

2.2   Summary of Paper C

Title: *Beyond the Probability Map: Representation of Posterior Facies Probability*
Authors: Yanhui Zhang, Dean S. Oliver and Yan Chen

In Paper C, we present another data assimilation method for the problem with categorical variables. The methodology consists of two major steps, an assimilation step and a post-processing step. At the assimilation step, the model variables are updated directly with observations using the EnKF-like methods without honoring the discreteness of the model variables. This is similar to the methodology proposed in Papers A and B. At the post-processing step, however, a penalty term forcing model variables to take discrete values is jointly minimized with the distance to the posterior realizations to solve for facies models that match data. The distance to posterior realizations is quantified using the ensemble approximation of the posterior covariance, which represents the joint probability of model parameters. The proposed method is compared with the probability map approach in terms of the preservation of data conditioning. Paper C was presented at 14th European Conference on the Mathematics of Oil Recovery in Catania, Sicily, Italy.

In this work, we considered a binary field (e.g., 2 facies types) and utilized the prior information of the discrete nature of model variables by adding a penalty term into the objective function to enforce the binary constraints during the optimization. Using the Gauss-Newton method, we derived an update equation for the minimization of the objective function, which is similar to the EnKF update equations. In the standard EnKF updating formulas, the posterior realization is a linear combination of the realizations from the initial ensemble. In a similar way, we observed that the updated realization with the derived update equation is actually a linear combination of the posterior realizations. If no localization was used in the original data assimilation, then the updated realization after the post-processing step must be a linear combination of initial realizations. As a result, localization of the covariance was used in the minimization step. In this case, localization is not necessary to prevent ensemble collapse, but only necessary to provide sufficient degrees of freedom to incorporate the binary constraints.

In another type of post-processing approach, the probability map is inferred from the updated mean of the ensemble using a linear transformation function and then incorporated into a resampling process attempting to retain the data match of the regenerated realizations. Data matches obtained in this method are generally poor, however,

because the probability map neglects important joint probabilities of model parameters imposed by flow data. To illustrate the importance of preserving the correlation information captured by the posterior ensemble, the proposed method is compared with the probability map approach in three synthetic cases with a focus on the aspect of the data match. All the results in the three examples show better data matches for the proposed post-processing approach which attempts to capture the data information not only from the marginal probabilities for each model variable but also the important joint probabilities between model variables.

2.3   Summary of Paper D

Title: *Ensemble Updating of Geologic Faceis with Curvelet Regularization*
Authors: Yanhui Zhang, Dean S. Oliver, Hervé Chauris and Daniela Donno

In Paper D, we focus on the use of the curvelet transform to reduce the noise from the posterior realizations after the assimilation of production data with ensemble-based methods. Curvelets provide an almost optimal sparse representation of objects with edges, making them well-suited for denoising estimated geologic facies distributions. The denoising of the updated model variables is implemented in the curvelet domain by minimizing an objective function which promotes the sparsity of curvelet coefficients. Because preservation of the data match is an important measure of the performance of the denoising method, the role of the approximation of the inverse posterior covariance is examined in the minimization.

An iterative ensemble smoother (LM-EnRML) is used to condition model variables to production data in a straightforward way. The posterior covariance is estimated from the updated ensemble as a measure of posterior PDF. In order to examine the ability to maintain a good data match while denoting the image with a curvelet representation two different objective functions are evaluated. The first objective function, without the inverse posterior covariance, neglects the correlations in model variables that result from assimilation of production data, and is analogous to the one used in the basis pursuit denoising problem in image processing. This approach allows us to look into the denoising behavior of curvelets from the relationship between the sparseness of curvelet coefficients and the data mismatch of corresponding reconstructed realizations. The ensemble approximation of the inverse posterior covariance is added into the objective function in the second approach to measure the deviation from the posterior realizations. Here we focus on the improvement of the preservation of data match after the curvelet shrinkage. An iterative shrinkage scheme is then implemented on the curvelet coefficient domain of the updated model variables to minimize the objective functions to reduce the incurred noise from the data-integration process.

We examined the proposed approach in a synthetic 2D binary channelized facies model. The first experiment was focused on inspecting the efficiency of curvelets in representing the facies fields with channels. The second example considered a history-matching problem in which the curvelet transform was used to reduce the noise from the posterior realizations obtained by integration of production data with LM-EnRML. The results showed that curvelets are useful for denoising of posteriori images of channels after history matching, but lose data match unless the covariance is included. In that

case, the data match remains relatively good, but not as good as achieved at the end of history matching.

Bibliography

[1] AANONSEN, S. I., NÆVDAL, G., OLIVER, D. S., REYNOLDS, A. C., AND VAL-
LÈS, B. 2009. Ensemble Kalman filter in reservoir engineering — a review. *SPE Journal 14,* 3, 393–412.

[2] AGBALAKA, C. C. AND OLIVER, D. S. 2008. Application of the EnKF and localization to automatic history matching of facies distribution and production data. *Math. Geosci. 40,* 4, 353–374.

[3] AGBALAKA, C. C. AND OLIVER, D. S. 2009. Automatic history matching of production and facies data with non-stationary proportions using EnKF, SPE 118916. In *Proceedings of the 2009 SPE Reservoir Simulation Symposium, The Woodlands, Texas*.

[4] AGBALAKA, C. C. AND OLIVER, D. S. 2011. Joint updating of petrophysical properties and discrete facies variables from assimilating production data using the EnKF. *SPE Journal 16,* 2, 318–330.

[5] ANDERSEN, T., ZACHARIASSEN, E., HOYE, T., MEISINGSET, H., OTTERLEI, C., VAN WIJNGAARDEN, A., HATLAND, K., MANGEROY, G., AND LIESTOL, F. 2006. Method for conditioning the reservoir model on 3D and 4D elastic inversion data applied to a fluvial reservoir in the north sea (SPE100190). In *68th EAGE Conference & Exhibition,12-15 June, Vienna, Austria*.

[6] ANDERSON, J. L. 2007. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena 230,* 1–2, 99–111.

[7] ANNAN, J. D. AND HARGREAVES, J. C. 2004. Efficient parameter estimation for a highly chaotic system. *Tellus A 56,* 5, 520–526.

[8] ARENAS, E., VAN KRUIJSDIJK, C., AND OLDENZIEL, T. 2001. Semi-automatic history matching using the pilot point method including time-lapse seismic data. *SPE 71634*.

[9] ARMSTRONG, M., GALLI, A., BEUCHER, H., LOC'H, G., RENARD, D., DOLIGEZ, B., ESCHARD, R., AND GEFFROY, F. 2011. *Plurigaussian Simulations in Geosciences*, 2nd revised edition ed. Springer Berlin Heidelberg.

[10] ARROYO-NEGRETE, E., DEVEGOWDA, D., DATTA-GUPTA, A., AND CHOE, J. 2008. Streamline-assisted ensemble Kalman filter for rapid and continuous reservoir model updating. *SPE Reservoir Evaluation & Engineering 11,* 6 (DEC), 1046–1060.

[11] ASTRAKOVA, A. AND OLIVER, D. S. 2015. Conditioning truncated pluri-Gaussian models to facies observations in ensemble-Kalman-based data assimilation. *Mathematical Geosciences 47,* 3, 345–367.

[12] BIANCO, A., COMINELLI, A., DOVERA, L., NÆVDAL, G., AND VALLÈS, B. 2007. History matching and production forecast uncertainty by means of the ensemble Kalman filter: A real field application (SPE-107161). In *SPE Europec/EAGE Annual Conference and Exhibition. Society of Petroleum Engineers. London, UK.*

[13] BURGERS, G., VAN LEEUWEN, P. J., AND EVENSEN, G. 1998. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review 126,* 6, 1719–1724.

[14] BUTALA, M., YUN, J., CHEN, Y., FRAZIN, R., AND KAMALABADI, F. 2008. Asymptotic convergence of the ensemble kalman filter. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. 825–828.

[15] CASTRO, S., CAERS, J., OTTERLEI, C., MEISINGSET, H., HOYE, T., GOMEL, P., AND ZACHARIASSEN, E. 2009. Incorporating 4d seismic data into reservoir models while honoring production and geologic data: A case study. *The Leading Edge 28,* 12, 1498–1505.

[16] CHANG, H., CHEN, Y., AND ZHANG, D. 2010. Data assimilation of coupled fluid flow and geomechanics using ensemble Kalman filter (SPE 118963). *SPE Journal 15,* 2, 382–394.

[17] CHEN, Y. AND OLIVER, D. S. 2010a. Cross-covariances and localization for EnKF in multiphase flow data assimilation. *Comput. Geosci. 14*, 579–601.

[18] CHEN, Y. AND OLIVER, D. S. 2010b. Ensemble-based closed-loop optimization applied to Brugge Field. *SPE Reservoir Evaluation & Engineering 13,* 1, 56–71.

[19] CHEN, Y. AND OLIVER, D. S. 2012a. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math. Geosci. 44*, 1–26.

[20] CHEN, Y. AND OLIVER, D. S. 2012b. Localization of ensemble-based control setting updates for production optimization. *SPE Journal 17,* 1, 122–136.

[21] CHEN, Y. AND OLIVER, D. S. 2013. Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computational Geosciences 17,* 4, 689–703.

[22] CHEN, Y. AND OLIVER, D. S. 2014. History matching of the Norne full-field model with an iterative ensemble smoother. *SPE Reservoir Evaluation & Engineering 17,* 2, 244–256.

[23] CROSS, T. A. AND HOMEWOOD, P. W. 1997. Amanz Gressly's role in founding modern stratigraphy. *Geological Society of America Bulletin 109,* 12, 1617–1630.

[24] DENNIS, J. E. AND SCHNABEL, R. B. 1983. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs.

[25] DEUTSCH, C. V. AND JOURNEL, A. G. 1998. *GSLIB: Geostatistical Software Library and User's Guide*, Second ed. Oxford University Press, New York.

[26] DEVEGOWDA, D., ARROYO-NEGRETE, E., DATTA-GUPTA, A., AND DOUMA, S. G. 2007. Efficient and robust reservoir model updating using ensemble Kalman filter with sensitivity-based covariance localization (SPE-106144). In *SPE Reservoir Simulation Symposium*.

[27] DOVERA, L. AND DELLA ROSSA, E. 2011. Multimodal ensemble kalman filtering using gaussian mixture models. *Computational Geosciences 15*, 307–323.

[28] EFENDIEV, Y., HOU, T., AND LUO, W. 2006. Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. *SIAM Journal on Scientific Computing 28,* 2, 776–803.

[29] EHRENDORFER, M. 2007. A review of issues in ensemble-based Kalman filtering. *Meteorologische Zeitschrift 16,* 6, 795–818.

[30] EMERICK, A. AND REYNOLDS, A. 2011. Combining sensitivities and prior information for covariance localization in the ensemble kalman filter for petroleum reservoir applications. *Computational Geosciences 15,* 2, 251–269.

[31] EMERICK, A. A. AND REYNOLDS, A. C. 2012. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computational Geosciences*, 1–21.

[32] EMERICK, A. A. AND REYNOLDS, A. C. 2013a. Ensemble smoother with multiple data assimilation. *Computers & Geosciences 55,* 0, 3 – 15.

[33] EMERICK, A. A. AND REYNOLDS, A. C. 2013b. History-matching production and seismic data in a real field case using the ensemble smoother with multiple data assimilation. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers.

[34] EVENSEN, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. 99,* C5, 10143–10162.

[35] EVENSEN, G. 1997. Advanced data assimilation for strongly nonlinear dynamics. *Monthly Weather Review 125,* 6, 1342–1354.

[36] EVENSEN, G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics 53,* 4, 343–367.

[37] EVENSEN, G. 2006. *Data Assimilation: The Ensemble Kalman Filter*. Springer.

[38] EVENSEN, G. 2009. *Data Assimilation: The Ensemble Kalman Filter*, Second ed. Springer Verlag.

[39] EVENSEN, G., HOVE, J., MEISINGSET, H. C., REISO, E., SEIM, K. S., AND ESPELID, Ø. 2007. Using the EnKF for assisted history matching of a North Sea reservoir model (SPE 106184). In *Proceedings of the 2007 SPE Reservoir Simulation Symposium,*.

[40] EVENSEN, G. AND VAN LEEUWEN, P. J. 2000. An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review 128,* 6, 1852–1867.

[41] FERTIG, E. J., HUNT, B. R., OTT, E., AND SZUNYOGH, I. 2007. Assimilating non-local observations with a local ensemble Kalman filter. *Tellus A 59,* 5, 719–730.

[42] FURRER, R. AND BENGTSSON, T. 2007. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivar. Anal. 98,* 2, 227–255.

[43] GALLI, A., BEUCHER, H., LE LOC'H, G., DOLIGEZ, B., AND GROUP, H. 1994. The pros and cons of the truncated Gaussian method. In *Geostatistical Simulations*. Kluwer Academic, Dordrecht, 217–233.

[44] GAO, G. AND REYNOLDS, A. C. 2006. An improved implementation of the LBFGS algorithm for automatic history matching. *SPE Journal 11,* 1, 5–17.

[45] GAO, G., ZAFARI, M., AND REYNOLDS, A. C. 2006. Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF. *SPE Journal 11,* 4, 506–515.

[46] GASPARI, G. AND COHN, S. E. 1999. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society 125,* 554, 723–757.

[47] GAVALAS, G. R., SHAH, P. C., AND SEINFELD, J. H. 1976. Reservoir history matching by Bayesian estimation. *SPE Journal 16,* 6, 337–350.

[48] GELMAN, A., CARLIN, J. B., STERN, H. S., AND RUBIN, D. B. 2003. Bayesian data analysis, (chapman & hall/crc texts in statistical science).

[49] GILKS, W. R., RICHARDSON, S., AND SPIEGELHALTER, D. J., Eds. 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, New York.

[50] GU, Y. AND OLIVER, D. S. 2005. History matching of the PUNQ-S3 reservoir model using the ensemble Kalman filter. *SPE Journal 10,* 2, 51–65.

[51] GU, Y. AND OLIVER, D. S. 2007. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE Journal 12,* 4, 438–446.

[52] HADAMARD, J. 1902. Sur les problémes aux dérivées partielles et leur signification physique. *Princeton University Bulletin 13*, 49–52.

[53] HAMILL, T. M., WHITAKER, J. S., AND SNYDER, C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review 129,* 11, 2776–2790.

[54] HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. 2009. *The Elements of Statistical Learning*, Second edition ed. Springer-Verlag.

[55] HASTINGS, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57,* 1, 97–109.

[56] HAUGEN, V., NÆVDAL, G., NATVIK, L.-J., EVENSEN, G., BERG, A. M., AND FLORNES, K. M. 2008. History matching using the ensemble Kalman filter on a North Sea field case. *SPE Journal 13,* 4, 382–391.

[57] HEGSTAD, B. K., OMRE, H., TJELMELAND, H., AND TYLER, K. 1994. Stochastic simulation and conditioning by annealing in reservoir description. In *Geostatistical Simulation*, M. Armstrong and P. A. Dowd, Eds. Kluwer Acad., 43–55.

[58] HORN, R. A. 1990. The Hadamard product. In *Matrix Theory and Applications*. American Mathematical Society, 87–170.

[59] HOUTEKAMER, P. L. AND MITCHELL, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review 126,* 3, 796–811.

[60] HOUTEKAMER, P. L. AND MITCHELL, H. L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review 129,* 1, 123–137.

[61] HOUTEKAMER, P. L., MITCHELL, H. L., PELLERIN, G., BUEHNER, M., CHARRON, M., SPACEK, L., AND HANSEN, B. 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review 133,* 3, 604–620.

[62] HU, L., ZHAO, Y., LIU, Y., SCHEEPENS, C., AND BOUCHARD, A. 2013. Updating multipoint simulations using the ensemble kalman filter. *Computers & Geosciences 51,* 0, 7 – 15.

[63] JAFARPOUR, B. 2011. Wavelet reconstruction of geologic facies from nonlinear dynamic flow measurements. *Geoscience and Remote Sensing, IEEE Transactions on 49,* 5, 1520–1535.

[64] JAFARPOUR, B. AND KHODABAKHSHI, M. 2011. A probability conditioning method (PCM) for nonlinear flow data integration into multipoint statistical facies simulation. *Math. Geosci. 43*, 133–164.

[65] JAFARPOUR, B. AND MCLAUGHLIN, D. B. 2009. Reservoir characterization with the discrete Cosine transform. *SPE Journal 14,* 1, 182–201.

[66] JAZWINSKI, A. H. 1970. *Stochastic Processes and Filtering Theory*. Academic Press, New York.

[67] KALMAN, R. E. 1960. A new approach to linear filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineering 82*, 35–45.

[68] KHODABAKHSHI, M. AND JAFARPOUR, B. 2013. A bayesian mixture-modeling approach for flow-conditioned multiple-point statistical facies simulation from uncertain training images. *Water Resources Research 49,* 1, 328–342.

[69] KITANIDIS, P. K. 1995. Quasi-linear geostatistical theory for inversing. *Water Resour. Res. 31,* 10, 2411–2419.

[70] LE, D. H., EMERICK, A. A., AND REYNOLDS, A. C. 2015. An adaptive ensemble smoother with multiple data assimilation for assisted history matching. Society of Petroleum Engineers.

[71] LE LOC'H, G., BEUCHER, H., GALLI, A., DOLIGEZ, B., AND GROUP, H. 1994. Improvement in the truncated Gaussian method: Combining several Gaussian Functions. In *Proceedings of ECMOR IV, Fourth European Conference on the Mathematics of Oil Recovery*.

[72] LI, G. AND REYNOLDS, A. C. 2009. Iterative ensemble Kalman filters for data assimilation. *SPE Journal 14,* 3, 496–505.

[73] LI, R., REYNOLDS, A. C., AND OLIVER, D. S. 2003a. History matching of three-phase flow production data. *SPE Journal 8,* 4, 328–340.

[74] LI, R., REYNOLDS, A. C., AND OLIVER, D. S. 2003b. Sensitivity coefficients for three-phase flow history matching. *Journal of Canadian Petroleum Technology 42,* 4, 70–77.

[75] LIEN, M. AND MANNSETH, T. 2014. Facies estimation through data assimilation and structure parameterization. *Computational Geosciences 18,* 5, 869–882.

[76] LIU, N. AND OLIVER, D. S. 2003. Evaluation of Monte Carlo methods for assessing uncertainty. *SPE Journal 8,* 2, 188–195.

[77] LIU, N. AND OLIVER, D. S. 2005a. Critical evaluation of the ensemble Kalman filter on history matching of geologic facies. *SPE Reservoir Evaluation & Engineering 8,* 6, 470–477.

[78] LIU, N. AND OLIVER, D. S. 2005b. Ensemble Kalman filter for automatic history matching of geologic facies. *Journal of Petroleum Science and Engineering 47,* 3–4, 147–161.

[79] LORENC, A. C. 2003. 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.

[80] LORENTZEN, R. J., FJELDE, K. K., FRØYEN, J., LAGE, A. C. V. M., NÆVDAL, G., AND VEFRING, E. H. 2001. Underbalanced drilling: Real time data interpretation and decision support. In *SPE/IADC Drilling Conference*.

[81] LORENTZEN, R. J., FLORNES, K. M., AND NÆVDAL, G. 2012. History matching channelized reservoirs using the ensemble Kalman filter. *SPE Journal 17,* 1, 137–151.

[82] LORENTZEN, R. J., NÆVDAL, G., AND SHAFIEIRAD, A. 2012. Estimating facies fields by use of the ensemble Kalman filter and distance functions–applied to shallow-marine environments. *SPE Journal 3,* 1, 146–158.

[83] MANNSETH, T. 2014. Relation between level set and truncated pluri-gaussian methodologies for facies representation. *Mathematical Geosciences 46,* 6, 711–731.

[84] MATHERON, G., BEUCHER, H., DE FOUQUET, C., GALLI, A., GUERILLOT, D., AND RAVENNE, C. 1987. Conditional simulation of the geometry of fluvio-deltaic reservoirs, SPE 16753. In *Proc. of 62nd Ann. Tech. Conf. of the SPE*. 123–131.

[85] MAYBECK, P. S. 1982. *Stochastic models, estimation, and control*. Vol. 3. Academic press.

[86] MEINHOLD, R. J. AND SINGPURWALLA, N. D. 1983. Understanding the Kalman filter. *The American Statistician 37,* 2, 123–127.

[87] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H., AND TELLER, E. 1953. Equations of state calculations by fast computing machines. *J. Chem. Physics 21*, 1087–1092.

[88] MORENO, D. AND AANONSEN, S. Stochastic facies modelling using the level set method. In *EAGE Petroleum Geostatistics* (2007). In: Petrolium Geostatistics, 10-14 September 2007, Cascais, Portugal, A16, Extended Abstracts Book, EAGE Publications BV, Utrecht, Netherlands.

[89] MORENO, D. AND AANONSEN, S. I. 2011. Continuous facies updating using the ensemble Kalman filter and the level set method. *Mathematical Geosciences 43*, 951–970.

[90] NÆVDAL, G., JOHNSEN, L. M., AANONSEN, S. I., AND VEFRING, E. H. 2005. Reservoir monitoring and continuous model updating using ensemble Kalman filter. *SPE Journal 10,* 1, 66–74.

[91] OLIVER, D. S. 1996. On conditional simulation to inaccurate data. *Math. Geology 28,* 6, 811–817.

[92] OLIVER, D. S. AND CHEN, Y. 2011. Recent progress on reservoir history matching: a review. *Comput. Geosci. 15,* 1, 185–221.

[93] OLIVER, D. S., CHEN, Y., AND NÆVDAL, G. 2010. Updating Markov chain models using the ensemble Kalman filter. *Comput. Geosci. In Press, posted online 17 December 2010*.

[94] OLIVER, D. S., CUNHA, L. B., AND REYNOLDS, A. C. 1997. Markov chain Monte Carlo methods for conditioning a permeability field to pressure data. *Math. Geol. 29,* 1, 61–91.

[95] OLIVER, D. S., HE, N., AND REYNOLDS, A. C. 1996. Conditioning permeability fields to pressure data. In *European Conference for the Mathematics of Oil Recovery, V*. 1–11.

[96] OLIVER, D. S., REYNOLDS, A. C., AND LIU, N. 2008. *Inverse Theory for Petroleum Reservoir Characterization and History Matching*, First ed. Cambridge University Press, Cambridge.

[97] OSHER, S. AND FEDKIW, R. 2003. *Level set methods and dynamic implicit surfaces*. Vol. 153. Springer Science & Business Media.

[98] OSHER, S. AND SETHIAN, J. A. 1988. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics 79,* 1, 12 – 49.

[99] PETERS, L., ARTS, R. J., BROUWER, G. K., GEEL, C. R., CULLICK, S., LORENTZEN, R., CHEN, Y., DUNLOP, K. N. B., VOSSEPOEL, F. C., XU, R., SARMA, P., ALHUTALI, A. H., AND REYNOLDS, A. C. 2010. Results of the Brugge benchmark study for flooding optimization and history matching. *SPE Reserv. Evalu. Eng. 13,* 3, 391–405.

[100] PYRCZ, M. J. AND DEUTSCH, C. V. 2014. *Geostatistical reservoir modeling*. Oxford university press.

[101] REYNOLDS, A. C., HE, N., AND OLIVER, D. S. 1999. Reducing uncertainty in geostatistical description with well testing pressure data. In *Reservoir Characterization—Recent Advances*, R. A. Schatzinger and J. F. Jordan, Eds. American Association of Petroleum Geologists, 149–162.

[102] RICHARDSEN, S. AND RANDEN, T. 2005. Mapping 3d geo-bodies based on level set and marching methods. In *Mathematical Methods and Modelling in Hydrocarbon Exploration and Production*, A. Iske and T. Randen, Eds. Mathematics in Industry, vol. 7. Springer Berlin Heidelberg, 247–265.

[103] ROGGERO, F. AND HU, L. Y. 1998. Gradual deformation of continuous geostatistical models for history matching. *SPE 49004, Annual Technical Conference*.

[104] SARMA, P. AND CHEN, W. 2009. Generalization of the ensemble Kalman filter using kernels for non-gaussian random fields, SPE-119177. In *SPE Reservoir Simulation Symposium, 2-4 February, The Woodlands, Texas, USA*.

[105] SARMA, P. AND CHEN, W. 2011. Robust and efficient handling of model constraint with the kernel-based ensemble Kalman filter, SPE-141948. In *SPE Reservoir Simulation Symposium, 21–23 February, The Woodlands, Texas, USA*.

[106] SARMA, P., DURLOFSKY, L. J., AND AZIZ, K. 2008. Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics. *Math. Geosci. 40,* 1, 3–32.

[107] SEBACHER, B., HANEA, R., AND HEEMINK, A. 2013. A probabilistic parametrization for geological uncertainty estimation using the ensemble Kalman filter (EnKF). *Computational Geosciences 17,* 5, 813–832.

[108] SKJERVHEIM, J.-A. AND EVENSEN, G. 2011. An ensemble smoother for assisted history matching. In *SPE Reservoir Simulation Symposium, 21–23 February, The Woodlands, Texas*.

[109] SKJERVHEIM, J.-A., EVENSEN, G., AANONSEN, S. I., RUUD, B. O., AND JOHANSEN, T. A. 2007. Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter. *SPE Journal 12,* 3, 282–292.

[110] STREBELLE, S. 2002. Conditional simulation of complex geological structures using multiple-point statistics. *Math. Geol. 34,* 1, 1–21.

[111] TARANTOLA, A. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial Mathematics.

[112] TAVAKOLI, R. AND REYNOLDS, A. C. 2011. Monte Carlo simulation of permeability fields and reservoir performance predictions with SVD parameterization in RML compared with EnKF. *Computational Geosciences 15*, 99–116.

[113] THULIN, K., LI, G., AANONSEN, S. I., AND REYNOLDS, A. C. 2007. Estimation of initial fluid contacts by assimilation of production data with EnKF. In *Proceedings of the 2007 SPE Annual Technical Conference and Exhibition*.

[114] TIERNEY, L. 1994. Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics 22*, 1701–1762.

[115] TJELMELAND, H. AND EIDSVIK, J. 2004. On the use of local optimizations within Metropolis-Hastings updates. *Journal of the Royal Statistical Society Series B-Statistical Methodology 66,* Part 2, 411–427.

[116] VÀLLES, B. AND NÆVDAL, G. 2009. Revisiting Brugge case study using a hierarchical ensemble Kalman filter (IPTC 14074). In *International Petroleum Technology Conference, 7–9 December 2009, Doha, Qatar*.

[117] VAN LEEUWEN, P. J. AND EVENSEN, G. 1996. Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review 124,* 12, 2898–2913.

[118] WANG, Y., LI, G., AND REYNOLDS, A. C. 2010. Estimation of depths of fluid contacts and relative permeability curves by history matching using iterative ensemble Kalman smoothers (SPE 119056). *SPE Journal 15,* 2, 509–525.

[119] WANG, Y. AND LI, M. 2011. Reservoir history matching and inversion using an iterative ensemble Kalman filter with covariance localization. *Petroleum Science 8*, 316–327. 10.1007/s12182-011-0148-7.

[120] WU, Z., REYNOLDS, A. C., AND OLIVER, D. S. 1999. Conditioning geostatistical models to two-phase production data. *SPE Journal 4,* 2, 142–155.

[121] ZAFARI, M. AND REYNOLDS, A. C. 2007. Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter. *SPE Journal 12,* 3, 382–391.

[122] ZHANG, Y. AND OLIVER, D. S. 2010. Improving the ensemble estimate of the Kalman gain by bootstrap sampling. *Math. Geosci. 42,* 3, 327–345.

[123] ZHANG, Y. AND OLIVER, D. S. 2011a. Evaluation and error analysis: Kalman gain regularization versus covariance regularization. *Comput. Geosci. 15,* 3, 489–508.

[124] ZHANG, Y. AND OLIVER, D. S. 2011b. History matching using the ensemble Kalman filter with multiscale parameterization: A field case study. *SPE Journal 16,* 2, 307–317.

[125] ZHAO, Y., REYNOLDS, A. C., AND LI, G. 2008. Generating facies maps by assimilating production data and seismic data with the ensemble Kalman filter, SPE-113990. In *Proceedings of the 2008 SPE Improved Oil Recovery Symposium, Tulsa, Oklahoma, April 21–23.*