Multilevel assimilation of inverted seismic data

Mohammad Nezhadali

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2023

UNIVERSITY OF BERGEN

Multilevel assimilation of inverted seismic data

Mohammad Nezhadali

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 14.04.2023

Year: 2023

Title: Multilevel assimilation of inverted seismic data

Name: Mohammad Nezhadali

Print: Skipnes Kommunikasjon / University of Bergen

Preface

This thesis is submitted as a partial fulfillment of the requirements for the degree of Philosophiae Doctor (Ph.D) in applied mathematics at the University of Bergen. The supervising committee has consisted of Professor Inga Berre (University of Bergen), Doctor Trond Mannseth (NORCE), Doctor Kristian Fossum (NORCE), and Doctor Tuhin Bhakta (NORCE).

The PhD project has been supported by NORCE Norwegian Research Centre AS through the Research Council of Norway (RCN no. 295002) and the industrial partners Aker BP ASA, Equinor Energy AS, Lundin Energy Norway AS, Repsol Norge AS, Shell Global Solutions International B.V., Total E&P Norge AS, and Wintershall Dea Norge AS.

Acknowledgements

People say “start projects in which you win even if you lose!”. This PhD project was definitely one of them right from the beginning. However, I feel no loss. The project accomplished as much as was planned or even more. I learned a lot, enjoyed a lot, and am very grateful to all those who were part of this experience.

I wish to thank my supervisors: Trond Mannseth, Kristian Fossum, Tuhin Bhakta, and Inga Berre, for guiding me all through the way in the past three years. In addition, I would like to thank Patrick, Andreas, Rolf, Nazanin, and all my colleagues at NORCE with whom I had fruitful and joyous discussions. NORCE offered me the best academic milieu I have ever experienced; friendly, educative, and scientific.

I would also like to express my special thanks to my family. I am very grateful to my parents, without whose support this work would have been impossible. My sister, three years my junior, has always been a source of inspiration. Her impact has been far beyond the scope of this PhD; very influential in my personal development. Thanks a lot Shadi!

Lastly, it would be unforgivable to forget about my dearest friends: the people with whom I celebrated success joyfully and could count on in moments of hardship. Thank you Mohammadreza and Alireza. I would also like to thank Babak, Diddy, William, and all my friends in Bergen and around the world who turn life into a better experience. Thank you very much!

Mohammad Nezhadali
Bergen, December 2022

English abstract

In ensemble-based data assimilation (DA), the ensemble size is usually limited to around one hundred. Straightforward application of ensemble-based DA can therefore result in significant Monte Carlo errors, often manifesting themselves as severe underestimation of parameter uncertainties. Assimilation of large amounts of simultaneous data enhances the negative effects of Monte Carlo errors. Distance-based localization is the conventional remedy for this problem. However, it has its own drawbacks, e.g. not allowing for true long-range correlations and difficulty in assimilation of data which do not have a specific physical location. Use of lower-fidelity models reduces the computational cost per ensemble member and therefore renders the possibility to reduce Monte Carlo errors by increasing the ensemble size, but it also adds to the modeling error. Multilevel data assimilation (MLDA) uses a selection of models forming hierarchies of both computational cost and computational accuracy, and tries to obtain a better balance between Monte Carlo errors and modeling errors.

In this PhD project, several MLDA algorithms were developed and their quality for assimilation of inverted seismic data was assessed in simplistic reservoir problems. Utilization of multilevel models entails introduction of some numerical errors (multilevel modeling error, MLME) to the problem in addition to the already existing numerical errors. Several computationally inexpensive methods were devised for partially accounting for MLME in the context of multilevel data assimilation. They were also investigated in simplistic reservoir history-matching problems. Finally, one of the novel MLDA algorithms was chosen and its performance was assessed in a realistic reservoir history-matching problem.

Norwegian abstract

I ensemble-basert data-assimilering (DA) er størrelsen på ensemblet vanligvis begrenset til hundre medlemmer. Rett frem bruk av ensemble-basert DA kan resultere i betydelig Monte Carlo-feil, som ofte viser seg som alvorlig undervurdering av parameterusikkerheter. Assimilering av store mengder samtidige data forsterker de negative effektene av Monte Carlo-feilen. Avstandsbasert lokalisering er det konvensjonelle middelet for å begrense dette problemet. Denne metoden har imidlertid sine egne ulemper. Den vil, f.eks., fjerne sanne korrelasjoner over lange distanser og det er svært vanskelig å benytte på data som ikke har en unik fysisk plassering. Bruk av modeller med lavere kvalitet reduserer beregningskostnadene per ensemble-medlem og gir derfor muligheten til å redusere Monte Carlo-feilen ved å øke ensemble-størrelsen. Men, modeller med lavere kvalitet øker også modelleringsfeilen. Data-assimilering på flere nivåer (MLDA) bruker et utvalg av modeller som danner hierarkier av både beregningskostnad og beregningsnøyaktighet, og prøver å oppnå en bedre balanse mellom Monte Carlo-feil og modelleringsfeil.

I dette PhD-prosjektet ble flere MLDA-algoritmer utviklet og deres kvalitet for assimilering av inverterte seismiske data ble vurdert på forenklede reservoarproblemer. Bruk av modeller på flere nivå innebærer introduksjon av noen numeriske feil (multilevel modeling error, MLME), i tillegg til de allerede eksisterende numeriske feilene. Flere beregningsmessig rimelige metoder ble utviklet for delvis å kompansere for MLME i gjennomføring av data-assimilering på flere nivåer. Metodene ble også undersøkt under historie tilpassing på forenklede reservoar problemer. Til slutt ble en av de nye MLDA-algoritmene valgt og ytelsen ble vurdert på et historie tilpassings problem med en realistisk reservoar modell.

Outline

The thesis consists of two parts. Part I is devoted to the scientific background required for the collection of research articles provided in Part II, and it is structured as follows.

Chapter 1 gives an introduction to the reservoir history-matching problem and motivates the investigations performed in this work. Chapter 2 defines the inverse problem and explains different approaches to it. Chapter 3 discusses ensemble-based data assimilation (DA) and presents several state-of-the-art DA algorithms. Chapter 4 is devoted to regularization techniques. It briefly discusses classical regularization techniques followed by a more detailed discussion on localization and ensemble-subspace inversion. Chapter 5 describes forward modeling, including reservoir flow modeling, petro-elastic modeling, and seismic modeling. It is followed by Chapter 6 on surrogate modeling which will cover both response-surface surrogates and lower-fidelity surrogates. Chapter 7 explains multilevel data assimilation. Finally in Chapter 8, I present summaries of the publications associated with this project and discuss the outlook of the work.

List of publications

A Mohammad Nezhadali, Tuhin Bhakta, Kristian Fossum, and Trond Mannseth. *A novel approach to multilevel data assimilation.* In ECMOR XVII, vol. 2020, no. **1**, pp. 1-13. European Association of Geoscientists & Engineers, 2020.

B Mohammad Nezhadali, Tuhin Bhakta, Kristian Fossum, and Trond Mannseth. *Multilevel assimilation of inverted seismic data with correction for multilevel modeling error.* Frontiers in Applied Mathematics and Statistics **7** (2021): 673077.

C Mohammad Nezhadali, Tuhin Bhakta, Kristian Fossum, and Trond Mannseth. *Iterative multilevel assimilation of inverted seismic data.* Computational Geosciences **26**, no. 2 (2022): 241-262.

D Mohammad Nezhadali, Tuhin Bhakta, Kristian Fossum, and Trond Mannseth. *Sequential multilevel assimilation of inverted seismic data.* accepted by the journal Computational Geosciences.

E Mohammad Nezhadali, Tuhin Bhakta, Kristian Fossum, and Trond Mannseth. *Towards Application of Multilevel Data Assimilation in Realistic Reservoir History-Matching Problems.* In ECMOR 2022, vol. 2022, no. **1**, pp. 1-12. European Association of Geoscientists & Engineers, 2022.

Contents

Part I

Scientific background

Chapter 1

Introduction

Reliability is one of the key concerns when it comes to energy supply. This key characteristic of fossil fuels together with their economical advantage over other energy sources has made them dominate the energy market. This dominance seems to continue in the foreseeable future. Oil and gas in particular are not only of interest as energy sources but also of crucial importance in a large array of industries. Many of the goods the modern human uses are partially made of hydrocarbon products. Hydrocarbons are well integrated in human's life.

Hydrocarbons are produced from subsurface reservoirs. In conventional reservoirs, extraction of oil and gas requires drilling several production wells. Drilling each well is, economically, very costly. Hence, any decision about it entails high risks. As a result, oil companies should ensure about the quality of their decisions before implementing reservoir development plans. However, due to involvement of several uncertainties in the process, the decisions are to be made under uncertainty.

The uncertainty can have political, economic, or physical cause. The political and economic sources of uncertainty include wars, taxation, oil and gas prices, limitations on production, etc. which are extremely difficult to model and quantify. However, modeling and quantification of physical uncertainties are feasible. Mathematical models are developed for prediction of hydrocarbon production based on characteristics of the subsurface reservoirs. These characteristics are unknown and constitute the sources of physical uncertainty.

Mathematical models parameterize the subsurface flow problem based on the characteristics of the reservoir and yield formulations connecting the parameters and model forecasts. The mathematical models for reservoir flow normally entail complex mathematical equations. Except for basic cases, the model forecasts cannot be obtained

analytically. Therefore, the mathematical equations are discretized and are solved numerically. A model run consists of giving a set of parameters to the model as input, and obtaining the model forecasts as output. Each model run has an associated computational cost. In certain methods, e.g. Monte Carlo (MC) sampling methods, the model is run many times for different sets of inputs. Therefore, the total computational cost of the process has a linear correlation with the the number of model runs.

Bayesian approach is a robust method for quantification of uncertainty. It can be thought of as a learning process. Initially, there exists a very limited knowledge about the hydrocarbon reservoir. By the passage of time, some data are acquired, e.g. by drilling exploration wells or conducting seismic surveys. These data are then assimilated into the mathematical models which results in better estimation of the unknown parameters and better quantification of the associated uncertainty. This process is known as data assimilation (or in the context of hydrocarbon reservoir problems, history-matching).

Ensemble-based data assimilation (DA) algorithms are examples of MC methods which utilize Bayesian methodology and are widely used for uncertainty quantification in hydrocarbon reservoir problems. In these methods both the data and the unknown parameters are treated as random variables. The uncertainty in the parameters is represented using samples from them. The numerical models are then run on all members of the sample, and an ensemble of model forecasts is formed which represents the uncertainty in the model forecasts. Finally by comparing the model forecasts and the observation data, the ensemble is updated and the posterior uncertainty is obtained.

DA methods have several advantages including computational and storage efficiency, particularly in the problems where the number of uncertain parameters are large. However, they have their own limitations. Monte Carlo approximations play a crucial role in ensemble-based DA. Due to computational-cost limitations, the ensemble size is limited to roughly one hundred. Using straightforward ensemble-based DA, the degrees of freedom of the problem would equal the ensemble size, and such an approach would result in significant Monte Carlo errors. The negative effects of Monte Carlo errors are enlarged if large amounts of data are assimilated simultaneously, e.g. assimilation of inverted seismic data. This normally results in underestimation of the variance of the unknown parameters, and in more severe cases ensemble collapse.

The most widely used treatment for Monte Carlo errors is a regularization technique called distance-based localization. As the name suggests the underlying assumption of this method is that the unknown parameters should be updated only based on the data at their locality. In spite of its being the standard method of regularization, it has its own drawbacks. It does not allow for the parameters to be updated by the data

acquired away from their vicinity, although there may exist true correlations between them. Additionally, this method has difficulty assimilating the data which do not belong in any specific spatial location.

Simply increasing the ensemble size will of course reduce Monte-Carlo errors, but it will also increase computational cost. Utilization of surrogate models which have lower computational cost comparing to fine mathematical models renders the possibility of increasing the ensemble size without increasing the total computational cost. Use of a surrogate reservoir model will, however, introduce modeling errors in addition to those already present in fine simulation results.

Multilevel simulations utilize a selection of surrogate models that constitute hierarchies in both fidelities and computational costs (multilevel models). The idea is to decrease Monte Carlo errors without increasing numerical errors too much. Multilevel data assimilation (MLDA) utilizes multilevel simulations in the forecast step of the DA. In this PhD work, several MLDA algorithms were developed and investigated for assimilation of inverted seismic data in simplistic reservoir models (Papers A, C, and D). Multilevel simulations, similar to lower fidelity simulations, introduce modeling errors to the problem, in addition to the modeling errors already existing. Calling these secondary errors multilevel modeling errors (MLME), several computationally inexpensive methods were devised for partially addressing MLME and their quality was assessed for assimilation of inverted seismic data in simplistic reservoir models (Paper B). Finally, one of the developed MLDA algorithms was investigated for assimilation of inverted seismic data in a realistic reservoir model (Paper E).

Chapter 2

Inverse problems

In order for us to understand the natural phenomena, we study physical systems. There are three different steps to study a physical system [Tarantola, 2005]. The first step is parameterization, in which a minimum number of model quantities are found such that they completely characterize the system. The second step is forward modeling, in which we try to discover the physical laws governing the system. In doing so, we formulate mathematical models using which we can predict some observable phenomena in the system for any given set of model quantities. Since neither the parameterization nor the forward modeling are exact, these predictions come with associated errors. The third step is inverse modeling. In this step the actual values of some measurements of observable phenomena are used to infer some information about the model quantities. As opposed to the first two steps which are inductive, inverse modeling is deductive. By this we mean that unlike the first two steps where the process of thinking and inference cannot be explicitly explained, in the third step, given a set of assumptions, the inference is made mostly using the principles of logic and probability theory.

2.1 Inverse problem formulation

Consider the equation relating the observation data, d, and the model quantities, z, using the forward model $\mathcal{M}$,

$$d = \mathcal{M}(z)\,. \tag{2.1}$$

As mentioned parameterization and forward modeling entail errors. Therefore, a correct formulation of the problem is

$$d = \mathcal{M}(z_{true}) + \epsilon\,, \tag{2.2}$$

where z_{true} represents "true" model quantities obtained with accurate model, and ϵ represents the unknown error term. Considering ϵ as an unknown variable, without loss of generality one can assume it has zero mean. This error entails both error in the measurements and modeling. However, since the error in modeling is hard to quantify, it is normally neglected.

A mathematical problem is considered well-posed if the following three conditions hold [Hadamard, 1902].

(i) There must exit a solution.

(ii) The solution must be unique.

(iii) The solution must be stable.

If any of these conditions does not hold, the corresponding problem is called an ill-posed problem. In the case of inverse problems, due to presence of noise in the observations and modeling error, the first condition normally does not hold. The second condition does not hold necessarily, since many sets of model quantities may equally satisfy (2.1). The last condition may be violated in certain inverse problems, meaning that small perturbations in the observations result in drastic changes in the solution. Therefore, solving the inverse problems, normally being ill-posed, is a hard task.

In this work, if z and d are fields, we assume that they are discretized into sets of model quantities and data. For study of continuous inverse problems, see e.g. [Bagchi and Borkar, 1984; Bergemann and Reich, 2012; Borkar and Bagchi, 1982]. There are two approaches to inverse problem: the classical approach and the stochastic approach. I will discuss these two approaches, respectively.

2.2 Classical approach

Based on the ill-posed nature of the problem, it is assumed that there exists no solution z such that (2.1) is fully satisfied. Considering (2.2) instead, in the classical approach no assumptions are made on the nature of the error term, ϵ. A cost function is defined as the least squares of the error terms given by

$$J(z) = ||\mathcal{M}(z) - d||^2 \,, \tag{2.3}$$

which will be equivalent of the Euclidean norm of the total error; the most commonly used norm for such problems. The solution of the inverse problem will reduce to solving

the minimization problem given by,

$$z^* = \arg\min_z J(z)\,. \tag{2.4}$$

Assuming continuous functionality for $\mathcal{M}(z)$, a necessary condition for the minima of the problem, z^*, (both local and global) is

$$\nabla J(z) = 0\,. \tag{2.5}$$

As $\mathcal{M}$ is normally non-linear, the solution of (2.5) is found using iterative optimization algorithms, including Newton algorithm, Gauss-Newton algorithm, Levenberg-Marquardt algorithm, or Quasi-Newton methods. For details of these optimization algorithms, see e.g. [Bonnans et al., 2006].

Some problems that may arise when using classical approach to inverse problem are presence of multiple local minima, divergence of the minimization sequence of the optimization algorithms, and instability of the solution with respect to data. A technique that can alleviate these problems is regularization. Tikhonov regularization method [Tikhonov, 1966] is one of the most popular methods in regularization. It will be discussed briefly in Section 4.

2.3 Stochastic approach

As opposed to the classical approach in which the nature of error terms does not play any role in formulation of the problem, in stochastic approach the error terms are considered as random variables with known statistics. The non-exact forward model is given by (2.2), similar to the classical approach. However, since the data are considered as stochastic variables, for each of them there exists a probability distribution function (PDF). Assuming independence between the data, the conditional probability of the data given the model quantities can be formulated as

$$f_{D|Z}(d|z) = f_{D_1|Z}(d_1|z)\ldots . f_{D_{N_D}|Z}(d_{N_D}|z) \tag{2.6}$$

where N_D is the number of data, and $f_{D_k|Z}$ denotes the conditional probability of datum k given the model quantities, $1 \leq k \leq N_D$. Since the data are given and the problem is about finding models that match the data well, by a change of perspective, we can write the likelihood function, $\mathcal{L}(z, d)$ as

$$\mathcal{L}(z, d) = f(d|z) = f(d_1|z)\ldots . f(d_{N_D}|z)\,. \tag{2.7}$$

In (2.7) I have used f for all the PDFs in the equation, and assumed it is clear in the context. This abuse of notation will be used from this point on. Assuming Gaussian model for error, we will have

$$\mathcal{L}(z,d) \propto \prod_{k=1}^{N_D} \exp\left(-\frac{(d_k - (\mathcal{M}(z))_k)^2}{2\sigma_{\mathrm{k}}}\right), \tag{2.8}$$

where σ_k denotes the standard deviation of the k^{th} error term, and $(\mathcal{M}(z))_k$ denotes the k^{th} element of the model forecast vector. Defining the negative log-likelihood as the cost function,

$$J(z) = \sum_{k=1}^{N_D} \frac{(d_k - (\mathcal{M}(z))_k)^2}{2\,\sigma_{\mathrm{k}}}, \tag{2.9}$$

and minimizing it, results in Maximum Likelihood (ML) estimate of the model quantities. If the error terms are independent and share the same Gaussian distribution, ML estimate reduces to the classical approach. However, in the more general case, when the errors are correlated, the minimization can be formulated as,

$$z^* = \arg\min_z ||\mathcal{M}(z) - d||^2_{\mathbb{C}_D^{-1}}, \tag{2.10}$$

where $\mathbb{C}_D$ is the data error covariance matrix, and $||.||_{\mathbb{C}_D^{-1}}$ is the norm induced by this matrix.

2.3.1 Bayesian formulation

A sub-category of stochastic formulation of the inverse problem is Bayesian approach. In this approach a prior knowledge about Z is considered, and it is updated using the data. In some sense, Bayesian formulation of the inverse problem is a learning algorithm. Considering (2.1), the problem that we try to solve is estimation of Z using observed measurements of D. We assume Z and D as random vectors, and z and d are realizations from them. Denoting PDF of Z as $f(z)$, and the conditional PDF of D given Z as $f(d|z)$, using Bayes' rule we can calculate the conditional PDF Z given D, $f(z|d)$, as

$$f(z|d) = \frac{f(d|z)\,f(z)}{f(d)}. \tag{2.11}$$

Using this formulation, the prior distribution of Z is conditioned to D and results in the posterior distribution of Z. The denominator is the PDF of D which acts as a normalizing factor. In practice, however, this term is hard to quantify. Hence, the proportionality,

$$f(z|d) \propto f(d|z)\,f(z), \tag{2.12}$$

is utilized instead of (2.11).

2.4 Parameter estimation

The model quantities are divided into two categories. The first category consists of the model quantities which are variant with time. These are called model states. The second category consists of the coefficients in the mathematical equations of the model which do not change with time. These are called model parameters. In the context of reservoir history-matching, most of the time, we want to estimate parameters like permeabilities and porosities which are considered constant with time. Hence, we formulate a parameter estimation problem.

The methods developed for parameter estimation are historically based on the state estimation techniques developed for studying the stochastic processes. Assuming Z is variant with time. There are two approaches in conditioning the time-variant model quantities to the data: smoothing and filtering. In smoothing the model quantities at each time step are conditioned to all the data. On the other hand, in filtering, at each time step, the model quantities are conditioned to the data in that time step and the data in the past time steps [Särkkä, 2013]. In other words filtering assumes a causal relationship between the model quantities and the data, while smoothing does not. Here we consider smoothing as the approach of data assimilation.

Considering the smoothing formulation, assuming a dynamic model we have,

$$f(z_1, z_2, ..., z_K|d_1, d_2, ..., d_K) \propto f(d_1, d_2, ..., d_K|z_1, z_2, ..., z_K)\, f(z_1, z_2, ..., z_K)\,. \tag{2.13}$$

In (2.13), z_k and d_k denote the model quantities and the data at time step k, $1 \leq k \leq K$. Since we want to formulate the parameter estimation problem, the model quantities are assumed to be the same in all the time steps. Hence, (2.13) reduces to

$$f(z|d_1, d_2, ..., d_K) \propto f(d_1, d_2, ..., d_K|z)\, f(z)\,. \tag{2.14}$$

If the data at different time steps are independent, the posterior parameters distribution is given by,

$$f(z|d_1, d_2, ..., d_K) \propto f(d_1|z)\, f(d_2|z)\, ...\, f(d_K|z)\, f(z)\,. \tag{2.15}$$

2.4.1 Sampling the posterior distribution

In spite of straightforward derivation of the parameter update formulae, these equations normally involve complex multi-dimensional integrals which are difficult to solve analytically. For realistic inverse problems, due to exponential growth of the computational cost with increase in dimensionality of the problem, discretizing (2.14) and numerical integration of it is not feasible either. However, there are algorithms which can sample from the posterior correctly under certain conditions. The most popular algorithm for sampling from the posterior is Markov Chain Monte Carlo (MCMC).

The MCMC algorithm relies on the Markov chain property, meaning that starting from an arbitrary sample from the distribution z, a long chain of realizations are sampled from the parameter space, such that the probability of sampling each point only depends on the previous sample point, not the rest of the chain before that point. This chain can sample a distribution of interest given certain conditions are satisfied.

The well-known Metropolis-Hasting algorithm [Hastings, 1970; Metropolis et al., 1953] can be summarized as follows. Consider $p(z^*|z^\times)$ as the proposal probability of moving from $z^\times$ to z^* at a certain point in the chain.

1. Generate an initial sample from the parameter distribution z^1 and set $k = 1$.
2. Generate a sample z^* from the distribution $p(z^*|z^k)$
3. Calculate the acceptance ratio $a(z^*, z^k)$ given by
$$a(z^*, z^k) = \min\left(1\,,\, \frac{f(z^*|d)\; p(z^k|z^*)}{f(z^k|d)\; p(z^*|z^k)}\right)\,.$$
4. Generate a random sample, u, from the uniform distribution $\mathcal{U}[0, 1]$
5. If $a(z^*, z^k) > u$ then $z^{k+1} = z^*$, else $z^{k+1} = z^k$
6. If a certain chain length is not reached, set $k \leftarrow k + 1$, and go to step 2.

For sufficiently long chain lengths, this algorithm will sample the posterior distribution correctly if $p(z^*|z^\times)$ is designed such that the chain will be aperiodic, meaning that same pieces of chain will not repeat in fixed intervals, and moving from a point to any other point is positive recurrent, meaning that the number of steps needed for the move is finite [Robert and Casella, 1999].

One of the main issues with MCMC algorithms is that they normally need exceedingly large chain lengths to obtain a correct sample from a distribution. Due to high computational costs for each of the sample points, e.g. one simulation run in the reservoir history-matching problem, achieving long chains is not a practical possibility.

2.4.2 Linear-Gaussian problems

Another way of tackling the difficulty of solving (2.14) is imposing certain assumptions on the distributions and the relation between the model parameters and the data, such that (2.14) will be simplified. There exists a closed form solution for this equation, if the following two conditions hold.

(i) Both the error in modeling and measurements, and the uncertainty in the prior estimates of the models are Gaussian.

(ii) There is a linear relation between the parameters and the data.

This case is presented here. We assume the uncertainty in the prior model estimate is Gaussian. Hence, we can write,

$$f(z) = \text{const. } \exp\left(-\frac{1}{2}(z-\overline{z}^{pri})^T \mathbb{C}_{Z^{pri}}{}^{-1}(z-\overline{z}^{pri})\right) , \tag{2.16}$$

where $\overline{z}^{pri}$ is the mean of the prior parameter estimates, $\mathbb{C}_{Z^{pri}}$ is the prior parameters covariance matrix, const is a normalizing constant, and T denotes the transpose operator. Similarly, we assume that the observation d_o with observation error covariance matrix $\mathbb{C}_d$ is given, and the modeling error defined as $d - \mathcal{M}(z)$ has a Gaussian error given by $\mathcal{N}(0, \mathbb{C}_M)$. Given these assumptions, the likelihood function is given by

$$f(d_o|z) = \text{const. } \exp\left(-\frac{1}{2}\left(d_\text{o} - \mathcal{M}(z)\right)^\text{T} \mathbb{C}_D^{-1}\left(d_\text{o} - \mathcal{M}(z)\right)\right) , \tag{2.17}$$

where exp is the exponential function, and $\mathbb{C}_D = \mathbb{C}_d + \mathbb{C}_M$. This is derived by convolution of the two Gaussian distributions associated with $\mathbb{C}_d$ and $\mathbb{C}_M$ [Tarantola, 2005]. Hence, for a given observation d_o, the posterior parameters distribution is given by

$$\begin{aligned} f(z|d_o) &= \text{const} \,.\, f(d_o|z)\, f(z) \\ &= \text{const.exp}[-\frac{1}{2}(z-\overline{z}^{pri})^T \mathbb{C}_{Z^{pri}}{}^{-1}(z-\overline{z}^{pri}) - \frac{1}{2}\left(d_o - \mathcal{M}(z)\right)^T \mathbb{C}_D^{-1}\left(d_o - \mathcal{M}(z)\right)] \\ &= \text{const} \,.\, \exp(-J(z)) \,, \end{aligned} \tag{2.18}$$

where $J(z)$ is,

$$J(z) = \frac{1}{2}(z - \overline{z}^{pri})^T \mathbb{C}_{Z^{pri}}{}^{-1}(z - \overline{z}^{pri}) + \frac{1}{2}\left(d_o - \mathcal{M}(z)\right)^T \mathbb{C}_D^{-1}\left(d_o - \mathcal{M}(z)\right) \tag{2.19}$$

Assuming $\mathcal{M}$ is linear,

$$\mathcal{M}(z) = \mathrm{M}\, z\ , \tag{2.20}$$

where M maps the parameters space into the data space, we will have,

$$J(z) = \frac{1}{2}(z - \overline{z}^{pri})^T \mathbb{C}_{Z^{pri}}{}^{-1}(z - \overline{z}^{pri}) + \frac{1}{2}\left(d_o - \mathrm{M}\, z\right)^T \mathbb{C}_D^{-1}\left(d_o - \mathrm{M}\, z\right)\ . \tag{2.21}$$

Then the posterior distribution of the model parameters will also be Gaussian, and mean and variance will be given by,

$$\begin{aligned} \overline{z}^{pos} &= \left(\mathrm{M}^{\mathrm{T}}\mathbb{C}_D^{-1}\mathrm{M} + \mathbb{C}_{Z^{\mathrm{pri}}}{}^{-1}\right)^{-1}\left(\mathrm{M}^{\mathrm{T}}\mathbb{C}_D^{-1}\, d_{\mathrm{o}} + \mathbb{C}_{Z^{\mathrm{pri}}}{}^{-1}\,\overline{z}^{\mathrm{pri}}\right) \\ &= \overline{z}^{pri} + \left(\mathrm{M}^{\mathrm{T}}\mathbb{C}_D^{-1}\mathrm{M} + \mathbb{C}_{Z^{\mathrm{pri}}}{}^{-1}\right)^{-1}\mathrm{M}^{\mathrm{T}}\mathbb{C}_D^{-1}\left(d_{\mathrm{o}} - \mathrm{M}\,\overline{z}^{\mathrm{pri}}\right) \\ &= \overline{z}^{pri} + \mathbb{C}_{Z^{pri}}\mathrm{M}^{\mathrm{T}}\left(\mathrm{M}\,\mathbb{C}_{Z^{\mathrm{pri}}}\,\mathrm{M}^{\mathrm{T}} + \mathbb{C}_D\right)^{-1}\left(d_{\mathrm{o}} - \mathrm{M}\,\overline{z}^{\mathrm{pri}}\right)\ , \end{aligned} \tag{2.22}$$

and

$$\begin{aligned} \mathbb{C}_Z^{pos} &= \left(\mathrm{M}^{\mathrm{T}}\,\mathbb{C}_D^{-1}\,\mathrm{M} + \mathbb{C}_Z^{pos\,-1}\right)^{-1} \\ &= \mathbb{C}_{Z^{pri}} - \mathbb{C}_{Z^{pri}}\,\mathrm{M}^{\mathrm{T}}\left(\mathrm{M}\,\mathbb{C}_{Z^{\mathrm{pri}}}\,\mathrm{M}^{\mathrm{t}} + \mathbb{C}_D\right)^{-1}\,\mathrm{M}\,\mathbb{C}_{Z^{\mathrm{pri}}}\ , \end{aligned} \tag{2.23}$$

where the second an third equality in (2.22) and the second equality in (2.23) are given by Woodbury matrix identity [Woodbury, 1950]. For details of derivation see [Tarantola, 2005]. The original formulation of the explained method was derived for the state estimation problem by Kalman [Kalman, 1960]. The represented method for parameter estimation heavily draws on his work.

Chapter 3

Ensemble-based methods

Sampling the posterior field is a challenging task in high-dimensional problems. It is proven that MCMC method samples the posterior in (2.14) correctly, given the Markov chain is long enough. However, due to costly runs of the forward model, attaining sufficiently long Markov chains is not a possibility for the reservoir history-matching problem. Adding the Linear-Gaussian assumption, Evensen [1994] formulated an alternative Monte Carlo sampling method for solving (2.14) based on Kalman filter [Kalman, 1960]. This algorithm, known as ensemble Kalman filter (EnKF), is considerably cheaper than MCMC in terms of the computational cost and gives a good estimation of the posterior distribution if the problem is not far from the Linear-Gaussian assumptions. In essence, it can be described as a Monte Carlo approximation of the Kalman filter. The novelty of this algorithm comes in utilization of an ensemble for computation of the statistics of the model quantities and the model forecasts. This enables EnKF to obtain the statistics of the prior model quantities as well as the statistics of the model forecasts by running the forward model on all the ensemble members. Van Leeuwen and Evensen [1996] formulated Ensemble smoother (ES) using the same idea, and ever since ensemble-based methods have been widely used in oceanography, atmospheric sciences, and oil reservoir problem, to name a few, see e.g. [Evensen et al., 2022; Hamill, 2006; Jung et al., 2018; Oliver and Chen, 2011; Ruiz et al., 2013] .

There have been several improvements to the ensemble-based algorithms discussed above, and several algorithms have been developed based on them, see e.g. [Burgers et al., 1998; Chen and Oliver, 2013; Emerick and Reynolds, 2013; Gu and Oliver, 2007; Raanes et al., 2019; Sakov et al., 2012]. Here I discuss three of these methods, i.e. ES [Van Leeuwen and Evensen, 1996], ensemble randomized maximum likelihood (EnRML) [Chen and Oliver, 2012], and a general form of ensemble smoother with multiple data assimilation (ESMDA) [Emerick and Reynolds, 2013; Mannseth, 2020].

3.1 Definition of general terms

I start with definition of the generic terms for all the ensemble-based methods that are going to be discussed.

The distribution of the model parameters is assumed to have mean $\overline{z}$ and covariance matrix $\mathbb{C}_Z$. This distribution is sampled, and its realizations are concatenated into the ensemble parameters matrix, Z (Roman and boldface, not to be mistaken with the italic parameters random vector Z),

$$\mathrm{Z} = (z_1, z_2, ..., z_{N_e}) \ , \tag{3.1}$$

where Z is a $N_Z \times N_e$ matrix; N_Z and N_e being the number of unknown parameters and the ensemble size, respectively. If the distribution is Gaussian, the ensemble members can be sampled using,

$$z = \overline{z} + \mathrm{L}\, x \ , \tag{3.2}$$

where L is any matrix such that $\mathrm{L}\,\mathrm{L}^{\mathrm{T}} = \mathbb{C}_Z$, e.g. Cholesky decomposition, and x is a vector of the same size as z whose elements are independent samples from standard normal distribution. We also define the projection matrix Π,

$$\Pi = \frac{1}{\sqrt{N_e - 1}} \left(\mathrm{I}_{N_e} - 1\, 1^{\mathrm{T}}\right) \ , \tag{3.3}$$

where I_{N_e} is the identity matrix of size N_e, and 1 is a vector of ones of the required size. Multiplying Z from right to Π will make the ensemble centered (zero-mean), and scaled with a factor of $1/\sqrt{N_e - 1}$. Using the projection, we define the scaled perturbation, B_Z, as

$$\mathrm{B}_Z = \mathrm{Z}\Pi \tag{3.4}$$

such that the ensemble covariance matrix will be given as,

$$\mathrm{C}_Z = \mathrm{B}_Z \mathrm{B}_Z^{\mathrm{T}} \, . \tag{3.5}$$

Similarly, defining the model forecast, y, associated with each ensemble member as

$$y = \mathcal{M}(z) \, . \tag{3.6}$$

The $N_Y \times N_e$ matrix of model forecasts, N_Y being the dimension of the model forecasts, is defined similar to Z based on concatenation of the ensemble of the forecasts as

$$\mathrm{Y} = (y_1, y_2, ..., y_{N_e}) \ . \tag{3.7}$$

Defining B_Y as

$$\mathrm{B}_Y = \mathrm{Y}\Pi\,, \tag{3.8}$$

the ensemble covariance matrix of the model forecasts, C_Y, can be formulated as

$$\mathrm{C}_Y = \mathrm{B}_Y\mathrm{B}_Y^{\mathrm{T}}\,. \tag{3.9}$$

Similarly, the cross-covariance between the parameters and the model forecasts is given as

$$\mathrm{C}_{Z,Y} = \mathrm{B}_Z\mathrm{B}_Y^{\mathrm{T}}\,. \tag{3.10}$$

3.2 Ensemble-based data assimilation algorithms

3.2.1 Ensemble smoother

Ensemble smoother (ES) [Van Leeuwen and Evensen, 1996] was formulated for problems which are not too nonlinear, and consists of one DA step, which is described as follows.

Firstly, the matrix of prior parameters ensemble is sampled as,

$$\mathrm{Z} = \left(z_1^{pri}, z_2^{pri}, ..., z_{N_e}^{pri}\right)\,. \tag{3.11}$$

After running the forward model on the prior ensemble, the prior matrix of parameter forecasts, Y, the perturbation matrices B_Z and B_Y, and the sample covariance matrices C_Z, C_Y, and $\mathrm{C}_{Z,Y}$ are computed based on the guidelines in Section 3.1.

Corresponding to each of the ensemble members, one realization of the perturbed data, d, is sampled from D; $\{d_j|1 \leq j \leq N_e, D \sim \mathcal{N}\left(\overline{d}, \mathbb{C}_D\right)\}$.

The updated ensemble of parameters realizations are given by,

$$z_j^{pos} = z_j^{pri} + \mathrm{K}\left(d_j - y_j\right)\,, \tag{3.12}$$

where the Kalman gain matrix, K is given by

$$\mathrm{K} = \mathrm{C}_{Z,Y}\left(\mathrm{C}_Y + \mathbb{C}_D\right)^{-1}\,. \tag{3.13}$$

The limiting distribution given by the updated realizations of the model parameters, in the infinite ensemble size, samples the posterior distribution of the model parameters

correctly, given the problem is Linear-Gaussian.

3.2.2 Ensemble randomized maximum likelihood

Ensemble randomized maximum likelihood (EnRML) was introduced in [Chen and Oliver, 2012] as an iterative ensemble smoother. This algorithm was developed to handle more non-linear inverse problems.

Randomizing the objective function in (2.19), the cost function associated with each ensemble member is given by

$$J_j(z_j) = \frac{1}{2}||z_j - z_j^{pri}||^2_{\mathbb{C}_{Z^{pri}}{}^{-1}} + \frac{1}{2}||\mathcal{M}(z_j) - d_j||^2_{\mathbb{C}_D^{-1}} \,. \tag{3.14}$$

In the linear-Gaussian case J_j is minimized by the update in (3.12). However, assuming non-linearity in $\mathcal{M}$, this will not be the case. Therefore, an iterative scheme should be used to minimize J_j for all realizations.

At each iteration, i, the forward model is run on the ensemble, the required statistics and matrices are calculated based on definitions in Section 3.1. (I add superscript i to the notations so as to discriminate them from the single-iteration ES case. Iteration 0 corresponds to the prior ensemble.)

EnRML uses Gauss-Newton scheme for minimization of the cost function. This scheme requires the gradient and the Hessian of J_j for each of the ensemble members. In EnRML, these are computed using approximations based on the ensemble. Hence, the gradient, ∇J_j^i, and the Hessian, H^i, for the parameter realizations at iteration i, z_j^i, are given as

$$\nabla J_j^i \approx \mathbb{C}_{Z^{pri}}{}^{-1} \left(z_j^i - z_j^{pri}\right) + \mathrm{M^i}^T \mathbb{C}_D^{-1} \left(\mathcal{M}(z_j^i) - d_j\right) \,, \tag{3.15}$$

$$H^i \approx \mathbb{C}_{Z^{pri}}{}^{-1} + \mathrm{M^i}^T \mathbb{C}_D^{-1} \mathrm{M^i} \,, \tag{3.16}$$

respectively. In these formulae, $\mathrm{M^i}$ denotes the approximation to Jacobian of $\mathcal{M}(z)$ in the vicinity of the ensemble of parameters in that iteration, and H^i is approximated by neglecting the derivative of $\mathrm{M^i}$. $\mathrm{M^i}^T$ denotes the transpose (or more generally, the adjoint) of $\mathrm{M^i}$. In IES , the approximation,

$$\mathrm{M^i} \approx \mathrm{C}_{Z,Y}{}^{\mathrm{i}} \mathrm{C}_Z{}^{\mathrm{i}+} \,, \tag{3.17}$$

is used for calculation of $\mathrm{M^i}$, where superscript + denotes Moore-Penrose pseudo inverse. A mathematical justification for this can be found in [Raanes et al., 2019]. The realization

z_j^i is then updated as

$$z_j^{i+1} = z_j^i - \beta \left(H^{i^{-1}} \nabla J_j^i \right) , \tag{3.18}$$

where β is the step length which is defined based on "restricted-step algorithm" in [Oliver et al., 2008]. The update equation can be written as

$$z_j^{i+1} = z_j^i + \beta \left(\Delta_j^{i,pri} + \Delta_j^{i,lik} \right) , \tag{3.19}$$

where $\Delta_j^{i,pri}$ and $\Delta_j^{i,lik}$ are given by

$$\Delta_j^{i,pri} = (\mathrm{I}_{N_Z} - \mathrm{K^i M^i})[z_j^{pri} - z_j^i] , \tag{3.20}$$

$$\Delta_j^{i,lik} = \mathrm{K^i}[d_j - \mathcal{M}(z_j^i)] . \tag{3.21}$$

In these Equations, I_{N_Z} is the identity matrix of the parameter vector dimension, and the Kalman gain $\mathrm{K^i}$ is given by

$$\mathrm{K^i} = \mathrm{C}_{Z^{\mathrm{pri}}} \mathrm{M^i}^{\mathrm{T}} \left(\mathrm{M^i C}_{Z^{\mathrm{pri}}} \mathrm{M^i}^{\mathrm{T}} + \mathbb{C}_D \right)^{-1} . \tag{3.22}$$

The iterations are continued until convergence is obtained.

3.2.3 Assimilation of multiple linearly dependent data

In this section, I present an algorithm for assimilation of a set of linearly dependent data vectors $\{D^l\}_{l=1}^L$ where $\{D^l = \mathrm{U}^l D^L\}$, and U^l is a linear transformation. These types of data can be found in the context of multilevel data assimilation.

Consider a set of forward models $\{\mathcal{M}_l\}_{l=1}^L$ corresponding to each of the data vectors. An ensemble of prior realizations is generated, and at assimilation step l the forward model is run on all the ensemble members,

$$y_j^l = \mathcal{M}_l(z_j^l) . \tag{3.23}$$

The required matrices and distributions are generated based on definitions in Section 3.1. Here we add the superscript l to the notations to distinguish between different assimilation steps.

Assuming the data vector at assimilation step l is distributed as $D^l \sim \mathcal{N}(\overline{d}^l, \mathbb{C}_D^l)$, and d_j^l are realizations of it, each of the ensemble members is updated as

$$z_j^{l+1} = z_j^l + \mathrm{K}_l (d_j^l - y_j^l) \tag{3.24}$$

where the Kalman gain K_l is given by

$$\mathrm{K}_l = \mathrm{C}_{Z,Y}{}^l \left(\mathrm{C}_Y{}^l + \mathbb{C}_D^l\right)^{-1} . \tag{3.25}$$

In the asymptotic ensemble-size condition and Gaussian assumption for the distributions, the described algorithm samples correctly from the posterior parameters distribution given the following conditions hold,

(i) $\mathcal{M}_l(z) = \mathrm{U}^l \, \mathcal{M}_\mathrm{L}(z) = \mathrm{U}^l \, \mathrm{M} \, z$

(ii) $\mathbb{C}_D^l = \mathrm{A}^l \, \mathrm{U}^l \, \mathbb{C}_D^L \, \mathrm{U}^{l\mathrm{T}} \, \mathrm{A}^{l\mathrm{T}}$

(iii) $\sum_{l=1}^L \mathrm{U}^{l\,T} \mathbb{C}_D^l{}^{-1} \mathrm{U}^l = \mathbb{C}_D^L{}^{-1}$

The first condition states that $\{\mathcal{M}_l\}_{l=1}^L$ should all be linear and the same linear transformation relating different data levels should be relating the model forecasts. The second condition states that the transformed data error covariance matrix at level l, $\mathrm{U}^l \, \mathbb{C}_D^L \, \mathrm{U}^{l\mathrm{T}}$, should be inflated using the arbitrary inflation matrix A^l such that the third condition holds.

These conditions are known as the partially multiple data assimilation (PMDA) conditions. They are introduced in [Mannseth, 2020].

3.2.4 Ensemble smoother with multiple data assimilation

Ensemble smoother with multiple data assimilation (ESMDA) [Emerick and Reynolds, 2013] was another effort to address the problem of assimilation of data with stronger non-linearity. In this algorithm the data error covariance matrix is inflated, and the assimilation of data is performed in several steps. Substitution of one large update with several small updates helps to direct the ensemble towards the solution of the inverse problem.

Consider the algorithm in Section 3.2.3. If $\{D^l = D^L\}_{l=1}^L$ and $\{\mathcal{M}_l = \mathcal{M}_L\}_{l=1}^L$, i.e. the linear transformations relating the data and forward models at different levels are identity matrices, the algorithm introduced in Section 3.2.3 will be reduced to the ESMDA algorithm. Defining $\mathrm{A}^l = \alpha_l \, \mathrm{I}_L$, where α_l is a scalar value and I_L is the identity matrix with the size of the data vector, the PMDA condition will also be simplified such that it can be summarized as

$$\sum_{l=1}^{L} \frac{1}{\alpha_l} = 1 . \tag{3.26}$$

This condition is known as the multiple data assimilation (MDA) condition. It is proved that satisfying this condition in the linear-Gaussian case and asymptotically large ensemble sizes, ESMDA samples the posterior correctly. However, it is unclear how important MDA condtion is in the general non-linear problem.

Chapter 4

Regularization

As mentioned in Section 2 many of the inverse problems are ill-posed, i.e. negating one or more of the assumptions of a well-posed problem. Since solving an ill-posed problem is a challenging task, some strategies should be used to reformulate the problem such that its solution becomes more stable. A broad group of these techniques which stabilize the solutions of inverse problems are known as regularization techniques. In general regularization is the process of approximating the solution of an ill-posed problem by a family of similar well-posed problems.

In the classical approach, Tikhonov regularization [Tikhonov, 1966] is one of the standard formulations. Considering (2.3), using Tikhonov regularization, this equation can be reformulated as

$$J(z) = ||\mathcal{M}(z) - d||^2 + ||\Gamma(z - z^0)|| \,. \tag{4.1}$$

In this formulation by manipulating Γ, and solving the corresponding minimization problem, one can limit the solutions of the inverse problem to only those solutions which are inside a multi-dimensional ellipsoid centering around z^0. This will reduce the chance of obtaining non-physical solutions due to instability of the ill-posed problem. For other regularization techniques in classical approach, see e.g. [Benning and Burger, 2018; Engl et al., 1996]

In ensemble-based data assimilation (DA) the ensemble size is limited. In the reservoir history-matching problem it is normally around one hundred [Emerick and Reynolds, 2011; Yin et al., 2019; Zhang et al., 2021]. Since the number of unknown parameters in the problem is normally much larger, the degree of freedom of the problem is determined by the ensemble size. A problem that often arises in the prime form of the DA algorithms is presence of spurious correlations when computing the covariances. Since computation of the covariance matrices used in the update step of the DA is a Monte-Carlo process,

the limited ensemble size can result in spurious correlations both between parameters themselves, and between model forecasts and the parameters. One of the common consequences of the Monte-Carlo errors associated with the DA process is the ensemble of parameters realizations to fit for some of the noise, thereby reducing the uncertainty in the ensemble too much and consequently underestimating it. In case of simultaneous assimilation of large amounts of data, e.g. inverted seismic data, the problem is even more aggravated, and sometimes the DA results in ensemble collapse.

Increasing the ensemble size would be the most effective way of resolving this problem, but due to limited computational resources, this is not a possibility. Several strategies are devised to address this problem. I cover two of the main general strategies for regularization of the DA methods, namely localization and ensemble sub-space inversion.

4.1 Localization

The idea behind localization is defining some form of locality for each parameter such that the update equation will be mostly affected by the data in that locality. Localization was originally formulated based on the physical location of the data, as the term implies. Hence, distance-based localization is known to be the standard method for localization [Agbalaka and Oliver, 2008; Bannister, 2017; Chen and Oliver, 2010; Soares et al., 2018]. However, another method for localization, known as correlation-based localization [Luo and Bhakta, 2020; Luo et al., 2018], is also gaining interest in recent years. In this section, I present both of these methods. In addition, I present three of the main techniques by which the update formulations of the DA algorithms are manipulated, i.e. local analysis, Kalman gain localization, and covariance localization.

4.1.1 Local analysis

Local analysis, introduced in [Brusdal et al., 2003], is a regularization method which is very popular in the atmospheric sciences, weather and oceanography community [Oke et al., 2013; Ott et al., 2004; Sakov and Bertino, 2011]. This technique has also been used in the reservoir history-matching problem [Chen and Oliver, 2017; Fahimuddin et al., 2010]. In this method, both the unknown parameters and the data are grouped into certain groups, normally based on their location, and consequently the update equation is performed for each group of parameters using their corresponding group of data, independently of other groups. For example, in the reservoir history-matching problem, the permeabilities and porosities associated with each column of grid cells can be con-

sidered as one group. If the DA algorithm chosen is ES, for example, the Kalman gain will be of the form,

$$\mathrm{K}^* = \mathrm{C}_{Z,Y}{}^* \left(\mathrm{C}_Y{}^* + \mathbb{C}_D^*\right)^{-1} , \tag{4.2}$$

where $\mathrm{C}_{Z,Y}{}^*$ has only the rows and columns from $\mathrm{C}_{Z,Y}$ which are pertaining to the parameters and model forecasts which are being updated or assimilated in that specific group. Similar rule applies to $\mathrm{C}_Y{}^*$ and $\mathbb{C}_D^*$, and $*$ is a wild-card notation for each of the groups. By doing so, each group of parameters are conditioned to only a part of the data, and as a result the issue of balancing the degrees of freedom of the problem and the information content in the data is alleviated.

One of the issues with local analysis is deciding on the groups. There are no obvious guidelines for choice of the groups, and it normally needs expert knowledge of the field. This reduces the robustness of the method. Another problem with local analysis is that instead of one update equation, the process may need thousands of them. This can be computationally very demanding. However, since the updates are performed independently, parallelization is a possibility which can reduce the computation time drastically.

4.1.2 Kalman gain localization

One of the most straightforward methods of localization is the Kalman-gain localization. The Kalman gain matrix has the row size equal to the number of unknown parameters, and the column size equal to the number of data points to be assimilated. Hence, each of the elements of the Kalman gain matrix represents how much a specific parameter is affected by a specific datum. The idea in Kalman-gain localization is that each of the elements of the Kalman gain matrix is tapered, i.e. multiplied by a real number between 0 and 1, such that each parameter is mostly affected by the data at its locality. Hence the data at the locality of each parameter are tapered close to 1, and the data far from the locality of that parameter are tapered close to 0. Therefore, the Kalman-gain for this method can be written as,

$$\mathrm{K}^* = \mathrm{T} \circ \mathrm{K} , \tag{4.3}$$

where K is the Kalman gain from an arbitrary ensemble-based algorithm, T is a tapering matrix whose elements are between 0 and 1, and $\circ$ is the element-wise multiplication (or the Schur product) of the two matrices.

Matrix T is essentially designed to remove the spurious correlations between the parameters and the data. There are several ways to design this matrix both based on the spatial distance of the parameters and the data, and the correlations between them which will

be discussed in Sections 4.1.4 and 4.1.5, respectively.

As a direct consequence of its formulation, Kalman-gain localization can only be used in the DA algorithms for which a Kalman gain update can be formulated, e.g. it cannot be used coupled with algorithm proposed by [Evensen et al., 2019]. There are some advantages with using this technique. Implementation-wise it can be simply added to the final step of many DA algorithms, e.g. ES and ESMDA. In the original formulation of the Kalman gain the updated ensemble will always be in the sub-space spanned by the prior ensemble. By use of Kalman gain localization rank of the Kalman-gain matrix will increase, and the inverse problem is solved in a larger space.

4.1.3 Covariance localization

An alternative to Kalman-gain localization is to localize the covariances in the update equation [Devegowda et al., 2007; Emerick and Reynolds, 2011; Sakov and Bertino, 2011]. The difference between this method and Kalman-gain localization is tapering the parameters-forecasts cross-covariance and model-forecast covariance matrices instead of tapering the entire Kalman gain matrix. Considering the Kalman gain formula for ES, after implementing the covariance localization, it will transform into

$$\mathrm{K}^* = \left(\mathrm{T}_{Z,Y} \circ \mathrm{C}_{Z,Y}\right)\left(\mathrm{T}_{Y,Y} \circ \mathrm{C}_Y + \mathbb{C}_D\right)^{-1} , \tag{4.4}$$

where $\mathrm{T}_{Z,Y}$ is the tapering matrix for the cross-covariance between the parameters and the model forecasts, $\mathrm{T}_{Y,Y}$ is the tapering matrix for the covariance of the model forecasts, and $\circ$ denotes Schur product. In this equation, $\mathrm{T}_{Y,Y}$ has the following form,

$$\mathrm{T}_{Y,Y} = \mathrm{T}_Y \mathrm{T}_Y^{\mathrm{t}} , \tag{4.5}$$

where superscript t denotes transpose of the matrix. This is done to assure that the tapering matrix has the correct form of a covariance matrix. Designing the tapering matrices is similar to their design for Kalman-gain localization and will be discussed in Sections 4.1.4 and 4.1.5.

A possible limitation of this method is the size of $\mathrm{T}_{Y,Y}$. Since the dimension of the data can be very large, computation and storage of $\mathrm{T}_{Y,Y}$ may not be feasible. However, comparison of Kalman-gain localization and covariance localization in multi-phase flow DA shows that covariance localization is more robust and avoids artifacts when the size of the localization area is small [Chen and Oliver, 2010]. On the other hand, Kalman-gain localization normally results in higher ensemble variability after the update [Chen

and Oliver, 2010].

4.1.4 Distance-based tapering

The underlying assumption in distance-based tapering is that there exist spatial correlations between the data and parameters such that they are highest when the distance between a datum and a parameter is small, and the correlation decreases by increase in distance. Finally after reaching a critical distance this correlation should be zero. Based on this underlying assumption, several formulations are suggested for the tapering function calculating these spatial correlations. One of the most popular formulations, Gaspari-Cohn formulation [Gaspari and Cohn, 1999], for the tapering function, $\tau(h)$, is given as

$$\tau(h) = \begin{cases} -\frac{1}{4}\left(\frac{h}{\Lambda}\right)^5 + \frac{1}{2}\left(\frac{h}{\Lambda}\right)^4 + \frac{5}{8}\left(\frac{h}{\Lambda}\right)^3 - \frac{5}{3}\left(\frac{h}{\Lambda}\right)^2 + 1 & 0 \leq h \leq \Lambda \\ \frac{1}{12}\left(\frac{h}{\Lambda}\right)^5 - \frac{1}{2}\left(\frac{h}{\Lambda}\right)^4 + \frac{5}{8}\left(\frac{h}{\Lambda}\right)^3 + \frac{5}{3}\left(\frac{h}{\Lambda}\right)^2 - 5\frac{h}{\Lambda} + 4 - \frac{2}{3}\left(\frac{h}{\Lambda}\right)^{-1} & \Lambda \leq h \leq 2\Lambda \\ 0 & h > 2\Lambda \end{cases} \quad (4.6)$$

where h is the spatial distance between a datum and an unknown parameter, and Λ is the Gaspari-Cohn critical distance.

Another well-known formulation for the tapering function, known as Furrer-Bengtsson formulation [Furrer and Bengtsson, 2007], is given as

$$\tau(\rho) = \frac{1}{1 + (1 + c^2/\rho^2)/N_e}\,, \quad (4.7)$$

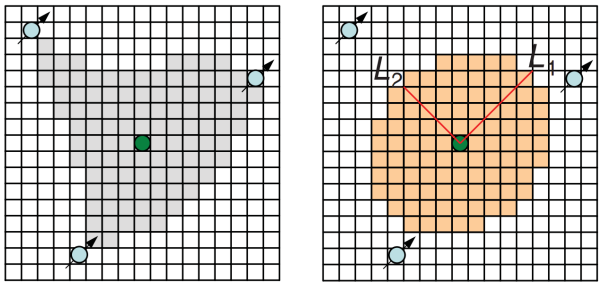
where ρ is obtained using a correlation function, e.g. exponential function given as

$$\rho(h) = c\exp(-3h/\Lambda)\,, \quad (4.8)$$

c is the value of the correlation function at $h = 0$, N_e is the ensemble size, and Λ is the critical distance.

The elements of T in Sections 4.1.2 and 4.1.3 are calculated by these formulations or similar ones, based on the distances of the corresponding data and the unknown parameters. In the case of local analysis, the parameters and the data can be grouped by choosing the rows and columns of the covariance matrices in (4.2) such that $\tau > 0$.

Defining critical distances and choice of correlation functions in distance-based localization is not a straightforward task and requires domain expertise. In case of assimilation of well data, these quantities can be defined using the drainage area of each of the wells



(a) (b)

Figure 4.1: Illustration of ellipse of influence generated based on the drainage region, (a) drainage region, (b) matched ellipse of influence [Emerick and Reynolds, 2011]

[Emerick and Reynolds, 2011; Soares et al., 2018]. As can be seen in Figure 4.1, first the drainage area of a well is approximated and then the critical distances for localization and anisotropy are chosen such that the ellipse of influence matches the drainage area well. Similarly, for the inverted seismic data, one can compute empirical correlation functions between the model forecasts and parameters as a function of their distance and choose critical distances and anisotropy accordingly.

4.1.5 Correlation-based tapering

There are several issues with distance-based tapering including, difficulty in choice of critical distances, elimination of true non-local correlations, necessity for data to have physical locations associated with them, and requirement of expertise and domain knowledge. Use of correlation-based tapering was suggested in [Evensen et al., 2009]. Afterwards, a robust method for correlation based tapering was introduced in [Luo and Bhakta, 2018, 2020]. Using this method, it is possible to perform localization on data for which spatial location cannot be defined, e.g. data represented by wavelet coefficients.

This technique treats the Kalman gain matrix based on the mutual parameter-forecast correlation, rather than distance. The idea here is that due to statistical errors, the sample correlations whose absolute values are smaller than a critical threshold are more influenced by spurious correlations comparing to the sample correlations whose absolute values are large. There are two approaches to this correlation-based localization technique: hard-threshold and soft-threshold. In hard-threshold approach, the elements of the tapering matrix T are defined as $t_{ij} = I(|\rho_{ij}| > \theta_{ij})$, where I is the indicator function whose output is 1 if the condition is satisfied and 0 otherwise, $|.|$ is the absolute value function, ρ_{ij} is the correlation coefficient between parameter i and model forecast

j, and θ_{ij} is the cutoff threshold.

For any of the model forecasts the threshold values are defined jointly as $\{\theta_{ij}\}_{i=1}^{N_Z} = \theta_j$ for all the model parameters. By assuming that random assignment of each set of the model forecasts to each realization of the model parameters results in spurious correlations with Gaussian distribution and zero mean, the threshold for datum j is computed based on all these spurious correlations, ϵ_{ij}, obtained by shuffling and random assignment of the model forecasts as [Luo and Bhakta, 2018]

$$\theta_j = \frac{\text{median}(|\epsilon_{\text{ij}}|)}{0.6745} \,. \tag{4.9}$$

As an alternative, in another procedure called soft thresholding, one can use Gaspari-Cohn formulation in (4.6) to avoid sharp changes in the tapering matrix. In this formulation, instead of h/Λ, a new variable is defined to act as pseudo-distance,

$$\zeta = \frac{1 - |\rho_{ij}|}{1 - \theta_j} \,. \tag{4.10}$$

Use of hard threshold can result in abrupt changes and sharp updates in certain areas of the parameter field [Luo and Bhakta, 2020]. Use of soft threshold can alleviate this problem. However, correlation-based localization has its own limitations, including suffering from Monte-Carlo errors for defining thresholds and not taking into account small but true correlations.

4.2 Ensemble subspace inversion

Another method that can balance the information content in the data and the degree of freedom of the problem is ensemble-subspace inversion [Evensen, 2004]. In this method the data are projected onto the ensemble subspace, and as a result lose some of their information content. However, this method normally results in more stable updates. Here, I briefly explain ensemble-subspace inversion.

Consider the total covariance term in (3.13),

$$\mathrm{C_T} = \mathrm{C}_Y + \mathbb{C}_D \,, \tag{4.11}$$

where C_Y is given by

$$\mathrm{C}_Y = \mathrm{B}_Y \mathrm{B}_Y^{\mathrm{T}} \,, \tag{3.9}$$

as discussed in 3.1. We consider the singular value decomposition of B_Y and truncate it

with its largest r singular values,

$$\mathrm{B}_Y = \mathrm{U}\Sigma\mathrm{V}^{\mathrm{T}} \,. \tag{4.12}$$

This truncation is done to remove the smallest singular values that cause instability in inversion. Matrix $\mathrm{C_T}$ can then be approximated as,

$$\mathrm{C_T} \approx \mathrm{U}\,\Sigma\left(\mathrm{I_r} + \Sigma^{-1}\,\mathrm{U}^{\mathrm{T}}\,\mathbb{C}_D\,\mathrm{U}\,\Sigma^{-1}\right)\Sigma\,\mathrm{U}^{\mathrm{T}} \tag{4.13}$$

where $\mathrm{I_r}$ is the identity matrix of size r and

$$\mathrm{U}\,\Sigma\Sigma^{-1}\,\mathrm{U}^{\mathrm{T}}\,\mathbb{C}_D\,\mathrm{U}\,\Sigma^{-1}\Sigma\,\mathrm{U}^{\mathrm{T}} = \mathrm{U}\,\mathrm{U}^{\mathrm{T}}\,\mathbb{C}_D\,\mathrm{U}\,\mathrm{U}^{\mathrm{T}} \tag{4.14}$$

denotes orthogonal projection of $\mathbb{C}_D$ into the $N_D \times N_D$ space spanned by truncated singular value decomposition of ensemble perturbations. Defining X as

$$\mathrm{X} = \Sigma^{-1}\,\mathrm{U}^{\mathrm{T}}\,\mathbb{C}_D\,\mathrm{U}\,\Sigma^{-1}\,, \tag{4.15}$$

we will have

$$\mathrm{C_T} \approx \mathrm{U}\,\Sigma\,(\mathrm{I_r} + \mathrm{X})\,\Sigma\,\mathrm{U}^{\mathrm{T}}\,. \tag{4.16}$$

Since X is a symmetric matrix with real elements, its eigenvalue decomposition can be given as,

$$\mathrm{X} = \mathrm{W}\,\Omega\,\mathrm{W}^{\mathrm{T}}\,, \tag{4.17}$$

where W is the matrix of eigenvectors, and Ω is the matrix of eigenvalues. Using these notations, $\mathrm{C_T}$ can be reformulated as

$$\mathrm{C_T} \approx (\mathrm{U}\,\Sigma\,\mathrm{W})\,(\mathrm{I_r} + \Omega)\,(\mathrm{U}\,\Sigma\,\mathrm{W})^{\mathrm{T}}\,. \tag{4.18}$$

Finally inversion of $\mathrm{C_T}$ in (3.13) is substituted with Moore-Penrose pseudo-inverse of the approximate total error covariance as,

$$\mathrm{C_T}^{-1} \approx (\mathrm{U}\,\Sigma\,\mathrm{W})\,(\mathrm{I_r} + \Omega)^{-1}\,(\mathrm{U}\,\Sigma\,\mathrm{W})^{\mathrm{T}}\,. \tag{4.19}$$

In case $N_e \ll N_D$, in addition to stability, use of ensemble-subspace inversion also has computational cost advantages, since inversion of a $N_D \times N_D$ matrix ($\mathrm{C_T}$) is substituted with inversion of an $r \times r$ matrix ($\mathrm{I_r} + \Omega$) plus some inexpensive matrix multiplications.

Chapter 5

Forward Modeling

Computation of the model forecasts is a crucial part of data assimilation. The process of computing model forecasts from a complete set of model inputs which describe the physical system is called forward modeling. This process normally involves four stages, namely establishing physical models, formulating mathematical models, discretization of the mathematical models, and development of computer algorithms to compute the model forecasts based on the discretized models [Chen, 2007]. Even though all of the steps are crucial and rigorously studied in the research community, the main focus of this chapter will be on the mathematical modeling step.

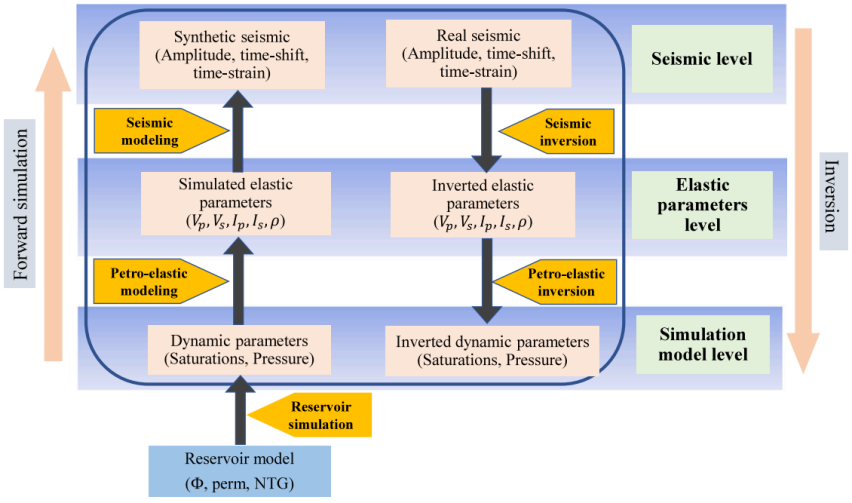


Figure 5.1: Levels of integration of seismic data [Oliver et al., 2021]

In the context of assimilation of seismic data in the reservoir history-matching problem,

the forward modeling output depends on the level of integration of data. As shown in Figure 5.1, the seismic data are originally given in the seismic level, i.e. amplitude, time-shifts, time-strains, while the reservoir simulation model outputs are obtained using a reservoir flow model which provides saturations and pressures. Therefore, if one wants to compare the observation data and the simulated output, they have to decide on a proper level of data integration. As can be seen in Figure 5.1, there are three levels for data integration; i.e. (i) seismic level, in which seismic data are kept in their original level, and the model forecasts obtained from reservoir simulation go through both petro-elastic modeling and seismic modeling; (ii) elastic parameters level, in which the seismic data go through seismic inversion, and the reservoir modeling outputs are converted to elastic parameters using petro-elastic modeling; and (iii) simulation model level, in which the model forecasts from flow simulators are kept in their original level, and the seismic data goes through both seismic inversion and petro-elastic inversion processes. The most widely used level for assimilation of 4-dimensional seismic data is level (ii) which avoids limitations such as sensitivity to porosity and pore-pressure changes in the reservoir (pertaining to level (i)) and complexity of modeling the uncertainty associated with the pressure and saturation fields (pertaining to level (iii)).

In this section I briefly discuss the different sections of a forward model used in assimilation of seismic data. This forward model can be thought as constituting three serial sections: (i) reservoir flow model which takes the reservoir rock properties, initial, and boundary conditions, and outputs pressure and saturation fields, (ii) petro-elastic model taking the reservoir rock properties and pressure and saturation fields and giving the elastic parameters as outputs, and (iii) seismic model that takes the elastic parameters and outputs simulated seismic data.

5.1 Reservoir flow model

Historically, several methods have been used for modeling the behavior of oil reservoirs, including material balance methods, which consider the entire reservoir as one simulation grid block and use mass balance on that block to model behavior of the reservoir, statistical methods, which typically use the data acquired from similar reservoirs to predict behavior of the reservoir being investigated, and analytical methods which normally use mass, momentum, and energy-balance equations to come up with analytical formulations for predicting the behavior of reservoirs in simplistic problems.

The mainstream method for predicting behavior of the reservoir is, however, formulating systems of partial differential equations (PDEs) coupled with some auxiliary equations

for the reservoir parameters, and then discretizing these equations in order to be able to solve them in more general settings. Based on the characterizations of the problem and the simplifications performed on it, the problem can be understood as single-phase flow, two-phase flow, the black-oil model, compositional flow, non-isothermal flow, etc. One of the most widely used models in reservoir simulation is the black-oil model. This method will be discussed in more details in this section.

5.1.1 The Black-oil model

The black-oil model assumes there exist three phases in the reservoir, i.e. water, oil, and gas. In addition it assumes gas can be partially soluble in oil; this solubility is assumed to be a function of pressure and temperature. This assumption obviously disregards different components of oil and gas, and only considers their collective behavior. However, this model relaxes the assumption that there is no mass transfer between phases, as opposed to two-phase flow for example.

Before starting description of the black oil model, it is worth mentioning that since there exist standard notations for the variables in reservoir modeling, we stick to them. The result is that many of these notations will be the same as those used in other sections, but should not be mistaken.

Writing the continuity equation for each of the phases present in the black oil model, and combining them with the well known Darcy's law gives [Chen, 2007],

$$\frac{\partial}{\partial t}\left(\frac{\phi S_w}{B_w}\right) = \nabla.\left(T_w\left[\nabla p_w - \gamma_w \nabla z\right]\right) + \frac{q_w}{B_w}\,, \tag{5.1}$$

for water,

$$\frac{\partial}{\partial t}\left(\frac{\phi S_o}{B_o}\right) = \nabla.\left(T_o\left[\nabla p_o - \gamma_o \nabla z\right]\right) + \frac{q_o}{B_o}\,, \tag{5.2}$$

for oil, and

$$\frac{\partial}{\partial t}\left[\phi\left(\frac{S_g}{B_g} + \frac{R_{so}S_o}{B_o}\right)\right] = \nabla.\left(T_g\left[\nabla p_g - \gamma_g \nabla z\right] + R_{so}\,T_o\left[\nabla p_o - \gamma_o \nabla z\right]\right) + \frac{q_g}{B_g} + \frac{R_{so}\,q_o}{B_o}\,, \tag{5.3}$$

for gas. In these equations t denotes time variable, ϕ is the porosity of reservoir rock indicating the ratio of porous space to the total volume of the rock, S_* is the saturation of phase $*$ which is the ratio of the porous space occupied by that phase, B_* is the formation volume factor of phase $*$ which is an indicator of how much a unit volume of

that phase expands when it is brought from reservoir condition to standard atmospheric conditions, p_* is the pressure of phase $*$, ∇z is the gradient of the coordinate system with respect to the direction of gravitational force, and R_{so} is the amount of dissolved gas in a unit volume of oil computed in standard atmospheric conditions. In addition γ_* and T_* are computed as

$$\gamma_* = \rho_* \mathcal{G} \,, \tag{5.4}$$

where ρ_* is the density of that phase, and $\mathcal{G}$ is the gravitational constant, and

$$T_* = \frac{K_*}{\mu_* B_*} K, \tag{5.5}$$

where K is the absolute permeability of the rock which is a measure of how well fluids flow in that medium, K_* is the relative permeability of phase $*$ which is between 0 and 1 for each of the phases, and μ_* is the viscosity of that phase.

In addition to these PDEs there exist some auxiliary equations,

$$S_w + S_o + S_g = 1 \,, \tag{5.6}$$

$$p_{cow} = p_o - p_w \,, \tag{5.7}$$

$$p_{cgo} = p_g - p_o \,, \tag{5.8}$$

where the capillary pressures p_{cow} and p_{cgo} are known functions of S_w and S_o, respectively.

As for the initial conditions, the reservoir is normally considered in equilibrium with its surrounding environment, e.g. pressure at a certain depth to be equal to the hydro-static pressure. For the boundary conditions, depending on the geology, one may consider for example no flow boundary if it is isolated, or constant pressure boundary if there exists a strong aquifer which supports the pressure. After setting these conditions and knowing the reservoir parameters, the problem will be determined and can be solved to obtain pressures and saturations by time.

In the next step, the equations are discretized and solved numerically. If the grid is Cartesian, the finite difference scheme is normally used for solving the problem, or else in irregular grids, finite volume method can be used, since it conserves mass which is a very important constraint in the porous media flow problems.

5.2 Petro-elastic modeling

In petro-elastic modeling the fluid and rock properties as well as saturations and pressures are taken as input, and the elastic properties of the medium are computed as output. One of the standard models in rock physics which is widely used for calculating the effects of fluid change, e.g. induced by production of petroleum in a reservoir, is Gassmann's model for fluid substitution [Gassmann, 1951]. The low-frequency Gassmann's theory formulates the effective bulk modulus of the saturated rock, $\mathbf{K}_s$ as [Smith et al., 2003]

$$\mathbf{K}_s = \mathbf{K}_d + \frac{\left(1 - \frac{\mathbf{K}_d}{\mathbf{K}_m}\right)^2}{\frac{\phi}{\mathbf{K}_f} + \frac{1-\phi}{\mathbf{K}_m} - \frac{\mathbf{K}_d}{{\mathbf{K}_m}^2}} , \tag{5.9}$$

where ϕ is the rock porosity, and $\mathbf{K}_d$, $\mathbf{K}_m$, and $\mathbf{K}_f$ are the bulk moduli of the dry rock, the mineral the rock consists of, and the fluid which saturates the rock, respectively. In addition, since the fluids do not bear shear stress, the shear modulus of the saturated rock, $\mathbf{G}_s$, is simply formulated as

$$\mathbf{G}_s = \mathbf{G}_d , \tag{5.10}$$

where $\mathbf{G}_d$ is the shear modulus of the dry rock. These equations are particularly valid for the low frequency waves propagating in the medium. In case the frequencies are high, there will not be an equilibrium for pore pressures throughout the pore space, which complicates the problem. In low frequencies the fluid will have enough time to flow in the pore and there will not be pressure gradients in the pores induced by the seismic waves.

One of the most well-known models for computation of elastic parameters $\mathbf{K}_d$ and $\mathbf{G}_d$ based on the moduli of the mineral, porosity, and effective pressure is Hertz-Mindlin contact theory [Mindlin and Deresiewicz, 1953]. Based on this theory, for sandstone, first bulk modulus and shear modulus of the dry well-sorted rock at critical porosity, $\mathbf{K}_{HM}$ and $\mathbf{G}_{HM}$, are calculated by [Fahimuddin, 2010]

$$\mathbf{K}_{HM} = \left(\frac{n^2(1-\phi_c)^2{\mathbf{G}_m}^2}{18\pi(1-\nu_m)^2}p_{\text{eff}}\right)^{\frac{1}{3}} , \tag{5.11}$$

$$\mathbf{G}_{HM} = \frac{5-4\nu_m}{5(2-\nu_m)}\left(\frac{3n^2(1-\phi_c)^2{\mathbf{G}_m}^2}{2\pi^2(1-\nu_m)^2}p_{\text{eff}}\right)^{\frac{1}{3}} \tag{5.12}$$

where ϕ_c is the critical porosity (which for sandstone is assumed to be 0.4), p_{eff} is the effective pressure (the difference between the overburden pressure and pore pressure), ν_m is Poisson's ratio for the mineral (or if the rock is a mixture of minerals it is considered as Poisson's ratio of the solid phase), and n is the average contact points of a grain in

the rock. The moduli of the dry rock are then computed as

$$\mathbf{K}_d = \left(\frac{\frac{\phi}{\phi_c}}{\mathbf{K}_{HM} + \frac{4}{3}\mathbf{G}_{HM}} + \frac{1 - \frac{\phi}{\phi_c}}{\mathbf{K}_m + \frac{4}{3}\mathbf{G}_{HM}} \right)^{-1} - \frac{4}{3}\mathbf{G}_{HM} \,, \tag{5.13}$$

$$\mathbf{G}_d = \left(\frac{\frac{\phi}{\phi_c}}{\mathbf{G}_{HM} - a} + \frac{1 - \frac{\phi}{\phi_c}}{\mathbf{G}_m + a} \right)^{-1} - a \,, \tag{5.14}$$

where

$$a = \frac{\mathbf{G}_{HM}}{6} \left(\frac{9\mathbf{K}_{HM} + 8\mathbf{G}_{HM}}{\mathbf{K}_{HM} + 2\mathbf{G}_{HM}} \right) \,. \tag{5.15}$$

In addition the bulk modulus of mixture of fluids is given by harmonic averaging (Wood's equation) as

$$\frac{1}{\mathbf{K}_f} = \frac{S_w}{\mathbf{K}_w} + \frac{S_o}{\mathbf{K}_o} + \frac{S_g}{\mathbf{K}_g} \,. \tag{5.16}$$

The density of the saturated rock, ρ_s, can be computed as

$$\rho_s = \phi \left(S_w \rho_w + S_o \rho_o + S_g \rho_g \right) + (1 - \phi)\rho_m \,, \tag{5.17}$$

where ρ_* and S_* are the density and saturation of phase $*$ (water, oil, and gas), and ρ_m is the density of the mineral constituting the rock. Using the above-mentioned quantities, the acoustic properties of the medium can be calculated. Hence, isotropic compressional velocity, V_P, shear velocity, V_S, acoustic impedance (bulk impedance), I_P, and Poisson's ratio, ν_s, can be formulated as

$$V_P = \sqrt{\frac{\mathbf{K}_s + \frac{4}{3}\frac{\mathbf{G}_s}{}}{\rho_s}} \,, \tag{5.18}$$

$$V_S = \sqrt{\frac{\mathbf{G}_s}{\rho_s}} \,, \tag{5.19}$$

$$I_P = \rho_s V_P \,, \tag{5.20}$$

$$\nu_s = \frac{V_P^2 - 2\,V_S^2}{2\,V_P^2 - V_S^2} \,. \tag{5.21}$$

In 4D inverted seismic data, the quantities of interest are compared to their value in a baseline in time and their differences are considered.

The discrete form of the formula presented for petro-elastic modeling is straightforward computation of the formula for each grid cell of the grid. Therefore, the computational cost is linear with grid size, and it is not an expensive operation comparing to the flow model.

5.3 Seismic modeling

In seismic modeling propagation of seismic waves through earth are modeled. In seismic acquisition process, seismic waves are sent to the earth (e.g. using explosions or air guns), and after the waves travel through layers of the geological field, their reflections from different layers are recorded (e.g. using hydrophones or geophones). These are the actual measurements that we record as seismic data. Knowing the governing physical laws (wave equations), the layering of the geological field, the petro-elastic parameters of different layers, and the form of the seismic waves sent, one can model the response recorded at each receiver by time. It will be the superposition of all the reflection waves that are detected by the receivers. It can also be considered as a convolution of a source wave function and a reflectivity function.

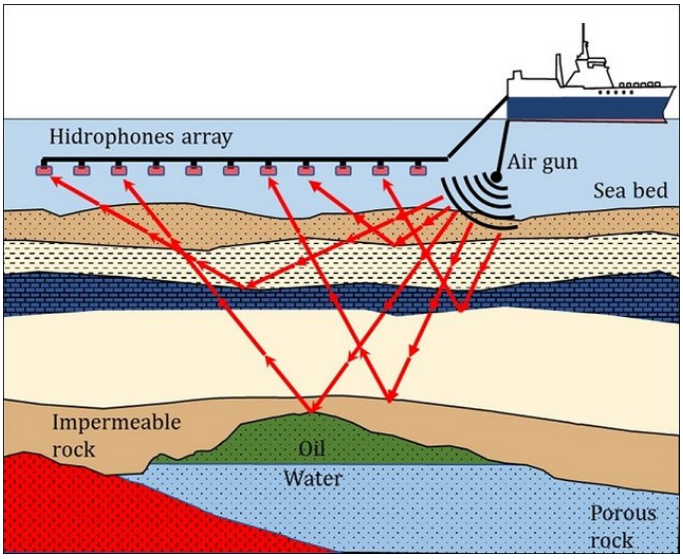


Figure 5.2: Representation of seismic waves [Saraiva et al., 2021]

In seismic inverse modeling, the layering of geological field and the elastic properties are estimated using the seismic data acquired. In the deterministic inversion, the convoluted signal recorded at the receivers is deconvoluted to obtain elastic properties and layering.

Chapter 6

Surrogate modeling

Real world systems are governed by physical laws. Systems of interest are parameterized, and the physical law are discovered and tuned in the process of forward modeling. The obtained forward models are abstractions of the real physical world. These forward models, known as high fidelity models are run to to obtain model responses for any given set of model inputs. In this context, fidelity of a model refers to the degree of realism of a model. These high fidelity forward models are normally computationally expensive. Hence, given a fixed amount of computational resources, affordability of such models is very limited. Surrogate modeling can be assumed as a second level of abstraction [Razavi et al., 2012]. Surrogate models are used to approximate the model responses of the high-fidelity model, via use of cheaper-to-run "surrogates". Hence, the main aim of surrogate modeling is reduction of the computational cost per run, without too much loss of accuracy. Surrogate modeling has been used in wide range of applications including but not limited to sensitivity analysis, prediction, optimization, design space exploration, and uncertainty quantification.

Assuming the original response of the model is given by (2.1), the response from the surrogate can be assumed to have the following form,

$$\widetilde{\mathcal{M}}(z) \approx \mathcal{M}(z)\,. \tag{6.1}$$

Consider the surface of model responses with respect to any set of model inputs. This surface typically forms a non-linear hyper-plane. Based on the way surrogate models approximate this response surface, they are divided into two general groups: response surface surrogates and lower-fidelity physically-based surrogates. Response surface surrogates approximate this non-linear hyper-plane by use of statistical or empirical data-driven models. Lower-fidelity physically-based model run simplifications of the original

high-fidelity model to approximate the response surface. Normally each of the elements of the model response is approximated separately in the case of response surface surrogate, while they are obtained jointly when using lower-fidelity physically-based models. Here, I briefly discuss both of the approaches to surrogate modeling.

6.1 Response surface surrogates

There is no emulation of any internal component of the original high-fidelity model in response surface surrogates which are also known as "metamodels" in the literature [Kyprioti and Taflanidis, 2021; Morimoto et al., 2021; Roman et al., 2020]. These surrogates approximate the relationship between the input variables and model responses. The origins of response surface surrogates date back to the use of classic non-linear optimization techniques, e.g. Taylor series expansion. Taylor series expansion and similar methods are based on approximation of the function and its derivatives in one point, whereas the state of the art metamodels normally use a collection of points in the model input space to approximate the response surface. There are several methods for sampling from the model input space, including full factorial, central composite design, and Latin hyper-cube sampling [Oehlert, 2010]. There are several function approximation models in the literature. In the rest of this section I cover some of the state of the art function approximation models: polynomials, radial basis function, Kriging, support vector machines, and artificial neural networks.

6.1.1 Polynomials

Polynomial model is one of the simplest function approximation models. The second-degree polynomials are very popular for response approximation. Assuming the input model space has dimension, N_Z, the total number of coefficients of the polynomial will be $(N_Z + 1)(N_Z + 2)/2$. The formulation of the second-degree polynomial model can be written as

$$\widetilde{\mathcal{M}}(z) = \sum_{i=0}^{N_Z} \sum_{j=i}^{N_Z} \gamma_{i,j} z_i z_j \ , \tag{6.2}$$

where z_i and z_j denote the i^{th} and j^{th} parameters (0^{th} being a constant), and $\gamma_{i,j}$ is their corresponding coefficient. These coefficients are normally determined using least squares regression method.

Polynomials are very successful in local approximation of the original response when it is locally unimodal. However, as it is clear in the functional form of the first and second-

degree polynomials, they can only be used as global surrogate models when the original response surface is linear, quadratic, or similar to them. Application of higher order polynomials is not practical for large N_Z. Firstly, specifying the degree of polynomial is not straightforward, and secondly the number of coefficients to be determined will grow exponentially with the degree of the polynomial, so that the estimated coefficients will have very high variances.

Polynomials have been reported to perform poorly as surrogate models compared to other function approximation models [Hussain et al., 2002; Regis and Shoemaker, 2004]. However, when the form of original response function is known a priori, and it is similar to a polynomial, this function approximation model has shown successful [Fen et al., 2009].

6.1.2 Radial basis functions

Radial basis function (RBF) model is formulated based on a weighted summation of radial basis functions. The name of the model gives a short description of how it works. It is radial meaning that it obtains its value based on distances from certain points in the parameter space. It uses basis functions meaning that summation of several functions define the surrogate model response in the entire domain. Several forms of standard basis functions can be found in the literature, including Gaussian (which results in "the smoothest interpolation"), multiquadric, inverse quadratic, and thin plate spline, to name a few. Some of the the basis functions are parametric meaning that the sensitivity of the response function on its input domain can be tuned, e.g. Guassian basis funcion, while some others are not, e.g. thin plate spline. RBF based on Gaussian basis functions can be formulated as

$$\widetilde{\mathcal{M}}(z) = \sum_{k=1}^{K} w_k \exp(-\gamma ||z - \mu_k||^2) \,, \tag{6.3}$$

where μ_k is one of the K representative points chosen in the parameter space (they are normally obtained using K-means algorithm), w_k is a weight factor, and γ is a parameter for tuning the smoothness of the response surrogate.

After choosing the radial basis function, the problem of approximating the model response boils down to finding the proper weights for each of the basis functions and tuning the RBF parameters. In order to find the proper weights, a linear regression problem is formulated and the weights are determined using least squares regression. As for the parameters, either they are found using try and error or they are determined jointly with the weights using Expectation Maximization (EM) algorithm [Moon, 1996].

One of the problems with RBFs is that the sensitivity of the response to the change in all the N_Z directions is assumed to be the same. This means that changes in all the parameters are assumed to have similar impact on the response function, which is not normally true.

6.1.3 Kriging

Kriging is a statistically driven function approximation model. The idea in Kriging is similar to that of RBF. The unknown value of the response function is approximated using a weighted sum at each locality. However, in kriging, instead of basis functions, the actual values of the response function at other points in the parameter space are utilized. Kriging can be formulated as

$$\widetilde{\mathcal{M}}(z) = \sum_{i=1}^{N_e} w_i \mathcal{M}(z_i), \tag{6.4}$$

where $\mathcal{M}(z_i)$ is an available response from z_i in the parameter space, w_i is its weighting factor, and N_e is the total number of points at which the response function is known.

In Kriging the possibility that the model response is more sensitive to change in certain directions of the parameter space is taken into account. This is done using variograms and spatial correlations, adding to model flexibility. In simple Kriging, the weights of different points for calculation of the unknown response function is computed by formulating a set of linear equations where the coefficients are the covariances between the known response functions obtained at different points in the parameter space (which accounts for redundancy of the data) and the covariances between response functions obtained at known points and the unknown point (which takes into account the closeness). The statistical basis of this method helps to formulate an uncertainty level for the approximated response function at any given point.

6.1.4 Support vector machines

Here I briefly explain support vector machines (SVM) which are widely used for both regression and classification problems. They are known to be among the best "out of the box" classifiers [James et al., 2021]. In order for approximation of the model response, as opposed to polynomial model, RBF, and kriging, SVM uses only a subset of the known data set. This subset are known as support vectors. SVM is a non-linear extension of the support vector classifier. As seen in Figure 6.1, in a linearly separable space,

SVM gives the hyper-plane which maximizes the margin between the support vectors pertaining to different classes in classification. In the case depicted in Figure 6.1 the two

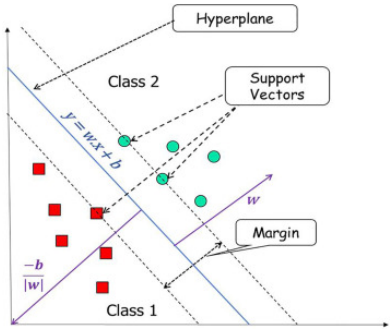


Figure 6.1: Support vector machine in linearly separable space [Rani et al., 2022]

classes are linearly separable and one can use a hard margin to separate two classes. In general soft margins are used which do not strictly classify all the points correctly, but improve the out-of-sample prediction power of the method. In addition the parameter space is transformed into spaces of higher dimensions using kernels, e.g. Gaussian kernel explained in Section 6.1.2, so that the model will become more flexible.

6.1.5 Artificial neural networks

Inspired by the neural networks of live organisms, artificial neural networks (ANN) are designed to approximate model responses for any set of model inputs, and they are assumed to be one of the strongest methods in function approximation. According to the universal approximation theorem, given enough amount of data, and enough number of nodes, a feed-forward neural network with only one hidden layer can approximate any response function that maybe of interest [Hornik et al., 1989]. There are many architectures for ANNs including, single-layer and multi-layer feed-forward networks, single-layer and multi-layer recurrent networks, and convolutional neural networks. Figure 6.2 represents a multi-layer feed-forward ANN. In the input layer the values from the parameter space are inserted into the network. At each node the hidden layers inputs are summed, a bias is added, and a non-linear function of the total sum is given as output. A formulation of the output of an arbitrary node at a hidden layer, f, can be given as

$$f = g\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_k\right), \tag{6.5}$$

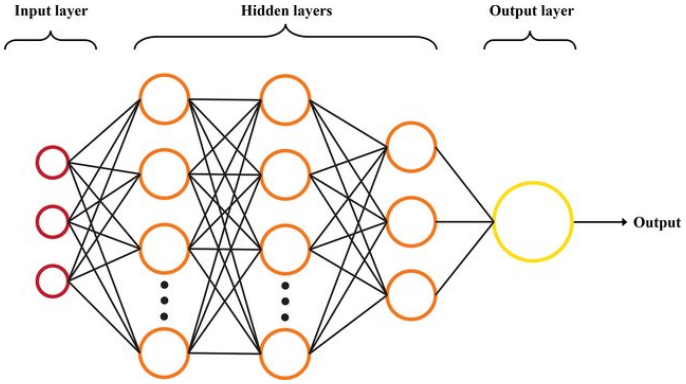


Figure 6.2: A multi-layer feed-forward ANN [Vassallo et al., 2020]

where β_0 is the bias for that node, $\{x_k\}_{k=1}^{K}$ and $\{\beta_k\}_{k=1}^{K}$ are inputs and weights of the inputs for the particular node, and g is a non-linear activation function. Sigmoid function is one the most popular activation functions whose formulation is given as

$$g(x) = \frac{\exp(x)}{1 + \exp(x)} \,. \tag{6.6}$$

However, use of Sigmoid function results in slow convergence rates when the parameters of the network are optimized. Hence, activation functions like the rectified linear unit function are used as an alternative. By doing the calculations on the entire network, the approximation of the response function is obtained in the output layer.

The parameters of the neural network, biases and weights, are normally initialized arbitrarily, and then by use of the data they are optimized using back propagation [Rumelhart et al., 1995] on stochastic gradient decent algorithm [Amari, 1993].

ANNs are extensively used in a wide array of applications, due to their robustness. Particularly, they have been of interest in reservoir history-matching [Brantson et al., 2018; Chai et al., 2021; Costa et al., 2014]. However, training neural networks is a computationally expensive process, and the trained neural networks are normally suitable for particular cases and cannot be generalized well.

6.2 Lower fidelity physically-based surrogates

Unlike response surface surrogates being data-driven methods for approximating the high-fidelity simulations based on model responses at a certain number of points in the

parameter space, lower fidelity physically-based surrogates are computationally less expensive simulation models which are less faithful to the physical equations governing the phenomena. In some publications lower-fidelity models are also referred to as "coarse" models, while high-fidelity models are referred to as "fine" models [Bandler et al., 1994].

Lower fidelity models are assumed to have at least two advantages over response surface surrogates: (i) they are assumed to give better approximations of the original response function in the areas of the parameter space which are less explored, (ii) the problem of flexibility of the surrogate model due to high-dimensionality of the parameter space is avoided because of use of domain-specific equations for approximating the response function. The basic assumption in lower-fidelity modeling is that fine models and coarse models have the basic features of the problem in common and are correlated in a way [Kennedy and O'Hagan, 2000]. Based on this assumption, the response from the coarse model is reasonably close to the response from the fine model everywhere in the parameter space, and as a consequence the model outputs from the coarse model are reliable in the unexplored areas of the space. If this assumption does not hold, the advantage of using lower-fidelity models would be minimal.

One of the main applications of the lower-fidelity models is in optimization problems where using some strategies both fine and coarse models are used adaptively [Dixit and Elsheikh, 2022; Forrester et al., 2007; Gano et al., 2006]. Another domain in which lower-fidelity surrogate models are used extensively is uncertainty quantification [Bomers et al., 2019; Palar and Shimoyama, 2017; Yamazaki and Mavriplis, 2013].

6.2.1 Types of lower fidelity models

Depending on the type of the original high-fidelity model, there are at least three different categories of low-fidelity models. The first class are essentially the same as the high-fidelity model, but with higher numerical error. In simulation models based on partial differential equations, for example, the lower fidelity models can be, for example, the original model with a coarser spatial or temporal grid [Leary et al., 2003; Thokala and Martins, 2007; Zhonghua et al., 2020]. Use of simpler basis functions for finite element methods can also be a lower fidelity model for finite element methods using more complex basis functions. Utilization of less strict convergence criteria for models based on iterative solvers can also be a lower fidelity model for the high-fidelity models involving fully converged iterative solvers [Forrester et al., 2006].

Particularly coarsening the spatial grid in the simulation models based on partial differential equations is a very efficient method for reduction of computational cost. The

dominant cost of the simulations in large problems is known to be pertaining to the iterative linear solvers. The computational cost associated with one simulation is assumed to be proportional to G^{γ}, where G is the number of the active grid cells in the in the simulation grid, with $\gamma \in [1.25, 1.5]$, [Axelsson, 1996]. This proportionality being greater than linear, can result in drastic reduction in the computational cost with slight reductions in the number of grid cells via coarsening. This idea has also been used in petroleum reservoir simulation models [Fossum and Mannseth, 2017].

The second class of lower-fidelity models are based on model order reduction (MOR) methods. MOR techniques were initially developed in systems and control theory for studying dynamical systems. The main aim was to reduce the model complexity while keeping the input-output structure. In MOR, firstly the most crucial characteristics of the model are captured, and then the model is simplified using rigorous mathematical techniques, with a minimal use of the knowledge about the underlying system [Schilders et al., 2008]. Some of the most widely used MOR techniques are truncated balanced realizations method [Moore, 1981], Hankel-norm reduction [Glover, 1984], proper orthogonal decomposition [Sirovich and Kirby, 1987], and Krylov subspace methods [Liesen and Strakos, 2013].

The third class of lower-fidelity surrogate models are formulated based on a simpler model of the reality. In these surrogates the physics of the problem is either partially ignored or approximated. Considering less complex geometry, simpler boundary conditions, or utilization of two-dimensional models instead of three-dimensional models are examples of strategies to come up with lower fidelity models of this class. In the domain of fluid dynamics, Navier-Stokes equations can be assumed of high fidelity, while Euler equations are of lower fidelity. In the petroleum reservoir flow problem, considering the compositional model with all the oil and gas components can be assumed the model with the highest fidelity, while black oil model consisting of only three phases; namely, oil, gas, and water; can be assumed as the lower fidelity model. In general any world system can be modeled in a hierarchy of models with different fidelities.

6.2.2 Correction of lower fidelity response

There normally exist a discrepancy between the model response from the lower-fidelity models and the high-fidelity model. There exists several approaches to address this discrepancy, two of the most common ways are explained here, i.e. correction functions, and space-mapping.

The most common method is formulation of correction functions which are applied to

the lower-fidelity model response and its value is improved to align better with the high-fidelity model response. Two of the general ways to formulate these methods are the additive approach, e.g. see [Viana et al., 2009], and the multiplicative approach, e.g. see [Alexandrov and Lewis, 2001]. The additive and multiplicative correction functions are given as

$$f_a(z) = \mathcal{M}(z) - \widetilde{\mathcal{M}}(z)\,, \tag{6.7}$$

$$f_m(z) = \mathcal{M}(z)/\widetilde{\mathcal{M}}(z)\,, \tag{6.8}$$

respectively (the formulation of multiplicative correction function here assumes a one-dimensional model response). After formulating these functions the corrected low fidelity model response, $\widetilde{\mathcal{M}}_c$ is given using the following formulations,

$$\widetilde{\mathcal{M}}_c(z) = \widetilde{\mathcal{M}}(z) + f_a(z)\,, \tag{6.9}$$

$$\widetilde{\mathcal{M}}_c(z) = \widetilde{\mathcal{M}}(z) f_m(z)\,, \tag{6.10}$$

depending on the approach selected.

The process of formulation of these correction functions is essentially very similar to the process of approximating a response surface surrogate. However, the models that are used should be simpler due to the fact that the lower-fidelity model response is already considerably close to the high-fidelity model response. Linear regression, e.g. see [Vitali et al., 2002], and quadratic polynomials, e.g. see [Sun et al., 2010] are two of the common simple models used for the correction functions.

The second approach for addressing the discrepancy between the fine and coarse model responses is space-mapping [Bakr et al., 2001; Bandler et al., 1994]. In this approach, the variables of the original parameter space are mapped to the lower-fidelity model parameter space such that using the mapped parameters as the input to the lower-fidelity model will yield model responses which are close enough to the original high-fidelity model responses. In other words, for any z in the original parameter space, $\hat{z}$ is found in the lower fidelity model parameter space, such that $|\widetilde{\mathcal{M}}(\hat{z}) - \mathcal{M}(z)|$ will be minimized. Using this mapping lower fidelity model responses are assumed to have less error. However, several complexities may arise in implementation of this technique. Non-uniqueness of the mapping can result in difficulties in convergence of the optimization sub-problem that is solved for generating the mapping. Since generating the mapping entails solving a large number of optimization sub-problems, if the lower-fidelity model is not considerably cheaper than the original high-fidelity model, implementation of this method may not be computationally efficient.

6.3 Surrogate model selection

Selection of the surrogate models is normally done based on the specifications of the problem, level of familiarity of the scientists with the methods, and availability of the software. However, some general comments can be made for surrogate model selection.

One of the decisive factors in selection between the response surface surrogates and lower-fidelity physically-based models is the dimensionality of the parameter space. Response surface surrogates normally have limitations when the dimension of the parameter space is large, but lower-fidelity models are normally less affected by this limitation. Another advantage of lower-fidelity models is in emulating functions with multiple outputs or in the multi-objective optimization problems, since they inherently (due to physical laws governing their equations) account for the correlations between different model outputs or different objectives.

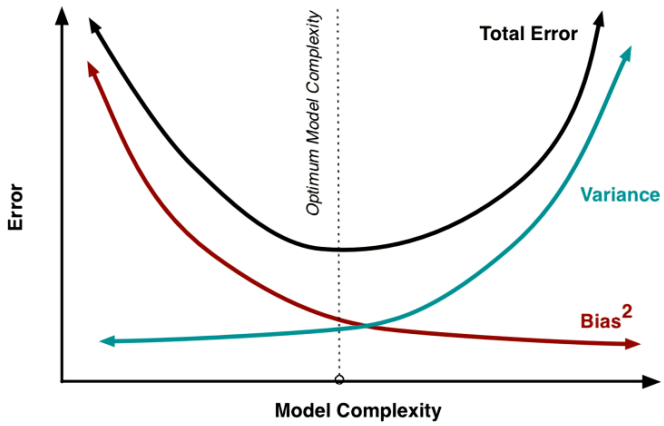


Figure 6.3: Representation of bias-variance trade-off [Neal et al., 2018]

When response surface surrogates are chosen normally the complexity of the original model should decide the complexity of the surrogate model. In general there exists a well-known bias-variance trade-off for choosing the optimal response surface surrogate model. Choice of simpler models has high bias, meaning that the model disregards many of the features of the original response function, but it has few parameters to be determined resulting in lower variance. On the other hand, complex models are less biased, but one has to determine many model parameters, meaning that the model is too conformable with respect to the available sample of model responses resulting high variance in estimation of the model parameters. The optimal level of model complexity normally resides somewhere in between. Figure 6.3 presents this trade-off.

Chapter 7

Multilevel data assimilation

Multilevel data assimilation (MLDA) can refer to several different concepts. Depending on what we mean by the term multilevel, it can refer to at least three different concepts. The first concept is utilization of multilevel forward models in the forecast step of the data assimilation. This is also known as multi-fidelity data assimilation [Peherstorfer et al., 2018]. The second concept is assimilation of multilevel data, e.g. several levels of upscaling for spatially distributed data [Mannseth, 2020]. The third concept that MLDA can refer to is assimilation of data which are not in the same resolution as the model forecasts. This is also known as multi-scale data assimilation. The focus of this chapter is on the first two concepts and they will be discussed, respectively. For the third concept, the reader is referred to, e.g., [de Moraes et al., 2020; Montzka et al., 2012].

7.1 Multi-fidelity data assimilation

Monte Carlo approximations play a crucial role in ensemble-based data assimilation (DA). Due to computational-cost limitations, the ensemble size is limited to roughly one hundred. Using straightforward ensemble-based DA, the degrees of freedom of the problem would equal the ensemble size, and such an approach would result in significant Monte Carlo errors. The negative effects of Monte Carlo errors are enlarged in the presence of large amounts of simultaneous data, e.g. inverted seismic data, resulting in underestimation of variance of the unknown parameters, and in more severe cases ensemble collapse.

The most widely used treatment for Monte Carlo errors is use of a regularization technique called localization. This was discussed in Chapter 4. In spite of its advantages, this technique has limitations including ignoring true long-range correlations in distance-

based localization or true small correlations in correlation-based localization.

Simply increasing the ensemble size will reduce Monte-Carlo errors, but it will also increase computational cost. Utilization of a lower-cost and lower-fidelity model renders the possibility of increasing the ensemble size without increasing the total computational cost. DA using various types of lower fidelity models has been applied to several inverse problems, e.g., within petroleum reservoir modeling [Fossum and Mannseth, 2017; He et al., 2013; Tarrahi et al., 2016] and atmospheric science [Hatfield et al., 2018]. Use of a lower-fidelity reservoir model will, however, introduce modeling errors in addition to those already present in conventional-fidelity simulation results.

Multilevel simulations utilize a selection of models for the the forecast step of the DA. This results in hierarchies in both fidelities and computational costs. The idea is to decrease Monte Carlo errors without increasing numerical errors too much. There are various ways to construct multilevel models. One can use several of the lower-fidelity models explained in Chapter 6 and construct a multilevel model, for a detailed discussion the reader is referred to e.g. [Peherstorfer et al., 2018]. Several MLDA formulations are introduced in the literature based on multilevel simulations [Fossum and Mannseth, 2018; Fossum et al., 2020; Hoel et al., 2016, 2020; Moldovan et al., 2021; Popov et al., 2021; Xu et al., 2018] . I will explain multilevel ensemble Kalman filter (MLEnKF) and multilevel hybrid ensemble Kalman filter (MLHEnKF) in this Chapter.

7.2 Multilevel data

In Section 3.2.3 an algorithm for assimilation of linearly dependent data was explained. A special case of linearly dependent data is multilevel data which is obtained by sequential coarsening of a set of spatially distributed data. There can be advantages in assimilation of this kind of multilevel data instead of the original single-level spatially distributed data.

For a class of problems (including the problem considered here) where the model forecast can be seen as a spatially integrated response to a spatially varying parameter field, there exists a correlation between small-scale oscillations in the parameter domain and the degree of non-linearity of the mapping from parameter field to model forecast, see, e.g., [Chavent and Liu, 1989; Grimstad and Mannseth, 2000]. This correlation is such that coarsening the simulation grid and upscaling the associated parameters will generally result in weaker nonlinearity in the coarser forward models compared to the finer ones. Hence, considering several levels of coarseness for the multilevel model corresponding to

the different levels of data, and assimilating the data sequentially from the coarsest to the finest is expected to result in a robust DA algorithm [Fossum, 2015]. By utilization of coarse models in the first steps of the DA followed by more non-linear fine models in the next steps the algorithm is expected to avoid convergence to local minima and gradually zoom in on the solution of the inverse problem.

7.3 Multilevel Monte Carlo methods

Consider estimation of $\mathbb{E}[Y]$, where Y is a random variable in its corresponding probability space, and $\mathbb{E}$ denotes expectation operator. Monte Carlo formulation for the sample estimate of the expectation, $\mathrm{E}[Y]$, is simply given as

$$\mathrm{E}[Y] = \frac{1}{N}\sum_{i=1}^{N} Y_i \,, \tag{7.1}$$

where Y_i are independent samples from Y and N is the sample size. The variance of the estimate will be $N^{-1}\mathbb{V}[Y]$, where $\mathbb{V}$ is the population variance of Y. This results in the root mean square error to be $O(N^{-1/2})$ and reaching an error level of ϵ to require $O(N^2)$ samples. This can be computationally very demanding.

Consider estimation of $\mathbb{E}[Y^1]$, but it is cheaper to sample Y^0 which approximates Y^1. In this case, a two-level Multilevel Monte Carlo (MLMC) formulates an unbiased estimator for $\mathbb{E}[Y^1]$ using the identity,

$$\mathbb{E}[Y^1] = \mathbb{E}[Y^0] + \mathbb{E}[Y^1 - Y^0] \,. \tag{7.2}$$

The Monte Carlo formulation of this identity is

$$\mathrm{E}[Y^1] = \frac{1}{N_0}\sum_{i=1}^{N_0} Y_i^0 + \frac{1}{N_1}\sum_{i=1}^{N_1} \left(Y_i^1 - Y_i^0\right) \,. \tag{7.3}$$

Defining C_0 and C_1 as the computational costs pertaining to one simulation of Y^0 and $Y^1 - Y^0$, the total computational cost will be $N_0 C_0 + N_1 C_1$. Additionally, defining V_0 and V_1 as the variances of Y^0 and $Y^1 - Y^0$, and assuming the samples are independent, the total variance will be $N_0^{-1}\mathrm{V}_0 + N_1^{-1}\mathrm{V}_1$. Under fixed computational cost constraint and treating both N_0 and N_1 as real variables, the minimum variance will be obtained if

$$\frac{N_1}{N_0} = \sqrt{\frac{\mathrm{V}_1 C_0}{\mathrm{V}_0 C_1}} \,. \tag{7.4}$$

Generalization of MLMC to more levels is natural. Given $Y^0, ..., Y^{L-1}$ approximate Y^L with increasing accuracy and computational cost, the identity,

$$\mathbb{E}[Y^L] = \mathbb{E}[Y^0] + \sum_{l=1}^{L} \mathbb{E}\left[Y^l - Y^{l-1}\right] , \tag{7.5}$$

is an unbiased estimator for $\mathbb{E}[Y^L]$. The Monte Carlo formulation of this identity is given as

$$\mathrm{E}[Y^L] = \frac{1}{N_0}\sum_{i=1}^{N_0} Y_i^0 + \sum_{l=1}^{L}\left(\frac{1}{N_l}\sum_{i=1}^{N_l}\left(Y_i^l - Y_i^{l-1}\right)\right) . \tag{7.6}$$

Defining C_l and V_l as the computational cost and variance pertaining to $Y^l - Y^{l-1}$ (where $Y^{-1} := 0$), the minimum variance for the estimator will be obtained by minimizing $\sum_{l=0}^{L} N_l^{-1}\mathrm{V}_l$ given the constraint $\sum_{l=0}^{L} N_l C_l = Cte$.

Assuming geometrical growth in the computational cost and numerical error, there exists a theorem connecting these entities to each other for the multilevel estimator [Giles, 2015]. Let Y be a random variable and Y^l its level l numerical approximation. If there exist $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \min(\beta, \gamma)$ and

(i) $|\mathbb{E}[Y^l - Y]| \leq c_1 2^{-\alpha l}$

(ii) $\mathrm{V}_l \leq \mathrm{c}_2 2^{-\beta l}$

(iii) $C_l \leq c_3 2^{\gamma l}$

then there exists a positive constant c_4 such that for any ϵ there are L values $\{N_l\}_{l=1}^{L}$ for which the multilevel estimator

$$\widehat{Y} = \frac{1}{N_0}\sum_{i=1}^{N_0} Y_i^0 + \sum_{l=1}^{L}\left(\frac{1}{N_l}\sum_{i=1}^{N_l}\left(Y_i^l - Y_i^{l-1}\right)\right) \tag{7.7}$$

results in mean squared error (MSE) with a bound

$$\mathbb{E}[\left(Y - \widehat{Y}\right)^2] < \epsilon^2 \tag{7.8}$$

and the computational cost has a bound

$$\mathbb{E}[C] \leq \begin{cases} c_4\epsilon^{-2} & \beta > \gamma\,, \\ c_4\epsilon^{-2}\left(\log\epsilon\right)^2 & \beta = \gamma\,, \\ c_4\epsilon^{-2-(\gamma-\beta)/\alpha} & \beta < \gamma. \end{cases} \tag{7.9}$$

Other crucial advancements in MLMC domain are works including formulation of randomized MLMC, which uses the multilevel simulators in a probabilistic way [Rhee and Glynn, 2012], and introduction of multi-dimensional levels instead of a hierarchy of levels [Haji-Ali et al., 2016].

7.3.1 Multilevel ensemble Kalman filter

Hoel et al. [2016] introduced a formulation for ensemble-based data assimilation based on MLMC. Multilevel Ensemble Kalman filter (MLEnKF) uses MLMC for computation of the statistics (mean and covariance) of the model forecasts. As this algorithm was originally introduced for state estimation problem, I present its original formulation, rather than parameter estimation formulation.

Consider the evolution of random vector Z with time,

$$Z_{t+1} = \mathcal{M}(Z_t)\,, \tag{7.10}$$

and a set of noisy observations

$$D_t = HZ_t + \epsilon \tag{7.11}$$

where H is the observation matrix, and ϵ is the observation error assumed to be normally distributed, i.e. $\epsilon \sim \mathcal{N}(0, \mathbb{C}_D)$.

Assuming $\{\mathcal{M}_l\}_{l=0}^{L}$ to be a hierarchy of approximations to $\mathcal{M} := \mathcal{M}_\infty$ with increasing accuracy and computational cost, formulation of the multilevel mean and multilevel covariance can be summarized as,

$$\mu_t^{\mathrm{ML}} = \sum_{l=0}^{L} \mathrm{E}[\mathcal{M}_l(Z_t) - \mathcal{M}_{l-1}(Z_t)]\,, \tag{7.12}$$

$$\mathrm{C}_t^{\mathrm{ML}} = \sum_{l=0}^{L} \mathrm{C}[\mathcal{M}_l(Z_t)] - \mathrm{C}[\mathcal{M}_{l-1}(Z_t)] \tag{7.13}$$

where $\mathrm{E}[\mathcal{M}_l(Z_t)]$ and $\mathrm{C}[\mathcal{M}_l(Z_t)]$ denote the sample mean and sample covariance computed based on model forecasts at level l, with $\mathcal{M}_{-1} := 0$. This formulation of covariance keeps the covariance symmetric, but it will not be necessarily positive-definite. Hence, an additional treatment, e.g. singular value decomposition of the multilevel covariance matrix and keeping only the singular vectors associated with positive singular values, is needed for stability of the algorithm.

For realization j at level l, $z_{j,l}$, whose model forecasts are obtained using $\mathcal{M}_l$, the update

equation will be of the form

$$z_{t,j,l}^{pos} = z_{t,j,l}^{pri} + \mathrm{K}^{\mathrm{ML}} \left(d_{t,j,l} - H z_{t,j,l} \right) \ , \tag{7.14}$$

where $d_{t,j,l}$ is a perturbed realization from the observation data at time t distributed as $D \sim \mathcal{N}(\overline{d}_t, \mathbb{C}_D)$. The multilevel Kalman gain, K^{ML}, is given as

$$\mathrm{K}^{\mathrm{ML}} = \mathrm{C}_t^{\mathrm{ML}} \mathrm{H}^{\mathrm{T}} \left(\mathrm{H} \mathrm{C}_t^{\mathrm{ML}} \mathrm{H}^{\mathrm{T}} + \mathbb{C}_D \right)^{-1} \ . \tag{7.15}$$

Based on certain assumptions on the form of $\{\mathcal{M}_l\}_{l=0}^{L}$ and characteristics of the multilevel estimators, similar theorems to the one given in Section 7.3 are formulated, connecting the computational cost and computational errors, for MLEnKF. For details see [Hoel et al., 2016]. One of these assumptions is geometric reduction in the variance of difference between model forecasts obtained at two consecutive levels. This requires small difference between the levels, thereby requiring large L. As the computational resources are limited, and the main aim of use of multilevel models is reduction of computational costs, this is not always a possibility in reservoir history-matching problems.

7.3.2 Multilevel hybrid ensemble Kalman filter

Fossum et al. [2020] introduced another formulation for multilevel statistics and introduced Multilevel hybrid ensemble Kalman filter. Similar to MLEnKF, this algorithm can be thought of as performing EnKF Evensen [1994] with the covariances in the update equation being formulated based on a multilevel model.

This algorithm can be briefly summarized as follows. Consider (7.10) and (7.11). Inspired by Bayesian model averaging, see e.g. [Fragoso et al., 2018], hybrid formulations were introduced for the statistics of the model forecasts [Fossum et al., 2020]. Specifying weights for each of the levels, the multilevel hybrid mean is given as

$$\mu_t^{\mathrm{MH}} = \sum_{l=0}^{L} w_l \left\{ \mathrm{E}[\mathcal{M}_l(Z_t)] \right\} , \tag{7.16}$$

where w_l is the weight specified for level l, with the condition,

$$\sum_{l=1}^{L} w_l = 1 \ . \tag{7.17}$$

The multilevel hybrid covariance of the model forecasts is formulated based on the same

idea and the law of total covariance as

$$\mathrm{C}_t^{\mathrm{MH}} = \sum_{\mathrm{l}=0}^{\mathrm{L}} w_l \left\{ \mathrm{C}[\mathcal{M}_l(Z_t)] \; + \left(\mathrm{E}[\mathcal{M}_l(Z_t)] - \mu_t^{\mathrm{MH}} \right) \left(\mathrm{E}[\mathcal{M}_l(Z_t)] - \mu_t^{\mathrm{MH}} \right)^T \right\} . \tag{7.18}$$

This formulation of the multilevel covariance has the advantage that it will always be positive-definite (sum of several positive-definite matrices is positive-definite). Hence, unlike the estimator of covariance used in MLEnKF, it does not need any additional treatment. This formulation is not an unbiased estimator for $\mathrm{C}[\mathcal{M}_L(Z_t)]$. However, it is a useful technique for solving the problems in which variance error dominates bias, which is often the case in reservoir history-matching problems [Fossum et al., 2020]. Additionally, for the estimator of the mean to be unbiased Fossum et al. [2020] proposed that the model forecasts be corrected as

$$\widetilde{\mathcal{M}}_l(Z_t) = \mathcal{M}_l(Z_t) + \mathrm{E}[\mathcal{M}_L(Z_t)] - \mathrm{E}[\mathcal{M}_{\mathrm{l}}(Z_t)] \, , \tag{7.19}$$

and (7.16) and (7.18) be calculated using $\{\widetilde{\mathcal{M}}_l\}_{l=0}^{L}$ instead of $\{\mathcal{M}_l\}_{l=0}^{L}$. The update equation for realization $z_{t,j,l}$ in MLHEnKF is given as

$$z_{t,j,l}^{pos} = z_{t,j,l}^{pri} + \mathrm{K}^{\mathrm{MH}} \left(d_{t,j,l} - H z_{t,j,l} \right) \, , \tag{7.20}$$

where the multilevel hybrid Kalman gain, K^{MH}, is formulated as

$$\mathrm{K}^{\mathrm{MH}} = \mathrm{C}_t^{\mathrm{MH}} \mathrm{H}^{\mathrm{T}} \left(\mathrm{H} \mathrm{C}_t^{\mathrm{MH}} \mathrm{H}^{\mathrm{T}} + \mathbb{C}_D \right)^{-1} \, . \tag{7.21}$$

This algorithm was tested on a selection of reservoir history-matching problems and consistently performed better than the widely used EnKF. It also performed better than MLEnKF in the problems where variance error dominated the bias [Fossum et al., 2020].

Chapter 8

Summaries of Papers and Outlook

A total of five papers are included in this thesis. In this chapter I summarize each of them separately and conclude the chapter with discussing the outlook of this work.

8.1 Summary of Paper A

Title: *A novel approach to multilevel data assimilation*
Authors: M. Nezhadali, T. Bhakta, K. Fossum, and T. Mannseth.

In Paper A, a multilevel ensemble-based data assimilation method, named multilevel hybrid ensemble smoother (MLHES), was introduced and was compared with two standard data assimilation techniques, i.e. two variants of ensemble smoother (ES) with localization. As the analysis step of MLHES is similar to that of ES, to compare it with the asymptotic results of the ES (which is the best vanilla ES can get), the assimilation outcome was also compared with a reference result which was generated based on ES with a large ensemble size. The numerical experiment was performed on a two-dimensional reservoir model. Fixing the computational power between different methods in the experiments, we assimilated the time-lapse bulk impedance data using the three schemes. The results from MLHES were superior to the two variants of ES with localization and were also reasonably similar to the results obtained from the reference case. This superiority was consistent across both the mean and the variance of the posterior parameters. In addition, the cumulative density function of the ensemble of updated model parameters was evaluated in three random cells in the field. In all the three cells, MLHES showed a closer mean, variance, and distribution function to those of the reference case comparing to the standard DA schemes. Furthermore, MLHES showed consistency in quality of history-matching by variation in observation data error, in estimation of both

the mean and the variance of the posterior parameters. This Paper was presented in the fully online event ECMOR XVII.

8.2 Summary of Paper B

Title: *Multilevel assimilation of inverted seismic data with correction for multilevel modeling error*

Authors: M. Nezhadali, T. Bhakta, K. Fossum, and T. Mannseth.

In Paper B, four schemes were presented using which one can partially account for the modeling error introduced to the system by use of multilevel models. multilevel data assimilation (MLDA) utilizes multilevel simulations in the forecast step. MLDA therefore renders the possibility of decreasing Monte Carlo (MC) errors without increasing the total computational cost, but MLDA will also introduce multilevel modeling errors (MLME) that are not present in conventional simulation results. The underlying assumption is therefore that the gain in reducing the MC error is larger than the loss in introducing the MLME. If the MLME could be approximately accounted for, however, MLDA performance could be further improved.

Four computationally inexpensive approximate MLME correction schemes were considered. They were denoted mean bias correction, stochastic correction, deterministic correction, and telescopic correction. The abilities of the four schemes to correct for the MLME were assessed in two ways, utilizing numerical experiments with three selected reservoir models.

Firstly, a measure for success of addressing the MLME was considered, i.e., normalized correction ratio (NCR). Then the statistics of the NCR for different levels using the four schemes were compared. The results showed that the correction schemes were capable of reducing the MLME, but the amount of reduction depended on the case and on the level. In general, the MLME correction schemes were more successful in correcting the MLME in the coarser levels compared to the finer ones, with telescopic correction showing the best performance followed by deterministic correction, stochastic correction, and mean bias correction.

Secondly, we assessed the performances of the different MLME-corrected model forecasts in assimilation of inverted seismic data, using the Multilevel Hybrid Ensemble Smoother (MLHES). The resulting posterior mean and variance fields with and without MLME correction were visually compared to results obtained from conventional Ensemble Smoother (ES) with localization, utilizing results obtained with conventional ES with

an unrealistically large ensemble size as the gold standard. It was found that MLHES with and without MLME correction outperformed conventional ES with localization.

Use of all four MLME correction schemes, and in fact also MLDA without MLME correction, mostly resulted in posterior parameter estimates with similar quality. For each example, we also ran a computationally much more costly MLDA variant where the MLME had been exactly accounted for in the model forecasts. No differences in quality between results obtained with this method and results obtained with several of the computationally inexpensive MLDA variants were found in all three examples. Altogether, these results indicate that the MLHES is reasonably robust towards MLMEs.

8.3 Summary of Paper C

Title: *Iterative multilevel assimilation of inverted seismic data*
Authors: M. Nezhadali, T. Bhakta, K. Fossum, and T. Mannseth.

In Paper C, a recently devised MLDA algorithm (MLHES) for assimilation of spatially distributed data was discussed, and an iterative version of it (IMLHES) was introduced. In addition, performance of these algorithms were evaluated in comparison with two standard DA algorithms, i.e. ensemble smoother (ES) with localization and EnRML (iterative ensemble smoother, IES) with localization. In doing so, three experiments were conducted. Each experiment was performed on a reservoir model and consisted of six algorithm runs: conventional ES with localization (ES-LOC), MLHES, ES with an exceedingly large ensemble size (ES-REF), conventional IES with localization (IES-LOC), IMLHES, and IES with an exceedingly large ensemble (IES-REF).

In order to assess the numerical results, firstly, the mean and variance fields of posterior unknown parameters (log permeability) were generated and assessed visually. The assessments showed that in all experiments the results from MLHES were more similar to those of ES-REF than the results from ES-LOC were. Except for one experiment, where both IES-LOC and IES-REF did not converge to the proximity of global optimum, the same conclusion was true about the iterative algorithms. The exception suggests an additional advantage of IMLHES over IES. It was also observed that iterations resulted in all the mean posterior fields obtained by IMLHES to be closer to the permeability field from which the observation data were generated than the mean posterior field obtained by MLHES.

Secondly, fine-scale simulations were conducted for all the posterior ensembles of all algorithms. Plots of simulated time-lapse bulk impedance means and variances were

compared to plots of observed time-lapse bulk impedance. Visual analysis of these plots showed that in all the applicable cases, the multilevel algorithms performed more similar to the reference cases than the conventional algorithms did. Additionally, the means of model forecasts obtained from IMLHES were closer to the observation data than the means of model forecasts obtained from MLHES were.

8.4 Summary of Paper D

Title: *Sequential multilevel assimilation of inverted seismic data*
Authors: M. Nezhadali, T. Bhakta, K. Fossum, and T. Mannseth.

In Paper D, two variants of a novel sequential multilevel data assimilation (SeMLDA) algorithm, SMLES-S and SMLES-H, were introduced. In addition, performances of these algorithms were assessed in comparison with two conventional DA algorithms and a simultaneous MLDA algorithm. In doing so, three experiments were performed on three reservoir models. The three experiments were designed such that the performance of the algorithms were evaluated in different settings for the prior parameter fields (different variogram types; different anisotropies; and various correlation lengths including long-range correlation, short-range correlation, and mixture of long-range and short-range correlations) and different ranges for the variograms used for the data-error covariance. Each of the experiments consisted of seven algorithm runs: SMLES-S, SMLES-H, ESMDA with localization (ESMDA-LOC), vanilla ESMDA with an exceedingly large ensemble (ESMDA-REF), an iterative simultaneous MLDA algorithm (IMLHES), iterative ensemble smoother with localization (IES-LOC), and vanilla iterative ensemble smoother with an exceedingly large ensemble (IES-REF). Results of the experiments were assessed both qualitatively and quantitatively.

In order for qualitative evaluation of the numerical results, firstly, the mean and the variance of posterior parameter fields were generated and assessed visually. The relative performances of the different methods were similar for all three experiments. The assessments showed that both SMLES-S and SMLES-H performed more similar to ESMDA-REF than ESMDA-LOC did in estimation of the posterior parameter mean field. Regarding estimation of the variance fields, SMLES-H overestimated the variance while SMLES-S underestimated it. The superiority of performance of both SeMLDA algorithms over ESMDA-LOC was evident, also for the variance fields. Among the iterative algorithms, IMLHES performed more similar to IES-REF than IES-LOC did. There was no indication of superior performance of either SMLES-H or IMLHES over each other in any of the experiment when their performances were compared to IES-REF re-

sults. However, IMLHES used more computational resources than either of the SeMLDA algorithms. Both IMLHES and SMLES-H performed slightly better than SMLES-S.

Additionally, fine-scale simulations were run for all the posterior ensembles obtained by the different algorithms in all the experiments. Plots of the mean and the variance of model forecasts from the different algorithms were compared to each other. Visual analysis of these plots showed that in all the experiments the ML algorithms performed better than their conventional DA counterparts. Among the ML algorithms, SMLES-S consistently performed better than both SMLES-H and IMLHES in estimation of the variance of the posterior model forecasts. Either of the other two MLDA algorithms did not have a clear advantage over each other.

Two metrics were adopted for quantitative comparison of the DA results obtained by different algorithms for estimation of both mean and variance of the posterior parameters and model forecasts. The metrics indicated that the ML algorithms generally performed better than the conventional DA algorithms in estimation of both mean and variance of the posterior parameters. They also indicated that SMLES-H and IMLHES performed slightly better than SMLES-S in estimation of the variance of the posterior parameters, and that all the MLDA algorithms performed better than IES-LOC in estimation of mean and variance of the posterior model forecasts. SMLES-S also performed consistently superior to ESMDA-LOC in estimation of mean and variance of the posterior model forecasts, while this was not observed for IMLHES and SMLES-H. Among the ML algorithms, SMLES-S clearly performed best when it came to estimation of the variance of the model forecasts. The other two algorithms did not consistently perform better than one another.

8.5 Summary of Paper E

Title: *Towards Application of Multilevel Data Assimilation in Realistic Reservoir History-Matching Problems*

Authors: M. Nezhadali, T. Bhakta, K. Fossum, and T. Mannseth.

In Paper E, performance of a novel multilevel ensemble-based data assimilation algorithm (SMLES) was assessed for assimilation of inverted seismic data in realistic oil reservoir problems. In doing so, Reservoir Model UNISIM-I generated by the state university of Campinas (UNICAMP) was chosen. A multilevel model was defined based on coarsening of the spatial grid of this reservoir. A numerical experiment was designed to compare this novel algorithm with standard ESMDA with localization (ESMDA-LOC). The numerical experiment was performed by estimation of logarithmic permeability field,

by assimilating the bulk impedance as observation data. The history-matching results were compared against the observation data and the synthetic truth. Qualitative assessments of the mean fields of the posterior parameters and model forecasts showed that both SMLES and ESMDA-LOC are equally successful in recovering the mean synthetic truth field and mean observation data. Even though there was no criteria for comparison of the posterior standard deviation fields of parameters and model forecasts, ESMDA-LOC resulted in ensemble collapse. There was a possibility to improve the results from ESMDA-LOC with optimizing by tuning the localization scheme; however, this would have added to the computational cost. SMLES, on the other hand, performed well without need for such tuning. This work was presented in ECMOR 2022, Hague, the Netherlands.

8.6 Outlook

In this project, we developed several algorithms for multilevel data assimilation (MLDA) of spatially distributed data and schemes for partially addressing the multilevel modeling error (MLME).

The research on this field can be continued on several fronts. The multilevel algorithms were devised and tested successfully, but not optimized. Several of the internal hyperparameters of the multilevel algorithms (number of levels, allocation of resources between the levels, types of surrogate models used, etc.) can be investigated more rigorously and optimized.

Most of the reservoir models investigated in this research were two-dimensional simple models. The performance of MLDA algorithms and MLME correction schemes can also be assessed in a larger number of realistic reservoir models, and real petroleum reservoir models. This work focused on estimation of permeability field based on assimilation of bulk impedance data. Use of other parameters and data types, and combination of them with the parameters and the data considered in this research can be investigated in the context of multilevel data assimilation.

Finally, as real reservoir models are more complex than the fields investigated in this work, extensive coarsening of the grid may result in large numerical errors and model bias. This limitation may result in decrease in the advantage of use of multilevel models. However, since multilevel models are utilized in the forecast step of the DA, and localization techniques are performed in the analysis step, these two techniques can be simply combined. Investigation of coupling these two techniques is also an intriguing

possibility for research.

Bibliography

C. C. Agbalaka and D. S. Oliver. Application of the EnKF and localization to automatic history matching of facies distribution and production data. *Mathematical Geosciences*, 40(4):353–374, 2008.

N. M. Alexandrov and R. M. Lewis. An overview of first-order model management for engineering optimization. *Optimization and Engineering*, 2(4):413–430, 2001.

S.-i. Amari. Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4-5):185–196, 1993.

O. Axelsson. *Iterative solution methods*. Cambridge university press, 1996.

A. Bagchi and V. Borkar. Parameter identification in infinte dimensional linear systems: Parameter identification. *Stochastics: An International Journal of Probability and Stochastic Processes*, 12(3-4):201–213, 1984.

M. H. Bakr, J. W. Bandler, K. Madsen, and J. Søndergaard. An introduction to the space mapping technique. *Optimization and Engineering*, 2(4):369–384, 2001.

J. W. Bandler, R. M. Biernacki, S. H. Chen, P. A. Grobelny, and R. H. Hemmers. Space mapping technique for electromagnetic optimization. *IEEE Transactions on microwave theory and techniques*, 42(12):2536–2544, 1994.

R. Bannister. A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(703): 607–633, 2017.

M. Benning and M. Burger. Modern regularization methods for inverse problems. *Acta Numerica*, 27:1–111, 2018.

K. Bergemann and S. Reich. An ensemble Kalman-Bucy filter for continuous data assimilation. *Meteorologische Zeitschrift*, 21(3):213, 2012.

A. Bomers, R. M. Schielen, and S. J. M. H. Hulscher. Application of a lower-fidelity surrogate hydraulic model for historic flood reconstruction. *Environmental modelling & software*, 117:223–236, 2019.

J.-F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. *Numerical optimization: theoretical and practical aspects*. Springer Science & Business Media, 2006.

V. Borkar and A. Bagchi. Parameter estimation in continuous-time stochastic processes. *Stochastics: An International Journal of Probability and Stochastic Processes*, 8(3): 193–212, 1982.

E. T. Brantson, B. Ju, B. O. Omisore, D. Wu, A. E. Selase, and N. Liu. Development of machine learning predictive models for history matching tight gas carbonate reservoir production profiles. *Journal of Geophysics and Engineering*, 15(5):2235–2251, 2018.

K. Brusdal, J.-M. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, P. J. van Leeuwen, E. Dombrowsky, and J. Verron. A demonstration of ensemble-based assimilation methods with a layered ogcm from the perspective of operational ocean forecasting systems. *Journal of Marine Systems*, 40:253–289, 2003.

G. Burgers, P. Jan van Leeuwen, and G. Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly weather review*, 126(6):1719–1724, 1998.

Z. Chai, A. Nwachukwu, Y. Zagayevskiy, S. Amini, and S. Madasu. An integrated closed-loop solution to assisted history matching and field optimization with machine learning techniques. *Journal of Petroleum Science and Engineering*, 198:108204, 2021.

G. Chavent and J. Liu. Multiscale parametrization for the estimation of a diffusion coefficient in elliptic and parabolic problems. *IFAC Proceedings Volumes*, 22(4):193–202, 1989.

Y. Chen and D. S. Oliver. Cross-covariances and localization for EnKF in multiphase flow data assimilation. *Computational Geosciences*, 14(4):579–601, 2010.

Y. Chen and D. S. Oliver. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Mathematical Geosciences*, 44(1):1–26, 2012.

Y. Chen and D. S. Oliver. Levenberg–marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computational Geosciences*, 17(4):689–703, 2013.

Y. Chen and D. S. Oliver. Localization and regularization for iterative ensemble smoothers. *Computational Geosciences*, 21(1):13–30, 2017.

Z. Chen. *Reservoir simulation: mathematical techniques in oil recovery*. SIAM, 2007.

L. A. N. Costa, C. Maschio, and D. J. Schiozer. Application of artificial neural networks in a history matching process. *Journal of Petroleum Science and Engineering*, 123: 30–45, 2014.

R. J. de Moraes, H. Hajibeygi, and J. D. Jansen. A multiscale method for data assimilation. *Computational Geosciences*, 24(2):425–442, 2020.

D. Devegowda, E. Arroyo, A. Datta-Gupta, and S. G. Douma. Efficient and robust reservoir model updating using ensemble Kalman filter with sensitivity-based covariance localization. In *SPE Reservoir Simulation Symposium*. OnePetro, 2007.

A. Dixit and A. Elsheikh. Robust well-production control using surrogate assisted reinforcement learning. In *ECMOR 2022*, volume 2022, pages 1–9. European Association of Geoscientists & Engineers, 2022.

A. A. Emerick and A. C. Reynolds. History matching a field case using the ensemble Kalman filter with covariance localization. *SPE Reservoir Evaluation & Engineering*, 14(04):443–452, 2011.

A. A. Emerick and A. C. Reynolds. Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, 55:3–15, 2013.

H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.

G. Evensen. Sampling strategies and square root analysis schemes for the EnKF. *Ocean dynamics*, 54(6):539–560, 2004.

G. Evensen, P. N. Raanes, A. S. Stordal, and J. Hove. Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching. *Frontiers in Applied Mathematics and Statistics*, 5:47, 2019.

G. Evensen, F. C. Vossepoel, and P. J. van Leeuwen. *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*. Springer Textbooks in Earth Sciences, Geography and Environment. Springer Nature, 2022.

G. Evensen et al. *Data assimilation: the ensemble Kalman filter*, volume 2. Springer, 2009.

A. Fahimuddin. *4D seismic history matching using the ensemble Kalman filter (EnKF): possibilities and challenges*. PhD thesis, The University of Bergen, 2010.

A. Fahimuddin, S. I. Aanonsen, and J.-A. Skjervheim. 4D seismic history matching of a real field case with EnKF: use of local analysis for model updating. In *SPE Annual Technical Conference and Exhibition*. OnePetro, 2010.

C.-S. Fen, C. Chan, H.-C. Cheng, et al. Assessing a response surface-based optimization approach for soil vapor extraction system design. *Journal of Water Resources Planning and Management*, 135(3):198, 2009.

A. I. Forrester, N. W. Bressloff, and A. J. Keane. Optimization using surrogate models and partially converged computational fluid dynamics simulations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 462(2071):2177–2204, 2006.

A. I. Forrester, A. Sóbester, and A. J. Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the royal society a: mathematical, physical and engineering sciences*, 463(2088):3251–3269, 2007.

K. Fossum. Assessment of sequential and simultaneous ensemble-based history matching methods for weakly non-linear problems. 2015.

K. Fossum and T. Mannseth. Coarse-scale data assimilation as a generic alternative to localization. *Computational Geosciences*, 21(1):167–186, 2017.

K. Fossum and T. Mannseth. A novel multilevel method for assimilating spatially dense data. In *ECMOR XVI-16th European Conference on the Mathematics of Oil Recovery*, volume 2018, pages 1–12. European Association of Geoscientists & Engineers, 2018.

K. Fossum, T. Mannseth, and A. S. Stordal. Assessment of multilevel ensemble-based data assimilation for reservoir history matching. *Computational Geosciences*, 24(1): 217–239, 2020.

T. M. Fragoso, W. Bertoli, and F. Louzada. Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review*, 86(1):1–28, 2018.

R. Furrer and T. Bengtsson. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis*, 98(2): 227–255, 2007.

S. E. Gano, J. E. Renaud, J. D. Martin, and T. W. Simpson. Update strategies for kriging models used in variable fidelity optimization. *Structural and Multidisciplinary Optimization*, 32(4):287–298, 2006.

G. Gaspari and S. E. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, 1999.

F. Gassmann. Uber die elastizitat poroser medien. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zurich*, 96:1–23, 1951.

M. B. Giles. Multilevel Monte Carlo methods. *Acta numerica*, 24:259–328, 2015.

K. Glover. All optimal hankel-norm approximations of linear multivariable systems and their l,∞-error bounds. *International journal of control*, 39(6):1115–1193, 1984.

A. A. Grimstad and T. Mannseth. Nonlinearity, scale, and sensitivity for parameter estimation problems. *SIAM Journal on Scientific Computing*, 21(6):2096–2113, 2000.

Y. Gu and D. S. Oliver. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *Spe Journal*, 12(04):438–446, 2007.

J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pages 49–52, 1902.

A.-L. Haji-Ali, F. Nobile, and R. Tempone. Multi-index Monte Carlo: when sparsity meets sampling. *Numerische Mathematik*, 132(4):767–806, 2016.

T. M. Hamill. Ensemble-based atmospheric data assimilation. *Predictability of weather and climate*, 124:156, 2006.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. 1970.

S. Hatfield, A. Subramanian, T. Palmer, and P. Düben. Improving weather forecast skill through reduced-precision data assimilation. *Monthly Weather Review*, 146(1):49–62, 2018.

J. He, P. Sarma, and L. J. Durlofsky. Reduced-order flow modeling and geological parameterization for ensemble-based data assimilation. *Computers & Geosciences*, 55: 54–69, 2013.

H. Hoel, K. J. Law, and R. Tempone. Multilevel ensemble Kalman filtering. *SIAM Journal on Numerical Analysis*, 54(3):1813–1839, 2016.

H. Hoel, G. Shaimerdenova, and R. Tempone. Multilevel ensemble Kalman filtering based on a sample average of independent EnKF estimators. *Foundations of Data Science*, 2(4):351, 2020.

K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

M. F. Hussain, R. R. Barton, and S. B. Joshi. Metamodeling: radial basis functions, versus polynomials. *European Journal of Operational Research*, 138(1):142–154, 2002.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2021.

S. Jung, K. Lee, C. Park, and J. Choe. Ensemble-based data assimilation in reservoir characterization: A review. *Energies*, 11(2):445, 2018.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

M. C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.

A. P. Kyprioti and A. A. Taflanidis. Kriging metamodeling for seismic response distribution estimation. *Earthquake Engineering & Structural Dynamics*, 50(13):3550–3576, 2021.

S. J. Leary, A. Bhaskar, and A. J. Keane. A knowledge-based approach to response surface modelling in multifidelity optimization. *Journal of Global Optimization*, 26(3): 297–319, 2003.

J. Liesen and Z. Strakos. *Krylov subspace methods: principles and analysis*. Oxford University Press, 2013.

X. Luo and T. Bhakta. Towards automatic and adaptive localization for ensemble-based history matching. In *ECMOR XVI-16th European Conference on the Mathematics of Oil Recovery*, volume 2018, pages 1–26. European Association of Geoscientists & Engineers, 2018.

X. Luo and T. Bhakta. Automatic and adaptive localization for ensemble-based history matching. *Journal of Petroleum Science and Engineering*, 184:106559, 2020.

X. Luo, T. Bhakta, G. Naevdal, et al. Correlation-based adaptive localization with applications to ensemble-based 4D-seismic history matching. *SPE Journal*, 23(02): 396–427, 2018.

T. Mannseth. Assimilation of multiple linearly dependent data vectors. *Computational Geosciences*, 24(1):349–354, 2020.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

R. D. Mindlin and H. Deresiewicz. Elastic spheres in contact under varying oblique forces. 1953.

G. Moldovan, G. Lehnasch, L. Cordier, and M. Meldi. A multigrid/ensemble Kalman filter strategy for assimilation of unsteady flows. *Journal of Computational Physics*, page 110481, 2021.

C. Montzka, V. R. Pauwels, H.-J. H. Franssen, X. Han, and H. Vereecken. Multivariate and multiscale data assimilation in terrestrial systems: A review. *Sensors*, 12(12): 16291–16333, 2012.

T. K. Moon. The expectation-maximization algorithm. *IEEE Signal processing magazine*, 13(6):47–60, 1996.

B. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE transactions on automatic control*, 26(1):17–32, 1981.

M. Morimoto, K. Fukami, K. Zhang, A. G. Nair, and K. Fukagata. Convolutional neural networks for fluid flow analysis: toward effective metamodeling and low dimensionalization. *Theoretical and Computational Fluid Dynamics*, 35(5):633–658, 2021.

B. Neal, S. Mittal, A. Baratin, V. Tantia, M. Scicluna, S. Lacoste-Julien, and I. Mitliagkas. A modern take on the bias-variance tradeoff in neural networks. *arXiv preprint arXiv:1810.08591*, 2018.

G. W. Oehlert. *A first course in design and analysis of experiments*. 2010.

P. R. Oke, P. Sakov, M. L. Cahill, J. R. Dunn, R. Fiedler, D. A. Griffin, J. V. Mansbridge, K. R. Ridgway, and A. Schiller. Towards a dynamically balanced eddy-resolving ocean reanalysis: Bran3. *Ocean Modelling*, 67:52–70, 2013.

D. S. Oliver and Y. Chen. Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15(1):185–221, 2011.

D. S. Oliver, A. C. Reynolds, and N. Liu. *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. Cambridge University Press, 2008. doi: 10.1017/CBO9780511535642.

D. S. Oliver, K. Fossum, T. Bhakta, I. Sandø, G. Nævdal, and R. J. Lorentzen. 4D seismic history matching. *Journal of Petroleum Science and Engineering*, 207:109119, 2021.

E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. Patil, and J. A. Yorke. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 56(5):415–428, 2004.

P. S. Palar and K. Shimoyama. Multi-fidelity uncertainty analysis in CFD using hierarchical kriging. In *35th AIAA applied aerodynamics conference*, page 3261, 2017.

B. Peherstorfer, K. Willcox, and M. Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *Siam Review*, 60(3):550–591, 2018.

A. A. Popov, C. Mou, A. Sandu, and T. Iliescu. A multifidelity ensemble Kalman filter with reduced order control variates. *SIAM Journal on Scientific Computing*, 43(2): A1134–A1162, 2021.

P. N. Raanes, A. S. Stordal, and G. Evensen. Revising the stochastic iterative ensemble smoother. *Nonlinear Processes in Geophysics*, 26(3):325–338, 2019.

A. Rani, N. Kumar, J. Kumar, J. Kumar, and N. K. Sinha. Chapter 6 - machine learning for soil moisture assessment. In R. C. Poonia, V. Singh, and S. R. Nayak, editors, *Deep Learning for Sustainable Agriculture*, Cognitive Data Science in Sustainable Computing, pages 143–168. Academic Press, 2022. ISBN 978-0-323-85214-2. doi: https://doi.org/10.1016/B978-0-323-85214-2.00001-X. URL `https://www.sciencedirect.com/science/article/pii/B978032385214200001X`.

S. Razavi, B. A. Tolson, and D. H. Burn. Review of surrogate modeling in water resources. *Water Resources Research*, 48(7), 2012.

R. G. Regis and C. A. Shoemaker. Local function approximation in evolutionary algorithms for the optimization of costly functions. *IEEE transactions on evolutionary computation*, 8(5):490–505, 2004.

C.-h. Rhee and P. W. Glynn. A new approach to unbiased estimation for SDE's. In *Proceedings of the 2012 Winter Simulation Conference (WSC)*, pages 1–7. IEEE, 2012.

C. P. Robert and G. Casella. *Monte Carlo statistical methods*, volume 2. Springer, 1999.

N. D. Roman, F. Bre, V. D. Fachinotti, and R. Lamberts. Application and characterization of metamodels based on artificial neural networks for building performance simulation: A systematic review. *Energy and Buildings*, 217:109972, 2020.

J. J. Ruiz, M. Pulido, and T. Miyoshi. Estimating model parameters with ensemble-based data assimilation: A review. *Journal of the Meteorological Society of Japan. Ser. II*, 91(2):79–99, 2013.

D. E. Rumelhart, R. Durbin, R. Golden, and Y. Chauvin. Backpropagation: The basic theory. *Backpropagation: Theory, architectures and applications*, pages 1–34, 1995.

P. Sakov and L. Bertino. Relation between two common localisation methods for the EnKF. *Computational Geosciences*, 15(2):225–237, 2011.

P. Sakov, D. S. Oliver, and L. Bertino. An iterative EnKF for strongly nonlinear systems. *Monthly Weather Review*, 140(6):1988–2004, 2012.

M. Saraiva, A. Forechi, J. D. O. Neto, A. DelRey, and T. Rauber. Data-driven full-waveform inversion surrogate using conditional generative adversarial networks. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021.

S. Särkkä. *Bayesian Filtering and Smoothing*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2013. doi: 10.1017/CBO9781139344203.

W. H. Schilders, H. A. Van der Vorst, and J. Rommes. *Model order reduction: theory, research aspects and applications*, volume 13. Springer, 2008.

L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Josa a*, 4(3):519–524, 1987.

T. M. Smith, C. H. Sondergeld, and C. S. Rai. Gassmann fluid substitutions: A tutorial. *Geophysics*, 68(2):430–440, 2003.

R. V. Soares, C. Maschio, and D. J. Schiozer. Applying a localization technique to Kalman gain and assessing the influence on the variability of models in history matching. *Journal of Petroleum Science and Engineering*, 169:110–125, 2018.

G. Sun, G. Li, M. Stone, and Q. Li. A two-stage multi-fidelity optimization procedure for honeycomb-type cellular materials. *Computational Materials Science*, 49(3):500–511, 2010.

A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, 2005. doi: 10.1137/1.9780898717921. URL `https://epubs.siam.org/doi/abs/10.1137/1.9780898717921`.

M. Tarrahi, S. H. Elahi, and B. Jafarpour. Fast linearized forecasts for subsurface flow data assimilation with ensemble Kalman filter. *Computational Geosciences*, 20(5): 929–952, 2016.

P. Thokala and J. R. Martins. Variable-complexity optimization applied to airfoil design. *Engineering optimization*, 39(3):271–286, 2007.

A. N. Tikhonov. On the stability of the functional optimization problem. *USSR Computational Mathematics and Mathematical Physics*, 6(4):28–33, 1966.

P. J. Van Leeuwen and G. Evensen. Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review*, 124(12):2898–2913, 1996.

D. Vassallo, R. Krishnamurthy, and H. Fernando. Decreasing wind speed extrapolation error via domain-specific feature extraction and selection. *Wind Energy Science*, 5: 959–975, 07 2020. doi: 10.5194/wes-5-959-2020.

F. A. Viana, V. Steffen, S. Butkewitsch, and M. de Freitas Leal. Optimization of aircraft structural components by using nature-inspired algorithms and multi-fidelity approximations. *Journal of Global Optimization*, 45(3):427–449, 2009.

R. Vitali, R. T. Haftka, and B. V. Sankar. Multi-fidelity design of stiffened composite panel with a crack. *Structural and Multidisciplinary Optimization*, 23(5):347–356, 2002.

M. A. Woodbury. Inverting modified matrices. In *Memorandum Rept. 42, Statistical Research Group*, page 4. Princeton Univ., 1950.

W. Xu, Q. Zheng, J. Zhang, L. Wu, and L. Zeng. Adaptive multi-fidelity data assimilation for nonlinear subsurface flow problems. In *AGU Fall Meeting Abstracts*, volume 2018, pages H23M–2133, 2018.

W. Yamazaki and D. J. Mavriplis. Derivative-enhanced variable fidelity surrogate modeling for aerodynamic functions. *AIAA journal*, 51(1):126–137, 2013.

Z. Yin, T. Feng, and C. MacBeth. Fast assimilation of frequently acquired 4D seismic data for reservoir history matching. *Computers & Geosciences*, 128:30–40, 2019.

K. Zhang, J. Zhang, X. Ma, C. Yao, L. Zhang, Y. Yang, J. Wang, J. Yao, and H. Zhao. History matching of naturally fractured reservoirs using a deep sparse autoencoder. *SPE Journal*, 26(04):1700–1721, 2021.

H. Zhonghua, X. Chenzhou, L. Zhang, Y. Zhang, K. Zhang, and S. Wenping. Efficient aerodynamic shape optimization using variable-fidelity surrogate models and multilevel computational grids. *Chinese Journal of Aeronautics*, 33(1):31–47, 2020.

Part II

Scientific results

Chapter 9

Paper A

A novel approach to multilevel data assimilation

Chapter 10

Paper B

Multilevel assimilation of inverted seismic data with correction for multilevel modeling error

frontiers in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 01 June 2021
doi: 10.3389/fams.2021.673077

Multilevel Assimilation of Inverted Seismic Data With Correction for Multilevel Modeling Error

Mohammad Nezhadali[1,2]*, *Tuhin Bhakta*[1], *Kristian Fossum*[1] and *Trond Mannseth*[1]

[1]NORCE Norwegian Research Center, Bergen, Norway, [2]University of Bergen, Bergen, Norway

With large amounts of simultaneous data, like inverted seismic data in reservoir modeling, negative effects of Monte Carlo errors in straightforward ensemble-based data assimilation (DA) are enhanced, typically resulting in underestimation of parameter uncertainties. Utilization of lower fidelity reservoir simulations reduces the computational cost per ensemble member, thereby rendering the possibility of increasing the ensemble size without increasing the total computational cost. Increasing the ensemble size will reduce Monte Carlo errors and therefore benefit DA results. The use of lower fidelity reservoir models will however introduce modeling errors in addition to those already present in conventional fidelity simulation results. Multilevel simulations utilize a selection of models for the same entity that constitute hierarchies both in fidelities and computational costs. In this work, we estimate and approximately account for the multilevel modeling error (MLME), that is, the part of the total modeling error that is caused by using a multilevel model hierarchy, instead of a single conventional model to calculate model forecasts. To this end, four computationally inexpensive approximate MLME correction schemes are considered, and their abilities to correct the multilevel model forecasts for reservoir models with different types of MLME are assessed. The numerical results show a consistent ranking of the MLME correction schemes. Additionally, we assess the performances of the different MLME-corrected model forecasts in assimilation of inverted seismic data. The posterior parameter estimates from multilevel DA with and without MLME correction are compared to results obtained from conventional single-level DA with localization. It is found that multilevel DA (MLDA) with and without MLME correction outperforms conventional DA with localization. The use of all four MLME correction schemes results in posterior parameter estimates with similar quality. Results obtained with MLDA without any MLME correction were also of similar quality, indicating some robustness of MLDA toward MLME.

Keywords: data assimilation, multilevel, modeling error, seismic data, reservoir model

OPEN ACCESS

Edited by:
Behnam Jafarpour, University of Southern California, Los Angeles, United States

Reviewed by:
Sanjay Srinivasan, Pennsylvania State University (PSU), United States
Lu Xiong, Middle Tennessee State University, United States

***Correspondence:**
Mohammad Nezhadali
mone@norceresearch.no

Specialty section:
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

Received: *26 February 2021*
Accepted: *10 May 2021*
Published: *01 June 2021*

Citation:
Nezhadali M, Bhakta T, Fossum K and Mannseth T (2021) Multilevel Assimilation of Inverted Seismic Data With Correction for Multilevel Modeling Error. Front. Appl. Math. Stat. 7:673077. doi: 10.3389/fams.2021.673077

1 INTRODUCTION

Sound decision-making in petroleum reservoir management depends on provision of reliable production forecasts from petroleum reservoir models, including provision of the uncertainty in the forecasts. The reliability is increased by utilization of available data for calibration of the models. Ensemble-based data assimilation (DA) methods, using statistically correct formulations, have accordingly become popular for automated reservoir history matching [1–4].

Monte Carlo approximations play crucial roles in ensemble-based DA. Due to computational cost limitations, the ensemble size is limited to roughly one hundred. Using straightforward ensemble-based DA, the degrees of freedom of the problem would equal the ensemble size, and such an approach would result in significant Monte Carlo errors. The negative effects of Monte Carlo errors are enlarged in the presence of large amounts of data to be assimilated simultaneously, for example, inverted seismic data, resulting in underestimation of variance of the unknown parameters and, in more severe cases, even in ensemble collapse.

The most widely used treatment for Monte Carlo errors is distance-based localization. The basic assumption underlying this method is that true correlations between a parameter and a datum decrease when the distance between their respective locations increases, and disappear if the distance exceeds a critical distance. This assumption may not always hold for subsurface problems. Different correlation functions and their utilization in DA can be found in References [5–7]. A proper choice of correlation function, as well as the critical distance in particular, depends on parameter and data types as well as on other problem settings. This reduces the robustness of distance-based localization, also for problems where its basic assumption does hold.

Simply increasing the ensemble size will, of course, reduce Monte Carlo errors, but it will also increase computational cost. Utilization of a lower cost and lower fidelity model renders the possibility of increasing the ensemble size without increasing the total computational cost. The use of a lower fidelity reservoir model will however introduce modeling errors in addition to those already present in conventional fidelity simulation results. The underlying assumption when applying lower fidelity models in DA is therefore that the gain in reducing Monte Carlo errors is larger than the loss in numerical simulation accuracy. If the abovementioned additional modeling errors could be approximately accounted for, utilization of lower fidelity models would be even more attractive. DA using various types of lower fidelity models has been applied to several inverse problems, for example, within petroleum reservoir modeling [8–10] and atmospheric science [11]. Note that since lower fidelity simulations are applied to the forecast step and localization is applied to the analysis step, the two techniques can be combined, if desired.

Multilevel simulations utilize a selection of models for the same entity that constitute hierarchies in both fidelities and computational costs (multilevel models). The idea is to decrease Monte Carlo errors without increasing numerical errors too much. There are a number of ways to realize multilevel models. We choose to construct them by spatial coarsening of the conventional simulation grid to several levels of coarseness and correspondingly upscale the associated grid-based parameter functions. Multilevel data assimilation (MLDA) [12–16] utilizes multilevel simulations in the forecast step. Since inverted seismic data are given on the conventional grid (denoted as the fine grid from now on), MLDA with such data must be able to handle differences in grid levels between data and model forecasts.

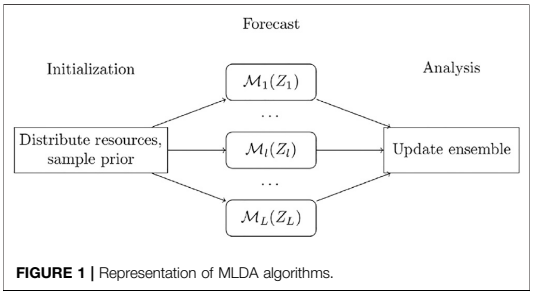


FIGURE 1 | Representation of MLDA algorithms.

Modeling errors are ubiquitous in all numerical simulations in the geosciences. In the context of a coarse-grid numerical model, three types of modeling errors can be envisioned: Type 1: the discrepancy between the physical reality and the solution obtained with a mathematical model attempting to model the physical reality; Type 2: the discrepancy between the solution obtained with that mathematical model and the model forecast from a numerical model resulting from discretization of the mathematical model; and Type 3: the discrepancy between the model forecast from that numerical model and the model forecast from a numerical model with a coarser simulation grid.

Assuming a normal distribution for the errors, a Bayesian framework for jointly accounting for Type 1 and Type 2 modeling errors in DA was presented in Reference [17]. The effect of Type 2 modeling errors on the solution to linear Gaussian inverse problems was analyzed in Reference [18]. (Neither Ref. [17] nor Ref. [18] were concerned with Type 3 modeling errors.) In this study, we quantify and approximately account for Type 3 modeling errors for each level in multilevel assimilation of spatially distributed data, such as inverted seismic data. (We will use the term multilevel modeling error (MLME) to denote this error from now on.) To this end, three computationally inexpensive approximate MLME correction schemes are developed, and their abilities to correct multilevel model forecasts for reservoir models with different types of MLME are assessed and compared to a previously proposed (also computationally inexpensive) MLME correction scheme. Additionally, we assess the performances of the different MLME-corrected model forecasts in assimilation of bulk impedance data. The posterior parameter estimates from MLDA with and without MLME correction are compared to results obtained from conventional single-level DA with localization.

The rest of this article is framed as follows. **Section 2** is devoted to explaining MLDA and introducing a recently devised algorithm for it. **Section 3** introduces MLME and the four proposed schemes for addressing it. **Section 4** explains the reservoir models used for assessment of the performance of MLME correction schemes in MLDA. In **Section 5**, we describe the numerical investigations, which are followed by their results in **Section 6**. Finally, in **Section 7**, we summarize and conclude the study.

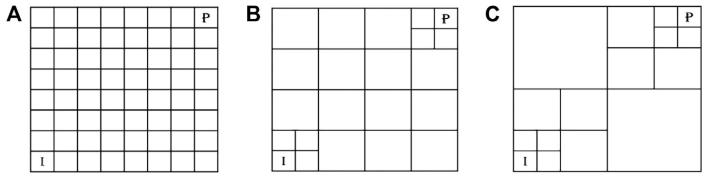


FIGURE 2 | Grid coarsening proposed by [15] performed on an 8 × 8 grid **(A)** finest level and **(B, C)** coarser levels.

2 MULTILEVEL DATA ASSIMILATION

The forecast step in ensemble-based DA takes the initial states and parameters as input and generates the model forecasts. In this work, the forecast step of MLDA is performed using a hierarchy of nested forward models, $\{\mathcal{M}_l\}_{l=0}^{L}$. After sampling from the prior distribution, the ensemble of prior state vectors is divided into L sub-ensembles. Hence, each of the sub-ensembles are modeled using corresponding forward models, as seen in **Figure 1**, where Z is the random vector of parameters and subscripts denote the sub-ensemble number.

In order to give a description for a full cycle of MLDA of spatially distributed data, multilevel models should be discussed. Additionally, since the MLDA uses the ensemble approximations of the mean and covariance of the model forecasts, which are in different resolutions for different levels, a robust transformation scheme should be devised for converting the model forecasts from one resolution to another. These two topics will be discussed in **Sections 2.1** and **2.2**, respectively. They will be followed by sections on upscaling of the observation data (**Section 2.3**) and formulation of multilevel statistics for mean and covariance of the model forecasts (**Section 2.4**). Afterward, a description of a recently devised method for MLDA of spatially distributed data, which will be used in our numerical experiments, is given in **Section 2.5**.

2.1 Multilevel Models

Let $\{\mathcal{M}_l\}_{l=0}^{L}$ be a set of deterministic models, where the accuracy and computational cost increase with an increase in l. Accordingly, they will form hierarchies of both accuracy and cost. One can think of several schemes to devise the hierarchy including but not limited to coarsening the spatial grid of the forward model together with upscaling the associated parameters, coarsening the temporal grid of the forward model, and relaxing the convergence criteria in the iterative linear solvers. All of these methods reduce the computational cost of the models and increase their numerical error. Coarsening the spatial grid and upscaling the associated parameters are chosen for the current work. The techniques presented in this work are however robust enough so that with minor manipulations, they can be used for other lower fidelity models.

As for coarsening the grid of the forward models, the authors of Reference [15] proposed a robust technique, which was also used in Reference [16]. This technique occurs in a sequence of steps. In each step, neighboring cells of the grid at the previous step are merged into a coarser cell, unless they are to be kept fine deliberately. A representation of the grid coarsening process for an 8 × 8 sample grid can be found in **Figure 2**. As it can be seen in the figure, coarsening has occurred in a uniform manner across both directions, except for the vicinity of two opposite corners, where the grid cells are kept in fine scale to boost the local numerical accuracy around the two wells, producer (P) and injector (I).

The parameters associated with the grid are upscaled in such a way that the physics of the problem do not change drastically. Upscaling of the parameters will be further discussed in **Section 4**.

2.2 Transformation of Model Forecasts

The discrepancy in coarseness of the multilevel grids results in the spatially distributed model forecasts to be in different resolutions for different levels. Therefore, in order to be able to compute the multilevel sample statistics of the model forecast, a robust transformation scheme should be devised for converting a model forecast from the resolution at one level to another.

In the problem at hand, transformation of the model forecast requires either upscaling or downscaling. To this end, a standard volume-weighted arithmetic averaging technique is used. Accordingly, we define a set of linear transformations, $\{U_p^q : \mathbb{R}^{S_p} \mapsto \mathbb{R}^{S_q} \big| 1 \le p, q \le L\}$, where S_p and S_q denote the dimension of the model forecast vector at arbitrary levels p and q, respectively, and U_p^q transforms the model forecast vector from level p to be compatible with level q.

Figure 3 gives two examples of transformation of a spatially distributed model forecast, one from a coarser grid to a finer grid and one vice versa. Each model forecast component is represented in its corresponding spatial grid cell.

As can be seen in **Figure 3**, in the upscaling procedure, the arbitrarily named model forecast components $\{a_i\}_{i=1}^{4}$ in the northwest zone from the finer grid (p) are averaged to form their corresponding model forecast component, $\overline{a}$, in the coarsened grid. Similar procedure has been performed for the rest of model forecast components, shown by *. In the downscaling procedure, on the other hand, the model forecast components in the coarse grid are simply copied to their corresponding components of the finer grid.

2.3 Upscaling of Observation Data

As part of the DA process, the mismatch between the model forecasts and observation data is to be calculated. Here, it is assumed that inverted seismic data are given in the resolution of

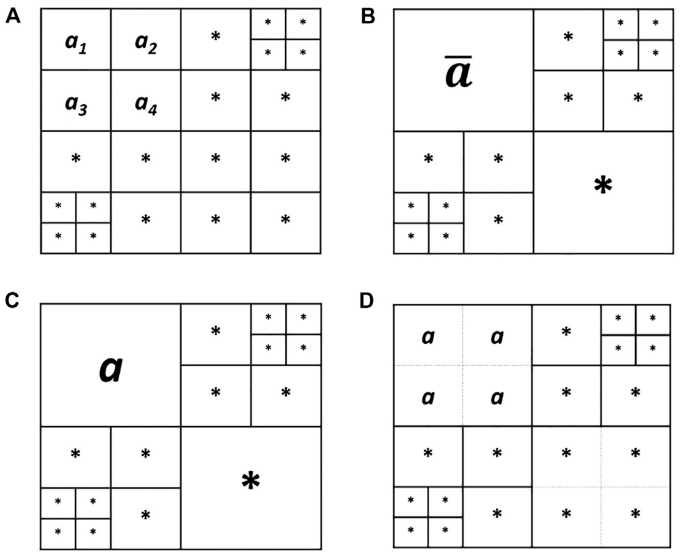


FIGURE 3 | Transformation of model forecast between two levels p (finer) and q (coarser) **(A)** model forecast in resolution of level p, **(B)** transformation of model forecast from resolution of level p to the resolution of level q, **(C)** model forecast in resolution of level q, and **(D)** transformation of model forecast from resolution of level q to the resolution of level p.

the finest simulation grid. Accordingly, for each of the levels, either the observation data should be upscaled to the resolution of model forecast or the model forecast should be downscaled to the resolution of the observation data. In this study, we take the former approach. Since the observation data are in the resolution of the finest model, using the same transformation functions as those designed for model forecasts on the fine observation data will result in upscaling of observation data into the preferred resolution.

2.4 Multilevel Statistics

Assuming we have approximations of the model forecasts, Y, being a function of the unknown parameter vector Z in several levels, a statistically correct scheme for approximation of multilevel statistics for Y is required. As for MLDA, the first two central moments of the model forecast are of foremost interest. Accordingly, formulations for these multilevel statistics are proposed.

Assuming the model with the highest fidelity, $\mathcal{M}_L$, to be exact, the authors of Reference [13] proposed an unbiased formulation for approximation of multilevel statistics for DA under certain conditions. Under these sets of conditions, the proposed method outperformed its alternatives [12]. For reservoir problems, however these conditions typically do not hold, and another formulation based on Bayesian model averaging (BMA) was proposed [12]. In this formulation, the statistics are computed based on reliability weights w_l for each of the levels l. This formulation is, by definition, a biased scheme for computation of multilevel moments; however, it will be a useful technique for problems in which variance error dominates bias, which is often the case for petroleum reservoir problems [12]. Using this formulation and transformations of the forecast, the authors of Reference [16] proposed a formulation of multilevel statistics for spatially distributed model forecasts, which is used in the current work. According to this scheme, the multilevel mean of the model forecast at level l is given as follows:

$$E_{ML}(Y_l) = \sum_{k=1}^{L} w_k U_k^l E(Y_k), \tag{1}$$

where $E(Y_k)$ denotes the sample mean of the model forecast at level k. Using the law of total variance, the multilevel approximation of covariance of the model forecast at level l is formulated as follows:

$$C_{\mathrm{ML}}(Y_l) = \sum_{k=1}^{L} w_k \Big\{ C\big(U_k^l Y_k\big) + \big(E\big(U_k^l Y_k\big) - E_{ML}(Y_l)\big)\big(E\big(U_k^l Y_k\big) - E_{ML}(Y_l)\big)^T \Big\}. \tag{2}$$

In addition, the parameter forecast cross-covariance can be written as follows:

$$\mathrm{C}_{\mathrm{ML}}(Z, Y_l) = \sum_{k=1}^{L} w_k \mathrm{C}\left(Z_k, U_k^l Y_k\right). \tag{3}$$

Using these ML moments enables us to utilize the classic DA formulations for updating the ensemble as will be presented in **Section 2.5**.

2.5 Multilevel Hybrid Ensemble Smoother

Utilizing statistics given by **Eqs. 1–3**, the authors of Reference [16] formulated an MLDA algorithm that rendered the possibility of assimilation of spatially distributed data, for example, inverted seismic data, in a multilevel manner. This DA algorithm was called multilevel hybrid ensemble smoother (MLHES). In this work, we briefly explain MLHES and utilize it in our numerical experiments. An iterative version of MLHES has also been developed[1].

Initially, a total of N_t realizations from the prior random parameter vector Z^{pr} are generated and divided into L sub-ensembles, that is, Z_l^{pr}, $1 \le l \le L$. Note that, regardless of l, all Z_l^{pr} are on the fine scale and the subscript denotes the model where they are used. Accordingly, prior realization j in sub-ensemble l, where $1 \le j \le N_l$ and $1 \le l \le L$, is called $z_{l,j}^{pr}$. Hence, there are N_l ensemble members in sub-ensemble l. Likewise, the model forecast pertaining to simulation of $z_{l,j}^{pr}$ by the forward model $\mathcal{M}_l$ is named $\widehat{y}_{l,j}$ and is given by

$$\widehat{y}_{l,j} = \mathcal{M}_l\left(z_{l,j}^{pr}\right). \tag{4}$$

The correction for MLME then would be performed as

$$y_{l,j} = \widehat{y}_{l,j} + \varepsilon_{l,\times}\left(z_{l,j}^{pr}\right), \tag{5}$$

where $y_{l,j}$ is the model forecast $\widehat{y}_{l,j}$ corrected for its MLME and $\varepsilon_{l,\times}$ is the generic term for correction of MLME. In Reference [16], the authors utilized the mean bias correction for addressing the MLME. This correction scheme is one of the MLME correction schemes investigated in this work and will be explained and discussed in more detail in **Section 3**.

After MLME correction, there will be a separate analysis step for each of the levels. The updated parameter vector of an arbitrary ensemble member is given by

$$z_{l,j}^{a} = z_{l,j}^{pr} + K_l\left(d_{l,j} - y_{l,j}\right), \tag{6}$$

where the observation data realization, $d_{l,j}$, is a random pick from $\mathbb{N}\left(U_L^l\mu_D, U_L^l\mathbb{C}_D U_L^{lT}\right)$, and μ_D and $\mathbb{C}_D$ are the original observation data mean and observation data error covariance in the finest level, respectively. The level-specific Kalman gain, K_l, is then given as

$$K_l = C_{\mathrm{ML}}(Z, Y_l)\left(C_{\mathrm{ML}}(Y_l) + U_L^l\mathbb{C}_D\left(U_L^l\right)^T\right)^{-1}, \tag{7}$$

where the multilevel statistics $C_{\mathrm{ML}}(Y_l)$ and $C_{\mathrm{ML}}(Z, Y_l)$ are given by **Eqs. 2** and **3**, respectively.

After the analysis step, the estimates of mean and covariance of the posterior parameter field are computed based on a pool composed of all realizations $z_{l,j}^{a}$ at all the levels as follows:

$$E(Z^a) = \frac{1}{N_t}\sum_{l=1}^{L}\sum_{j=1}^{N_l} z_{l,j}^{a} \tag{8}$$

$$C(Z^a) = \frac{1}{N_t - 1}\sum_{l=1}^{L}\sum_{j=1}^{N_l}\left(z_{l,j}^{a} - E(Z^a)\right)\left(z_{l,j}^{a} - E(Z^a)\right)^T \tag{9}$$

A pseudo-code of MLHES is presented in Appendix 1.

3 MULTILEVEL MODELING ERROR

Let R denote some spatially varying physical property, and let W denote the forecast of a mathematical model attempting to model R. Furthermore, let W_l denote the forecast of that mathematical model discretized at an arbitrarily selected level, l, and let $x_{l,n}$ denote the location of an arbitrarily selected point in the simulation grid at that level. The total modeling error in $W_l(x_{l,n})$ when attempting to model $R(x_{l,n})$ is then

$$\phi_l(x_{l,n}) = R(x_{l,n}) - W_l(x_{l,n}), \tag{10}$$

which can be rewritten as

$$\begin{aligned}\phi_l(x_{l,n}) = {} & \left(R(x_{l,n}) - W(x_{l,n})\right) + \left(W(x_{l,n}) - W_L(x_{l,n})\right) \\ & + \left(W_L(x_{l,n}) - W_l(x_{l,n})\right).\end{aligned} \tag{11}$$

The first term on the right-hand side of **Eq. 11** represents the error in the mathematical model's forecast of physical reality, and the second term represents the discretization error when simulating with a numerical model on the fine grid. We will consider the last term, which represents the error due to numerical simulation on the level-l grid, instead of on the fine grid.

The expression $W_L(x_{l,n})$ is precise only if $x_{l,n}$ coincides with a point in the simulation grid at level L, which will not be the case for the grid-coarsening procedure applied in the current work. To make this expression precise, we utilize the linear transformations defined in **Section 2.2**, and let $W_L(x_{l,n}) \stackrel{\mathrm{def}}{=} \left(U_L^l W_L(x_L)\right)_n$. We then define component n of the multilevel modeling error (MLME) as

$$\zeta_{l,n} = \left(U_L^l W_L(x_L)\right)_n - W_l(x_{l,n}), \tag{12}$$

and develop techniques for estimating $\zeta_l = (\zeta_{l,1} \ldots \zeta_{l,G_l})^T$ in model forecasts and approximately correcting for the MLME before assimilating data.

3.1 Multilevel Modeling Error Correction

Assuming fine model forecasts to be sufficiently accurate, ideally, the model forecasts at each level should be upscaled fine model forecasts to the resolution of that level, that is, the correction in **Eq. 5** should be $\varepsilon_{l,\times} = \zeta_l$, but due to computational limits, this is not a possibility. Accordingly, we try to approximate ζ_l using the discrepancies between the model forecasts at level l and the finest level, L. This will be done using the ensemble itself without any additional simulations.

The techniques developed here are therefore computationally cheap adjunctions which can be added to many MLDA algorithms with minor modifications. The four schemes formulated and investigated in this work are named as mean

[1]Nezhadali, M., Bhakta, T., Fossum, K., and Mannseth, T. Iterative multilevel assimilation of inverted seismic data. Submitted

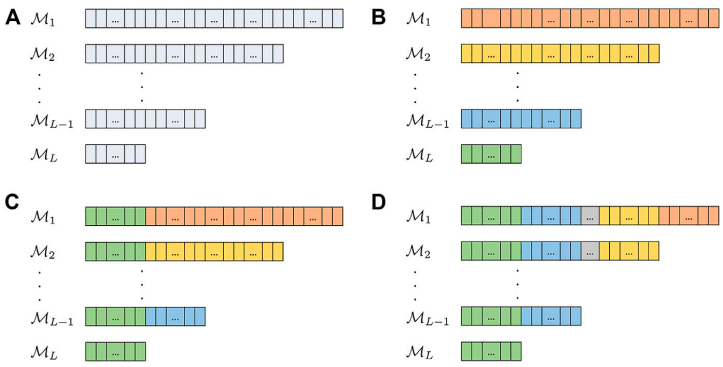


FIGURE 4 | (A) Division of the prior ensemble for Z into L sub-ensembles, **(B)** mean bias Correction prior setting, **(C)** stochastic and deterministic corrections prior settings, and **(D)** telescopic correction prior setting.

bias correction (MB), stochastic correction (ST), deterministic correction (DE), and telescopic correction (TE).

Figure 4A depicts general formation of the sub-ensembles from the prior ensemble for Z. The realizations in each sub-ensemble are put in the same line as the forward model that is used for their simulation; consider each of the unit cells in the rows as a realization of the prior and each row as a sub-ensemble. **Figures 4B–D** describe the requirements on prior realizations for different MLME correction schemes; if parts of two sub-ensembles are of the same color, those realizations are shared between those sub-ensembles and are to be simulated using both corresponding forward models.

3.1.1 Mean Bias Correction

This technique was used in Reference [12] for correction of the production data for their mean bias. There, the correction was formulated as

$$\varepsilon_{l,MB,P} = E(Y_L) - E(Y_l), \tag{13}$$

where $\varepsilon_{l,MB,P}$ is the mean bias correction term for production data. Here, we generalize this correction to be used also for spatially distributed data. Accordingly, $\varepsilon_{l,MB}$ is defined as

$$\varepsilon_{l,MB} = U_L^l E(Y_L) - E(Y_l), \tag{14}$$

where $E(Y_L)$ and $E(Y_l)$ are sample means of the model forecasts at levels L and l, respectively. Using this correction, the mean of the corrected forecast at every level would be equal to the upscaled mean of the forecast given by the most accurate (finest) model.

As can be seen in **Figure 4B**, in mean bias correction, as opposed to the rest of MLME correction schemes, the prior realizations are run on each of the forward models without any requirement to be run by other forward models.

3.1.2 Stochastic Correction

Simulating the sub-ensemble $\{z_{l,j}^{pr}\}_{j=1}^{N_L}$ using all the forward models, we can calculate ζ_l for those realizations as follows:

$$\zeta_l\left(z_{l,j}^{pr}\right) = U_L^l \mathcal{M}_L\left(z_{l,j}^{pr}\right) - \mathcal{M}_l\left(z_{l,j}^{pr}\right). \tag{15}$$

In the stochastic formulation, assuming a normal distribution for ζ_l, the realization of correction term is formulated as

$$\varepsilon_{l,ST} \sim \mathcal{N}\left(E(\zeta_l), C(\zeta_l)\right), \tag{16}$$

where $E(\zeta_l)$ and $C(\zeta_l)$ are the sample mean and covariance of realizations of ζ_l, respectively. As the ensemble size is often relatively small in comparison to the parameter vector size, the distribution defined in **Eq. 16** would not cover the full span of the probability space for ζ_l. The realizations of $\varepsilon_{l,ST}$, accordingly, would be in the sub-space spanned by the ensemble $\{\zeta_l(z_{l,j}^{pr})\}_{j=1}^{N_L}$

As seen in **Figure 4C** and **Eq. 15**, this correction scheme requires the realizations at sub-ensemble L to be simulated using all the forward models.

3.1.3 Deterministic Correction

Assume that ζ_l is a continuous function of Z. Furthermore, we assume local linearity for ζ_l and write the first two terms of its Taylor expansion around the population expectation of the parameter vector, $\mathbb{E}(Z)$ as

$$\zeta_l(Z) \approx \zeta_l(\mathbb{E}(Z)) + \frac{\partial \zeta_l}{\partial Z}\Big|_{\mathbb{E}(Z)} (Z - \mathbb{E}(Z)). \tag{17}$$

Under linearity assumption, we would have

$$\zeta_l(\mathbb{E}(Z)) \approx \mathbb{E}(\zeta_l(Z)). \tag{18}$$

To calculate the Jacobian of ζ_l, we use another approximation. Writing Stein's lemma gives,

$$\mathbb{C}(\zeta_l, Z) = \mathbb{E}\left(\frac{\partial \zeta_l}{\partial Z}\right)\mathbb{C}(Z), \tag{19}$$

where $\mathbb{C}(Z)$ and $\mathbb{C}(\zeta_l, Z)$ are covariance of Z and cross-covariance between ζ_l and Z, respectively. Rearranging gives

$$\mathbb{E}\left(\frac{\partial \zeta_l}{\partial Z}\right) = \mathbb{C}(\zeta_l, Z)\mathbb{C}(Z)^{-1}, \tag{20}$$

and the local linearity assumption gives the approximation

$$\left.\frac{\partial \zeta_l}{\partial Z}\right|_{\mathbb{E}(Z)} \approx \mathbb{E}\left(\frac{\partial \zeta_l}{\partial Z}\right). \tag{21}$$

We would then use the ensemble for approximation of both mean and Jacobian of ζ_l and use them for formulating the deterministic correction as follows:

$$\varepsilon_{l,DE}(E(Z)) \approx E(\zeta_l), \tag{22}$$

$$\left.\frac{\partial \varepsilon_{l,DE}}{\partial Z}\right|_{E(Z)} \approx C(\zeta_l, Z)\, C(Z)^{+}, \tag{23}$$

$$\varepsilon_{l,DE}(Z) = \varepsilon_{l,DE}(E(Z)) + \left.\frac{\partial \varepsilon_{l,DE}}{\partial Z}\right|_{E(Z)} (Z - E(Z)), \tag{24}$$

where the superscript + in **Eq. 23** denotes the Moore–Penrose pseudo-inverse. As shown in **Figure 4C**, in deterministic correction, as in stochastic correction, the realizations in the sub-ensemble L are to be simulated using all the forward models.

3.1.4 Telescopic Correction

This scheme utilizes the idea in deterministic correction in a telescopic manner so that it can benefit from the multilevel structure of the problem. The MLME can be written as

$$\zeta_l(Z) = U_L^l \mathcal{M}_L(Z) - \mathcal{M}_l(Z) = \sum_{k=l}^{L-1} U_k^l \left(U_{k+1}^k \mathcal{M}_{k+1}(Z) - \mathcal{M}_k(Z)\right), \tag{25}$$

and **Eq. 25** holds because all the transformations are linear and from a finer level to a coarser level. Accordingly, we can write

$$\zeta_l(Z) = \sum_{k=l}^{L-1} U_k^l e_k, \tag{26}$$

where

$$e_k(Z) = U_{k+1}^k \mathcal{M}_{k+1}(Z) - \mathcal{M}_k(Z). \tag{27}$$

This reformulation of the error term renders the possibility to approximate ζ_l via a summation of smaller error terms, which are approximated based on bigger ensembles. Hence, using the idea in deterministic correction for level-wise errors, e_k, one can write

$$e_k(Z) \approx E(e_k) + C(e_k, Z_k)\, C(Z_k)^{+} (Z - E(Z_k)), \tag{28}$$

where $E(e_k)$ is the sample mean of the partial error, $C(e_k, Z_k)$ is the sample cross-covariance of ζ_l and Z, $C(Z_k)$ is the sample covariance of the parameter vector, and $E(Z_k)$ is the sample mean of the parameter vector, all based on the realizations in sub-ensemble k. The telescopic correction term then is

$$\varepsilon_{l,TE}(Z) = \sum_{k=l}^{L-1} U_k^l \left(E(e_k) + C(e_k, Z_k)\, C(Z_k)^{+} (Z - E(Z_k))\right). \tag{29}$$

The idea here is that even though the errors in approximation aggregate in the summation in **Eq. 26**, the increase in the ensemble size would reduce Monte Carlo errors and the approximation of ζ_l would be more accurate, and overall, it would help to account better for the MLME.

In order to be able to perform this correction, a nested structure in the prior realizations is necessary. In other words, as seen in **Figure 4D**, all the realizations simulated by a forward model should also be computed using all the less accurate forward models than that model.

TABLE 1 | Shared properties of the reservoir models.

Fine cell dimension	30 × 30 × 30 (m^3)	Porosity	0.2
Initial oil saturation	0.85	Injector (I)	P.C. (275 bar)
Initial pressure	200 bar	Producer (P)	P.C. (100 bar)

P.C., pressure-controlled.

4 TEST MODELS

We are interested in assessing the quality of MLME correction schemes in reservoir history matching of inverted seismic data using MLHES. In accordance with it, three different reservoir models are considered. These reservoir models have some shared properties. They are two-dimensional, with 64 × 64 Cartesian grids, two wells in the opposite corners, an injector in the northeast corner, and a producer in the southwest corner. Compressible two-phase flow (oil and water), no-flow boundary conditions, and standard equations for capillary pressure and relative permeability are considered. A description of the other shared general properties of the reservoir models is given in **Table 1**. As the seismic vintages are different in each experiment, they are discussed separately in **Sections 5.1–5.3**.

In each of the reservoir models used in this work, the general structure is modified with the aim of increasing the MLME. These reservoir models are explained separately in **Sections 4.1–4.3**, and samples from the prior distribution of Z for each of the reservoir models can be found in **Figure 5**.

The forward models used for forecast $\{\mathcal{M}\}_{l=1}^{L}$ each consists of two segments. A reservoir flow model is used to predict the state variables in time, and a petro-elastic model is utilized for computing the elastic rock properties from parameters and predicted state variables.

The flow segment of the forward models is performed using Eclipse 100 [19]. Coarsening the grid is done by using the Eclipse keyword COARSEN, which merges groups of predefined neighboring cells to form a coarser grid. The upscaling of permeabilities is performed using pore-volume–weighted arithmetic averaging, and transmissibilities between two neighboring coarse cells in each direction are calculated based on harmonic averaging in that direction and summing it in other directions [19]. As for the petro-elastic segment of the forward model, an in-house model based on standard rock physics [20], [21, Report 1] was used.

4.1 Reservoir Model I

A nonpermeable fault has been added to the field with its normal vector pointing north, its eastern most point 4 grid cells away

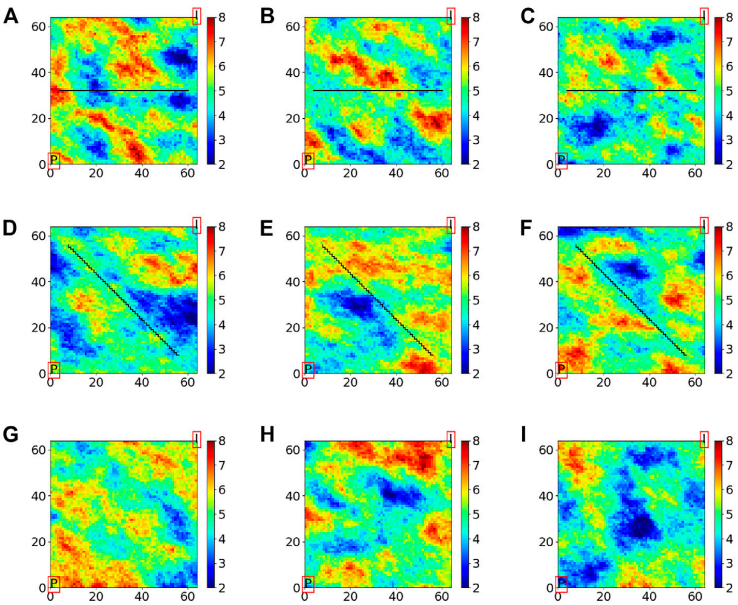


FIGURE 5 | Samples from the prior distributions for log permeability for the three experiments, **(A)–(C)** Experiment I, **(D)–(F)** Experiment II, and **(G)–(I)** Experiment III.

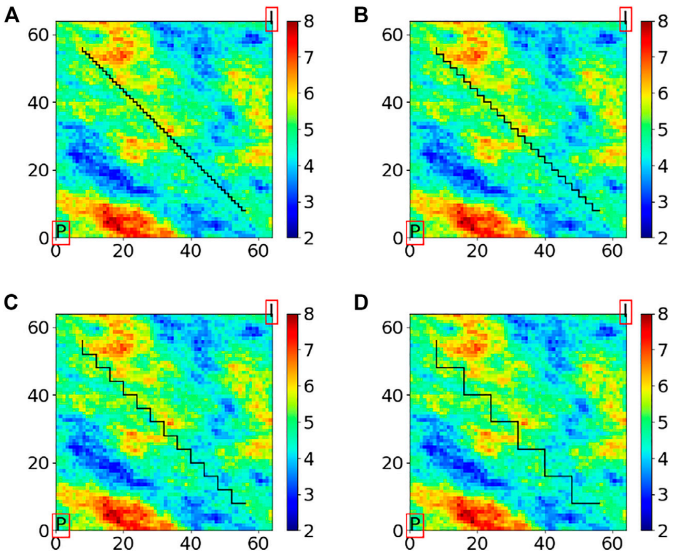


FIGURE 6 | Approximation of the fault for simulations, **(A)** original fault and **(B–D)** approximations at coarser levels.

TABLE 2 | Variogram used for observation data error; the unit for range is grid cells.

Variogram type	Range	Mean	Anisotropy ratio
Spherical	5	0	1

TABLE 3 | Specific characterizations of variograms of the prior distribution, the unit for range is grid cells.

	Range	Anisotropy ratio	Anisotropy angle
Experiment I	20	0.7	−30
Experiment II	30	0.5	−20
Experiment III	25	0.7	−30

TABLE 4 | Summary of resource allocation for the algorithms in Experiment I.

	Level 1	Level 2	Level 3	Level 4	Level 5
	$G_1 = 82$	$G_2 = 124$	$G_3 = 310$	$G_4 = 1060$	$G_5 = 4096$
	N_1	N_2	N_3	N_4	N_5
MLHES-NO	1,510	864	501	143	46
MLHES-MB	1,510	864	501	143	46
MLHES-ST	1,510	864	501	143	46
MLHES-DE	1,510	864	501	143	46
MLHES-TE	1,510	864	501	143	46
MLHES-EX	1,510	864	501	143	46
ES-LOC	–	–	–	–	100
ES-REF	–	–	–	–	10,000

TABLE 5 | Summary of resource allocation for the algorithms in Experiment II.

	Level 1	Level 2	Level 3	Level 4
	$G_1 = 124$	$G_2 = 310$	$G_3 = 1060$	$G_4 = 4096$
	N_1	N_2	N_3	N_4
MLHES-NO	1,404	652	170	40
MLHES-MB	1,404	652	170	40
MLHES-ST	1,404	652	170	40
MLHES-DE	1,404	652	170	40
MLHES-TE	1,404	652	170	40
MLHES-EX	1,404	652	170	40
ES-LOC	–	–	–	100
ES-REF	–	–	–	10,000

TABLE 6 | Summary of resource allocation for the algorithms in Experiment III.

	Level 1	Level 2	Level 3	Level 4
	$G_1 = 64$	$G_2 = 256$	$G_3 = 1024$	$G_4 = 4096$
	N_1	N_2	N_3	N_4
MLHES-NO	1,524	703	252	38
MLHES-MB	1,524	703	252	38
MLHES-ST	1,524	703	252	38
MLHES-DE	1,524	703	252	38
MLHES-TE	1,524	703	252	38
MLHES-EX	1,524	703	252	38
ES-LOC	–	–	–	100
ES-REF	–	–	–	10,000

from the east side of the field, and its western most point 4 grid cells away from the west side of the field. From **Figures 5A–C**, it is clear that flow from the injector to the producer has to pass through one of the narrow openings at the ends of the fault. Hence, there will be strong variations in the solution variables in these regions. Grid coarsening is therefore expected to produce stronger MLME for this example than for a similar example, but where no fault was present.

4.2 Reservoir Model II

An oblique fault stretching from 8 grid cells away from the northwest corner to 8 grid cells away from the southeast corner is added to the general reservoir model structure. In addition to the complexities similar to those associated with Reservoir Model I, as can be seen in **Figure 6**, the coarsening scheme in the presence of such a fault, which will be discussed in **Section 5.2**, results in some permeability values that are located on one side of the fault in the fine grid to contribute to an upscaled permeability value located on the other side of the fault in the coarsened grid. This introduces another type of MLME to the model.

4.3 Reservoir Model III

This reservoir model uses the general structure of the models without addition of faults and is used for investigation of a different type of MLME, which is formed by simplifying the grid coarsening scheme.

5 NUMERICAL INVESTIGATION

In order to assess the quality of the MLME correction schemes presented in this work, three experiments are conducted. The experiments are performed on the reservoir models discussed in **Section 4**.

The observation data are two sets of time-lapse bulk impedance data taken based on a baseline (day zero of production) and two vintages, which are different for each experiment and will be mentioned separately. These observation data are generated using the results of simulation from a random draw of the prior distribution. As the inverted seismic data are spatially correlated, we use a correlated covariance matrix for the data error. In doing so, a variogram with the specifications given in **Table 2** is considered. The marginal standard deviation of each observation value is given as

$$\sigma = r \max\{|D|, T\}, \tag{30}$$

where $r = 0.1$, D is the value of observation data at a certain location, and T is a threshold put to avoid too much certainty in the observation data whose absolute value is very small. This threshold is defined as the 1st smallest percentile of the absolute value of the observation data.

TABLE 7 | Experiment I: Mean of the elements of the median vector of NCR_l for different correction schemes.

	Level 1	Level 2	Level 3	Level 4
	$G_1 = 82$	$G_2 = 124$	$G_3 = 310$	$G_4 = 1060$
Mean bias	0.4256	0.9151	1.2069	2.5229
Stochastic	0.3503	0.4406	0.5384	0.4320
Deterministic	0.3395	0.4451	0.5473	0.4124
Telescopic	0.1240	0.2405	0.4097	0.4124

For each numerical experiment, we will compare plots of results (mean and variance fields) obtained with five versions of the MLHES to those obtained with ES with distance-based localization (ES-LOC). The five versions are MLHES with mean bias correction (MLHES-MB), MLHES with stochastic correction (MLHES-ST), MLHES with deterministic correction (MLHES-DE), MLHES with telescopic correction (MLHES-TE), and MLHES with no correction (MLHES-NO).

The gold standard (reference solution) for the comparison will be results obtained using ES with 10,000 ensemble members (ES-REF). By utilizing such an unrealistically large ensemble, we obtain results that are visually indistinguishable from the best results that can be achieved using ES.

Furthermore, we will show plots of results obtained with a scheme with exact correction for the MLME (MLHES-EX). These results are obtained by running fine-scale simulations with the same total ensemble size as for the multilevel simulations and thereafter upscaling model forecasts (with the appropriate sub-ensemble sizes) to the respective levels. Obviously, MLHES-EX is computationally much costlier than the rest of the MLHES variants, and it is included only to assess the effect of completely removing the MLME on posterior results. Finally, we will show plots of the log permeability realization used when generating the synthetic data ("truth").

A fixed computational power is considered for all runs (except for ES-REF and MLHES-EX). As the dominant cost of the DA process is pertaining to simulations of forward models, where iterative linear solvers dominate the computational costs for large problems, the computational cost pertaining to each ensemble member to be simulated using the forward model $\mathcal{M}_l$ is assumed to be proportional to G_l^{γ}, where G_l is the number of the active grid cells in the forward model at level l and $\gamma \in [1.25, 1.5]$ [22]. Here, we take $\gamma = 1.35$. Additionally, as usual for large-scale cases, the

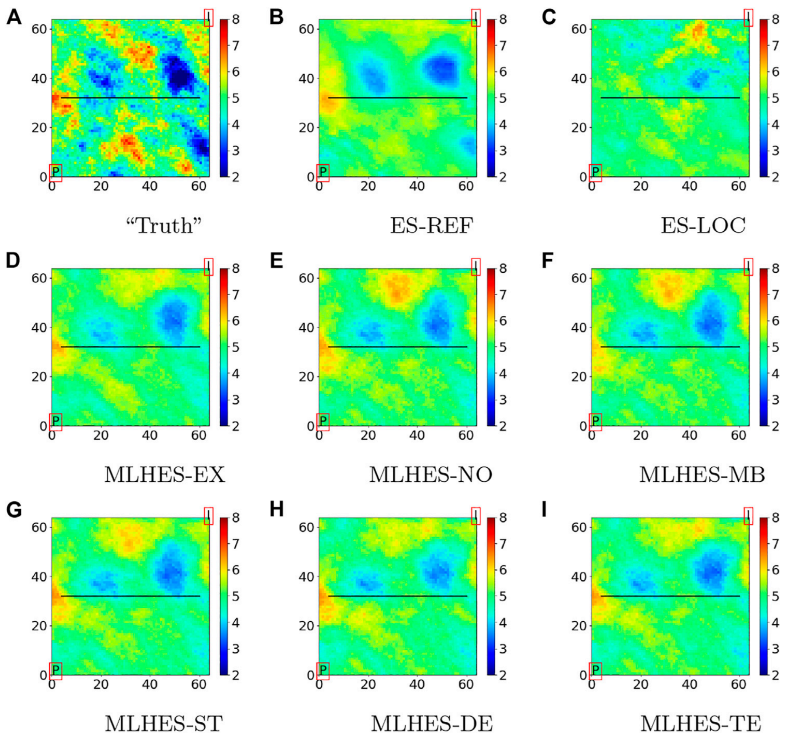


FIGURE 7 | Experiment I: Mean of posterior logarithmic permeability field.

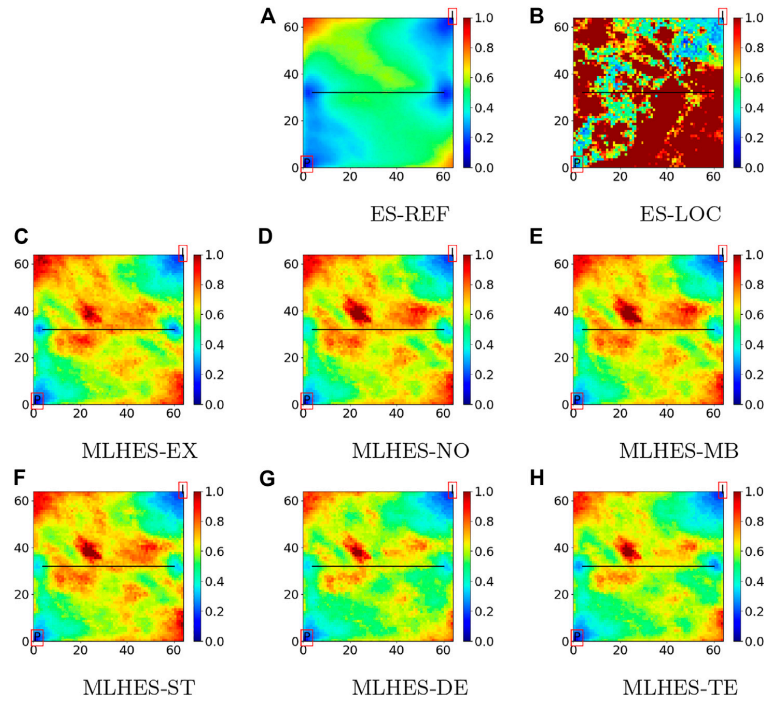


FIGURE 8 | Experiment I: Variance of posterior logarithmic permeability field.

ensemble size for standard single-level DA algorithms is set to be 100. Using this basis for calculations, the computational power allocated for all the runs would be equal if the following equation holds for all of them,

$$100G_L^{1.35} = \sum_{l=1}^{L} N_l G_l^{1.35}. \tag{31}$$

Considering this equation, we set N_l for different levels of the MLHES. There exist $L-1$ degrees of freedom for specification of the $\{N_l\}_{l=1}^{L}$. No optimization was performed for this specification; the only aim pursued was to keep the size of sub-ensembles ascending with a decrease in model fidelity. Several other similar settings that were tried resulted in similar DA outcomes.

For ES with distance-based localization, the tapering function for localization is based on Reference [23]. As for the MLHES, regarding weights of the hybrid mean and covariance matrices, $\{w_l\}_{l=1}^{L}$ in **Eqs. 1–3**, there is a possibility to improve the results by tuning the weights for specific cases, but here, we use the simplest choice—weights being all equal.

The unknown parameter vector in all the experiments is the logarithmic permeability field. The prior fields are based on three spherical variograms, all having mean 5 and variance 1, and specific characterizations given in **Table 3**. Samples from the prior distributions are given in **Figure 5**.

5.1 Experiment I

This experiment is conducted on Reservoir Model I with five levels corresponding to 82, 124, 310, 1,060, and 4,096 grid cells, respectively. A summary of the resource allocation for the different runs carried out in this experiment can be found in **Table 4**. The observation data for this experiment are generated based on seismic vintages at 4,000 and 8,000 days after production starts.

5.2 Experiment II

This experiment is conducted on Reservoir Model II. In this experiment, the presence of the oblique fault in the field interferes with coarsening the model. One way to handle this issue would be to avoid coarsening the grid around the fault area; however, this

TABLE 8 | Experiment II: Mean of the elements of the median vector of NCR_l for different correction schemes.

	Level 1	Level 2	Level 3
	$G_1 = 124$	$G_2 = 310$	$G_3 = 1060$
Mean bias	0.4097	0.7580	0.8284
Stochastic	0.3980	0.4622	0.4769
Deterministic	0.3716	0.4750	0.4381
Telescopic	0.1952	0.3461	0.4381

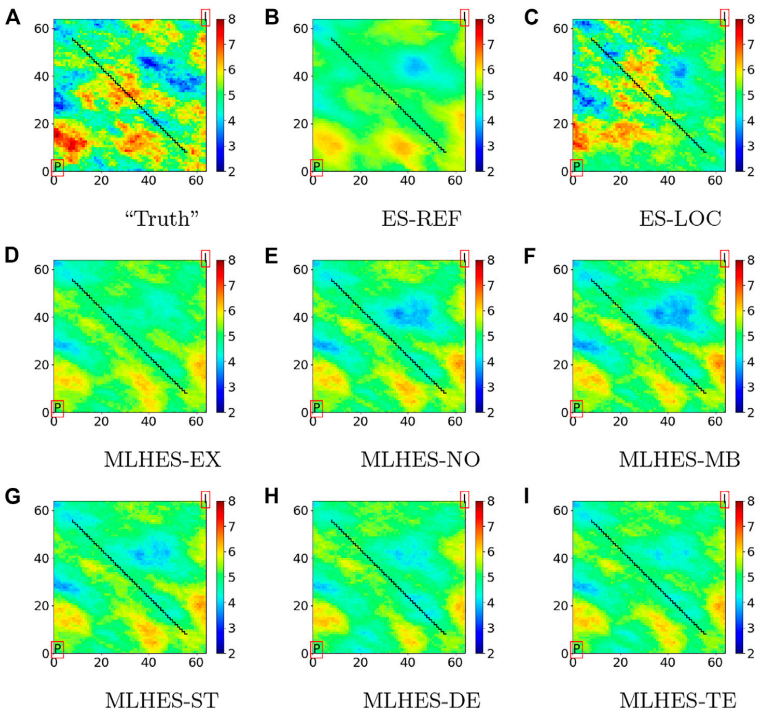


FIGURE 9 | Experiment II: Mean of posterior logarithmic permeability field.

would reduce the computational advantage of the multilevel scheme. In order to keep the grid coarsening as it is, the fault is approximated with bigger "zigzags," as depicted in **Figure 6**, for one realization of the logarithmic permeability field at different levels of coarseness. This makes the experiment to be more challenging than Experiment I, since in addition to coarsening the grid and upscaling the parameters, a structural characterization of the field (the fault) is also approximated.

MLHES is run with four levels corresponding to 124, 310, 1,060 and 4,096 grid cells, respectively. A summary of the resource allocation for the different algorithms carried out in this experiment can be found in **Table 5**. The observation data are generated based on seismic vintages at 5,000 and 10,000 days after beginning of production.

5.3 Experiment III

This experiment is conducted on Reservoir Model III. The coarsening of the grid is performed uniformly, so that also the regions near the wells are coarsened. Hence, a different type of MLME is generated. The number of grid cells is now complete powers of 2. Forming four levels of coarseness, the number of grid cells are 64, 256, 1,024, and 4,096. A summary of the resource allocation for the different algorithms carried out in this experiment can be found in **Table 6**. The observation data are generated based on seismic vintages at 4,000 and 8,000 days after production starts.

6 NUMERICAL RESULTS

The numerical results are assessed in two ways. First, we perform a quantitative analysis of the MLME-corrected model forecasts. Second, we perform a qualitative analysis of the results obtained when using the MLME-corrected forecasts in MLDA.

As for a quantitative analysis of success of MLME correction schemes, the normalized correction ratio for model forecasts at level l, NCR_l, defined as

$$NCR_l(Z) = \left|\left(Y_l - U_L^l Y_L\right) \oslash \left(\tilde{Y}_l - U_L^l Y_L\right)\right| = \left|\left(\varepsilon_{l,\times} - \zeta_l\right) \oslash \zeta_l\right|, \quad (32)$$

is considered. Here, $\oslash$ is the Hadamard division and $|^*|$ is the element-wise absolute value operator. If the correction scheme does not do any correction on a single element of NCR_l, it would result in that element to be equal to unity. Reduction in the error would result in the element moving toward zero, and an increase in the error would move that element toward infinity; hence, NCR_l is an indicator of the success of MLME correction schemes.

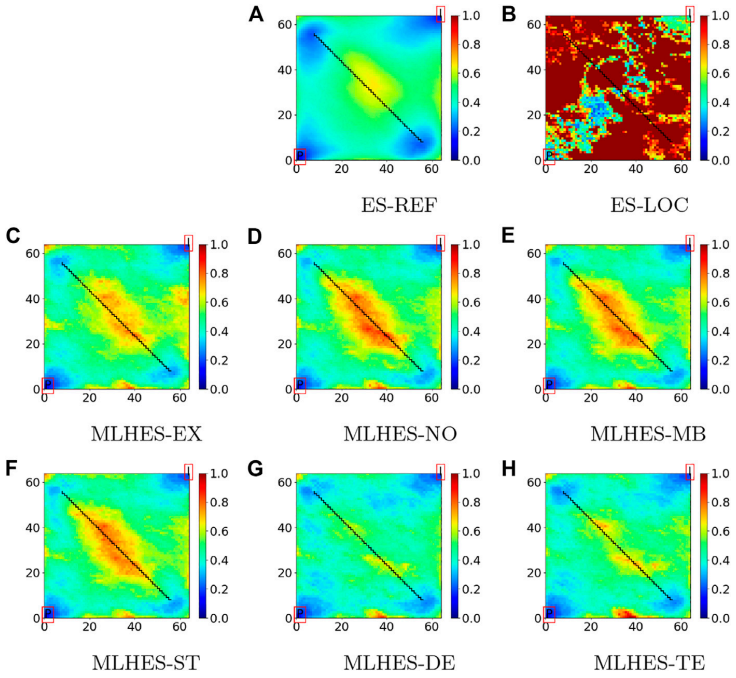


FIGURE 10 | Experiment II: Variance of posterior logarithmic permeability field.

NCR_l is different for different realizations. In order to assess the success of each of the correction schemes jointly for all realizations, the sample median of each of the elements of NCR_l is computed. The median is chosen since the mean of NCR_l is not a good indicator of success (NCR_l has a lower bound at zero but has no upper bound, and outliers would affect it disproportionately). Then, the mean of the elements of the sample median of NCR_l are reported for different levels in all MLME correction schemes for the three experiments.

As for the qualitative assessment of the DA results, the mean and the variance of the posterior unknown parameters are compared between different algorithms. We have not considered any specific formulation for computation of the final multilevel statistics. Accordingly, the simplest formulation is chosen, that is, reuniting all the sub-ensembles and treating them as one ensemble for computation of a posteriori mean and variance fields. ES-LOC was tested with several ranges for localization (critical distances), and the best results are presented for each of the experiments.

6.1 Results of Experiment I

As can be seen in **Table 7**, NCR_l is smaller in coarser models for all correction schemes. For a class of problems (including the problem considered here) where the model forecast can be seen as a spatially integrated response to a spatially varying parameter field, there exists a correlation between small-scale oscillations in the parameter domain and the nonlinearity strength of the mapping from parameter field to model forecast (see, e.g., [24, 25]). This correlation is such that coarsening the simulation grid and upscaling the associated parameters will generally result in weaker nonlinearity in the coarser forward models than the finer ones. The comparatively lower NCR_l in coarser levels than finer ones can be due to this decrease in nonlinearity by a decrease in l and also due to omission of local fluctuations in coarser model forecasts. In the case of the telescopic scheme, this can also be attributed to an increase in the ensemble size, which reduces the Monte Carlo errors associated with estimation of the MLME errors. All the schemes, except for mean bias correction, are, on average, successful in reduction of MLME. Telescopic correction for level 4 (level $L-1$ in general) reduces to deterministic correction, but in coarser levels, it has performed better than deterministic correction, which, in turn, performs slightly better than stochastic correction.

TABLE 9 | Experiment III: Mean of the elements of the median vector of NCR_l for different correction schemes.

	Level 1	Level 2	Level 3
	$G_1 = 124$	$G_2 = 310$	$G_3 = 1060$
Mean bias	0.6973	1.2892	2.7182
Stochastic	0.6112	0.6968	0.7517
Deterministic	0.7178	0.8782	0.8651
Telescopic	0.4268	0.6190	0.8651

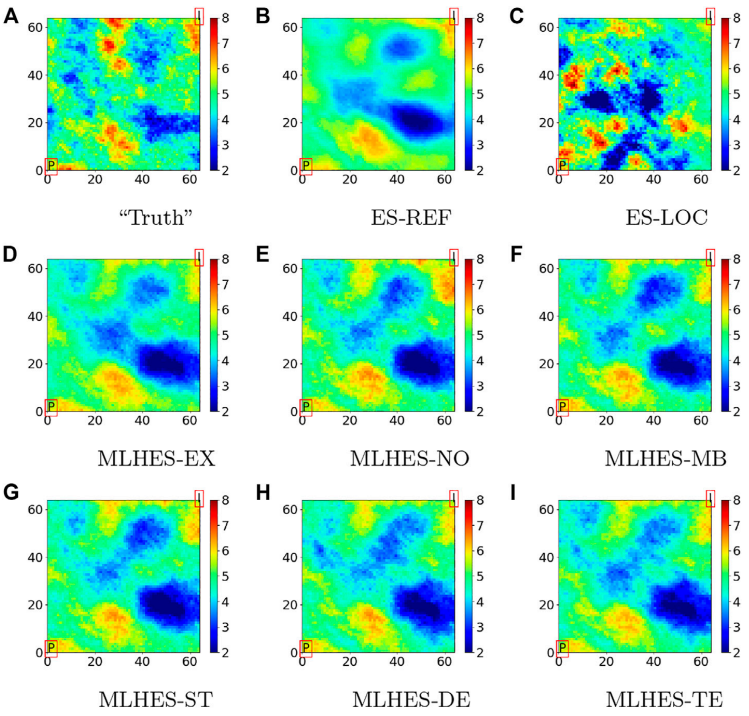


FIGURE 11 | Experiment III: Mean of posterior logarithmic permeability field.

Visual analysis of the mean permeability fields, given in **Figure 7**, shows that all MLHES variants are reasonably similar and more similar to ES-REF than ES-LOC is. This can be further confirmed by comparison of the variance fields given in **Figure 8**. The ES-LOC results presented here are based on the localization range of 40 grid cells. Additionally, no superiority of some MLME correction schemes over others is evident in visual assessment of the posterior mean and variance fields. The results from all MLHES variants are reasonably similar to the MLHES-EX results.

6.2 Results of Experiment II

Based on the trends in NCR_l, as can be seen in **Table 8**, the performance of the correction schemes has the same rank order as those of Experiment I, with telescopic correction showing the best performance, followed by deterministic, stochastic, and mean bias corrections. Visual analysis of the mean permeability fields, given in **Figure 9**, shows that all MLHES variants are reasonably similar and more similar to ES-REF than ES-LOC is. This can be further confirmed by comparison of the variance fields given in **Figure 10**. The ES-LOC results presented here are based on the localization range of 50 grid cells. Additionally, no superiority of some MLME correction schemes over others is evident in visual assessment of the posterior mean and variance fields. The results from all MLHES variants are reasonably similar to the MLHES-EX results.

6.3 Results of Experiment III

From **Table 9**, it is seen that NCR_l is comparatively higher in this experiment than the previous two experiments. The rank order of the performances stays the same, but the quality of correction has deteriorated for all the MLME correction schemes, except for the mean bias correction which has slightly improved.

Visual analysis of the mean permeability fields, given in **Figure 11**, shows that all MLHES variants are reasonably similar and more similar to ES-REF than ES-LOC. This can be further confirmed by comparison of the variance fields given in **Figure 12**. The ES-LOC results presented here are based on the localization range of 40 grid cells. Additionally, no superiority of some MLME correction schemes over others is evident in visual assessment of the posterior mean and variance fields. The results from all MLHES variants are reasonably similar to the MLHES-EX results.

7 SUMMARY AND CONCLUSION

With large amounts of simultaneous data, like inverted seismic data in reservoir modeling, negative effects of Monte Carlo errors in straightforward ensemble-based data assimilation (DA) are enhanced, typically resulting in underestimation of parameter uncertainties. Multilevel simulations utilize a selection of models for the same entity that constitute hierarchies both in fidelities and

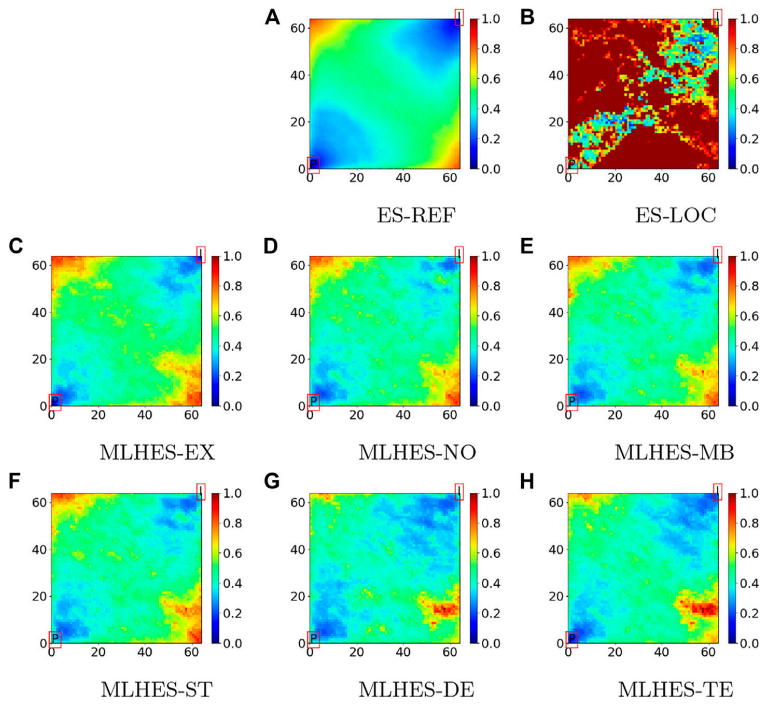


FIGURE 12 | Experiment III: Variance of posterior logarithmic permeability field.

computational costs. Multilevel data assimilation (MLDA) utilizes multilevel simulations in the forecast step. MLDA therefore renders the possibility of decreasing Monte Carlo errors without increasing the total computational cost, but MLDA will also introduce multilevel modeling errors (MLME) that are not present in conventional simulation results. The underlying assumption is therefore that the gain in reducing the Monte Carlo error is larger than the loss in introducing the MLME. If the MLME could be approximately accounted for, however MLDA performance could be further improved.

We have estimated and approximately accounted for the MLME. Four computationally inexpensive approximate MLME correction schemes have been considered. We have denoted these schemes mean bias correction, stochastic correction, deterministic correction, and telescopic correction. The three latter schemes have been developed in this work. The abilities of the four schemes to correct for the MLME have been assessed in two ways, utilizing numerical experiments with three selected reservoir models.

First, statistics for the normalized correction ratios for model forecasts at each level were compared. The results showed that the correction schemes were capable of reducing the MLME, but the amount of reduction depended on the case and on the level. In general, the MLME correction schemes were more successful in correcting the MLME in the coarser levels than the finer ones, with telescopic correction showing the best performance followed by deterministic correction, stochastic correction, and mean bias correction.

Second, we assessed the performances of the different MLME-corrected model forecasts in assimilation of inverted seismic data, using the multilevel hybrid ensemble smoother (MLHES). The resulting posterior mean and variance fields with and without MLME correction were visually compared to results obtained from conventional ensemble smoother (ES) with localization, utilizing results obtained with conventional ES with an unrealistically large ensemble size as the gold standard. It was found that MLHES with and without MLME correction outperformed conventional ES with localization.

The use of all four MLME correction schemes, and in fact also MLDA without MLME correction, mostly resulted in posterior parameter estimates with similar quality. For each example, we also ran a computationally much more costly MLDA variant where the MLME had been exactly accounted for in the model forecasts (termed MLHES-EX in **Sections 5–6**). No differences in quality between results obtained with MLHES-EX and results obtained with several of the computationally inexpensive MLDA variants were found in all three examples. We have run several examples in addition to those presented in the article. In none of these examples did the computationally inexpensive MLDA variants produce poor results, and straightforward MLHES (i.e., without any MLME correction) produced results of

similar quality. Altogether, these results indicate that the MLHES is reasonably robust toward MLMEs.

Further investigation concerning the robustness of MLHES with and without MLME correction can be conducted for additional reservoir models with different types of MLMEs than those considered here. On the other hand, the MLME correction techniques can be further developed. Their current versions address the MLME of spatially distributed data on their associated grid cells independently, that is, spatial correlations of the MLME are not considered. It would be interesting to consider spatial correlation of the MLME in future work.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MN is a Ph.D. student, and TB, KF, and TM are supervising him in his research. Accordingly, MN did the details of the research and most of writing with the help of his supervisors.

ACKNOWLEDGMENTS

The authors acknowledge financial support from the NORCE research project "Assimilating 4D Seismic Data: Big Data Into Big Models" which is funded by industry partners, Aker BP ASA, Equinor Energy AS, Lundin Energy Norway AS, Repsol Norge AS, Shell Global Solutions International B.V., Total E&P Norge AS, and Wintershall Dea Norge AS, as well as the Research Council of Norway (PETROMAKS2). We also thank Schlumberger for providing academic software licenses to ECLIPSE.

REFERENCES

1. Chen Y, and Oliver DS. Ensemble Randomized Maximum Likelihood Method as an Iterative Ensemble Smoother. *Math Geosci* (2012). 44(1):1–26. doi:10.1007/s11004-011-9376-z
2. Emerick AA, and Reynolds AC. Ensemble Smoother with Multiple Data Assimilation. *Comput Geosci* (2013). 55:3–15. doi:10.1016/j.cageo.2012.03.011
3. Luo X, Bhakta T, and Nævdal G. Correlation-based Adaptive Localization with Applications to Ensemble-Based 4d-Seismic History Matching. *SPE J* (2018). 23(02):396–427. doi:10.2118/185936-pa
4. Lorentzen RJ, Bhakta T, Grana D, Luo X, Valestrand R, and Nævdal G. Simultaneous Assimilation of Production and Seismic Data: Application to the Norne Field. *Comput Geosci* (2019). 12:1–14. doi:10.1007/s10596-019-09900-0
5. Gaspari G, and Cohn SE. Construction of Correlation Functions in Two and Three Dimensions. *Q J R Met. Soc.* (1999). 125(554):723–57. doi:10.1002/qj.49712555417
6. Chen Y, and Oliver DS. Cross-covariances and Localization for Enkf in Multiphase Flow Data Assimilation. *Comput Geosci* (2010). 14(4):579–601. doi:10.1007/s10596-009-9174-6
7. Emerick A, and Reynolds A. Combining Sensitivities and Prior Information for Covariance Localization in the Ensemble Kalman Filter for Petroleum Reservoir Applications. *Comput Geosci* (2011). 15(2):251–69. doi:10.1007/s10596-010-9198-y
8. He J, Sarma P, and Durlofsky LJ. Reduced-order Flow Modeling and Geological Parameterization for Ensemble-Based Data Assimilation. *Comput Geosciences* (2013). 55:54–69. doi:10.1016/j.cageo.2012.03.027
9. Tarrahi M, Elahi SH, and Jafarpour B. Fast Linearized Forecasts for Subsurface Flow Data Assimilation with Ensemble Kalman Filter. *Comput Geosci* (2016). 20(5):929–52. doi:10.1007/s10596-016-9570-7
10. Fossum K, and Mannseth T. Coarse-scale Data Assimilation as a Generic Alternative to Localization. *Comput Geosci* (2017). 21(1):167–86. doi:10.1007/s10596-016-9602-3
11. Hatfield S, Subramanian A, Palmer T, and Düben P. Improving Weather Forecast Skill through Reduced-Precision Data Assimilation. *Monthly Weather Rev* (2018). 146(1):49–62. doi:10.1175/mwr-d-17-0132.1
12. Fossum K, Mannseth T, and Stordal AS. Assessment of Multilevel Ensemble-Based Data Assimilation for Reservoir History Matching. *Comput Geosci* (2020). 24(1):217–39. doi:10.1007/s10596-019-09911-x
13. Hoel H, Law KJH, and Tempone R. Multilevel Ensemble Kalman Filtering. *SIAM J Numer Anal* (2016). 54(3):1813–39. doi:10.1137/15m100955x
14. de Moraes RJ, Hajibeygi H, and Jansen JD. A Multiscale Method for Data Assimilation. *Comput Geosciences* (2019). 12:1–18. doi:10.5194/egusphere-egu21-3880
15. Fossum K, and Mannseth T. "A Novel Multilevel Method for Assimilating Spatially Dense Data," in ECMOR XVI-16th European Conference on the Mathematics of Oil Recovery, Vol. 2018. London: European Association of Geoscientists & Engineers (2018). 1–12.
16. Nezhadali M, Bhakta T, Fossum K, and Mannseth T. "A Novel Approach to Multilevel Data Assimilation," in ECMOR XVII, Vol. 2020. London: European Association of Geoscientists & Engineers (2020). 1–13.
17. Evensen G. Accounting for Model Errors in Iterative Ensemble Smoothers. *Comput Geosci* (2019). 23(4):761–75. doi:10.1007/s10596-019-9819-z
18. Kaipio J, and Somersalo E. Statistical Inverse Problems: Discretization, Model Reduction and Inverse Crimes. *J Comput Appl Math* (2007). 198(2):493–504. doi:10.1016/j.cam.2005.09.027
19. Schlumberger L. *Eclipse Reservoir Simulation Software V2016* (2016). Technical Description Manual.
20. Batzle M, and Wang Z. Seismic Properties of Pore Fluids. *Geophys.* (1992). 57(11):1396–408. doi:10.1190/1.1443207
21. Fahimuddin A. *4D Seismic History Matching Using the Ensemble Kalman Filter (EnKF): Possibilities and challenges*. Ph.D. Thesis. The University of Bergen (2010). doi:10.2118/131453-ms
22. Axelsson O. *Iterative Solution Methods*. Cambridge: Cambridge University Press (1996).
23. Furrer R, and Bengtsson T. Estimation of High-Dimensional Prior and Posterior Covariance Matrices in Kalman Filter Variants. *J Multivar. Anal* (2007). 98(2):227–55. doi:10.1016/j.jmva.2006.08.003
24. Chavent C, and Liu J. Multiscale Parametrization for the Estimation of a Diffusion Coefficient in Elliptic and Parabolic Problems. *IFAC Proc Vol.* (1989). 22(4):193–202. doi:10.1016/s1474-6670(17)53542-9
25. Grimstad A-A, and Mannseth T. Nonlinearity, Scale, and Sensitivity for Parameter Estimation Problems. *SIAM J Sci Comput* (2000). 21(6):2096–113. doi:10.1137/s1064827598339104

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX A: MLHES PSEUDO-CODE

Algorithm 1: MLHES Algorithm

Define $\{\mathcal{M}_l\}_{l=1}^{L}$
Define $\{w_l\}_{l=1}^{L}$
for $l = 1, ..., L$ **do**
  for $j = 1, ..., N_l$ **do**
    $z_{l,j}^{pr} \leftarrow \mathcal{N}\left(\mathrm{E}(Z^{pr}), \mathrm{C}(Z^{pr})\right)$
    $\hat{y}_{l,j} \leftarrow \mathcal{M}_l(z_{l,j}^{pr})$
  end for
end for
for $l = 1, ..., L$ **do**
  for $j = 1, ..., N_l$ **do**
    $y_{l,j} \leftarrow \hat{y}_{l,j} + \varepsilon_{l,\times}\left(z_{l,j}^{pr}\right)$
  end for
end for
for $l = 1, ..., L$ **do**
  $\mathrm{E}_{\mathrm{ML}}(Y_l) \leftarrow \sum_{k=1}^{L} w_k U_k^l \,\mathrm{E}(Y_k)$
  $\mathrm{C}_{\mathrm{ML}}(Y_l) \leftarrow \sum_{k=1}^{L} w_k \{\mathrm{C}(U_k^l Y_k) + \left(\mathrm{E}(U_k^l Y_k) - \mathrm{E}_{\mathrm{ML}}(Y_l)\right) \left(\mathrm{E}(U_k^l Y_k) - \mathrm{E}_{\mathrm{ML}}(Y_l)\right)^T\}$
  $\mathrm{C}_{\mathrm{ML}}(Z, Y_l) \leftarrow \sum_{k=1}^{L} w_k \,\mathrm{C}(Z_k, U_k^l Y_k)$
  for $j = 1, ..., N_l$ **do**
    $d_{l,j} \leftarrow \mathcal{N}\left(U_L^l \mu_D, U_L^l \mathbb{C}_D {U_L^l}^T\right)$
    $z_{l,j}^{a} \leftarrow z_{l,j}^{pr} + \mathrm{C}_{\mathrm{ML}}(Z, Y_l)\left(\mathrm{C}_{\mathrm{ML}}(Y_l) + U_L^l \mathbb{C}_D {U_L^l}^T\right)^{-1} (d_{l,j} - y_{l,j})$
  end for
end for

Chapter 11

Paper C

Iterative multilevel assimilation of inverted seismic data

Computational Geosciences (2022) 26:241–262
https://doi.org/10.1007/s10596-021-10125-3

ORIGINAL PAPER

Iterative multilevel assimilation of inverted seismic data

Mohammad Nezhadali[1,2] · **Tuhin Bhakta**[2] · **Kristian Fossum**[2] · **Trond Mannseth**[2]

Received: 28 June 2021 / Accepted: 12 December 2021 / Published online: 16 February 2022

Abstract

In ensemble-based data assimilation (DA), the ensemble size is usually limited to around one hundred. Straightforward application of ensemble-based DA can therefore result in significant Monte Carlo errors, often manifesting themselves as severe underestimation of parameter uncertainties. Localization is the conventional remedy for this problem. Assimilation of large amounts of simultaneous data enhances the negative effects of Monte Carlo errors. Use of lower-fidelity models reduces the computational cost per ensemble member and therefore renders the possibility to reduce Monte Carlo errors by increasing the ensemble size, but it also adds to the modeling error. Multilevel data assimilation (MLDA) uses a selection of models forming hierarchies of both computational cost and computational accuracy, and tries to balance between Monte Carlo errors and modeling errors. In this work, we assess a recently developed MLDA algorithm, the Multilevel Hybrid Ensemble Smoother (MLHES), and introduce and assess an iterative version of this algorithm, the Iterative Multilevel Hybrid Ensemble Smoother (IMLHES). In our assessments, we compare these algorithms with conventional single-level DA algorithms with localization. To this end, a typical example of large amount of spatially distributed data, i.e. inverted seismic data, is considered and three data sets of this kind are assimilated in three different petroleum reservoir models. Qualitatively evaluating the DA outcomes, it is found that multilevel algorithms outperform their conventional single-level counterparts in obtaining the posterior statistics of both uncertain parameters and model forecasts. Additionally, it is observed that IMLHES performs better than MLHES in the same regard, and also successfully converges to the proximity of solution in a case where the considered iterative single-level algorithm did not converge to the global optimum.

Keywords Data assimilation · Multilevel methods · Iterative smoother · Ensemble-based history-matching

1 Introduction

Sound decision making in petroleum reservoir management depends on reliable production forecasts from reservoir models, including accurate estimates of uncertainty in the forecasts. The reliability is increased by utilization of available data for calibration of the models.

Ensemble-based Data Assimilation (DA) methods, using statistically correct formulations, have accordingly become popular for automated reservoir history-matching [7, 11, 25, 27, 30, 36, 37].

Monte Carlo approximations play a crucial role in ensemble-based DA. Due to computational-cost limitations, the ensemble size is limited to roughly one hundred. Using straightforward ensemble-based DA, the degrees of freedom of the problem would equal the ensemble size, and such an approach would result in significant Monte Carlo errors.

The negative effects of Monte Carlo errors are enlarged in the presence of large amounts of simultaneous data, e.g. inverted seismic data, resulting in underestimation of variance of the unknown parameters, and in more severe cases ensemble collapse. There have been several efforts on balancing the degrees of freedom of the problem and the amount of data [29, 31].

The most widely used treatment for Monte Carlo errors is distance-based localization [24]. The basic assumption underlying distance-based localization is that true correlations between a parameter and a datum decrease when the distance between their respective locations increase, and disappears if the distance exceeds a critical distance. This assumption may not always hold for subsurface problems. Different correlation functions and their utilization in DA can be found in [6, 9, 17]. A proper choice of correlation function, and the critical distance in

✉ Mohammad Nezhadali
mone@norceresearch.no

[1] University of Bergen, Nygårdsgaten 112, 5008 Bergen, Norway

[2] NORCE, Bergen, Norway

particular, depends on parameter and data types as well as on other problem settings. This reduces the robustness of distance-based localization, also for problems where its basic assumption does hold. Papers using ensemble-based methods for assimilation of seismic data [1, 10, 26], typically use localization methodologies developed originally for production data.

Simply increasing the ensemble size will of course reduce Monte-Carlo errors, but it will also increase computational cost. Utilization of a lower-cost and lower-fidelity model renders the possibility of increasing the ensemble size without increasing the total computational cost. Use of a lower-fidelity reservoir model will, however, introduce modeling errors in addition to those already present in conventional-fidelity simulation results. The underlying assumption when applying lower-fidelity models in DA is therefore that the gain in reducing Monte Carlo errors is larger than the loss in numerical simulation accuracy. DA using various types of lower fidelity models has been applied to several inverse problems, e.g., within petroleum reservoir modeling [13, 21, 40] and atmospheric science [20]. Note that since lower-fidelity simulations are applied to the forecast step and localization is applied to the analysis step, the two techniques can be combined, if desired.

Multilevel simulations utilize a selection of models for the same entity that constitute hierarchies in both fidelities and computational costs (multilevel models). The idea is to decrease Monte Carlo errors without increasing numerical errors too much. There are a number of ways to realize multilevel models. We choose to construct them by spatial coarsening of the conventional simulation grid to several levels of coarseness, and correspondingly upscale the associated grid-based parameter functions. Multilevel Data Assimilation (MLDA) [8, 14, 15, 22, 23, 32, 35] utilizes multilevel models in the forecast step. Since inverted seismic data are given on the conventional grid (denoted the fine grid from now on), MLDA with such data must be able to handle differences in grid levels between data and model forecasts.

As utilization of iterative methods helps to improve the quality of history matching in standard Ensemble Smoothers (ES) [7], a similar advancement in the domain of MLDA is possible. An MLDA smoother for assimilation of spatially distributed data, the Multilevel Hybrid Ensemble Smoother (MLHES), was developed in [32] and was assessed for assimilation of inverted seismic data. In this work we further investigate this algorithm and introduce an iterative version of MLHES, the Iterative Multilevel Hybrid Ensemble Smoother (IMLHES). We will also evaluate the performance of these algorithms in comparison with conventional DA algorithms, i.e. fine-level DA with localization, for assimilation of inverted seismic data in petroleum reservoir problems.

The rest of this paper is organized as follows. Section 2 is devoted to introducing some standard DA algorithms to establish a base for comparison. Section 3 introduces the MLDA algorithms. Section 4 explains the test models used for comparison of the performance of DA schemes. In Section 5 we describe the numerical investigations, which are followed by their results in Section 6. Finally, in Section 7 we summarize and conclude the paper.

2 Standard data assimilation schemes

Ensemble-based DA is a robust method for solving the parameter and state estimation problems using Bayesian methodology. We explain two widely used DA algorithms; Ensemble Smoother (ES), introduced in [41], and Iterative Ensemble Smoother (IES), introduced in [7]; both with localization [16], as conventional methods for DA.

2.1 Ensemble Smoother

Consider the prior ensemble $\{z_j^{pri}\}_{j=1}^{N_e}$, containing N_e realizations from the prior parameters random vector Z^{pri}. Their corresponding forecasts, $\{y_j\}_{j=1}^{N_e}$ (realizations from the random vector Y) are then obtained by running the forward model, $\mathcal{M}$, on each of the ensemble members as

$$y_j = \mathcal{M}(z_j^{pri}) \,. \tag{1}$$

Consequently, the empirical estimation of the mean and the covariance of the forecast random vector, $\mathrm{E}(Y)$ and $\mathrm{C}(Y)$, respectively, can be calculated as

$$\mathrm{E}(Y) = \frac{1}{N_e} \sum_{i=1}^{N_e} y_j \,, \tag{2}$$

$$\mathrm{C}(Y) = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} \big(y_j - \mathrm{E}(Y)\big) \big(y_j - \mathrm{E}(Y)\big)^T \tag{3}$$

Let d_j denote a realization of data to be assimilated, drawn from $D \sim \mathcal{N}(\mu_D, \mathbb{C}(D))$, where μ_D is the mean and $\mathbb{C}(D)$ is the data-error covariance. The linear-Gaussian assumption enables the possibility of formulating a closed form for the analysis step in the ES. Accordingly the analysis step for an arbitrary ensemble member z_j can be written as

$$z_j^a = z_j^{pri} + K(d_j - y_j) \,, \tag{4}$$

where K is the Kalman gain defined by

$$K = \mathrm{C}(Z, Y) \left(\mathrm{C}(Y) + \mathbb{C}(D)\right)^{-1} \,, \tag{5}$$

and $\mathrm{C}(Z, Y)$ is the empirical estimate of cross-covariance between parameters and model forecasts.

2.2 Iterative ensemble smoother

Iterative versions of ES are developed for improved performance of DA in nonlinear problems. Here we present the algorithm introduced in [7] with confined step length [34], and denote it Iterative Ensemble Smoother (IES) in the rest of this paper.

Writing the posterior logarithm of likelihood using Bayesian update equation and ignoring the constant terms, the objective function is given as [39]

$$\mathrm{J}(Z) = \frac{1}{2}||Z - \mathrm{E}(Z^{pri})||^2_{\mathrm{C}(Z^{pri})^{-1}} + \frac{1}{2}||\mathcal{M}(Z) - \mu_D||^2_{\mathbb{C}(D)^{-1}}. \tag{6}$$

The corresponding objective function for z_j is [7]

$$\mathrm{J}_j(z_j) = \frac{1}{2}||z_j - z_j^{pri}||^2_{\mathrm{C}(Z^{pri})^{-1}} + \frac{1}{2}||\mathcal{M}(z_j) - d_j||^2_{\mathbb{C}(D)^{-1}}. \tag{7}$$

In the linear-Gaussian case J_j is minimized by the update in Eq. 4. However, assuming nonlinearity in $\mathcal{M}$, this will not be the case. Accordingly, the Gauss-Newton scheme with confined step length is employed to minimize J_j for all realizations.

The Gauss-Newton scheme requires the gradient and the Hessian of J_j for each of the ensemble members. In this approach, these are computed using approximations based on the ensemble. Accordingly, the gradient, $\nabla \mathrm{J}_j^i$, and the Hessian, H^i, for the parameter vector realizations at iteration i, z_j^i, are given as

$$\nabla \mathrm{J}_j^i \approx \mathrm{C}(Z^{pri})^{-1}\left(z_j^i - z_j^{pri}\right) + \mathrm{M}^{i\,T}\mathbb{C}(D)^{-1}\left(\mathcal{M}(z_j^i) - d_{l,j}\right), \tag{8}$$

$$\mathrm{H}^i \approx \mathrm{C}(Z^{pri})^{-1} + \mathrm{M}^{i\,T}\mathbb{C}(D)^{-1}\,\mathrm{M}^i, \tag{9}$$

respectively. In these formulae, M^i denotes the approximation to Jacobian of $\mathcal{M}(Z)$, and H^i is approximated by neglecting the derivative of M^i. $\mathrm{M}^{i\,T}$ denotes the transpose (or more generally, the adjoint) of M^i. In IES [7, 19], the approximation,

$$\mathrm{M}^i \approx \mathrm{C}\left(Y^i, Z^i\right)\mathrm{C}\left(Z^i\right)^+, \tag{10}$$

is used for calculation of M^i, where superscript $+$ denotes Moore-Penrose pseudo inverse. A mathematical justification for this can be found in [36, Theorem 1].

The realization z_j^i is then updated as

$$z_j^{i+1} = z_j^i - \beta\left(\mathrm{H}^{i\,-1}\nabla \mathrm{J}_j^i\right), \tag{11}$$

where β is the step length which is defined based on "restricted-step algorithm" in [34]. The update equation can be written as

$$z_j^{i+1} = z_j^i + \beta\left(\Delta_j^{i,pri} + \Delta_j^{i,lik}\right), \tag{12}$$

where $\Delta_j^{i,pri}$ and $\Delta_j^{i,lik}$ are given by

$$\Delta_j^{i,pri} = (I_Z - K^i\,\mathrm{M}^i)[z_j^{pri} - z_j^i], \tag{13}$$

$$\Delta_j^{i,lik} = K^i[d_j - \mathcal{M}(z_j^i)]. \tag{14}$$

In these Equations, I_Z is the identity matrix of the parameter vector dimension, and the Kalman gain K^i is given by

$$K^i = \mathrm{C}(Z^{pri})\mathrm{M}^{i\,T}\left(\mathrm{M}^i\,\mathrm{C}(Z^{pri})\mathrm{M}^{i\,T} + \mathbb{C}(D)\right)^{-1}. \tag{15}$$

The iterations are continued until convergence is obtained.

3 Multilevel data assimilation

In MLDA the forecast step is performed using a set of models which have different costs and fidelities. Here, we define $\mathcal{M}_L := \mathcal{M}$, and $\{\mathcal{M}_l\}_{l=1}^{L-1}$ are approximations to $\mathcal{M}_L$ with increasing accuracy and computational cost as l increases. We will denote $\{\mathcal{M}_l\}_{l=1}^{L}$ a multilevel model. After sampling from the prior distribution, the ensemble of prior state vectors is divided into L sub-ensembles. Each of the sub-ensembles are modeled using its corresponding forward model, as seen in Fig. 1, where subscript l denotes the sub-ensemble index.

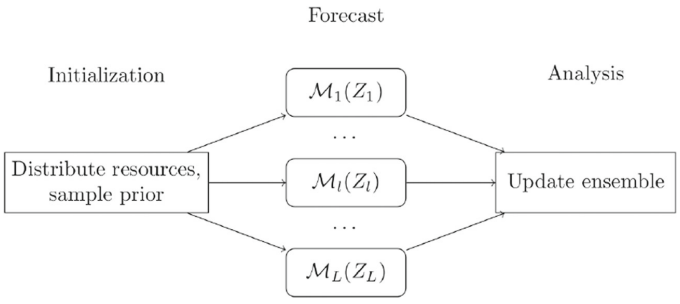


Fig. 1 Representation of a single forecast step of MLDA algorithms

3.1 Multilevel models

Multilevel models will form hierarchies of both accuracy and computational cost.

One can think of several schemes to devise the hierarchy, including but not limited to coarsening the spatial grid of the forward model, coarsening the temporal grid of the forward model, and relaxing the convergence criteria in the iterative linear solvers. All of these methods reduce the computational cost of the models and increase their numerical error. Coarsening the spatial grid and performing simulations on such grids is chosen for the current work (Note, however, that the parameters that we invert for are kept in the fine grid, meaning that upscaling the parameters is considered as part of the multilevel forward models). The techniques presented in this work are, however, robust enough so that with minor manipulations, they can be used for other lower fidelity models.

As for coarsening the grid of the forward models, [14] proposed a robust technique, which was also used in [32]. This technique occurs in a sequence of steps. In each step, neighboring cells of the grid at the previous step are merged into a coarser cell unless they are to be kept fine deliberately. A representation of the grid coarsening process for an 8×8 sample grid can be found in Fig. 2. As it can be seen in the figure, coarsening has occurred in a uniform manner across both directions, except for the vicinity of two opposite corners, where the grid cells are kept in fine scale to boost the local numerical accuracy around the two wells, producer (P) and injector (I). The aim is that the grid coarsening does not change the physics of the problem too much.

3.2 Transformation of model forecasts

The discrepancy in coarseness of the multilevel grids results in the spatially distributed model forecasts to be in different resolutions for different levels. Therefore, in order to be able to compute the multilevel sample statistics of model forecast, a robust transformation scheme should be devised for converting a model forecast from the resolution at one level to another.

In the problem at hand, transformation of the model forecast requires either upscaling or downscaling. Standard volume-weighted arithmetic averaging technique is used for upscaling.

Downscaled model forecasts are simply put equal to the corresponding coarse grid values. Accordingly, both upscaling and downscaling are linear transformations of model forecasts. Hence, we define a set of linear transformations, $\{U_f^c : \mathbb{R}^{\zeta_f} \mapsto \mathbb{R}^{\zeta_c} | 1 \leq c, f \leq L\}$, where ζ_f and ζ_c denote the dimension of model forecast vector at arbitrary levels f and c, respectively, and U_f^c transforms the model forecast vector from level f to be compatible with level c.

Figure 3 gives two examples of transformation of spatially distributed model forecast, one from a coarser grid to a finer grid, and one vice versa. Each model forecast component is represented in its corresponding spatial grid cell. As can be seen in Fig. 3, in the upscaling procedure, the arbitrarily named model forecast components $\{a_i\}_{i=1}^4$ in the northwest zone from the finer grid (level f), are averaged to form their corresponding model forecast component, $\bar{a}$, in the coarsened grid. Similar procedure has been performed for the rest of model forecast components, shown by *. In the downscaling procedure, on the other hand, the model forecast components in the coarse grid are simply copied to their corresponding components of the finer grid.

3.3 Upscaling of observation data

As part of the DA process, the mismatch between the model forecasts and observation data needs to be calculated. Here, it is assumed that inverted seismic data is given in the resolution of the finest simulation grid. Accordingly, for each of the levels, either the observation data should be upscaled to the resolution of model forecast or the model forecast should be downscaled to the resolution of the observation data. In this paper, we take the former approach.

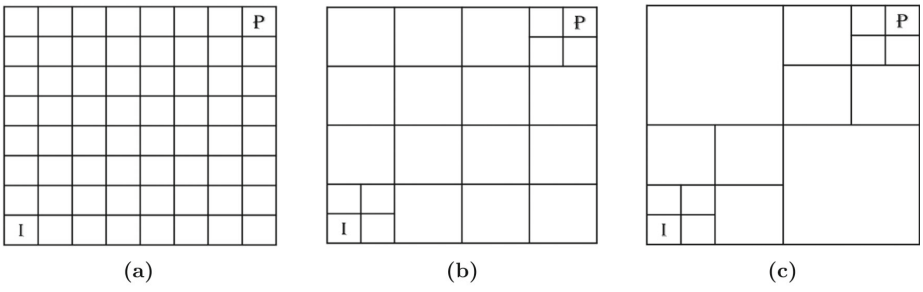


Fig. 2 Grid coarsening proposed by [14] performed on an 8×8 grid (**a**) Finest level (**b, c**) Coarser levels

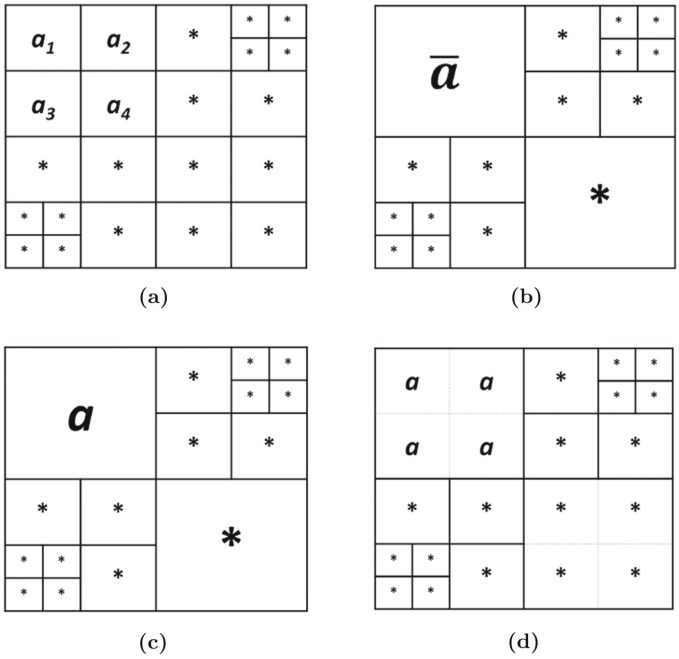


Fig. 3 Transformation of model forecast between two levels f (finer) and c (coarser) (**a**) model forecast in resolution of level f (**b**) transformation of model forecast from resolution of level f to the resolution of level c (**c**) model forecast in resolution of level c (**d**) transformation of model forecast from resolution of level c to the resolution of level f

Since the observation data is in the resolution of the finest model, using the same transformation functions as those designed for model forecasts on the fine observation data will result in upscaling of observation data into the preferred resolution. Accordingly, the transformed random vector of observation data at level l is given as

$$D_l = U_L^l D\,. \tag{16}$$

3.4 Multilevel statistics

Assuming we have approximations of the model forecasts, Y, being a function of the unknown parameter vector, Z, on several levels, a statistically correct scheme for approximation of multilevel statistics for Y is required. As for MLDA, the mean and the covariance of model forecast are of foremost interest. Accordingly, formulations for these multilevel statistics are proposed.

Assuming the model with the highest fidelity, $\mathcal{M}_L$, to be exact, [22] proposed an unbiased formulation for approximation of multilevel statistics for DA under certain conditions. Under these set of conditions, the proposed method outperformed its alternatives [15]. For reservoir problems, however, these conditions typically do not hold, and another formulation inspired by Bayesian Model Averaging (BMA) was proposed [15]. In this formulation the statistics are computed based on reliability weights w_l for each of the levels l. This formulation is, by definition, a biased scheme for computation of multilevel moments; however, it will be a useful technique for problems in which variance error dominates bias, which is often the case for petroleum reservoir problems [15]. Using this formulation and transformations of the forecast, [32] proposed a formulation of multilevel statistics for spatially distributed model forecasts which will be used in the current work. According to this scheme, the multilevel mean of the model forecast at level l is given as

$$\mathrm{E}_{\mathrm{ML},l}(Y) = \sum_{k=1}^{L} w_k U_k^l \,\mathrm{E}(Y_k)\,, \tag{17}$$

$$\sum_{k=1}^{L} w_k = 1\,, \tag{18}$$

where $\mathrm{E}(Y_k)$ denotes sample mean of the model forecast at level k. Using the law of total variance, the multilevel approximation of covariance of the model forecast at level l is formulated as

$$\mathrm{C}_{\mathrm{ML},l}(Y) = \sum_{k=1}^{L} w_k \{\mathrm{C}(U_k^l Y_k) + \Big(\mathrm{E}(U_k^l Y_k) - \mathrm{E}_{\mathrm{ML},l}(Y)\Big) \Big(\mathrm{E}(U_k^l Y_k) - \mathrm{E}_{\mathrm{ML},l}(Y)\Big)^T\}\,. \tag{19}$$

In addition, the parameter-forecast cross-covariance can be written as

$$\mathrm{C}_{\mathrm{ML},1}(Z,Y)=\sum_{k=1}^{L} w_k\{\mathrm{C}(Z_k,U_k^lY_k)+(\mathrm{E}(Z_k)-\mathrm{E}_{\mathrm{ML}}(\mathrm{Z}))\left(\mathrm{E}(U_k^lY_k)-\mathrm{E}_{\mathrm{ML},1}(Y)\right)^T\}, \quad (20)$$

where $\mathrm{E}_{\mathrm{ML}}(\mathrm{Z})$ is the multilevel formulation of the parameter-vector mean. This statistic is formulated using the same weights as in forecasts multilevel statistics, but since the parameters are in the same resolution for all levels, no transformation is needed for formulating it,

$$\mathrm{E}_{\mathrm{ML}}(\mathrm{Z})=\sum_{k=1}^{L} w_k\,\mathrm{E}(Z_k)\,. \quad (21)$$

Similarly, the multilevel covariance of the parameter vector is defined as

$$\mathrm{C}_{\mathrm{ML}}(\mathrm{Z})=\sum_{k=1}^{L} w_k\{\mathrm{C}(Z_k)+(\mathrm{E}(Z_k)-\mathrm{E}_{\mathrm{ML}}(\mathrm{Z}))\,(\mathrm{E}(Z_k)-\mathrm{E}_{\mathrm{ML}}(\mathrm{Z}))^T\}. \quad (22)$$

If all the Z_k, $1 \le k \le L$, share the same probability distribution function, which is the case in the first iteration of DA algorithms to be discussed in Section 3.5, Equations 21 and 22 will reduce to $\mathrm{E}(Z)$ and $\mathrm{C}(Z)$, respectively.

3.5 Multilevel data assimilation algorithms

Multilevel Hybrid Ensemble Smoother (MLHES) was proposed in [32] and tested on a petroleum reservoir model. Here, we firstly explain MLHES algorithm. Following that, we develop an iterative MLDA algorithm based on MLHES; i.e. IMLHES. The coarse model forecasts in both of these algorithms entail Multilevel Modeling Error (MLME), defined as the discrepancy between the upscaled fine model forecasts of a certain realization and the coarse model forecast of that realization. Several methods for correction of this error were assessed in [33]. Here, we utilize mean bias correction for addressing the MLME, which was also used in [32].

3.5.1 Multilevel hybrid ensemble smoother

Based on the decision on resource allocation, N_l ensemble members are simulated using $\mathcal{M}_l$. Running the forward simulator for every prior realization $z_{l,j}^{pri}$, where $1 \le l \le L$ and $1 \le j \le N_l$, we have

$$\widehat{y}_{l,j}=\mathcal{M}_l(z_{l,j}^{pri})\,, \quad (23)$$

where $\widehat{y}_{l,j}$ is the model forecast pertaining to realization $z_{l,j}^{pri}$. It should be mentioned that $z_{l,j}^{pri}$ vectors are essentially i.i.d samples from the prior distribution regardless of l and have the same dimension. The subscript l is used to denote that the forecasts pertaining to this realization are modeled using $\mathcal{M}_l$. The models forecasts, $\widehat{y}_{l,j}$, however, are in different resolutions for different l, and accordingly have different dimensions.

Next, the multilevel simulations are corrected for their mean bias by

$$y_{l,j}=\widehat{y}_{l,j}+\left(U_L^l\,\mathrm{E}(\widehat{Y}_L)-\mathrm{E}(\widehat{Y}_l)\right)\,. \quad (24)$$

This correction will help to account for part of the modeling error associated with coarser models, and results in the forecast mean to be unbiased.

In MLHES, multilevel approximation of the mean and covariance of the model forecasts are utilized for computation of the Kalman gain, and since the model forecasts are in different resolutions for different levels, this is done separately for each level. Accordingly, the updated parameters vector of an arbitrary ensemble member at level l is given by

$$z_{l,j}^{a}=z_{l,j}^{pri}+K_l(d_{l,j}-y_{l,j})\,, \quad (25)$$

where the observation data sample, $d_{l,j}$, is a random pick from $\mathcal{N}(U_L^l\mu_D, U_L^l\mathbb{C}(D){U_L^l}^T)$, and the level-specific Kalman gain, K_l, is given as

$$K_l=\mathrm{C}_{\mathrm{ML},1}(Z^{pri},Y)\left(\mathrm{C}_{\mathrm{ML},1}(Y)+U_L^l\mathbb{C}(D)(U_L^l)^T\right)^{-1}\,. \quad (26)$$

Here, the multilevel forecast covariance and the parameter-forecast cross-covariance are calculated by Eqs. 19 and 20, respectively.

A pseudo-code of MLHES is presented in Appendix A.

3.5.2 Iterative multilevel hybrid ensemble smoother

In order to be able to handle more nonlinear cases, an iterative version of MLHES is developed. In this algorithm, like with the MLHES, after the forecast step, we correct for the multilevel mean bias. To ease the notations, we consider the bias-correction as part of the new-defined forward models $\{\widetilde{\mathcal{M}}_l\}_{l=1}^{L}$ given as

$$\widetilde{\mathcal{M}}_l(Z)=\mathcal{M}_l(Z)+\left(U_L^l\,\mathrm{E}(\widehat{Y}_L)-\mathrm{E}(\widehat{Y}_l)\right)\,. \quad (27)$$

Similarly to how (6) was obtained in Section 2.2, we can write the objective function for level l as

$$\mathrm{J}_l(Z) = \frac{1}{2}||Z - \mathrm{E}(Z^{pri})||^2_{\mathrm{C}(Z^{pri})^{-1}} + \frac{1}{2}||\widetilde{\mathcal{M}}_l(Z) - U_L^l \mu_D||_{(U_L^l \mathbb{C}(D) U_L^{l\,T})^{-1}} . \tag{28}$$

Randomizing this objective function using the ensemble of realizations pertaining to level l, the objective function for realization j in sub-ensemble l can be written as

$$\mathrm{J}_{l,j}(z_{l,j}) = \frac{1}{2}||z_{l,j} - z_{l,j}^{pri}||^2_{\mathrm{C}(Z^{pri})^{-1}} + \frac{1}{2}||\widetilde{\mathcal{M}}_l(z_{l,j}) - d_{l,j}||_{(U_L^l \mathbb{C}(D) U_L^{l\,T})^{-1}} . \tag{29}$$

In order for minimization of $\mathrm{J}_{l,j}$ for all the realizations, the Gauss-Newton scheme with confined step length is employed. Accordingly, the gradient, $\nabla \mathrm{J}_{l,j}^i$, and the Hessian, H_l^i, for the parameter vector realizations at iteration i, $z_{l,j}^i$, are defined by

$$\nabla \mathrm{J}_{l,j}^i = \mathrm{C}(Z^{pri})^{-1}\left(z_{l,j}^i - z_{l,j}^{pri}\right) + \mathrm{M}_l^{i\,T}\left(U_L^l \mathbb{C}(D) U_L^{l\,T}\right)^{-1}\left(\widetilde{\mathcal{M}}_l(z_{l,j}^i) - d_{l,j}\right), \tag{30}$$

$$\mathrm{H}_l^i \approx \mathrm{C}(Z^{pri})^{-1} + \mathrm{M}_l^{i\,T}\left(U_L^l \mathbb{C}(D) U_L^{l\,T}\right)^{-1} \mathrm{M}_l^i , \tag{31}$$

respectively. Here, M_l^i denotes the ensemble approximation to Jacobian of $\widetilde{\mathcal{M}}_l$ at iteration i, and H_l^i is approximated by neglecting the derivative of M_l^i.

If the same procedure that was used in Section 2.2 to obtain (10) was followed for $\mathcal{M}_l$, the approximation,

$$\mathrm{M}_l^i \approx \mathrm{C}\left(Y_l^i, Z_l^i\right)\mathrm{C}\left(Z_l^i\right)^+ , \tag{32}$$

would be obtained. However, in order to be able to use the information also from other levels for approximation of M_l^i, the multilevel formulation of the parameter-forecast cross-covariance is utilized instead of its single-level formulation,

$$\mathrm{M}_l^i \approx \mathrm{C}_{\mathrm{ML},l}\left(Y^i, Z^i\right)\mathrm{C}_{\mathrm{ML}}(\mathrm{Z}^i)^+ . \tag{33}$$

The update formula for an arbitrary realization j of sub-ensemble l at iteration i, $z_{l,j}^i$, can be written as

$$z_{l,j}^{i+1} = z_{l,j}^i - \beta_l\left(\mathrm{H}_l^{i^{-1}} \nabla \mathrm{J}_{l,j}^i\right) , \tag{34}$$

where β_l is the step length at level l which is updated as in the restricted step algorithm, [34], at every iteration. The update equation can be written as

$$z_{l,j}^{i+1} = z_{l,j}^i + \beta_l\left(\Delta_{l,j}^{i,pri} + \Delta_{l,j}^{i,lik}\right) , \tag{35}$$

where by using Woodbury matrix lemma, $\Delta_{l,j}^{i,pri}$ and $\Delta_{l,j}^{i,lik}$ are given by

$$\Delta_{l,j}^{i,pri} = (I_Z - K_l^i \mathrm{M}_l^i)[z_{l,j}^{pri} - z_{l,j}^i] , \tag{36}$$

$$\Delta_{l,j}^{i,lik} = K_l^i[d_{l,j} - \widetilde{\mathcal{M}}(z_{l,j}^i)] , \tag{37}$$

and the level-specific Kalman gain, K_l^i, is obtained by

$$K_l^i = \mathrm{C}\left(Z^{pri}\right)\mathrm{M}_l^{i\,T}\left(\mathrm{M}_l^i \mathrm{C}(Z^{pri})\mathrm{M}_l^{i\,T} + U_L^l \mathbb{C}(D)(U_L^l)^T\right)^{-1} . \tag{38}$$

The iterations are then separately performed for each of the levels until convergence is obtained for all of them.

A pseudo-code of the IMLHES algorithm is presented in Appendix B.

4 Test models

Three different reservoir models are set up to assess the algorithms performances. These reservoir models have some shared properties. They are two-dimensional with Cartesian grids. For all of them, compressible two-phase flow (oil and water), no-flow boundary conditions, and standard equations for capillary pressure and relative permeability are considered. A description of the other shared general properties of the reservoir models is given in Table 1. Unique features of the reservoir models are explained separately in Sections 4.1–4.3.

The forward models used for forecasting each consist of two segments. A reservoir flow model is used to predict the state variables in time, and a petro-elastic model is utilized for computing the elastic rock properties from parameters and predicted state variables.

The flow segment of the forward model is derived by substitution of the Darcy's law into the mass conservation equation for each of the phases, resulting in [3]

$$\nabla.\left[\frac{k_{ro}}{\nu_o B_o} k(\nabla p_o - \rho_o g \nabla z)\right] = \frac{\partial}{\partial t}\left(\frac{\phi S_o}{B_o}\right) + q_o , \tag{39}$$

$$\nabla.\left[\frac{k_{rw}}{\nu_w B_w} k(\nabla p_w - \rho_w g \nabla z)\right] = \frac{\partial}{\partial t}\left(\frac{\phi S_w}{B_w}\right) + q_w , \tag{40}$$

where

$$S_o + S_w = 1 , \tag{41}$$

$$p_{cow} = p_o - p_w . \tag{42}$$

Table 1 Shared properties of the reservoir models

Fine cell dimensions:	$30 \times 30 \times 30$ (m^3)	Porosity:	0.2
Initial Oil saturation:	0.85	Initial Pressure:	200 bar

In these Equation, k denotes absolute permeability, and k_{r*} denotes the relative permeability of the corresponding phase. k_{r*} is a function of saturation of that phase, S_*. The pressure of a phase is denoted by p_*, and the capillary pressure, p_{cow}, is a function of S_w. Furthermore, g denotes the gravitational constant; v_*, B_*, and ρ_* are the viscosity, the formation volume factor, and density of their corresponding phases; and q_* denotes the sink or source term of its corresponding continuity equation.

The flow segment of the forward models is performed using Eclipse-100 [38]. Coarsening the grid is done by using the Eclipse keyword COARSEN, which merges groups of pre-defined neighboring cells to form a coarser grid. The upscaling of permeabilities is performed using pore-volume weighted arithmetic averaging, and transmissibilities between two neighboring coarse cells in each direction are calculated based on harmonic averaging in that direction and summing it in other directions [38].

As for the petro-elastic segment of the forward model, an in-house model based on standard rock-physics [4], [12, Report 1] was used.

4.1 Reservoir model I

This model has a 40 × 40 grid, and two wells, one producer (P) at southwest corner and one injector (I) at northeast corner. Both of the wells are pressure-controlled, the injector at 275 bar and the producer at 100 bar.

4.2 Reservoir model II

This model has a 64 × 64 grid, and two wells, one producer (P) at southwest corner, and one injector at northeast corner (I). Additionally, an oblique fault stretching from 8 grid cells distant from the northwest corner to 8 grid cells distant from southeast corner is added to the general reservoir model structure. As can be seen in Fig. 5, the coarsening scheme in the presence of such a fault, which will be discussed in Section 5.2, results in some permeability values that are located on one side of the fault in the fine grid to contribute to an upscaled permeability value located on the other side of the fault in the coarsened grid.

4.3 Reservoir model III

This model has a 70 × 70 grid and three wells, one injection well (I) in southeast corner and two producers (P1 and P2) in southwest and northwest corners. All the wells are pressure-controlled, I at 300 bar and P1 and P2 at 110 bar. This model has three zones, each of which contains one of the wells and has its own variogram for the permeability field, and there exist a smooth transition from one zone to another.

5 Numerical investigation

In order to compare the quality of the multilevel algorithms presented in this work with the standard DA methods, three experiments are conducted. Each experiment is performed on one of the three reservoir models discussed in Section 4.

The unknown parameter fields in all the experiments are logarithmic permeability fields, which have different distributions in each of the experiments.

The observation data are two sets of time-lapse bulk-impedance data taken based on a baseline (day zero of production) and two vintages, which are different for each experiment and will be described separately. These observation data are generated using the results of simulation of a random draw from the prior parameter distribution. As inverted seismic data typically are spatially correlated, we use a correlated covariance matrix for the data error. In doing so, a variogram with the specifications given in Table 2 is considered. The marginal standard deviation of each observation value is given as

$$\sigma = r \ \max\{|\delta|, \eta\}\,, \tag{43}$$

where $r = 0.1$, δ is the value of observation data at a certain location, and η is a threshold put to avoid too much certainty in the observation data whose absolute values are very small. This threshold is defined as the 1^{st} smallest percentile of the absolute value of the observation data.

For each numerical experiment we will compare plots of results obtained with multilevel algorithms, MLHES and IMLHES, with the results obtained from their standard DA counterparts, ES with localization (ES-LOC) and IES with localization (IES-LOC). The localization scheme used in ES-LOC and IES-LOC, is a distance-based localization based on the covariance structure given in [16] and spherical model for variogram.

The gold standards (reference solutions) for the comparison will be results obtained using ES with an exceedingly large ensemble (ES-REF) for smoothers, and results obtained using IES with an exceedingly large ensemble (IES-REF) for iterative algorithms. By utilizing such unrealistically large ensembles we obtain results that are visually indistinguishable from the best results that can be achieved using ES and IES. This was assured by running these algorithms with perturbations in their ensemble sizes.

Table 2 Variogram used for observation data error, the unit for range is grid cells

variogram type	range	mean	anisotropy ratio
spherical	5	0	1

Furthermore, we will show plots of the log permeability realizations used when generating the synthetic data ("Truth").

In each experiment, a fixed computational power is considered for each iteration of all algorithm runs (except for reference solutions). As the dominant cost of the DA process is pertaining to simulations of forward models, where iterative linear solvers dominate the computational costs for large problems, the computational cost pertaining to each ensemble member to be simulated using forward model, $\mathcal{M}_l$, is assumed to be proportional to G_l^γ, where G_l is the number of the active grid cells in the forward model at level l, and $\gamma \in [1.25, 1.5]$, [2]. Here, we take $\gamma = 1.35$. Additionally, as usual for large-scale cases, the ensemble size for standard single-level DA algorithms is set to be 100. Using this basis for calculations, the computational power allocated for all the runs will be equal if the following equation holds for all of them,

$$100 G_L^{1.35} = \sum_{l=1}^{L} N_l G_l^{1.35} \,. \tag{44}$$

Considering this equation, we set N_l for different levels of the MLHES. There exists $L-1$ degrees of freedom for specification of the $\{N_l\}_{l=1}^L$. No optimization was performed for this specification, the only aim pursued was to keep the size of sub-ensembles ascending with decrease in model cost. Several other similar settings that were tried resulted in similar DA outcomes.

For all experiments, the convergence criterion for the iterative algorithms was that improvements in the relative data mismatch should be smaller than 0.0001. The number of iterations required for convergence was approximately the same for IES-LOC and IMLHES. Accordingly, no adjustments were performed for equalizing the total computational cost used by these two algorithms.

For the MLHES and IMLHES, there is a possibility to improve the results by tuning the weights in Eqs. 17–22 for specific cases, but here we use the simplest choice–weights being all equal to $1/L$.

5.1 Experiment I

This experiment is conducted on Reservoir Model I. The multilevel algorithms have four levels, corresponding to 85, 154, 436, and 1600 grid cells, respectively. A summary of the resource allocation for the different runs carried out in this experiment can be found in Table 3. The observation data for this experiment are generated based on seismic vintages at 2000 and 4000 days after production starts.

The unknown logarithmic permeability field is based on a spherical variogram with mean and variance constant at 5 and 1, respectively, anisotropy angle and anisotropy ratio of -20 degrees and 0.33, and range 20 grid cells. Randomly selected realizations from this logarithmic permeability field can be found in Fig. 4.

Table 3 A summary of resource allocation for different runs in experiment I

	level 1	level 2	level 3	level 4
	$G_1 = 85$	$G_2 = 154$	$G_3 = 436$	$G_4 = 1600$
	N_1	N_2	N_3	N_4
ES-LOC	-	-	-	100
MLHES	525	471	231	30
ES-REF	-	-	-	8000
IES-LOC	-	-	-	100
IMLHES	525	471	231	30
IES-REF	-	-	-	8000

5.2 Experiment II

This experiment is conducted on Reservoir Model II. In this experiment, the presence of the oblique fault in the field interferes with coarsening the model. One way to handle this issue would be to avoid coarsening the grid around the fault area; however, this would reduce the computational efficiency of the multilevel scheme. In order to keep the grid coarsening as it is, the fault is approximated with bigger "zigzags" as depicted in Fig. 5 for one realization of the logarithmic permeability field at different levels of coarseness.

The multilevel algorithms have four levels, corresponding to 124, 310, 1060, and 4096 grid cells, respectively. A summary of the resource allocation for the different runs carried out in this experiment can be found in Table 4. The observation data for this experiment are generated based on seismic vintages at 5000 and 10000 days after production starts.

The unknown logarithmic permeability field is based on a spherical variogram with mean and variance constant at 5 and 1, respectively, anisotropy angle and anisotropy ratio of -30 degrees and 0.7, and range 25 grid cells. Randomly selected realizations from this logarithmic permeability field can be found in Fig. 6.

5.3 Experiment III

This experiment is conducted on Reservoir Model III.

The multilevel algorithms have four levels, corresponding to 163, 412, 1279, and 4900 grid cells, respectively. The observation data for this experiment are generated based on seismic vintages at 3000 and 6000 days after production starts.

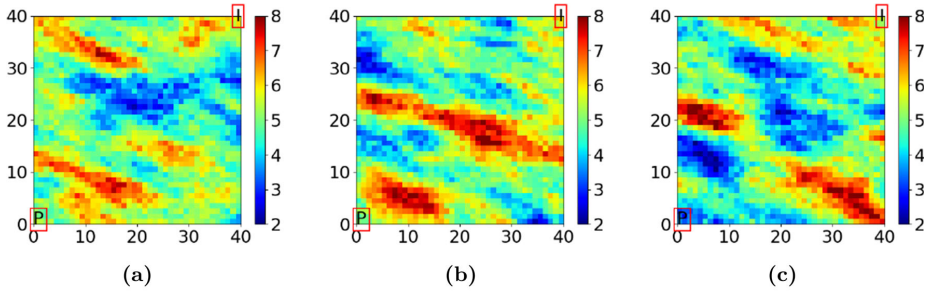


Fig. 4 Randomly selected realizations from prior distribution of the logarithmic permeability field of Reservoir model I

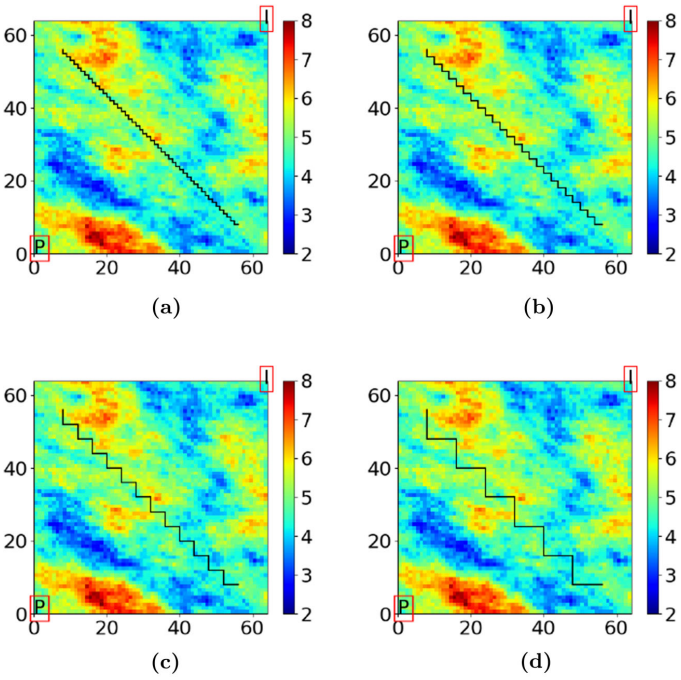


Fig. 5 Approximation of the fault for simulations, a: original fault, b-d: approximations at coarser levels

Table 4 A summary of resource allocation for different runs in experiment II

	level 1	level 2	level 3	level 4
	$G_1 = 124$	$G_2 = 310$	$G_3 = 1060$	$G_4 = 4096$
	N_1	N_2	N_3	N_4
ES-LOC	-	-	-	100
MLHES	1404	652	170	40
ES-REF	-	-	-	10000
IES-LOC	-	-	-	100
IMLHES	1404	652	170	40
IES-REF	-	-	-	10000

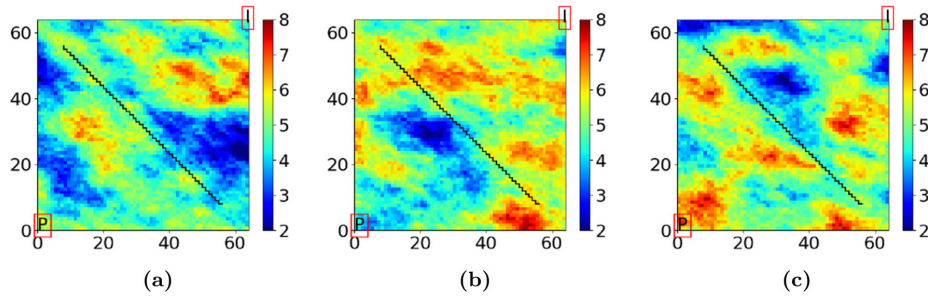


Fig. 6 Randomly selected realizations from prior distribution of the logarithmic permeability field of Reservoir model II

Table 5 A summary of resource allocation for different runs in Experiment III

	level 1	level 2	level 3	level 4
	$G_1 = 163$	$G_2 = 412$	$G_3 = 1279$	$G_4 = 4900$
	N_1	N_2	N_3	N_4
ES-LOC	-	-	-	100
MLHES	791	707	226	30
ES-REF	-	-	-	10000
IES-LOC	-	-	-	100
IMLHES	791	707	226	30
IES-REF	-	-	-	10000

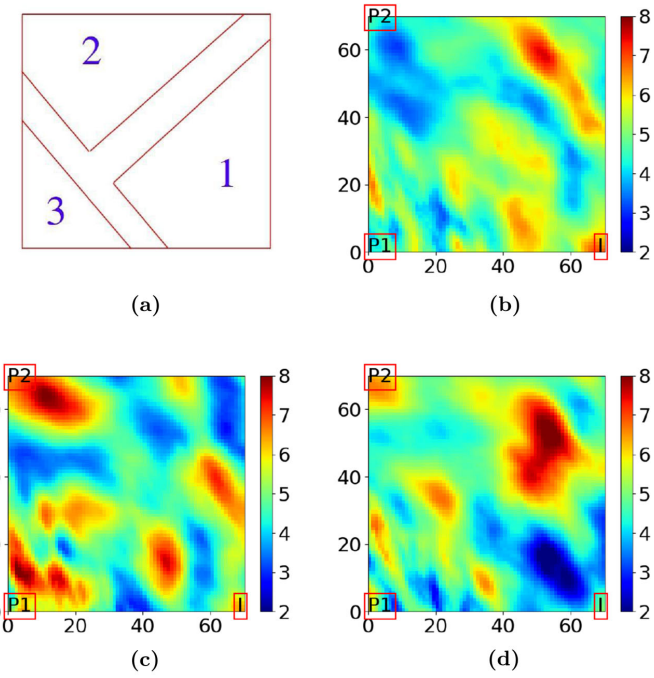


Fig. 7 Reservoir Model III, (**a**) three zones of the model (**b-d**) randomly selected realization from prior logarithmic permeability field

Table 6 The three variograms of Reservoir Model III

Zone	variance	mean	range	ratio	angle	type
1	1	5	20	0.4	-70	cubic
2	1	5	40	0.7	-30	cubic
3	1	5	30	0.6	-60	cubic

A summary of the resource allocation for different tests carried out in this experiment can be found in Table 5.

The unknown logarithmic permeability field is based on three different variograms in three zones of the field. The Zones, i.e. Zone 1, Zone 2, and Zone 3 can be seen in Fig. 7. The area without any assigned zone is a continuous transition from one zone to others. The details about the variograms based on which the distribution of the unknown parameters are defined can be found in Table 6. Randomly selected realizations from this logarithmic permeability field can be found in Fig. 7.

6 Numerical results

The results from the numerical experiments are assessed qualitatively, using the posterior parameters and forecasts.

Firstly, mean and variance of the posterior parameter fields obtained by different algorithms and reference cases are compared with each other. Additionally, since both the model forecasts and observation data are in different resolutions for different levels of multilevel algorithms, comparison of posterior forecasts as such is not a possibility. Instead, we run fine-scale simulations of the posterior

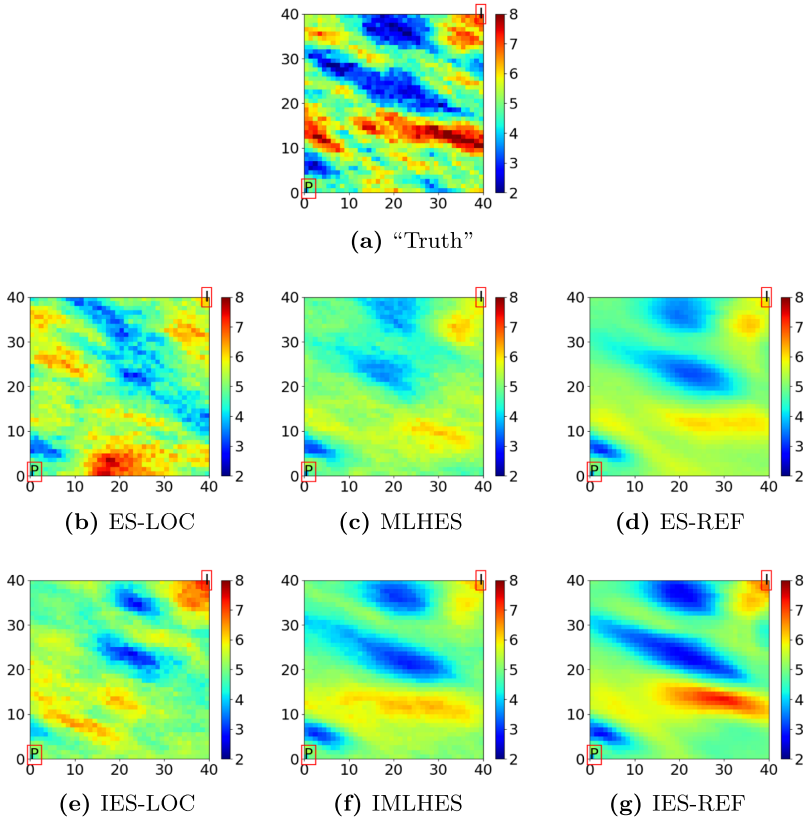


Fig. 8 Experiment I–Mean posterior logarithmic permeability field

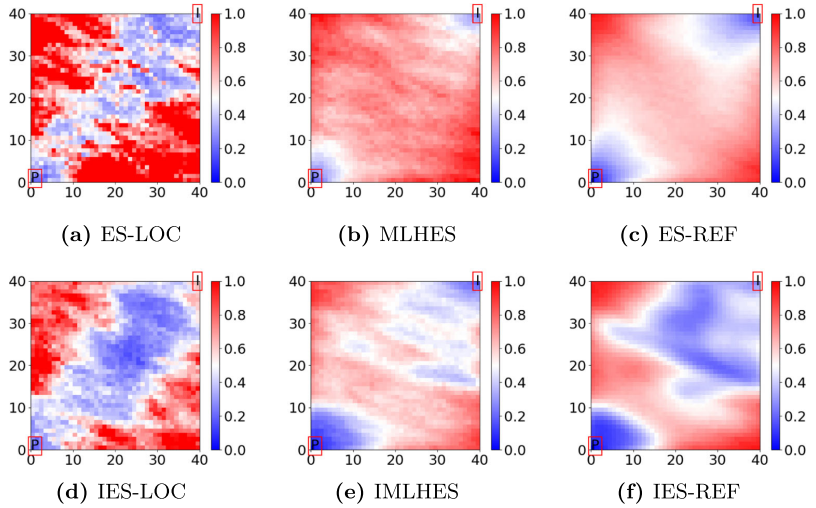


Fig. 9 Experiment I–Variance of posterior logarithmic permeability field

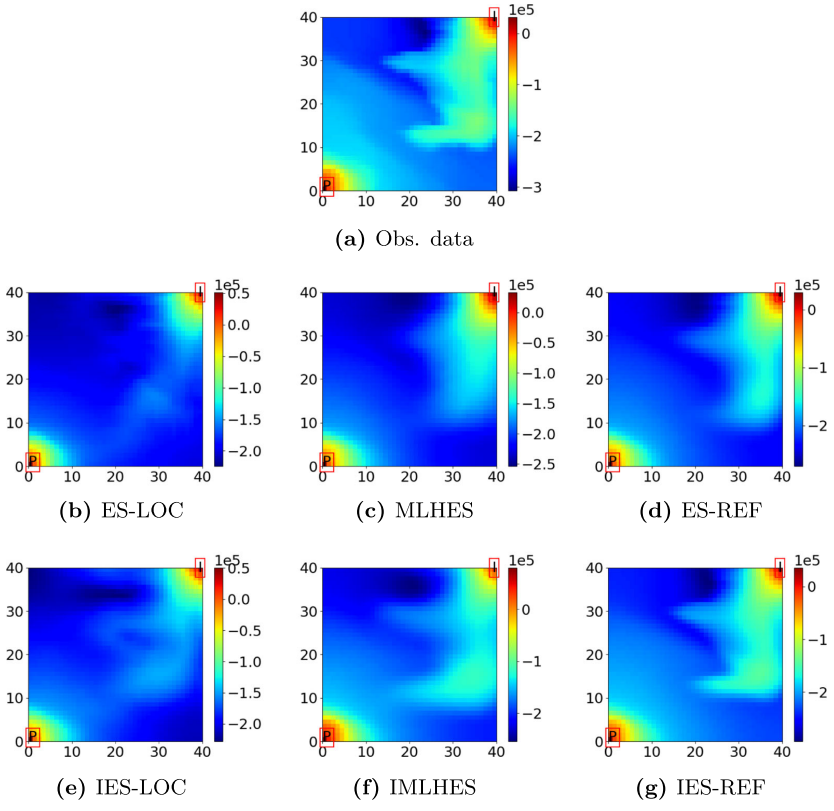


Fig. 10 Experiment I–Posterior time-lapse bulk impedance field ($\frac{m}{s}\frac{kg}{m^3}$) in comparison with observation data in the second vintage

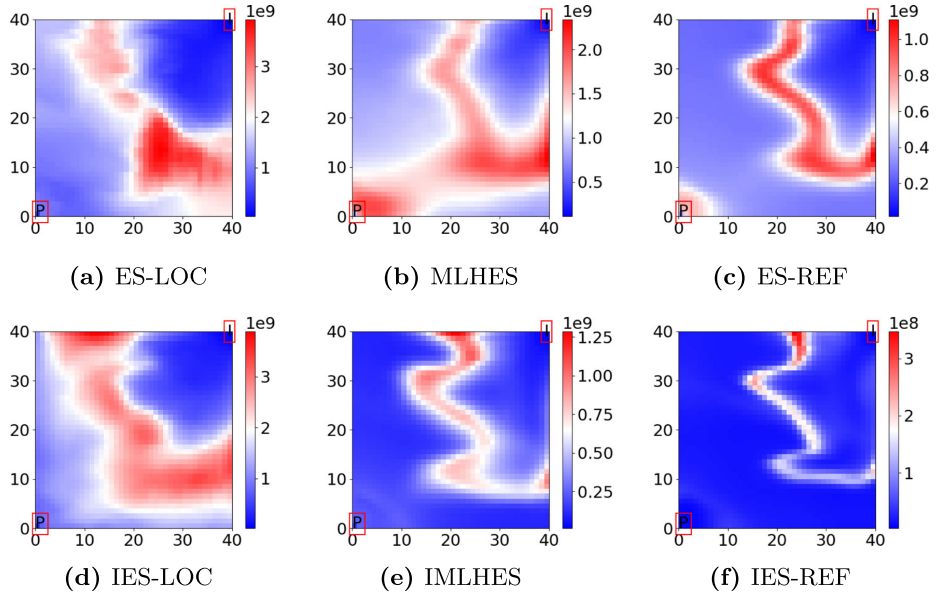


Fig. 11 Experiment I–Variance of posterior time-lapse bulk impedance field ($(\frac{m}{s}\frac{kg}{m^3})^2$), in the second vintage

ensemble for all algorithms and then plot the mean and variance of the model forecasts for all the algorithms and compare them together. The model forecasts of the second vintage are presented for all the experiments.

The simplest formulation is chosen for computation of posterior statistics of MLHES and IMLHES, i.e. re-uniting all the sub-ensembles in these algorithms and treating them as one ensemble for computation of posterior mean and variance fields.

ES-LOC was tested with several ranges for localization (critical distances), and the best results are presented for each of the experiments.

6.1 Results of Experiment I

Visual analysis of the mean permeability fields in Fig. 8 shows that MLHES and IMLHES results are more similar to ES-REF and IES-REF results, respectively, than ES-LOC and IES-LOC results are. This is confirmed by comparison of the variance fields in Fig. 9. In Fig. 8, improvements are seen in approximation of the "Truth" by utilization of IMLHES compared with MLHES.

From Figs. 10 and 11, it is seen that the statistics of posterior forecasts obtained by use of multilevel algorithms are more similar to those of reference cases than the results obtained by use of conventional algorithms. In Fig. 10, the mean of model forecasts obtained from IMLHES is more similar to the observation data than that of MLHES.

The ES-LOC and IES-LOC results presented here are based on the localization range of 30 grid cells.

6.2 Results of Experiment II

Visual analysis of the mean permeability fields in Fig. 12 shows that MLHES results are more similar to ES-REF results than ES-LOC results are. This can be further confirmed by comparison of the variance fields given in Fig. 13 and the statistics of model forecasts in Figs. 14 and 15.

In this experiment, IES-LOC does not converge to the proximity of global optimum. Since the same holds for IES-REF, there exists no reference for comparison of the iterative algorithms. Nonetheless, as can be seen in Fig. 12, IMLHES results are more similar to "True" permeability field than those of MLHES. Also, in Fig. 14, slight improvements in approximation of the observation data is seen in IMLHES results compared with MLHES results.

For a class of problems (including the problem considered here) where the model forecast can be seen as a spatially integrated response to a spatially varying parameter field, there exists a correlation between small-scale oscillations in the parameter domain and the nonlinearity strength

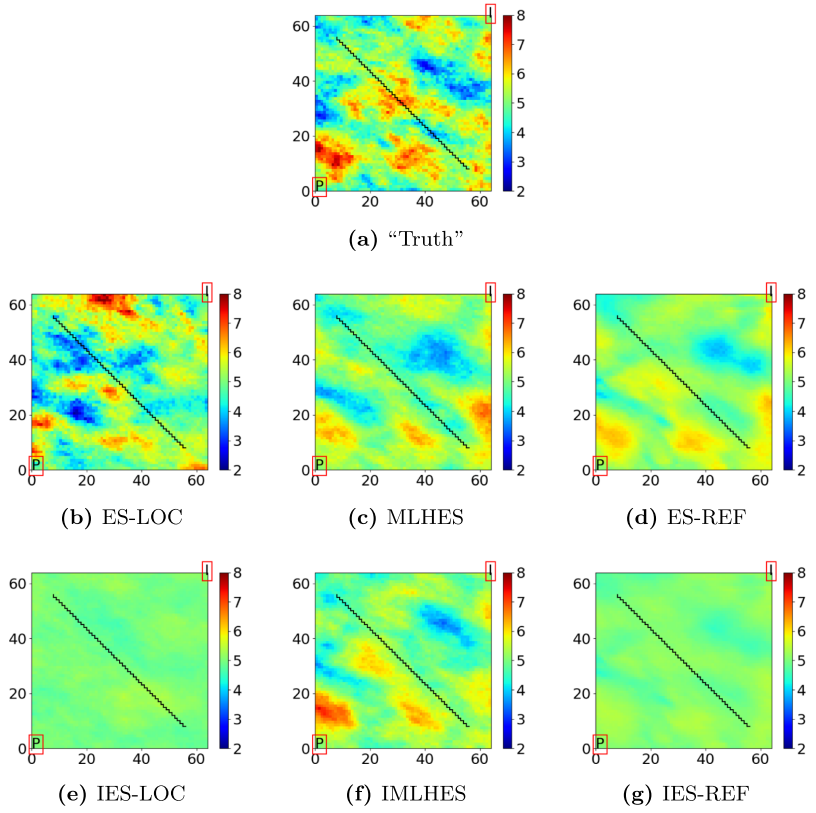


Fig. 12 Experiment II–Mean posterior logarithmic permeability field

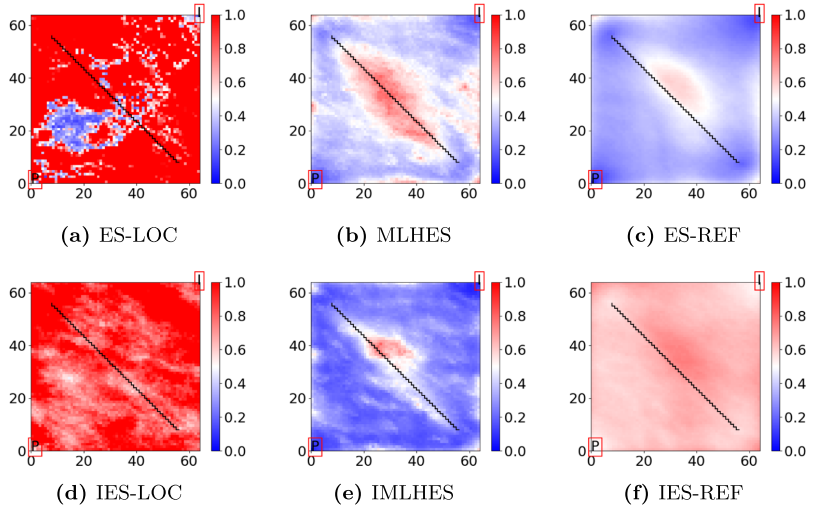


Fig. 13 Experiment II–Variance of posterior logarithmic permeability field

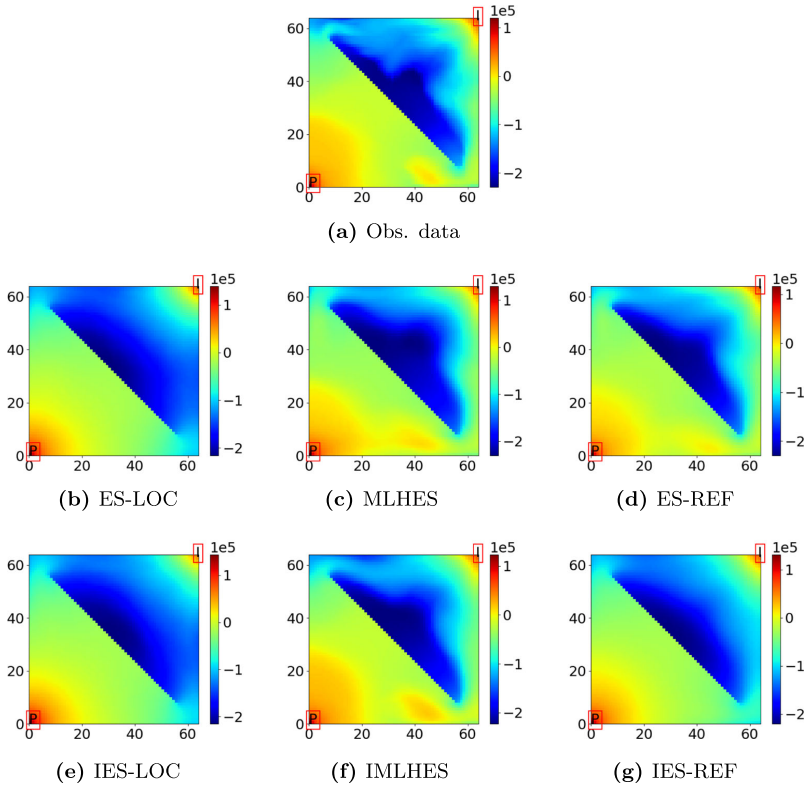


Fig. 14 Experiment II–Posterior time-lapse bulk impedance field ($\frac{m}{s}\frac{kg}{m^3}$) in comparison with observation data in the second vintage

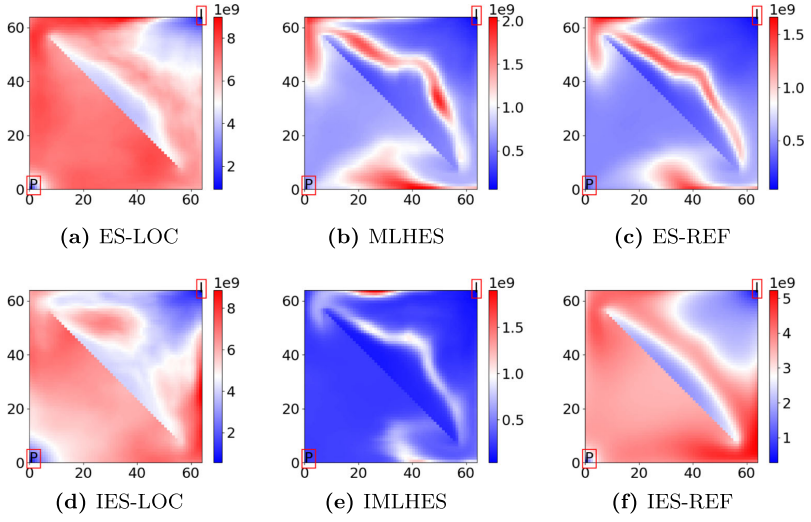


Fig. 15 Experiment II–Variance of posterior time-lapse bulk impedance field ($(\frac{m}{s}\frac{kg}{m^3})^2$), in the second vintage

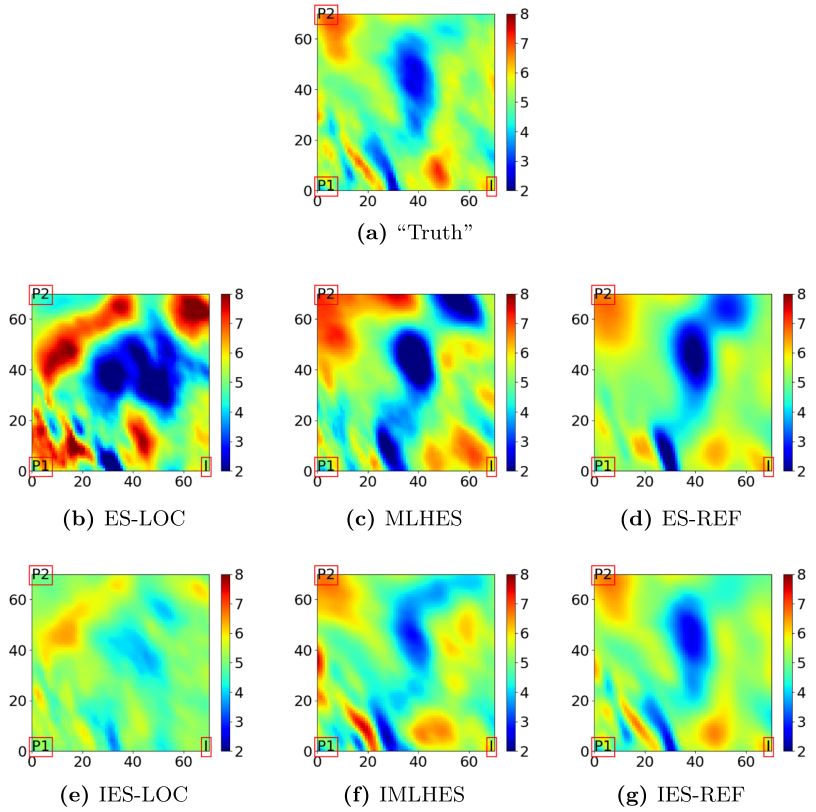


Fig. 16 Experiment III–Mean posterior logarithmic permeability field

of the mapping from parameter field to model forecast, see, e.g., [5, 18]. This correlation is such that coarsening the simulation grid and upscaling the associated parameters will generally result in weaker nonlinearity in the coarser forward models compared to the finer ones. Accordingly, use of IMLHES instead of IES could improve convergence, as was the case in experiment II.

The ES-LOC and IES-LOC results presented here are based on the localization range of 50 grid cells.

6.3 Results of Experiment III

Visual analysis of the mean permeability fields in Fig. 16 shows that MLHES and IMLHES results are more similar to ES-REF and IES-REF results, respectively, than ES-LOC and IES-LOC results are. This is confirmed by comparison of the variance fields in Fig. 17. In Fig. 16, improvements are seen in approximation of the "Truth" by utilization of IMLHES compared with MLHES.

From Figs. 18 and 19, it is seen that the statistics of posterior forecasts obtained by use of multilevel algorithms are more similar to those of reference cases than the results obtained by use of conventional algorithms. In Fig. 18, the mean of model forecasts obtained from IMLHES is more similar to the observation data than that of MLHES.

The ES-LOC and IES-LOC results presented here are based on the localization range of 60 grid cells.

7 Summary and Conclusions

In this work, a recently devised MLDA algorithm (MLHES) for assimilation of spatially distributed data was discussed, and an iterative version of it (IMLHES) was introduced. Both of these methods utilize generalizations of multilevel statistics introduced in [15] for Monte Carlo approximations of mean and covariance of model forecasts. In addition, performance of these algorithms were evaluated in comparison

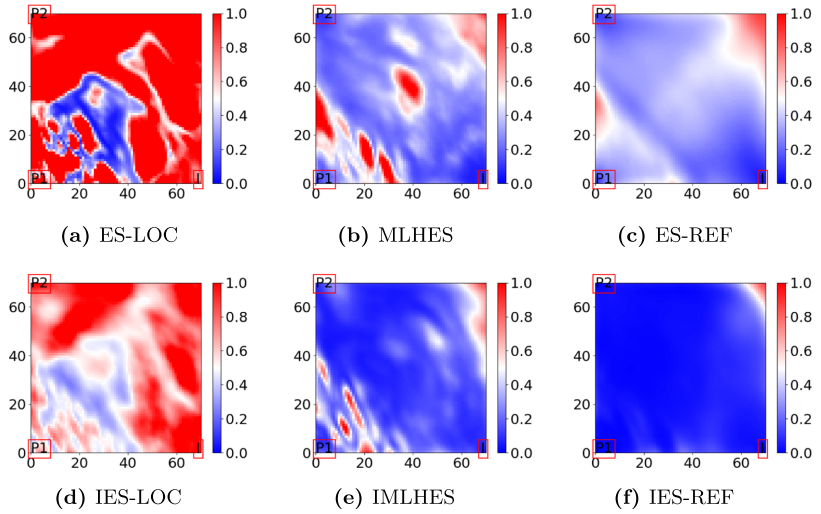


Fig. 17 Experiment III–Variance of posterior logarithmic permeability field

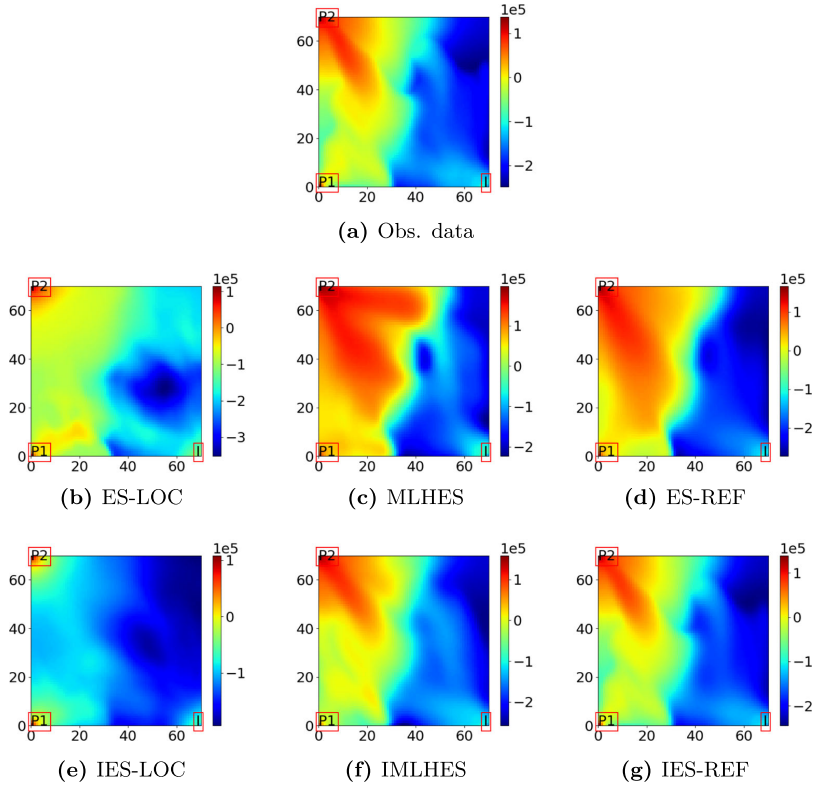


Fig. 18 Experiment III–Posterior time-lapse bulk impedance field ($\frac{m}{s}\frac{kg}{m^3}$) in comparison with observation data in the second vintage

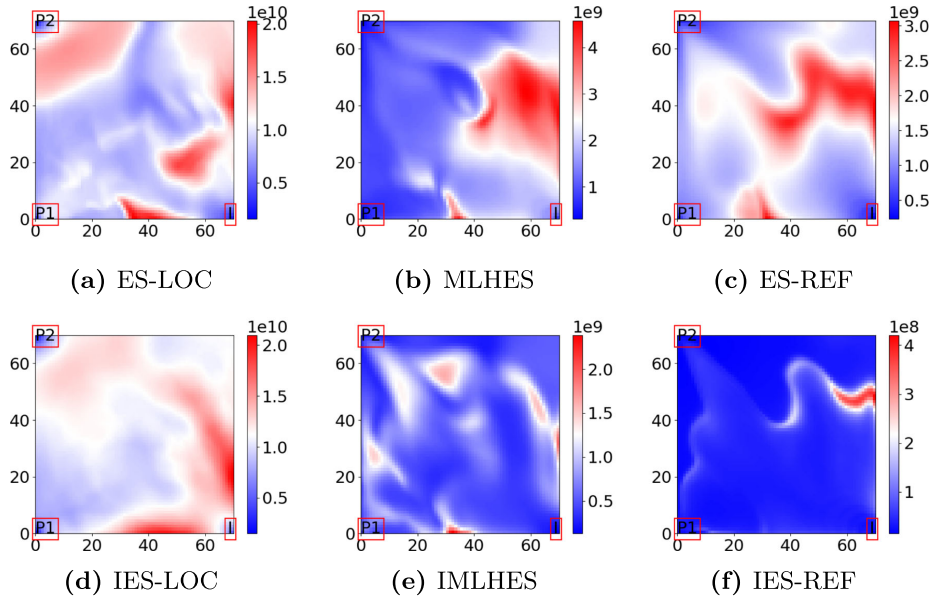


Fig. 19 Experiment III–Variance of posterior time-lapse bulk impedance field ($(\frac{m}{s}\frac{kg}{m^3})^2$), in the second vintage

with standard DA techniques. In doing so, three experiments were conducted. Each experiment was performed on a reservoir model and consisted of six algorithm runs: conventional Ensemble Smoother with localization (ES-LOC), MLHES, ES with an exceedingly large ensemble size (ES-REF), conventional Iterative Ensemble Smoother with localization (IES-LOC), IMLHES, and IES with an exceedingly large ensemble (IES-REF).

In order to assess the numerical results, firstly, the mean and variance fields of posterior unknown parameters (log permeability) were generated and assessed visually. The assessments suggest that in all experiments the results from MLHES were more similar to those of ES-REF than the results from ES-LOC were. Except for one experiment where both IES-LOC and IES-REF did not converge to the proximity of global optimum, the same conclusion was true about the iterative algorithms. The exception suggests an additional advantage of IMLHES over IES. It was also observed that iterations resulted in all the mean posterior fields obtained by IMLHES to be closer to the permeability field from which the observation data were generated than the mean posterior field obtained by MLHES.

Secondly, fine-scale simulations were conducted for all the posterior ensembles of all algorithms. Plots of simulated time-lapse bulk impedance means and variances were compared to plots of observed time-lapse bulk impedance. Visual analysis of these plots showed that in all the applicable cases, the multilevel algorithms performed more similar to the reference cases than the conventional algorithms did. Additionally, the means of model forecasts obtained from IMLHES were closer to the observation data than the means of model forecasts obtained from MLHES were.

In addition to the presented conventional DA algorithms, which utilize distance-based localization, ES and IES were used in conjunction with correlation-based localization [28] for assimilation of the same data, but this did not improve the conventional DA results. Furthermore, results from investigations that were not presented here suggest that MLDA algorithms show consistency in quality of history matching with respect to variation in observation data error.

There are several issues about the presented MLDA algorithms that can be investigated further. Firstly, the optimal extent of coarsening the grid was not discussed. The rule of thumb was to coarsen the grid until further coarsening results in marginal reduction in grid size, due to restrictions with coarsening around the wells. Additionally, the number of levels and allocation of resources between them to obtain the optimal result can also be further investigated. In this research, the weights for different

levels in the multilevel statistics was set to be all equal; however, there is no such constraint. Accordingly, tuning these weights can be a matter of investigation. Furthermore, the posterior statistics that was presented in this paper was based on a pool comprised of all the updated realizations of different levels put together. Since the accuracy and the computational power allocated per realization differs from level to level, the optimal choice of weights for presenting the posterior statistics needs further research. Moreover, the multilevel modeling error, which was partially accounted for by mean bias correction, can be studied in more detail. For this we refer to [33]. Finally, as realistic reservoir cases are more complex than the fields tested in this paper, the increase in the dimensionality of the parameters may call for a combination of localization and the proposed MLDA algorithms.

Appendix

A Multilevel hybrid ensemble smoother algorithm

Algorithm 1 MLHES algorithm.

Define $\{\mathcal{M}_l\}_{l=1}^{L}$
Define $\{w_l\}_{l=1}^{L}$
for $l = 1, ..., L$ **do**
  for $j = 1, ..., N_l$ **do**
    $z_{l,j} \leftarrow \mathcal{N}\left(\mathrm{E}(Z^{pri}), \mathrm{C}(Z^{pri})\right)$
    $\hat{y}_{l,j} \leftarrow \mathcal{M}_l(z_{l,j})$
  end for
end for
for $l = 1, ..., L$ **do**
  $y_{l,j} \leftarrow \widehat{y}_{l,j} + \{U_L^l\,\mathrm{E}(\widehat{Y}_L) - \mathrm{E}(\widehat{Y}_l)\}$
end for
for $l = 1, ..., L$ **do**
  $\mathrm{E}_{\mathrm{ML},l}(Y) \leftarrow \sum_{k=1}^{L} w_k U_k^l\,\mathrm{E}(Y_k)$
  $\mathrm{C}_{\mathrm{ML},l}(Y) \leftarrow \sum_{k=1}^{L} w_k\{\mathrm{C}(U_k^l Y_k) + \left(\mathrm{E}(U_k^l Y_k) - \mathrm{E}_{\mathrm{ML},l}(Y)\right)\left(\mathrm{E}(U_k^l Y_k) - \mathrm{E}_{\mathrm{ML},l}(Y)\right)^T\}$
  $\mathrm{C}_{\mathrm{ML},l}(Z, Y) \leftarrow \sum_{k=1}^{L} w_k\{\mathrm{C}(Z_k, U_k^l Y_k) + (\mathrm{E}(Z_k) - \mathrm{E}_{\mathrm{ML}}(Z))\left(\mathrm{E}(U_k^l Y_k) - \mathrm{E}_{\mathrm{ML},l}(Y)\right)^T\}$
  for $j = 1, ..., N_l$ **do**
    $d_{l,j} \leftarrow \mathcal{N}\left(U_L^l \mu_D, U_L^l \mathbb{C}(D) {U_L^l}^T\right)$
    $z_{l,j}^a \leftarrow z_{l,j}^{pri} + \mathrm{C}_{\mathrm{ML},l}(Z, Y)\left(\mathrm{C}_{\mathrm{ML},l}(Y) + U_L^l \mathbb{C}(D) {U_L^l}^T\right)^{-1}\left(d_{l,j} - y_{l,j}\right)$
  end for
end for

B Iterative Multilevel Hybrid Ensemble Smoother Algorithm

Algorithm 2 IMLHES algorithm.

Define $\{\mathcal{M}_l\}_{l=1}^{L}$
Define $\{w_l\}_{l=1}^{L}$
for $l = 1, ..., L$ **do**
  for $j = 1, ..., N_l$ **do**
    $d_{l,j} \leftarrow \mathcal{N}\left(U_L^l \mu_D, U_L^l \mathbb{C}(D) {U_L^l}^T\right)$
    $z_{l,j}^{pri} \leftarrow \mathcal{N}\left(\mathrm{E}(Z^{pri}), \mathrm{C}(Z^{pri})\right)$
    $z_{l,j}^0 \leftarrow z_{l,j}^{pri}$
  end for
end for
$i \leftarrow 0$
while not Converged **do**
  for $l = 1, ..., L$ **do**
    for $j = 1, ..., N_l$ **do**
      $\hat{y}_{l,j}^i \leftarrow \mathcal{M}_l(z_{l,j}^i)$
    end for
  end for
  for $l = 1, ..., L$ **do**
    for $j = 1, ..., N_l$ **do**
      $y_{l,j}^i = \hat{y}_{l,j}^i + \left(U_L^l\,\mathrm{E}(\widehat{Y}_L) - \mathrm{E}(\widehat{Y}_l)\right)$
    end for
  end for
  for $l = 1, ..., L$ **do**
    if Level has not converged **then**
      $\mathrm{E}_{\mathrm{ML},l}(Y^i) \leftarrow \sum_{k=1}^{L} w_k U_k^l\,\mathrm{E}(Y_k^i)$
      $\mathrm{E}_{\mathrm{ML}}(\mathrm{Z^i}) \leftarrow \sum_{k=1}^{L} w_k\,\mathrm{E}(Z_k^i)$
      $\mathrm{C}_{\mathrm{ML},l}(Z^i, Y_l^i) \leftarrow \sum_{k=1}^{L} w_k\{\mathrm{C}(Z_k^i, U_k^l Y_k^i) + \left(\mathrm{E}(Z_k^i) - \mathrm{E}_{\mathrm{ML}}(\mathrm{Z^i})\right)\left(\mathrm{E}(U_k^l Y_k^i) - \mathrm{E}_{\mathrm{ML},l}(Y^i)\right)^T\}$
      $\mathrm{C}_{\mathrm{ML}}(\mathrm{Z^i}) \leftarrow \sum_{k=1}^{L} w_k\{\mathrm{C}(Z_k^i) + \left(\mathrm{E}(Z_k^i) - \mathrm{E}_{\mathrm{ML}}(\mathrm{Z^i})\right)\left(\mathrm{E}(Z_k^i) - \mathrm{E}_{\mathrm{ML}}(\mathrm{Z^i})\right)^T\}$
      $\mathrm{M}_l^i \leftarrow \mathrm{C}_{\mathrm{ML},l}\left(Y^i, Z^i\right)\mathrm{C}_{\mathrm{ML}}(\mathrm{Z^i})^+$
      $K_l^i \leftarrow \mathrm{C}\left(Z^{pri}\right)\mathrm{M}_l^{iT}\left(\mathrm{M}_l^i\,\mathrm{C}(Z^{pri})\mathrm{M}_l^{iT} + U_L^l \mathbb{C}(D)(U_L^l)^T\right)^{-1}$
      for $j = 1, ..., N_l$ **do**
        $\Delta_{l,j}^{i,pri} \leftarrow (I_Z - K_l^i \mathrm{M}_l^i)[z_{l,j}^{pri} - z_{l,j}^i]$
        $\Delta_{l,j}^{i,lik} \leftarrow K_l^i[d_{l,j} - y_{l,j}^i]$
        $z_{l,j}^{i+1} \leftarrow z_{l,j}^i + \beta_l\left(\Delta_{l,j}^{i,pri} + \Delta_{l,j}^{i,lik}\right)$
      end for
      Update β_l
    end if
  end for
  $i \leftarrow i + 1$
end while

Acknowledgments The authors acknowledge financial support from the NORCE research project "Assimilating 4D Seismic Data: Big Data Into Big Models" which is funded by industry partners, Equinor Energy AS, Lundin Energy Norway AS, Repsol Norge AS, Shell Global Solutions International B.V., Total E&P Norge AS, and Wintershall Dea Norge AS, as well as the Research Council of Norway (PETROMAKS2). We also thank Schlumberger for providing academic software licenses to ECLIPSE.

Funding Open Access funding provided by NORCE Norwegian Research Centre AS.

References

1. Alfonzo, M., Oliver, D.S.: Seismic data assimilation with an imperfect model. Comput. Geosci. **24**(2), 889–905 (2020)
2. Axelsson, O.: Iterative solution methods. Cambridge University Press (1996)
3. Aziz, K., Settari, A.: Petroleum Reservoir Simulation. Applied Science Publ Ltd., London (1979)
4. Batzle, M., Wang, Z.: Seismic properties of pore fluids. Geophysics **57**(11), 1396–1408 (1992)
5. Chavent, C., Liu, J.: Multiscale parametrization for the estimation of a diffusion coefficient in elliptic and parabolic problems. IFAC Proc. Vol. **22**(4), 193–202 (1989)
6. Chen, Y., Oliver, D.S.: Cross-covariances and localization for EnKF in multiphase flow data assimilation. Comput. Geosci. **14**(4), 579–601 (2010)
7. Chen, Y., Oliver, D.S.: Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math. Geosci. **44**(1), 1–26 (2012)
8. de Moraes, R.J., Hajibeygi, H., Jansen, J.D.: A multiscale method for data assimilation. Comput. Geosci. **24**(2), 425–442 (2020)
9. Emerick, A., Reynolds, A.: Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. Comput. Geosci. **15**(2), 251–269 (2011)
10. Emerick, A.A.: Analysis of the performance of ensemble-based assimilation of production and seismic data. J. Pet. Sci. Eng. **139**, 219–239 (2016)
11. Emerick, A.A., Reynolds, A.C.: Ensemble smoother with multiple data assimilation. Comput. Geosci. **55**, 3–15 (2013)
12. Fahimuddin, A.: 4D Seismic History Matching Using the Ensemble Kalman Filter (EnKF): Possibilities and Challenges, PhD thesis. The University of Bergen (2010)
13. Fossum, K., Mannseth, T.: Coarse-scale data assimilation as a generic alternative to localization. Comput. Geosci. **21**(1), 167–186 (2017)
14. Fossum, K., Mannseth, T.: A novel multilevel method for assimilating spatially dense data. In: ECMOR XVI-16th European Conference on the Mathematics of Oil Recovery, vol. 2018, pp. 1–12. European Association of Geoscientists & Engineers (2018)
15. Fossum, K., Mannseth, T., Stordal, A.S.: Assessment of multilevel ensemble-based data assimilation for reservoir history matching. Comput. Geosci. **24**(1), 217–239 (2020)
16. Furrer, R., Bengtsson, T.: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. J. Multivar. Anal. **98**(2), 227–255 (2007)
17. Gaspari, G., Cohn, S.E.: Construction of correlation functions in two and three dimensions. Q. J. Roy. Meteorol. Soc. **125**(554), 723–757 (1999)
18. Grimstad, A.A., Mannseth, T.: Nonlinearity, scale, and sensitivity for parameter estimation problems. SIAM J. Sci. Comput. **21**(6), 2096–2113 (2000)
19. Gu, Y., Oliver, D.S.: An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. Spe J. **12**(04), 438–446 (2007)
20. Hatfield, S., Subramanian, A., Palmer, T., Düben, P.: Improving weather forecast skill through reduced-precision data assimilation. Mon. Weather. Rev. **146**(1), 49–62 (2018)
21. He, J., Sarma, P., Durlofsky, L.J.: Reduced-order flow modeling and geological parameterization for ensemble-based data assimilation. Comput Geosci. **55**, 54–69 (2013)
22. Hoel, H., Law, K.J., Tempone, R.: Multilevel ensemble Kalman filtering. SIAM J. Numer. Anal. **54**(3), 1813–1839 (2016)
23. Hoel, H., Shaimerdenova, G., Tempone, R.: Multilevel ensemble Kalman filtering based on a sample average of independent EnKF estimators. Found. Data Sci. **2**(4), 351 (2020)
24. Houtekamer, P.L., Mitchell, H.L.: A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather. Rev. **129**(1), 123–137 (2001)
25. Iglesias, M.A.: Iterative regularization for ensemble data assimilation in reservoir models. Comput. Geosci. **19**(1), 177–212 (2015)
26. Leeuwenburgh, O., Brouwer, J., Trani, M.: Ensemble-based conditioning of reservoir models to seismic data. Comput. Geosci. **15**(2), 359–378 (2011)
27. Lorentzen, R.J., Bhakta, T., Grana, D., Luo, X., Valestrand, R., Nævdal, G.: Simultaneous assimilation of production and seismic data: application to the Norne field. Comput. Geosci. **24**(2), 907–920 (2020)
28. Luo, X., Bhakta, T.: Towards automatic and adaptive localization for ensemble-based history matching. In: ECMOR XVI-16th European Conference on the Mathematics of Oil Recovery, vol. 2018, pp. 1–26. European Association of Geoscientists & Engineers (2018)
29. Luo, X., Bhakta, T., Jakobsen, M., Nævdal, G.: An ensemble 4D-seismic history-matching framework with sparse representation based on wavelet multiresolution analysis. SPE J. **22**(03), 985–1010 (2017)
30. Luo, X., Bhakta, T., Naevdal, G., et al.: Correlation-based adaptive localization with applications to ensemble-based 4D-seismic history matching. SPE J. **23**(02), 396–427 (2018)
31. Mannseth, T., Fossum, K.: Assimilating spatially dense data for subsurface applications–balancing information and degrees of freedom. Comput. Geosci. **22**(5), 1323–1349 (2018)
32. Nezhadali, M., Bhakta, T., Fossum, K., Mannseth, T.: A novel approach to multilevel data assimilation. In: ECMOR XVII, vol. 2020, pp. 1–13. European Association of Geoscientists & Engineers (2020)
33. Nezhadali, M., Bhakta, T., Fossum, K., Mannseth, T.: Multilevel assimilation of inverted seismic data with correction

for multilevel modeling error. Front. Appl. Math. Stat., 7. https://doi.org/10.3389/fams.2021.673077 (2021)
34. Oliver, D.S., Reynolds, A.C., Liu, N.s.: Inverse Theory for Petroleum Reservoir Characterization and History Matching. Cambridge University Press (2008)
35. Popov, A.A., Mou, C., Sandu, A., Iliescu, T.: A multifidelity ensemble Kalman filter with reduced order control variates. SIAM J. Sci. Comput. **43**(2), A1134–A1162 (2021)
36. Raanes, P.N., Stordal, A.S., Evensen, G.: Revising the stochastic iterative ensemble smoother. Nonlinear Process. Geophys. **26**(3), 325–338 (2019)
37. Sakov, P., Haussaire, J.-M., Bocquet, M.: An iterative ensemble Kalman filter in the presence of additive model error. Q. J. Roy. Meteorol. Soc. **144**(713), 1297–1309 (2018)
38. Schlumberger, L.: Eclipse reservoir simulation software v2016. Technical Description Manual (2016)
39. Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics (2005)
40. Tarrahi, M., Elahi, S.H., Jafarpour, B.: Fast linearized forecasts for subsurface flow data assimilation with ensemble Kalman filter. Comput. Geosci. **20**(5), 929–952 (2016)
41. Van Leeuwen, P.J., Evensen, G.: Data assimilation and inverse methods in terms of a probabilistic formulation. Mon. Weather. Rev. **124**(12), 2898–2913 (1996)

Chapter 12

Paper D

Sequential multilevel assimilation of inverted seismic data

Sequential Multilevel Assimilation of Inverted Seismic Data

Mohammad Nezhadali[1,2,*], Tuhin Bhakta[1], Kristian Fossum[1], and Trond Mannseth[1]

[1] *NORCE Norwegian Research Center*
[2] *University of Bergen*
[*] *mone@norceresearch.no, Nygårdsgaten 112, 5008 Bergen, Norway*

Abstract

We consider estimation of absolute permeability from inverted seismic data. Large amounts of simultaneous data, such as inverted seismic data, enhance the negative effects of Monte Carlo errors in ensemble-based Data Assimilation (DA). Multilevel (ML) models consist of a selection of models with different fidelities. Multilevel Data Assimilation (MLDA) attempts to obtain a better statistical accuracy with a small sacrifice of the numerical accuracy. Spatial grid coarsening is one way of generating an ML model. It has been shown that coarsening the spatial grid results in a problem with weaker nonlinearity, and hence, in a less challenging problem than the problem on the original fine grid. Accordingly, formulating a sequential MLDA algorithm which uses the coarser models in the first steps of the DA, followed by the finer models, helps to find an approximation to the solution of the inverse problem at the first steps and gradually converge to the solution. We present two variants of a sequential MLDA algorithm and compare their performance with both conventional DA algorithms and a simultaneous (i.e., using all the models on the different grids simultaneously) MLDA algorithm using numerical experiments. Both posterior parameters and posterior model forecasts are compared qualitatively and quantitatively. The results from numerical experiments suggest that all MLDA algorithms generally perform better than the conventional DA algorithms. In estimation of the posterior parameter fields, the simultaneous MLDA algorithm and one of the variants of sequential MLDA (SMLES-H) perform similarly and slightly better than the other variant (SMLES-S). While in estimation of the posterior model forecasts, SMLES-S clearly performs better than both the simultaneous MLDA algorithm and SMLES-H.

1 Introduction

Sound decision making in petroleum reservoir management depends on reliable production forecasts from reservoir models, including accurate estimates of uncertainty in the forecasts. The reliability is increased by utilization of available data for calibration of the models–through the process known as history-matching. Ensemble-based Data Assimilation (DA) methods, using statistically correct formulations, have become popular for automated reservoir history-matching [1, 2, 3, 4, 5, 6, 7]. We consider estimation of absolute permeability from inverted seismic data.

Ensemble-based methods have limited degree of freedom, which in conjunction with massive amounts of data, e.g. time-lapse seismic data, can result in over-fitting. There have been several efforts to balance the degrees of freedom of the problem and the information content in the data, including use of localization [8], reduction of data using machine learning techniques [9, 10], reduction in data size using the correlation between the data and wells' cumulative production [11], sparse representation of data using a wavelet transform [12], assimilation of only the saturation front or transformation of the data into position of fluid fronts [13, 14, 15], combination of coarsening the data and coarse model simulations [16], and projection of data into ensemble subspace in combination with local analysis [17].

Monte Carlo approximations play a crucial role in ensemble-based DA. Due to computational-cost limitations, the ensemble size is limited to roughly one hundred. Using straightforward ensemble-based DA, the degrees of freedom of the problem would equal the ensemble size, and such an approach would result in significant Monte Carlo errors. The negative effects of Monte Carlo errors are enlarged in the presence of large amounts of simultaneous data, such as inverted seismic data, resulting in underestimation of variance, and in more severe cases ensemble collapse.

The most widely used treatment for Monte Carlo errors is distance-based localization [18]. The basic assumption underlying distance-based localization is that true correlations between a parameter and a datum decrease when the distance between their respective locations increases, and disappears if the distance exceeds a critical distance. This assumption may not always hold for subsurface problems. Different localization functions and their utilization in DA can be found in [19, 20, 21]. A proper choice of localization function, and the critical distance in particular, depends on parameter and data types as well as on other problem settings. This reduces the robustness of distance-based localization, also for problems where its basic assumption does hold. Papers using ensemble-based methods for assimilation of seismic data [22, 23, 24], typically use localization methodologies developed originally for meteorological science which were later adapted to petroleum problems with production data.

Simply increasing the ensemble size will reduce Monte-Carlo errors, but it will also increase the computational cost. Utilization of a lower-cost and lower-fidelity model renders the possibility of increasing the ensemble size without increasing the total computational cost. Use of a lower-fidelity reservoir model will, however, introduce modeling errors in addition to those already present in conventional-fidelity simulation results. The underlying assumption when applying lower-fidelity models in DA is therefore that the gain in reducing Monte Carlo errors is larger than the loss in numerical simulation accuracy. DA using various types of lower-fidelity models has been applied to several inverse problems, e.g., within petroleum reservoir modeling [25, 26, 27] and atmospheric science [28]. Note that since lower-fidelity simulations are applied to the forecast step and localization is applied to the analysis step, the two techniques can be combined, if desired.

Lower-fidelity models can have high numerical errors. Additionally, choosing an optimal level of fidelity for these models is not straightforward. Multilevel (ML) simulations utilize a selection of models, which form hierarchies of both computational accuracy and computational costs (ML model). Multilevel Data Assimilation (MLDA) [29, 30, 31, 32, 33, 34, 35, 36, 37] utilizes an ML model in the forecast step of the DA. The levels of the multilevel model have different numerical accuracies. The MLDA, allocating some of the computational power to the models with lower fidelity, tries to achieve a low total error by keeping a balance between the numerical errors and statistical errors.

Some conventional DA methods (i.e., single-level ensemble-based DA methods) , like the ensemble smoother (ES) [38] , the ES with multiple DA (ESMDA) [2], and the iterative ES (IES) [1], assimilate data simultaneously, i.e. assimilate all data over a certain period at once, while other conventional methods, like the ensemble Kalman filter (EnKF) [39] and the EnKF

with multiple DA (EnKFMDA) [40], assimilate data sequentially. In the MLDA domain, a Simultaneous MLDA (SiMLDA) algorithm was developed for assimilation of inverted seismic data [33]. Numerical experiments show that this algorithm, the Multilevel Hybrid Ensemble Smoother (MLHES) whose formulation uses a hybrid Kalman gain based on several levels, outperforms its conventional DA counterparts with which it was compared [33, 41, 37]. However, strong nonlinearity can affect its performance negatively. Hence, another SiMLDA algorithm, the iterative version of MLHES (IMLHES) [37], was designed to handle this problem. Numerical experiments show that use of iterations improves performance of MLHES [37], but IMLHES is not without limitations. It is not obvious how to optimally formulate convergence criteria for the different levels of IMLHES. This could cause failed iterations and additional computational cost.

There are indications that utilization of sequential MLDA (SeMLDA) algorithms can benefit from certain properties of the problem of estimation of permeability from inverted seismic data. Firstly, analytical [42] and numerical [43] results show that sequential DA is expected to outperform simultaneous DA for weakly nonlinear problems. Secondly, for a class of problems (including the problem considered here) where the model forecast can be seen as a spatially integrated response to a spatially varying parameter field, there exists a correlation between small-scale oscillations in the parameter domain and the degree of nonlinearity of the mapping from parameter field to model forecast, see, e.g., [44, 45]. This correlation is such that coarsening the simulation grid and upscaling the associated parameters will generally result in weaker nonlinearity in the coarser forward models compared to the finer ones. Taking advantage of this, we consider several resolutions of inverted seismic data, and develop a SeMLDA scheme which first assimilates the coarsest resolution of the data corresponding to the coarsest forward model, followed by the data in higher resolution corresponding to more non-linear models. This construction corresponds to the optimal ordering of data as suggested by [42, 43]. Note that, here, the term 'sequential' pertains to a sequence of data with different resolutions, not data at different times.

In this work, we will introduce two variants of a SeMLDA algorithm and assess their performance in comparison with a conventional sequential DA algorithm, a conventional simultaneous DA algorithm, and a SiMLDA algorithm. This will be done with the help of numerical experiments.

The rest of this paper is organized as follows. Since the SeMLDA methods developed in this paper are partially inspired by the ESMDA, the ESMDA will be briefly presented in Section 2. Section 3 describes MLDA in general, and introduces two variants of a novel SeMLDA algorithm. Section 4 explains the test models used for numerical investigation. In Section 5 we describe the numerical investigations, which are followed by their results in Section 6. Finally, in Section 7 we summarize and conclude the paper.

2 Ensemble smoother with multiple data assimilation

The forward models used in the parameter estimation process are often nonlinear. In the case of assimilation of inverted seismic data, the nonlinearity comes from both fluid flow equations and rock physics modeling. This nonlinearity has motivated the development of several DA algorithms, including ESMDA [2]. This algorithm assimilates the same data N_a times while inflating the data-error covariance matrix. By doing so, several small assimilation steps are taken instead of one big assimilation step. This helps to better account for the nonlinearity of the problem.

At step i of ESMDA, an ensemble of N_e realizations $\{z_{i,j}^{\text{pri}}\}_{j=1}^{N_e}$ from the prior parameter vector Z_i^{pri} is considered. The ESMDA update consists of three steps.

Firstly, running the forward simulator for every realization, $z_{i,j}^{\text{pri}}$, $1 \leq j \leq N_e$, we have

$$y_{i,j} = \mathcal{M}(z_{i,j}^{\text{pri}}) \,, \tag{1}$$

where $y_{i,j}$ is a realization from the model forecasts random vector at assimilation step i, Y_i.

Secondly, the original observation data-error model, $D \sim \mathcal{N}(\mathbb{E}(D), \mathbb{C}(D))$, is slightly manipulated such that the random observation data vector at step i is given as $\widehat{D}_i \sim \mathcal{N}(\mathbb{E}(D), \alpha_i \mathbb{C}(D))$. The data-error covariance matrix is inflated using the scalar value α_i so that multiple assimilations of the same data does not violate Bayes' rule. Hence, the updated parameter vector of an arbitrary ensemble member is given by

$$z_{i,j}^{\text{pos}} = z_{i,j}^{\text{pri}} + K_i(\widehat{d}_{i,j} - y_{i,j}) \,, \tag{2}$$

where $\widehat{d}_{i,j}$ is a realization of $\widehat{D}_i$, and the Kalman gain, K_i, is given as

$$K_i = \mathrm{C}(Z_i^{\text{pri}}, Y_i) \left(\mathrm{C}(Y_i) + \alpha_i \mathbb{C}(D)\right)^{-1} \,. \tag{3}$$

The terms $\mathrm{C}(Y_i)$ and $\mathrm{C}(Z_i^{\text{pri}}, Y_i)$ denote the the covariance of Y_i and cross-covariance between Z_i^{pri} and Y_i, respectively.

Finally, while $i < N_a$, the prior ensemble at step $i+1$ is set equal to the posterior ensemble at step i;

$$z_{i+1,j}^{\text{pri}} \coloneqq z_{i,j}^{\text{pos}} \,, \tag{4}$$

for $1 \leq j \leq N_e$.

In general, the likelihood can be written as a product of inflated likelihoods. This process, known as tempering, fulfills Bayes theorem exactly. ESMDA is a special case of tempering. If the forward model, $\mathcal{M}$, is a linear model and the distributions of the prior parameters and the data errors are Gaussian, there exists a condition (denoted the MDA condition) for α_i which ensures that the ESMDA samples correctly from the posterior distribution. This condition is given as

$$\sum_{i=1}^{N_a} \frac{1}{\alpha_i} = 1 \,. \tag{5}$$

It is, however, unclear how important (5) is for problems where normality or linearity does not hold.

3 Multilevel Data Assimilation

In MLDA the forecast step is performed using a set of models which have different costs and fidelities. Here, we define $\mathcal{M}_L := \mathcal{M}$, and $\{\mathcal{M}_l\}_{l=1}^{L-1}$ where $\mathcal{M}_l$ is an approximation to $\mathcal{M}_L$ with increasing accuracy and computational cost as l increases. We will denote $\{\mathcal{M}_l\}_{l=1}^{L}$ an ML model.

3.1 Multilevel Model

Members of an ML model form hierarchies of both accuracy and computational cost. One can think of several schemes to devise the hierarchy including, but not limited to, coarsening the spatial grid of the forward model, increasing the length of the forward model time steps, and relaxing the convergence criteria in the iterative linear solvers. All of these methods reduce the computational cost of the models and increase their numerical error. Coarsening the spatial grid and performing simulations on such grids is chosen for the current work (Note, however,

that the parameters that we estimate are kept in the fine grid, meaning that upscaling the parameters is considered as part of the ML forward models).

Fossum and Mannseth [32] proposed a robust technique for grid coarsening, which was also used in [33, 41, 37]. In each coarsening step, neighboring cells of the grid at the previous step are merged into a coarser cell unless they are to be kept fine deliberately. A representation of the grid coarsening process for an 8×8 example grid can be found in Figure 1. The figure illustrates that coarsening has occurred in a uniform manner across both directions, except for the vicinity of two opposite corners, where the grid cells are kept in fine scale to boost the local numerical accuracy around the two wells, producer (P) and injector (I). The aim is that the grid coarsening should not change the physics of the problem too much.

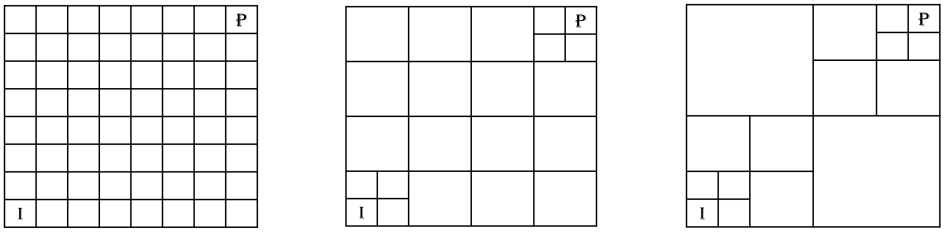


Figure 1: Grid coarsening proposed by [32] performed on an 8×8 grid (a) Finest level (b,c) Coarser levels

3.2 Transformation of Model Forecasts

The discrepancy in coarseness of the ML grids results in the spatially distributed model forecasts to be in different resolutions for different levels. Therefore, in order to be able to compute the ML sample statistics of model forecasts, a robust transformation scheme should be devised for converting a model forecast from the resolution at one level to another.

In the problem at hand, transformation of the model forecast requires either upscaling or downscaling. A standard volume-weighted arithmetic averaging technique is used for upscaling. Since the grids of the ML model used here have a nested structure, downscaled model forecasts are simply put equal to the corresponding coarse grid values. Accordingly, both upscaling and downscaling are linear transformations of model forecasts. Hence, we define a set of linear transformations, $\{U_f^c : \mathbb{R}^{\zeta_f} \mapsto \mathbb{R}^{\zeta_c} | 1 \leq c, f \leq L\}$, where ζ_f and ζ_c denote the dimension of model forecast vector at arbitrary levels f and c, respectively, and U_f^c transforms the model forecast vector from level f to level c.

3.3 Upscaling of Observation Data

As part of the DA process, the mismatch between the model forecasts and observation data needs to be calculated. Here, it is assumed that inverted seismic data is given in the resolution of the finest simulation grid, level L. Therefore, for each of the levels, either the observation data should be upscaled to the resolution of the model forecasts or the model forecasts should be downscaled to the resolution of the observation data. We take the former approach. Since the observation data is in the resolution of the finest model, using the same transformation functions as those designed for the model forecasts on the fine observation data will result in

upscaling of observation data into the required resolution. Hence, the transformed random vector of observation data at level l is given as

$$D_l = U_L^l D \,. \tag{6}$$

3.4 Sequential Multilevel Data Assimilation

The idea of sequential assimilation of spatially distributed data resonates well with the nature of the ML model used. Since coarsening the model is expected to result in weaker nonlinearity [44, 45], utilization of coarse models in the first steps of the DA followed by more non-linear fine models in the next steps can help to gradually zoom in on the solution of the inverse problem. Therefore, in SeMLDA the observation data are upscaled into several levels corresponding to the levels of the ML model. Afterwards, the data are assimilated sequentially starting from the coarsest resolution followed by the finer ones.

3.4.1 Sequential Multilevel Ensemble Smoother

In this section we discuss two versions of the Sequential Multilevel Ensemble Smoother (SMLES). This algorithm draws on the ESMDA algorithm [2], MLHES algorithm [33], and constraints associated with assimilation of linearly dependent data [46].

Initially, based on the available computational resources, the number of simulations performed on each level should be decided. Since the simulations are cheaper on the coarser levels, a higher number of simulations will be performed on those levels. Considering this, a decision is made on the resource allocation. Based on the resource allocation, a sample of N_1 ensemble members is generated based on the prior information. Afterwards, at step l the model pertaining to level l is used to assimilate the data transformed to match the resolution of the model forecasts at that level. Running the forward simulator for every realization, $z_{l,j}^{\text{pri}}$, $1 \leq j \leq N_l$, from the prior parameters random vector at level l, Z_l^{pri}, we have

$$y_{l,j} = \mathcal{M}_l(z_{l,j}^{\text{pri}}) \,, \tag{7}$$

where $y_{l,j}$ is a realization from the model forecasts random vector at step l, Y_l.

In the analysis step, realizations from the random observation data vector at level l, $\widetilde{D}_l \sim \mathcal{N}(U_L^l \mathbb{E}(D), \mathrm{C}(\widetilde{D}_l))$ are generated as $\{\widetilde{d}_{l,j}\}_{j=1}^{N_l}$. The form of $\mathrm{C}(\widetilde{D}_l)$ will be discussed in Section 3.4.2. The updated parameter vector of an arbitrary ensemble member is, then, given by

$$z_{l,j}^{\text{pos}} = z_{l,j}^{\text{pri}} + K_l^*(\widetilde{d}_{l,j} - y_{l,j}) \,. \tag{8}$$

Here, K_l^* is defined generically with * being a wildcard notation, and we introduce two flavors of the algorithm.

The straightforward flavor (SMLES-S) utilizes the data error and model at level l for formulation of the Kalman gain at step l. Accordingly, the Kalman gain is given as

$$K_l^S = \mathrm{C}(Z_l^{\text{pri}}, Y_l) \left(\mathrm{C}(Y_l) + \mathrm{C}(\widetilde{D}_l) \right)^{-1} \,, \tag{9}$$

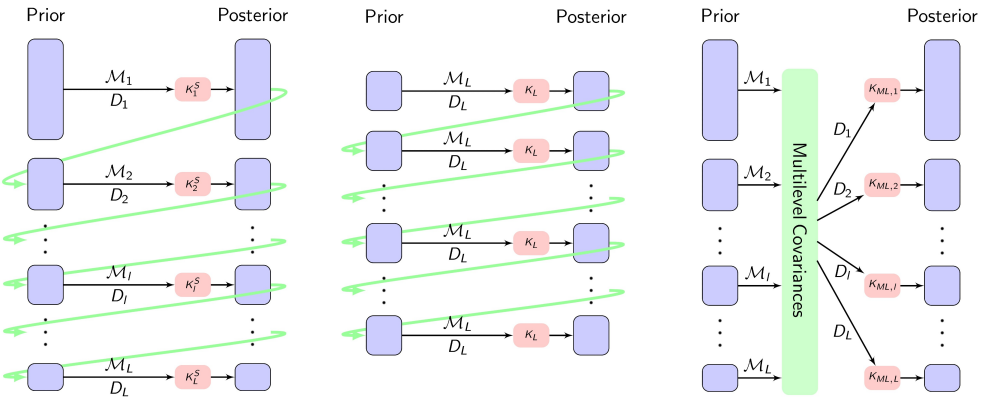
where $\mathrm{C}(Y_l)$ and $\mathrm{C}(Z_l^{\text{pri}}, Y_l)$ denote the covariance of Y_l and cross-covariance between Z_l^{pri} and Y_l, respectively.

Finally, while $l < L$, the prior for step $l+1$ is obtained from the posterior at level l as

$$z_{l+1,j}^{\text{pri}} \coloneqq z_{l,j}^{\text{pos}} \,, \tag{10}$$

for $1 \leq j \leq N_{l+1}$.

In SMLES-S only the realizations which have been simulated by all the models are considered for the final statistics. With a subtle manipulation of the data-error covariance in (9) for the realizations at level l which are not simulated by $\mathcal{M}_{l+1}$, the resulting posterior realizations in the linear-Gaussian case would sample correctly from the posterior distribution up to the data at level l. The condition that should hold at any level l is explained in Section 3.4.2.



(a) SMLES-S **(b)** ESMDA **(c)** MLHES

Figure 2: The schematic of two MLDA algorithms compared with ESMDA

Figure 2 represents the schematic of SMLES-S in comparison with ESMDA and MLHES algorithms. As can be seen from the figure, the ensemble size of SMLES-S shrinks as the level increases due to limited computational resources. Accordingly, at level L the ensemble size becomes small and it may end up in the situation that MLDA wants to avoid at the first place. A possible treatment would be utilization of localization in the finer levels, but that would reduce the robustness of the algorithm. Another alternative is to allow for transfer of information between different levels by utilization of ML statistics. By doing so, a hybrid version of the SMLES algorithm is formulated similar to MLHES and IMLHES. The hybrid Kalman gain can be formulated as

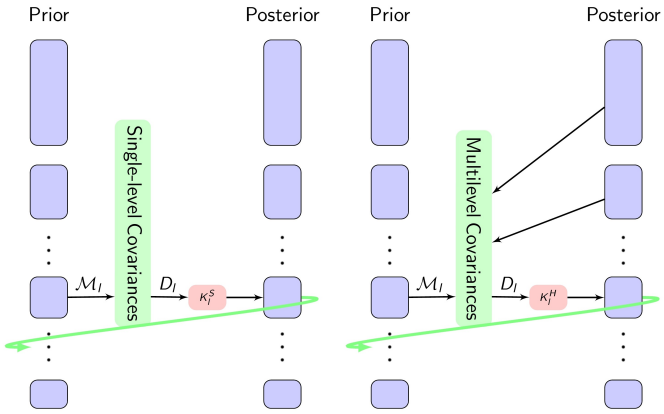
$$K_l^H = \mathrm{C}_l^H(Z,Y)\left(\mathrm{C}_l^H(Y_l) + \mathrm{C}(\widetilde{D}_l)\right)^{-1}, \tag{11}$$

where $\mathrm{C}_l^H(Y_l)$ and $\mathrm{C}_l^H(Z,Y)$ denote the ML covariance of Y_l and the ML covariance between Z and Y, respectively. The ML statistics will be discussed in Section 3.4.3.

The SMLES-H algorithm takes advantage of both conditioning all the realizations to the data up to the last level on which they are simulated and the ML statistics. A depiction of the difference between SMLES-S and SMLES-H is presented in Figure 3. Pseudo-codes of SMLES-S and SMLES-H are presented in Appendix A and Appendix B, respectively.

3.4.2 Partially Multiple Data Assimilation Condition

For the convenience of the reader, we present the main result from [46] (omitting the derivation). Similar to formulating a condition for assimilating a set of data multiple times such that the updated ensemble will sample correctly from the posterior in the linear-Gaussian case, there exists a condition which should hold if the data are linearly dependent, known as the Partially Multiple Data Assimilation (PMDA) condition [46].



(a) SMLES-S **(b)** SMLES-H

Figure 3: Step l of SMLES-S and SMLES-H

In the problem at hand, the data at any coarser level, D_c, is a linear transformation of the data at any finer level, D_f, such that for any $\{c, l, f\}$, $1 \leq c < l < f \leq L$, we have

$$U_f^c = U_l^c U_f^l \,. \tag{12}$$

Hence, we can define the PMDA condition for the data up to any arbitrary level $\mathcal{L}$, $1 \leq \mathcal{L} \leq L$.

Consider the data-error covariance matrix at level $l < \mathcal{L}$ to be given as

$$\mathrm{C}(\widetilde{D}_l) = A_l U_L^l \mathbb{C}(D) {U_L^l}^T {A_l}^T \,, \tag{13}$$

where A_l is an inflation matrix. Assuming the prior distribution is Gaussian, the forward model $\mathcal{M}_\mathcal{L}$ is linear, and for any coarser level c we have

$$\mathcal{M}_c = U_\mathcal{L}^c \mathcal{M}_\mathcal{L} \,, \tag{14}$$

the straightforward formulation of the SMLES algorithm, given in Section 3.4.1, samples correctly from the posterior distribution of the parameter random vector up to the data at level $\mathcal{L}$ if the following condition holds,

$$\sum_{l=1}^{\mathcal{L}} {U_\mathcal{L}^l}^T \mathrm{C}(\widetilde{D}_l)^{-1} U_\mathcal{L}^l = \left(U_L^\mathcal{L} \mathbb{C}(D) {U_L^\mathcal{L}}^T \right)^{-1} \,. \tag{15}$$

3.4.3 Multilevel statistics

Assuming we have approximations of the model forecasts, Y, being a function of the unknown parameter vector, Z, on several levels, a scheme for approximation of ML statistics for Y is required. As for MLDA, the mean and the covariance of model forecast are of foremost interest.

Assuming the model with the highest fidelity, $\mathcal{M}_L$, to be exact, and $\{\mathcal{M}_l\}_{l=1}^{L-1}$ to be approximations to $\mathcal{M}_L$ with a decreasing numerical error, [29] proposed a formulation for ML statistics. In this formulation, inspired by Bayesian Model Averaging, the statistics are computed based on reliability weights w_l for each of the levels l. This formulation is, by definition, a biased scheme for computation of ML moments; however, it will be a useful technique for problems in which variance error dominates bias, which is often the case for petroleum reservoir problems

[29]. Using this formulation and transformations of the forecast, [33] proposed a formulation of ML statistics for spatially distributed model forecasts. According to this scheme, the ML mean of the model forecast at level l is given as

$$\mathrm{E}_l^H(Y) = \sum_{k=1}^{L} w_k U_k^l \, \mathrm{E}(Y_k) \,, \tag{16}$$

$$\sum_{k=1}^{L} w_k = 1 \,, \tag{17}$$

where $\mathrm{E}(Y_k)$ denotes sample mean of the model forecast at level k. At step l the model forecasts pertaining to levels finer than l are not available, since the proposed algorithm is sequential. Hence, with a small change in the above-mentioned formulation, the formulation used in this work will be based on levels $1 \leq k \leq l$,

$$\mathrm{E}_l^H(Y) = \sum_{k=1}^{l} w_k U_k^l \, \mathrm{E}(Y_k) \,, \tag{18}$$

$$\sum_{k=1}^{l} w_k = 1 \,. \tag{19}$$

Using the law of total variance, the ML approximation of covariance of the model forecast at level l is formulated as

$$\mathrm{C}_l^H(Y) = \sum_{k=1}^{l} w_k \{ \mathrm{C}(U_k^l Y_k) + \big(\mathrm{E}(U_k^l Y_k) - \mathrm{E}_l^H(Y)\big) \big(\mathrm{E}(U_k^l Y_k) - \mathrm{E}_l^H(Y)\big)^T \} \,. \tag{20}$$

In addition, the parameter-forecast cross-covariance can be written as

$$\mathrm{C}_l^H(Z,Y) = \sum_{k=1}^{l} w_k \{ \mathrm{C}(Z_k, U_k^l Y_k) + \big(\mathrm{E}(Z_k) - \mathrm{E}^H(Z)\big) \big(\mathrm{E}(U_k^l Y_k) - \mathrm{E}_l^H(Y)\big)^T \} \,, \tag{21}$$

where $\mathrm{E}^H(Z)$ is the ML formulation of the parameter-vector mean. This statistic is formulated using the same weights as in forecasts ML statistics, but since the parameters are in the same resolution for all levels, no transformation is needed for formulating it,

$$\mathrm{E}^H(Z) = \sum_{k=1}^{l} w_k \, \mathrm{E}(Z_k) \,. \tag{22}$$

4 Test Models

Three different reservoir models are set up for numerical investigations. These reservoir models have some shared properties. They are two-dimensional with Cartesian grids. For all of them, compressible two-phase flow (oil and water), no-flow boundary conditions, and standard equations for capillary pressure and relative permeability, are considered. All the wells in these reservoir models are controlled by their bottom-hole pressure; the injectors at 300 bar, and the producers at 110 bar. A description of the other shared general properties of the reservoir models is given in Table 1. Unique features of the reservoir models are explained separately in Sections 4.1 - 4.3.

Table 1: Shared properties of the reservoir models

Fine cell dimensions:	$30 \times 30 \times 30\ (m^3)$	Porosity:	0.2
Initial Oil saturation:	0.85	Initial Pressure:	200 bar

The forward models consist of two segments. A reservoir flow model is used to predict the state variables, i.e. the pressure and saturation of the reservoir fluids, in time, and a petro-elastic model is utilized for computing the elastic rock properties from parameters and predicted state variables.

The flow segment of the forward model is derived by substitution of Darcy's law into the mass conservation equation for each of the fluid phases, resulting in [47]

$$\nabla.\left[\frac{k_{ro}}{\nu_o B_o} k(\nabla p_o - \rho_o g \nabla z)\right] = \frac{\partial}{\partial t}\left(\frac{\phi S_o}{B_o}\right) + q_o\ , \tag{23}$$

$$\nabla.\left[\frac{k_{rw}}{\nu_w B_w} k(\nabla p_w - \rho_w g \nabla z)\right] = \frac{\partial}{\partial t}\left(\frac{\phi S_w}{B_w}\right) + q_w\ , \tag{24}$$

where

$$S_o + S_w = 1\ , \tag{25}$$

$$p_{cow} = p_o - p_w\ . \tag{26}$$

In these equations, k denotes absolute permeability, and k_{r*} denotes the relative permeability of the corresponding phase, while $*$ is a wildcard notation. k_{r*} is a function of saturation of that phase, S_*. The pressure of a phase is denoted by p_*, and the capillary pressure, p_{cow}, is a function of S_w. Furthermore, g denotes the gravitational constant; ν_*, B_*, and ρ_* are the viscosity, the formation volume factor, and density of their corresponding phases; and q_* denotes the sink or source term of its corresponding continuity equation.

The flow segment of the forward models is performed using Eclipse-100 [48]. Coarsening the grid is done by using the Eclipse keyword COARSEN, which merges groups of pre-defined neighboring cells to form a coarser grid. The upscaling of permeabilities is performed using pore-volume weighted arithmetic averaging, and transmissibilities between two neighboring coarse cells in each direction are calculated based on harmonic averaging in that direction and summing it in other directions [48].

As for the petro-elastic segment of the forward model, an in-house model based on standard rock-physics equations [49], [50, Report 1] was used.

4.1 Reservoir Model I

This model has a 50×50 grid, and two wells, one producer (P) at center east and one injector (I) at center west. This model is designed to evaluate the performances of the different algorithms in parameter estimation of an oil reservoir with relatively long-range correlation in permeability field.

4.2 Reservoir Model II

This model has a 64×64 grid and five-spot well pattern, four injectors at the corners and a producer at the center of the field. This model is designed to assess the performances of the different algorithms in history-matching of a field with relatively short-range correlation length.

4.3 Reservoir Model III

This model has a 70×70 grid and two wells, an injection well in southwest corner and a production well in the northeast corner. This model has two permeability zones, one with a long-range correlation length and one with a short-range correlation length, and there exist a smooth transition from one zone to another.

5 Numerical Investigation

Three numerical experiments are conducted, one for each of the three reservoir models discussed in Section 4.

The unknown parameter field in all the experiments is the logarithmic permeability field, which has a different distribution in each of the experiments.

The observation data are two sets of time-lapse bulk-impedance data based on a baseline (day zero of production) and two vintages, which are different for each experiment and will be described separately. These observation data are generated using the results of simulation of a random draw from the prior parameter distribution. As inverted seismic data typically are spatially correlated, we use a non-diagonal covariance matrix for the data error, based on isotropic spherical variograms with mean 0. The ranges of the variograms are different for different experiments so that the robustness of the algorithms towards variogram range is assessed. The marginal standard deviation of each observation value is given as

$$\sigma = r \ \max\{|\delta|, \eta\} \,, \tag{27}$$

where $r = 0.1$, δ is the value of observation data at a certain location, and η is a threshold introduced to avoid too much certainty in observation data whose absolute values are very small. This threshold is defined as the 1^{st} smallest percentile of the absolute value of the observation data.

In order for comparison of SMLES with standard ESMDA, three algorithms are run; SMLES-S, SMLES-H, and ESMDA with localization (ESMDA-LOC). The gold standard for this comparison is the DA results obtained from vanilla ESMDA with an exceedingly large ensemble (ESMDA-REF). (Obviously, such an ensemble size would be computationally infeasible for a real application.)

In addition, in order for comparison of performance of SMLES with the iterative DA algorithms, the following algorithms are run: Iterative Ensemble Smoother (IES) [1] with localization (IES-LOC) and IMLHES. The gold standard for this comparison is the DA results obtained from vanilla IES with an exceedingly large ensemble (IES-REF).

What we want to estimate is the posterior distribution of the parameters and the model forecasts. The correct estimates would be given by Markov Chain Monte Carlo with an exceedingly large chain length. However, the focus of this work is comparing the novel MLDA algorithms with algorithms of the same class which are widely used in reservoir history matching, i.e. ensemble-based DA algorithms. Hence, ESMDA and IES with exceedingly large ensemble sizes are selected to remove the Monte Carlo errors and serve as the gold standards for comparison.

In order for these comparisons to be fair, there exists the "equal computational power" constraint. As the dominant cost of the DA process is pertaining to simulations of forward models, where iterative linear solvers dominate the computational costs for large problems, the computational cost relating to simulation of each ensemble member, using forward model $\mathcal{M}_l$, is assumed to be proportional to G_l^γ, where G_l is the number of the active grid cells in the forward model at level l, with $\gamma \in [1.25, 1.5]$, [51]. Here, we take $\gamma = 1.35$. Accordingly, the

computational power associated with each algorithm run can be written as

$$\Omega = \sum_{l=1}^{L} n_l G_l^{1.35} \, . \tag{28}$$

where n_l is the total number of simulations using $\mathcal{M}_l$. Using (28), the computational cost of the algorithm runs are set to be equal.

The iterative algorithms do not always have a fixed computational cost and many iterations are performed to satisfy the convergence criteria without considerable improvements in the objective function. Therefore more computational power (approximately twice the computational power of the other algorithms) was observed for these algorithms (IMLHES and IES-LOC).

Setting a fixed computational cost, there exists $L-1$ degrees of freedom for specification of the $\{N_l\}_{l=1}^{L}$ in the ML algorithms. No optimization was performed for this specification, the only aim pursued was to keep the size of sub-ensembles ascending with decreasing model cost. Several other similar settings that were tried resulted in similar DA outcomes.

The convergence criterion for the iterative algorithms was that improvements in the relative data mismatch should be smaller than 0.0001.

The localization scheme in ESMDA-LOC was based on covariance structure given in [52], spherical variogram, and the DA was performed using subspace inversion method proposed by [53]. As for IES-LOC, the localization scheme was based on covariance structure given in [52] and spherical variogram.

For the SMLES-H and IMLHES, there is a possibility to improve the results by tuning the weights in calculation of ML statistics for specific cases, but here we use the simplest choice–$\{w_l = \frac{1}{\mathcal{L}} \mid 1 \leq l \leq \mathcal{L}\}$.

Both ESMDA-LOC and ESMDA-REF are run with 6 steps, with $\alpha_i = 6$, $1 \leq i \leq 6$, in all the steps. Runs with higher number of steps were conducted but no major improvement in the DA results was observed.

As for SMLES-S, the inflation matrices are set as $A_l = L * I_{\zeta_l}$ for $1 \leq l < L$, where I_{ζ_l} is the identity matrix of size ζ_l. For $\mathcal{L} = L$, (15) is solved to calculate A_L. Unlike satisfying the MDA condition, which is computationally trivial, satisfying the PMDA condition is computationally very expensive for real field problems and becomes practically unfeasible. This is due to the presence of the costly inversions in (15). However, since the PMDA condition only secures correct sampling for linear-Gaussian problems, approximately satisfying this condition may suffice for the real field cases.

Regarding SMLES-H, the ensemble at level l is divided into two sub-ensembles; (I) the realizations which are simulated using $\mathcal{M}_{l+1}$ after the analysis step, $\{z_{l,j}^{\text{pri}} \mid 1 \leq j \leq N_{l+1}\}$, and (II) those that are not, $\{z_{l,j}^{\text{pri}} \mid N_{l+1} < j \leq N_l\}$. For Sub-ensemble (I) the inflation matrix is set as $A_l = L * I_{\zeta_l}$. For Sub-ensemble (II), (15) is solved for $\mathcal{L} = l$, so that $A_{\mathcal{L}}$ is calculated. This is done to assure that the PMDA condition is satisfied up the last level that each realization is simulated. In order to avoid additional computations, the ML statistics are calculated based on the posterior parameters and their corresponding linear update for the model forecasts.

5.1 Experiment I

This experiment is conducted on Reservoir Model I. The ML algorithms have four levels, corresponding to 157, 259, 685, and 2500 grid cells, respectively. A summary of the resource allocation for the different runs carried out in this experiment can be found in Table 2. All the numbers in the table are per assimilation step or iteration, except for SMLES-S and SMLES-H. For these two algorithms the total number of simulations performed at each of the levels are reported. This also holds for Tables 3 and 4 pertaining to Experiment II and Experiment III, respectively.

Table 2: A summary of resource allocation for different runs in experiment I

	level 1 $G_1 = 157$	level 2 $G_2 = 259$	level 3 $G_3 = 685$	level 4 $G_4 = 2500$
	N_1	N_2	N_3	N_4
ESMDA-REF	-	-	-	10000
SMLES-S	951	880	710	412
SMLES-H	1512	1134	756	378
ESMDA-LOC	-	-	-	100
IES-REF	-	-	-	10000
IMLHES	459	438	216	30
IES-LOC	-	-	-	100

The observation data for this experiment are generated based on seismic vintages at 2500 and 5000 days after production starts. The range of the variogram used for data-error covariance in this experiment is 15 grid cells.

The prior unknown logarithmic permeability field is based on an exponential variogram with mean and variance constant at 5 and 1, anisotropy angle and anisotropy ratio of 80 degrees and 0.7, and range 20 grid cells. Randomly selected realizations from this logarithmic permeability field can be found in Figure 4.

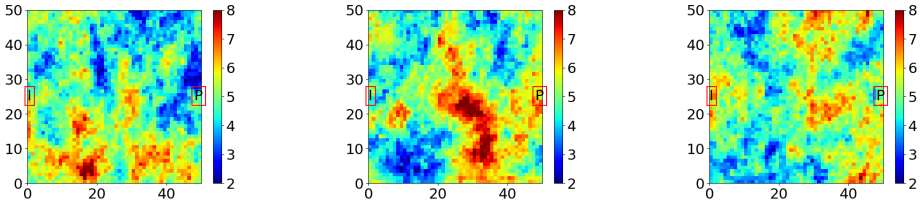


Figure 4: Randomly selected realizations from prior distribution of the logarithmic permeability field of Reservoir model I

5.2 Experiment II

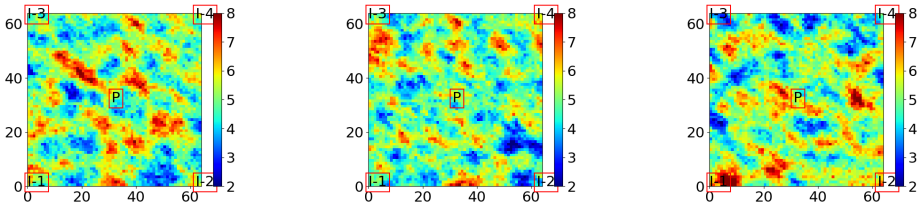
This experiment is conducted on Reservoir Model II. The ML algorithms have four levels, corresponding to 265, 433, 1147, and 4096 grid cells, respectively. A summary of the resource allocation for the different runs carried out in this experiment can be found in Table 3.

The observation data for this experiment are generated based on seismic vintages at 4000 and 8000 days after production starts. The range of the variogram used for data-error covariance in this experiment is 10 grid cells.

The prior unknown logarithmic permeability field is based on a spherical variogram with mean and variance constant at 5 and 1, the anisotropy angle and anisotropy ratio of -30 degrees and 0.7, and range 10 grid cells. Randomly selected realizations from this logarithmic permeability field can be found in Figure 5.

Table 3: A summary of resource allocation for different runs in experiment II

	level 1 $G_1 = 265$	level 2 $G_2 = 433$	level 3 $G_3 = 1147$	level 4 $G_4 = 4096$
	N_1	N_2	N_3	N_4
ESMDA-REF	-	-	-	10000
SMLES-S	1673	1380	1343	415
SMLES-H	1496	1122	748	374
ESMDA-LOC	-	-	-	100
IES-REF	-	-	-	10000
IMLHES	644	498	111	40
IES-LOC	-	-	-	100



(a) **(b)** **(c)**

Figure 5: Randomly selected realizations from prior distribution of the logarithmic permeability field of Reservoir model II

5.3 Experiment III

This experiment is conducted on Reservoir Model III. The ML algorithms have three levels, corresponding to 390, 1261, and 4900 grid cells, respectively. A summary of the resource allocation for the different runs carried out in this experiment can be found in Table 4.

Table 4: A summary of resource allocation for different runs in experiment III

	level 1 $G_1 = 390$	level 2 $G_2 = 1261$	level 3 $G_3 = 4900$
	N_1	N_2	N_3
ESMDA-REF	-	-	10000
SMLES-S	949	925	420
SMLES-H	1266	844	422
ESMDA-LOC	-	-	100
IES-REF	-	-	10000
IMLHES	768	246	32
IES-LOC	-	-	100

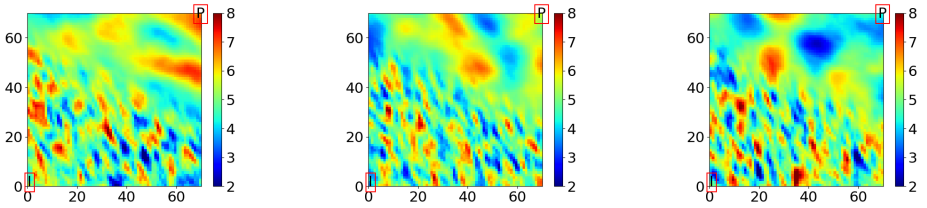
The observation data for this experiment are generated based on seismic vintages at 4000 and 8000 days after production starts. The range of the variogram used for data-error covariance in this experiment is 5 grid cells.

The unknown logarithmic permeability field is based on two different variograms in two zones of the field, one encompassing the northeastern part of the field and one encompassing

the southwestern part of the field with a smooth transition between them. The details about the variograms based on which the distribution of the unknown parameters are defined can be found in table 5. Randomly selected realizations from this logarithmic permeability field can be found in Figure 6.

Table 5: The two variograms of Reservoir Model III

	variance	mean	range	ratio	angle	type
Variogram 1	1	5	30	0.7	-30	cubic
Variogram 2	1	5	10	0.4	-70	cubic



(a) **(b)** **(c)**

Figure 6: Randomly selected realizations from prior distribution of the logarithmic permeability field of Reservoir model III

6 Numerical Results

The results from the numerical experiments are assessed both qualitatively and quantitatively, using the posterior parameters and forecasts.

As for qualitative analysis, firstly, mean and variance of the posterior parameter fields obtained by different algorithms are compared to the reference cases. Additionally, since the model forecasts are in different resolutions for different levels of the ML algorithms, comparison of the posterior model forecasts as such is not a possibility. Instead, we run fine-scale simulations of the posterior ensemble for all algorithms and then plot the mean and variance of the model forecasts for all the algorithms and compare them. The model forecasts of the second vintage are presented for all the experiments.

As for SMLES-H and IMLHES, the simplest formulation is chosen for computation of posterior statistics ($\mathrm{E}(Z^{\mathrm{pos}})$, $\mathrm{Var}(Z^{\mathrm{pos}})$, $\mathrm{E}(Y^{\mathrm{pos}})$, and $\mathrm{Var}(Y^{\mathrm{pos}})$). All posterior ensemble members at different levels are united into one ensemble and the empirical mean and variance are calculated.

The conventional DA algorithms (ESMDA-LOC and IES-LOC) were tested with several ranges for localization (critical distances), and the best results are presented for each of the experiments.

The quantitative analysis is performed using a measure suggested by [54] for comparison of DA results obtained by different schemes with a reference case. Consider ν to be a vector of interest for quantitative analysis, e.g. parameter estimate vector or model forecast vector. The following two metrics are computed for each of the algorithm runs [54],

$$\epsilon_{\mathrm{Mean}} = \frac{|| \left(\mathrm{Mean}(\nu_*) - \mathrm{Mean}(\nu^{\mathrm{pri}})\right) - \left(\mathrm{Mean}(\nu^{\mathrm{pos}}) - \mathrm{Mean}(\nu^{\mathrm{pri}})\right) ||_2}{|| \left(\mathrm{Mean}(\nu^{\mathrm{pos}}) - \mathrm{Mean}(\nu^{\mathrm{pri}})\right) ||_2}\,, \tag{29}$$

$$\epsilon_{\text{Var}} = \frac{||\text{Var}(\nu_*) - \text{Var}(\nu^{\text{pos}})||_2}{||\text{Var}(\nu^{\text{pos}})||_2}\,, \tag{30}$$

where Mean and Var represent the empirical mean and variance for different cases; superscripts pri and pos denote the prior and the reference posterior, respectively; the subscript $*$ is a wildcard notation for the algorithm of interest; and the distance is measured in 2-norm. Here, the reference posterior is calculated based on the results obtained by ESMDA-REF for all the experiments.

When estimation of ν is concerned, the ϵ_{Mean} metric has the property that it will be close to 0 for the algorithms that perform similar to the reference and will be close to 1 for the algorithms whose estimate of ν are similar to the prior estimate. Similarly, the ϵ_{Var} metric calculates the difference between the variance obtained by the algorithm of interest and the reference variance normalized by the norm of the reference variance, meaning that smaller values are preferred.

We will compute these two metrics for both posterior parameter estimates and the second vintage of the posterior model forecasts.

Table 6: Summary of quantitative analysis of the experiments for posterior parameter estimates

	Experiment I		Experiment II		Experiment III	
	ϵ_{Mean}	ϵ_{Var}	ϵ_{Mean}	ϵ_{Var}	ϵ_{Mean}	ϵ_{Var}
ESMDA-REF	0.00	0.00	0.00	0.00	0.00	0.00
SMLES-S	0.76	0.79	0.76	0.70	0.55	0.51
SMLES-H	0.51	0.67	0.87	0.39	0.54	0.21
ESMDA-LOC	1.34	9.09	1.21	0.65	1.07	0.95
IES-REF	0.26	0.09	0.34	0.11	0.22	0.08
IMLHES	0.60	0.58	0.80	0.36	0.61	0.20
IES-LOC	1.05	2.00	0.97	0.63	1.08	0.64

Table 7: Summary of quantitative analysis of the experiments for posterior model forecasts

	Experiment I		Experiment II		Experiment III	
	ϵ_{Mean}	ϵ_{Var}	ϵ_{Mean}	ϵ_{Var}	ϵ_{Mean}	ϵ_{Var}
ESMDA-REF	0.00	0.00	0.00	0.00	0.00	0.00
SMLES-S	0.16	0.58	0.076	0.46	0.037	0.30
SMLES-H	0.12	8.23	0.13	4.28	0.07	14.23
ESMDA-LOC	1.26	84.48	0.16	4.11	0.075	4.73
IES-REF	0.09	0.21	0.029	0.47	0.011	0.39
IMLHES	0.16	11.28	0.12	5.80	0.19	19.88
IES-LOC	1.27	316.33	0.21	13.12	0.31	97.79

6.1 Results of Experiment I

Visual analysis of the mean permeability fields in Figure 7 shows that the SMLES-S and SMLES-H results are more similar to the ESMDA-REF result than the ESMDA-LOC result is. It also shows that IMLHES performs more similar to IES-REF than IES-LOC does. There does not seem to be any considerable advantage in performance of a specific ML algorithm when the mean permeability fields obtained by the ML algorithms are compared with the IES-REF

result. However, due to convergence issues mentioned above, the SiMLDA algorithm used more computational resources than SeMLDA algorithms. This holds for all the experiments.

Checking the variance fields in Figure 8 confirms that both SMLES-S and SMLES-H perform better than ESMDA-LOC. It further confirms that IMLHES performs more similar to IES-REF than IES-LOC does. However, in this figure it is evident that SMLES-S underestimates the uncertainty in the posterior parameters while SMLES-H and IMLHES overestimate it. No indication of superiority of either SMLES-H or IMLHES over each other is noticeable in the variance fields, and both perform slightly better than SMLES-S. It is worth noting that, changing the color scale of the plot of the variance field for SMLES-S (denoted by SMLES-S* in the figure), shows that, in spite of underestimation of uncertainty, the structure of variance field is predicted accurately by this algorithm.

Superiority of performance of ML algorithms over the performance of the conventional DA algorithms is further confirmed by checking the statistics of the model forecasts in Figures 9 and 10. There is no clear indication of advantage of either SMLES-H or IMLHES over each other. However, SMLES-S performs better than both of them particularly in estimating the variance field of the model forecasts.

Quantitative measures given in Tables 6 and 7 also confirm that the ML algorithms generally perform more similar to ESMDA-REF than the conventional DA algorithms. Among the ML algorithms, SMLES-H performs slightly better in estimation of both the posterior parameters and model forecasts, while IMLHES performs slightly better in estimation of the variance of the posterior parameter field, and SMLES-S performs significantly better in estimation of the variance of the posterior model forecasts.

The ESMDA-LOC and IES-LOC results presented here are based on the localization range of 40 grid cells (1200 meters).

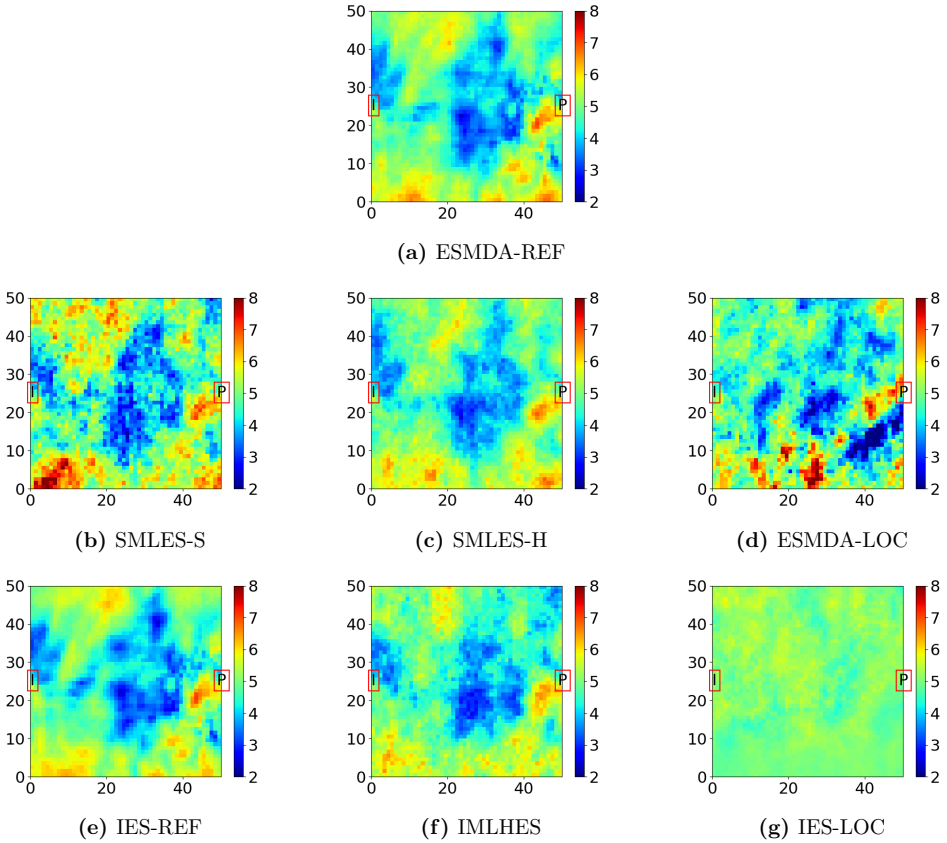
6.2 Results of Experiment II

Visual analysis of the mean permeability fields in Figure 11 shows that the SMLES-S and SMLES-H results are more similar to the ESMDA-REF result than the ESMDA-LOC result is. It also shows that IMLHES performs more similar to IES-REF than IES-LOC does. There does not seem to be any considerable advantage in performance of a specific ML algorithm when the mean permeability fields obtained by ML algorithms are compared with the IES-REF result.

Checking the variance fields in Figure 12 confirms that both SMLES-S and SMLES-H perform better than ESMDA-LOC. It further confirms that IMLHES performs more similar to IES-REF than IES-LOC does. However, in this figure it is evident that SMLES-S underestimates the uncertainty in the posterior parameters while SMLES-H and IMLHES overestimate it. No indication of superiority of either SMLES-H or IMLHES over each other is noticeable in the variance fields, and both perform slightly better than SMLES-S. As for Experiment I, changing the color scale of the plot of the variance field for SMLES-S (denoted by SMLES-S* in the figure), shows that, in spite of underestimation of uncertainty, the structure of variance field is predicted accurately by this algorithm.

Superiority of performance of the ML algorithms over the performance of the conventional DA algorithms is further confirmed by checking the statistics of the model forecasts in Figures 13 and 14. There is no clear indication of advantage of either SMLES-H or IMLHES over each other. However, SMLES-S performs better than both of them particularly in estimating the variance field of the model forecasts.

Quantitative measures given in Tables 6 and 7 also confirm that the ML algorithms generally perform more similar to ESMDA-REF than the conventional DA algorithms. All the MLDA algorithms perform reasonably similar in estimation of the mean of posterior parameters, while IMLHES and SMLES-H perform slightly better in estimation of the variance of the posterior



(a) ESMDA-REF

(b) SMLES-S **(c)** SMLES-H **(d)** ESMDA-LOC

(e) IES-REF **(f)** IMLHES **(g)** IES-LOC

Figure 7: Experiment I–Mean posterior logarithmic permeability field

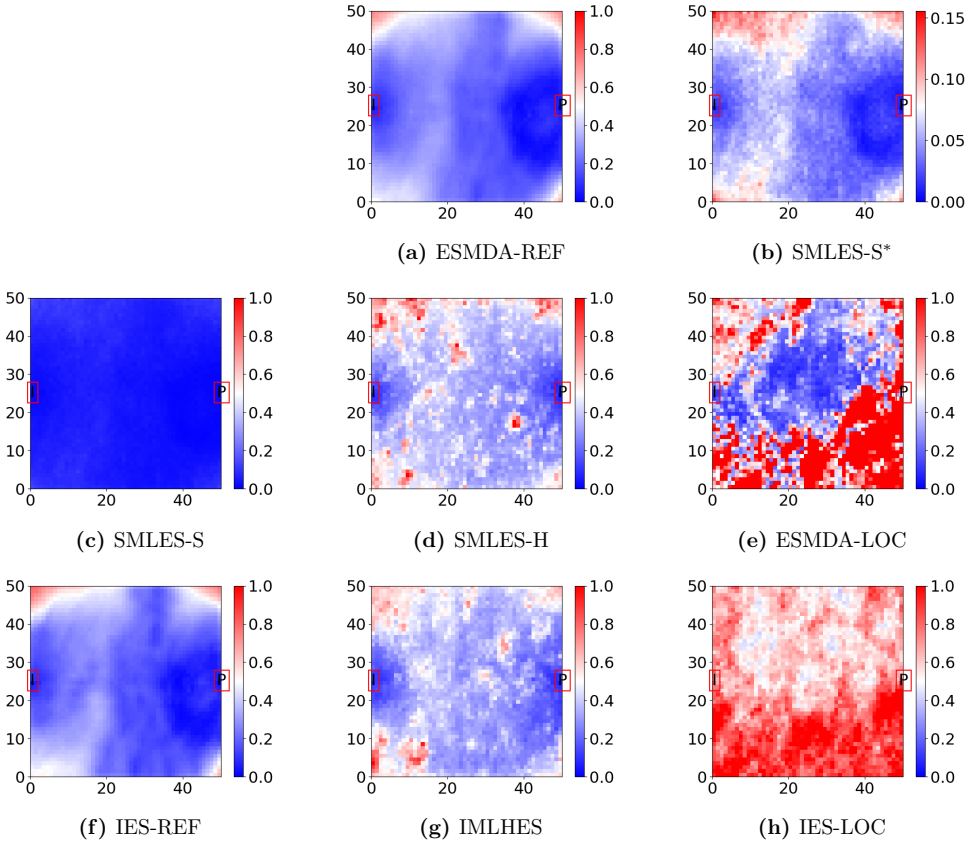
parameter field. As for estimation of the statistics of the posterior model forecasts, SMLES-S performs best, and its superiority is most evident in estimation of the variance of the posterior model forecasts.

The ESMDA-LOC and IES-LOC results presented here are based on the localization range of 60 grid cells (1800 meters).

6.3 Results of Experiment III

Visual analysis of the mean permeability fields in Figure 15 shows that the SMLES-S and SMLES-H results are more similar to the ESMDA-REF result than the ESMDA-LOC result is. It also shows that IMLHES performs more similar to IES-REF than IES-LOC does. There does not seem to be any considerable advantage in performance of a specific ML algorithm when the mean permeability fields obtained by ML algorithms are compared with the IES-REF result.

Checking the variance fields in Figure 16 confirms that both SMLES-S and SMLES-H perform better than ESMDA-LOC. It further confirms that IMLHES performs more similar to IES-REF than IES-LOC does. However, in this figure it is evident that SMLES-S underestimates the uncertainty in the posterior parameters while SMLES-H and IMLHES overestimate it. No indication of superiority of either SMLES-H or IMLHES over each other is noticeable in the variance fields, and both perform slightly better than SMLES-S. As for Experiments I and



(a) ESMDA-REF **(b)** SMLES-S* **(c)** SMLES-S **(d)** SMLES-H **(e)** ESMDA-LOC **(f)** IES-REF **(g)** IMLHES **(h)** IES-LOC

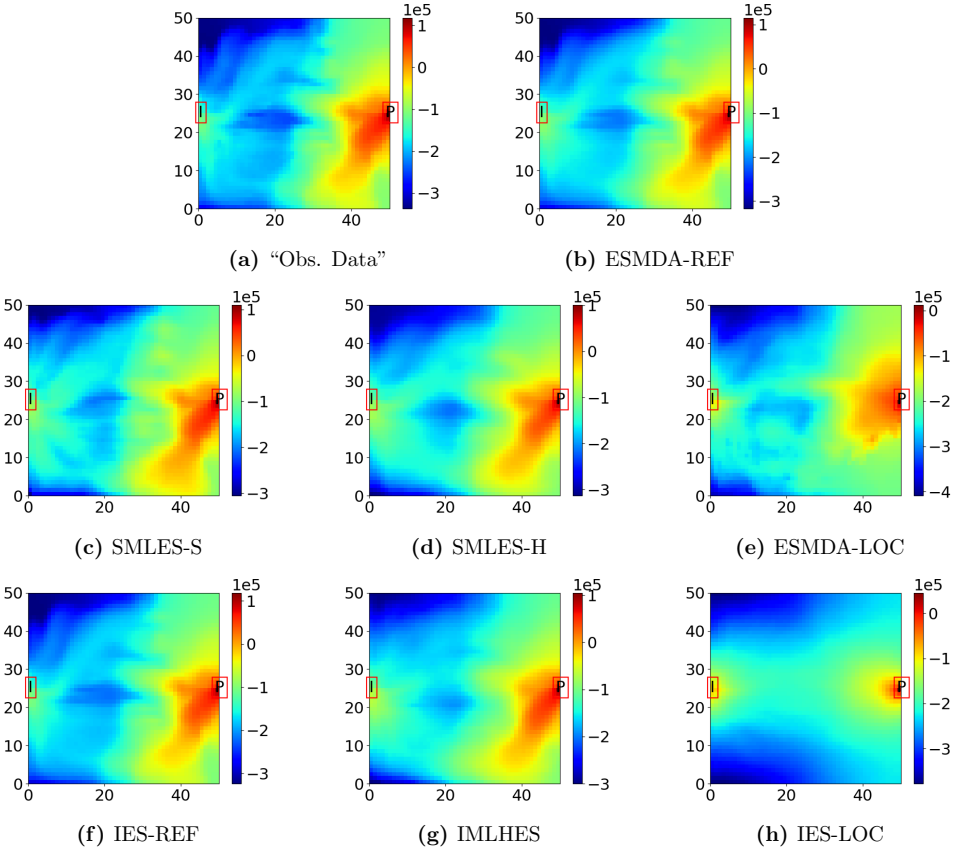
Figure 8: Experiment I–Variance of posterior logarithmic permeability field. The only difference between SMLES-S and SMLES-S* is in their scales.

II, changing the color scale of the plot of the variance field for SMLES-S (denoted by SMLES-S* in the figure), shows that, in spite of underestimation of uncertainty, the structure of variance field is predicted accurately by this algorithm.

Superiority of performance of the ML algorithms over the performance of the conventional DA algorithms is further confirmed by checking the statistics of the model forecasts in Figures 17 and 18. There is no clear indication of advantage of either SMLES-H or IMLHES over each other. However, SMLES-S performs better than both of them particularly in estimating the variance field of the model forecasts.

Quantitative measures given in Tables 6 and 7 also confirm that the ML algorithms generally perform more similar to ESMDA-REF than the conventional DA algorithms. All the MLDA algorithms perform reasonably similar in estimation of the mean of posterior parameters, while IMLHES and SMLES-H perform slightly better in estimation of the variance of the posterior parameter field. As for estimation of the statistics of the posterior model forecasts, SMLES-S performs best, and its superiority is most evident in estimation of the variance of the posterior model forecasts. SMLES-H performs slightly better than IMLHES in this regard.

The ESMDA-LOC and IES-LOC results presented here are based on the localization range of 60 grid cells (1800 meters).



(a) "Obs. Data" **(b)** ESMDA-REF

(c) SMLES-S **(d)** SMLES-H **(e)** ESMDA-LOC

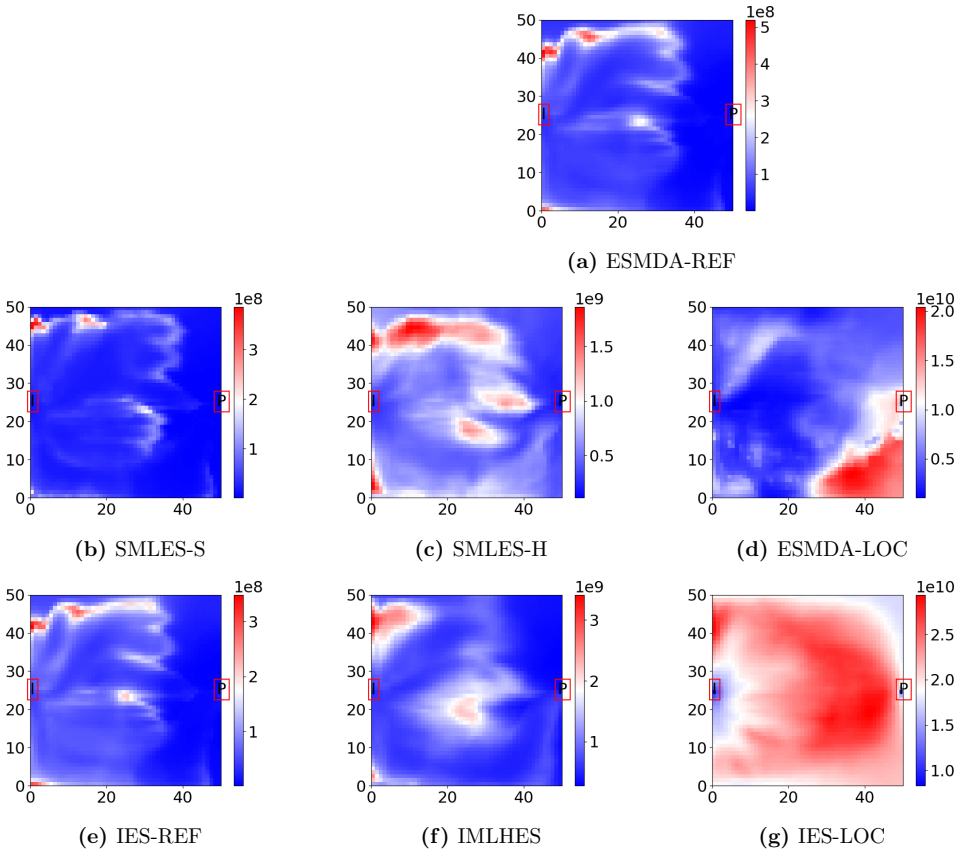
(f) IES-REF **(g)** IMLHES **(h)** IES-LOC

Figure 9: Experiment I–Mean of posterior time-lapse bulk impedance field ($\frac{m}{s}\frac{kg}{m^3}$) in comparison with observation data in the second vintage

7 Summary and Conclusions

In this work, two variants of a novel sequential MLDA algorithm, SMLES-S and SMLES-H, were introduced. In addition, performances of these algorithms were assessed in comparison with two conventional DA algorithms and a simultaneous MLDA algorithm. In doing so, three experiments were performed on three reservoir models. The three experiments were designed such that the performance of the algorithms were evaluated in different settings for the prior parameter fields (different variogram types; different anisotropies; and various correlation lengths including long-range correlation, short-range correlation, and mixture of long-range and short-range correlations) and different ranges for the variograms used for the data-error covariance. Each of the experiments consisted of seven algorithm runs: SMLES-S, SMLES-H, ESMDA with localization (ESMDA-LOC), vanilla ESMDA with an exceedingly large ensemble (ESMDA-REF), an iterative MLDA algorithm (IMLHES), iterative ensemble smoother with localization (IES-LOC), and vanilla iterative ensemble smoother with an exceedingly large ensemble (IES-REF). Results of the experiments were assessed both qualitatively and quantitatively.

In order for qualitative evaluation of the numerical results, firstly, the mean and the variance of posterior parameter fields were generated and assessed visually. The relative performances of the different methods were similar for all three experiments. The assessments showed that both



(a) ESMDA-REF

(b) SMLES-S **(c)** SMLES-H **(d)** ESMDA-LOC

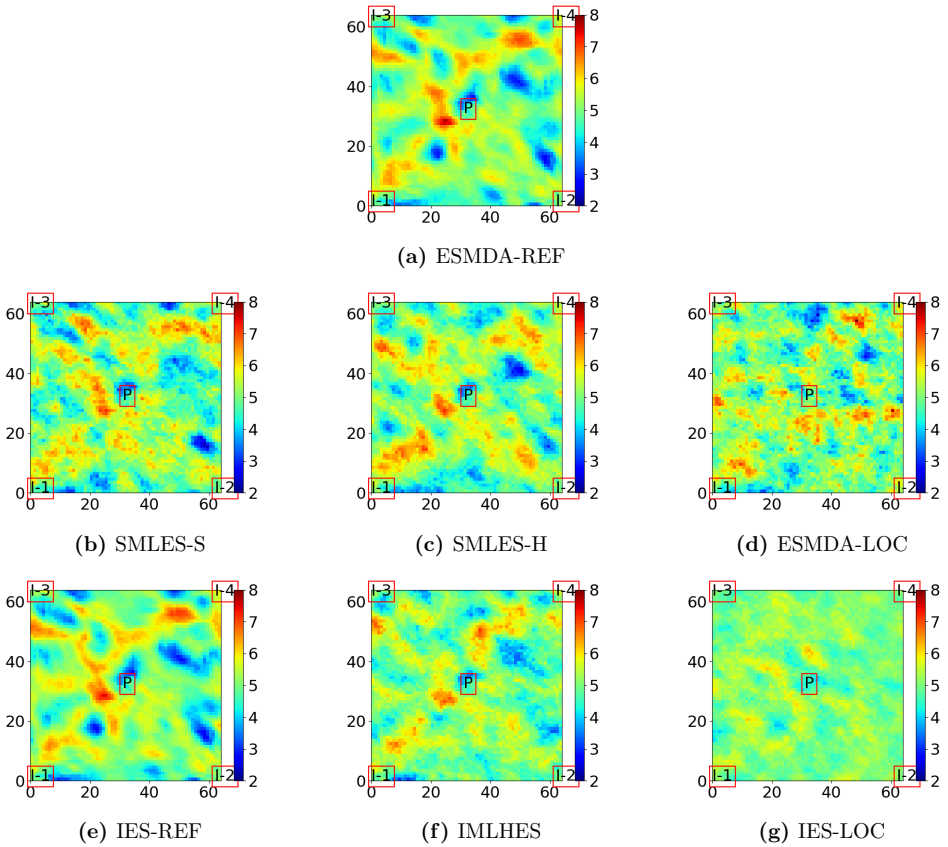
(e) IES-REF **(f)** IMLHES **(g)** IES-LOC

Figure 10: Experiment I–Variance of posterior time-lapse bulk impedance field ($(\frac{m}{s}\frac{kg}{m^3})^2$), in the second vintage. Note that the color bars have very different scales.

SMLES-S and SMLES-H performed more similar to ESMDA-REF than ESMDA-LOC did in estimation of the posterior parameter mean field. Regarding estimation of the variance fields, SMLES-H overestimated the variance while SMLES-S underestimated it. The superiority of performance of both SeMLDA algorithms over ESMDA-LOC was evident, also for the variance fields. Among the iterative algorithms, IMLHES performed more similar to IES-REF than IES-LOC did. There was no indication of superior performance of either SMLES-H or IMLHES over each other in any of the experiment when their performances were compared to IES-REF results. However, IMLHES used more computational resources than either of the SeMLDA algorithms. Both IMLHES and SMLES-H performed slightly better than SMLES-S.

Additionally, fine-scale simulations were run for all the posterior ensembles obtained by the different algorithms in all the experiments. Plots of the mean and the variance of model forecasts from the different algorithms were compared to each other. Visual analysis of these plots showed that in all the experiments the ML algorithms performed better than their conventional DA counterparts. Among the ML algorithms, SMLES-S consistently performed better than both SMLES-H and IMLHES in estimation of the variance of the posterior model forecasts. Either of the other two MLDA algorithms did not have a clear advantage over each other.

Two metrics were adopted for quantitative comparison of the DA results obtained by different algorithms for estimation of both mean and variance of the posterior parameters and



(a) ESMDA-REF

(b) SMLES-S **(c)** SMLES-H **(d)** ESMDA-LOC

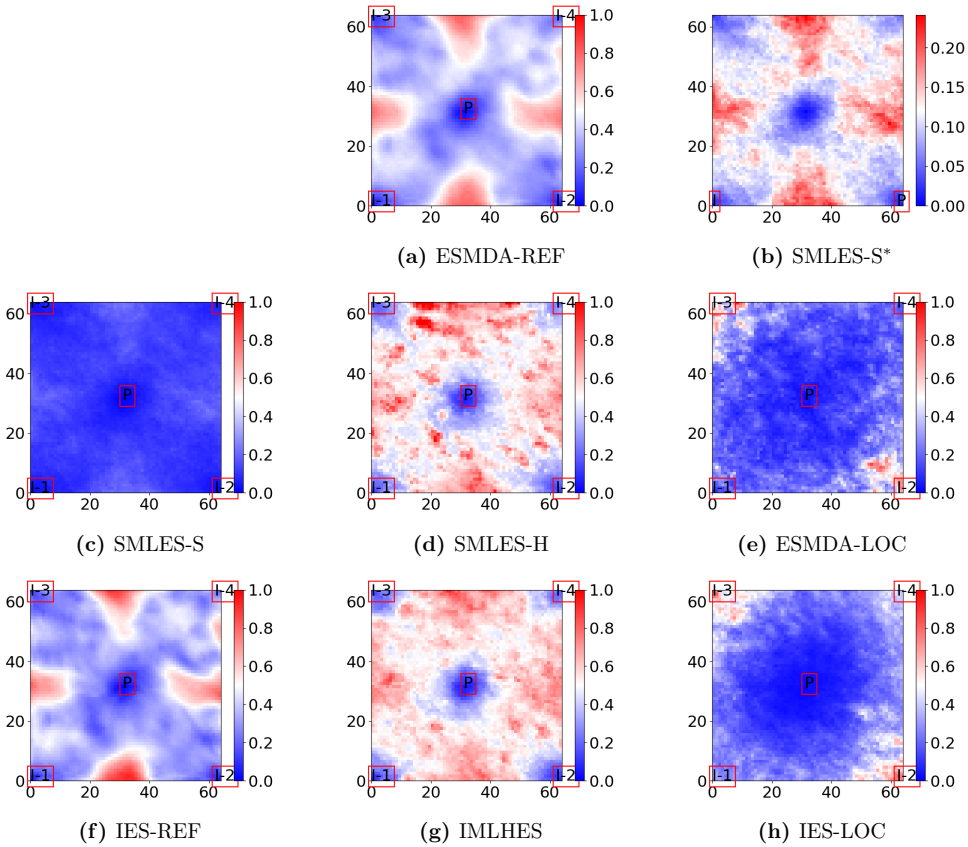
(e) IES-REF **(f)** IMLHES **(g)** IES-LOC

Figure 11: Experiment II–Mean posterior logarithmic permeability field

model forecasts. The metrics indicated that the ML algorithms generally performed better than the conventional DA algorithms in estimation of both mean and variance of the posterior parameters. They also indicated that SMLES-H and IMLHES performed slightly better than SMLES-S in estimation of the variance of the posterior parameters, and that all the MLDA algorithms performed better than IES-LOC in estimation of mean and variance of the posterior model forecasts. SMLES-S also performed consistently superior to ESMDA-LOC in estimation of mean and variance of the posterior model forecasts, while this was not observed for IMLHES and SMLES-H. Among the ML algorithms, SMLES-S clearly performed best when it came to estimation of the variance of the model forecasts. The other two algorithms did not consistently perform better than one another.

There were significant differences between the results from the MLDA algorithms and the results from the conventional DA algorithms in all the experiments. Simultaneous assimilation of large amounts of data into ensembles of small size partly explains the under-performance of the conventional algorithms. In the case of inverted seismic data we noticed true long-range correlations between the data and parameters. Regularization of the Kalman gain using distance-based localization, which disregards these true correlations and distorts the update formula, is another cause for the significant difference between the results.

Even though the results obtained from the experiments did not show any clear indication of superiority of SeMLDA over SiMLDA, they suggest that similar quality of DA results can be

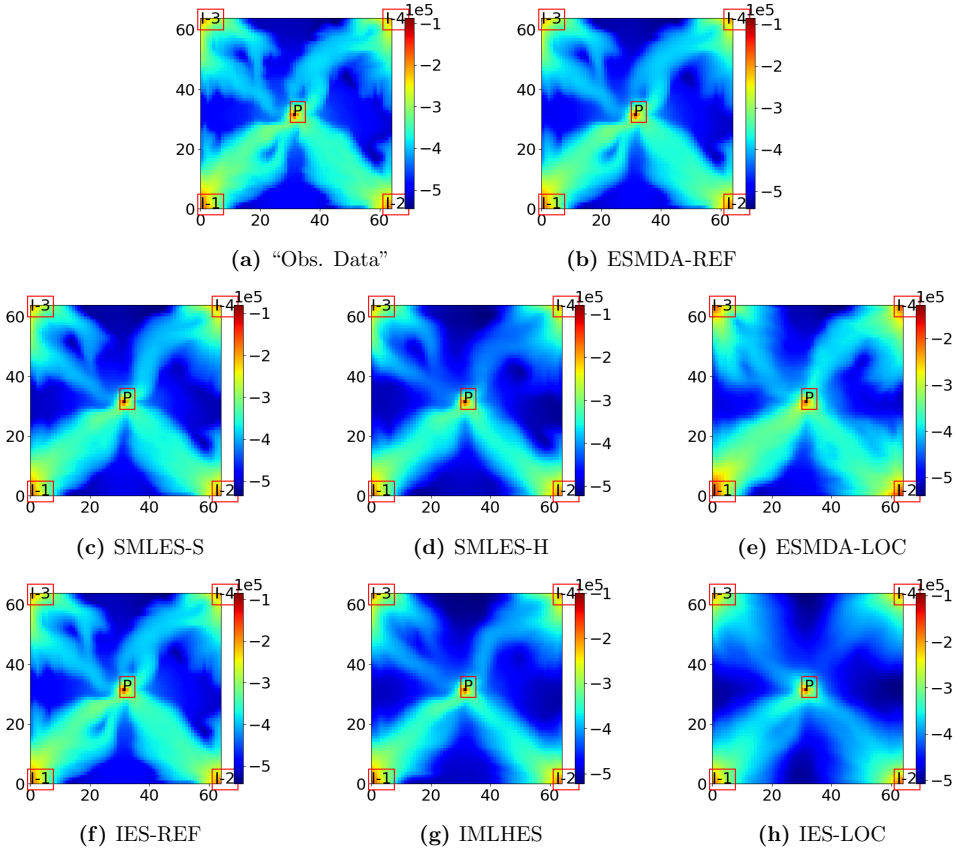


(a) ESMDA-REF **(b)** SMLES-S* **(c)** SMLES-S **(d)** SMLES-H **(e)** ESMDA-LOC **(f)** IES-REF **(g)** IMLHES **(h)** IES-LOC

Figure 12: Experiment II–Variance of posterior logarithmic permeability field. The only difference between SMLES-S and SMLES-S* is in their scales.

obtained by SeMLDA using a fixed and smaller computational power compared to SiMLDA.

There were several issues in this work that can be further investigated. As satisfying the PMDA condition is not a possibility in real field cases, devising robust techniques for approximately satisfying this condition can be studied. In order for optimization of both SeMLDA algorithms several of their characteristics can be further researched, e.g. the optimal extent of coarsening the grid, the number of levels and allocation of resources between them, the weights for different levels in the ML statistics, and the formulation of posterior statistics, to name a few. Additionally, generalizations of SMLES algorithms can also be studied, e.g. by assimilating the data more than once in some of the levels using more inflated data-error covariance matrices. Finally, as realistic reservoir cases are more complex than the fields investigated in this work, extensive coarsening of the grid may result in large numerical error and model bias. This limitation, and increase in the dimensionality of the parameters, may call for combination of localization and the proposed MLDA algorithms.



(a) "Obs. Data" **(b)** ESMDA-REF

(c) SMLES-S **(d)** SMLES-H **(e)** ESMDA-LOC

(f) IES-REF **(g)** IMLHES **(h)** IES-LOC

Figure 13: Experiment II–Mean of posterior time-lapse bulk impedance field ($\frac{m}{s}\frac{kg}{m^3}$) in comparison with observation data in the second vintage

Declarations

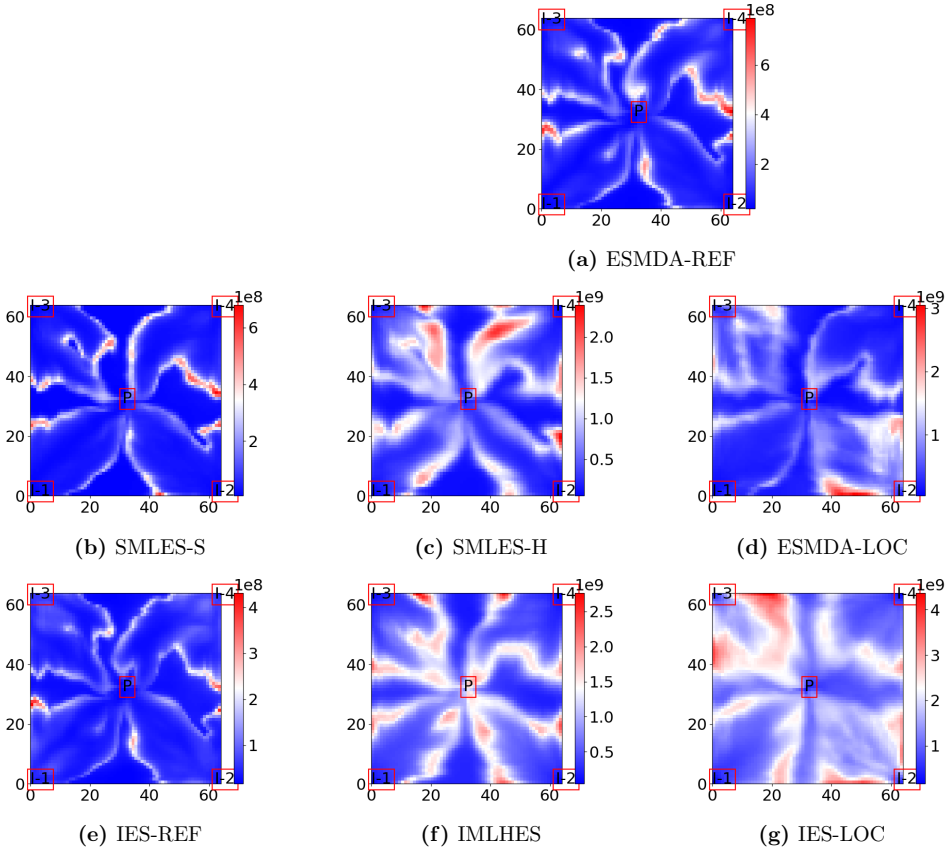
Compliance with Ethical Standards

Funding and competing interests

The authors acknowledge financial support from the NORCE research project "Assimilating 4D Seismic Data: Big Data Into Big Models" which is funded by industry partners, Equinor Energy AS, Lundin Energy Norway AS, Repsol Norge AS, Shell Global Solutions International B.V., TotalEnergies EP Norge AS, and Wintershall Dea Norge AS, as well as the Research Council of Norway (PETROMAKS2). We also thank Schlumberger for providing academic software licenses to ECLIPSE.

Financial interests

There are no financial interests to disclose.



(a) ESMDA-REF

(b) SMLES-S **(c)** SMLES-H **(d)** ESMDA-LOC

(e) IES-REF **(f)** IMLHES **(g)** IES-LOC

Figure 14: Experiment II–Variance of posterior time-lapse bulk impedance field ($(\frac{m}{s}\frac{kg}{m^3})^2$), in the second vintage. Note that the color bars have very different scales.

Non-financial interests

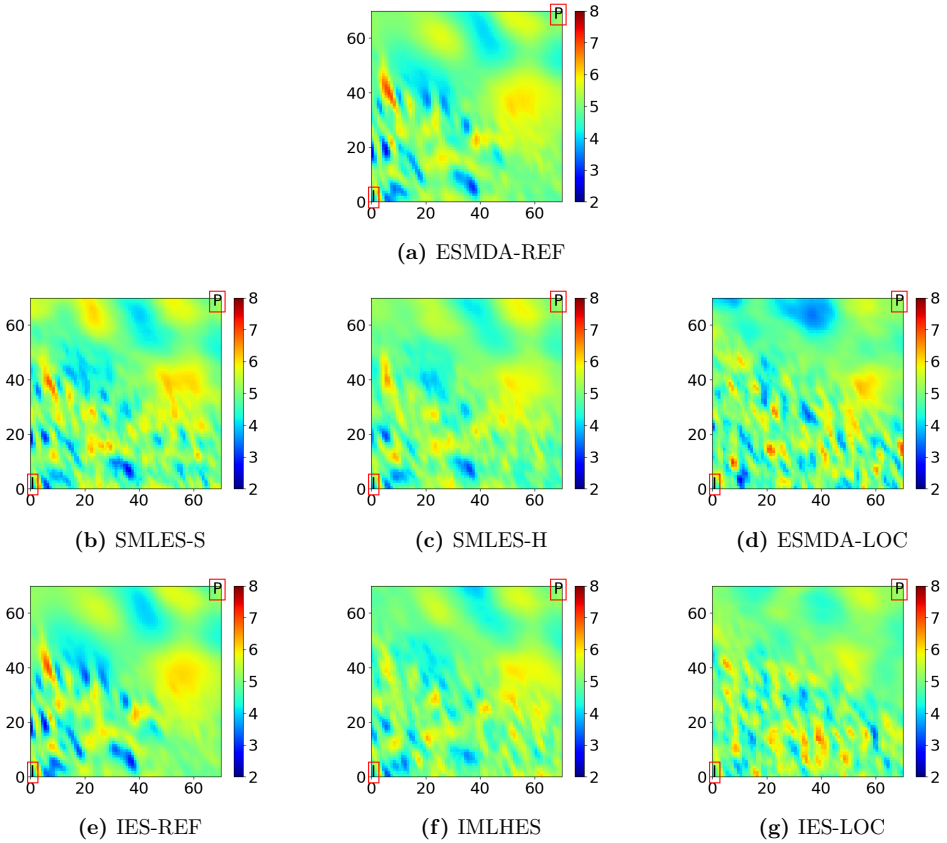
The fourth author is an editorial board member in Computational Geosciences. Otherwise, there are no non-financial interests.

A Algorithm. Sequential Multilevel Ensemble Smoother– Straightforward Formulation

B Algorithm. Sequential Multilevel Ensemble Smoother– Hybrid Formulation

References

[1] Yan Chen and Dean S Oliver. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Mathematical Geosciences*, 44(1):1–26, 2012.



(a) ESMDA-REF

(b) SMLES-S **(c)** SMLES-H **(d)** ESMDA-LOC

(e) IES-REF **(f)** IMLHES **(g)** IES-LOC

Figure 15: Experiment III–Mean posterior logarithmic permeability field

[2] Alexandre A Emerick and Albert C Reynolds. Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, 55:3–15, 2013.

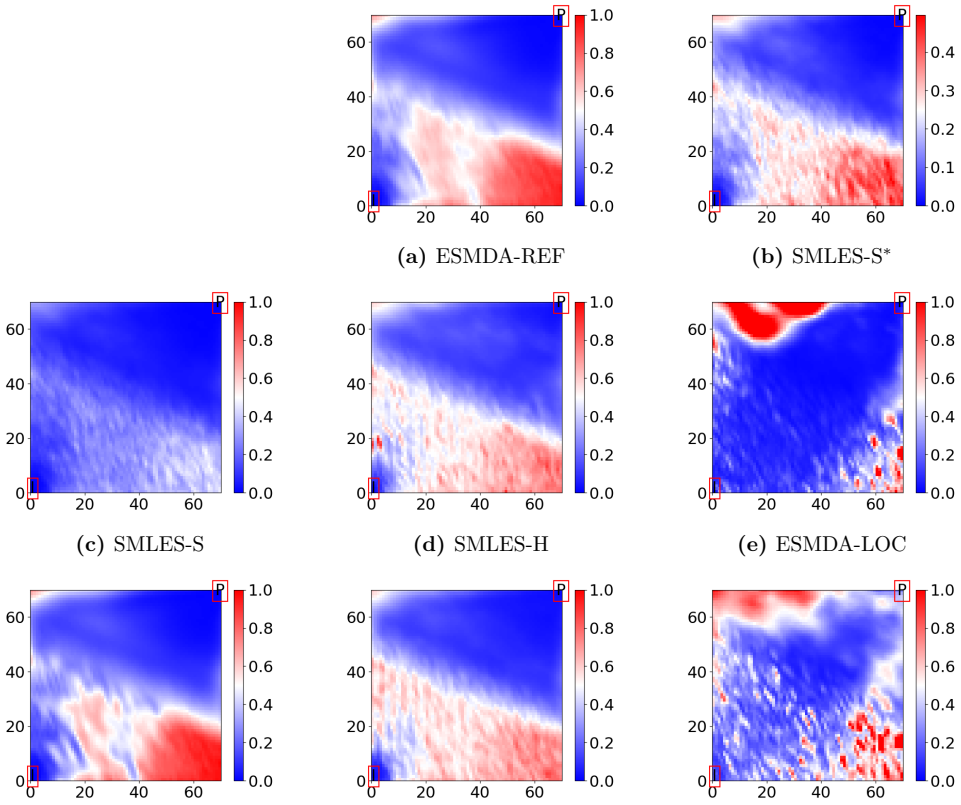
[3] Xiaodong Luo, Tuhin Bhakta, Geir Nævdal, et al. Correlation-based adaptive localization with applications to ensemble-based 4D-seismic history matching. *SPE Journal*, 23(02): 396–427, 2018.

[4] Rolf J Lorentzen, Tuhin Bhakta, Dario Grana, Xiaodong Luo, Randi Valestrand, and Geir Nævdal. Simultaneous assimilation of production and seismic data: application to the Norne field. *Computational Geosciences*, 24(2):907–920, 2020.

[5] Marco A Iglesias. Iterative regularization for ensemble data assimilation in reservoir models. *Computational Geosciences*, 19(1):177–212, 2015.

[6] Patrick Nima Raanes, Andreas Størksen Stordal, and Geir Evensen. Revising the stochastic iterative ensemble smoother. *Nonlinear Processes in Geophysics*, 26(3):325–338, 2019.

[7] Pavel Sakov, Jean-Matthieu Haussaire, and Marc Bocquet. An iterative ensemble Kalman filter in the presence of additive model error. *Quarterly Journal of the Royal Meteorological Society*, 144(713):1297–1309, 2018.



(a) ESMDA-REF **(b)** SMLES-S* **(c)** SMLES-S **(d)** SMLES-H **(e)** ESMDA-LOC **(f)** IES-REF **(g)** IMLHES **(h)** IES-LOC

Figure 16: Experiment III–Variance of posterior logarithmic permeability field. The only difference between SMLES-S and SMLES-S* is in their scales.

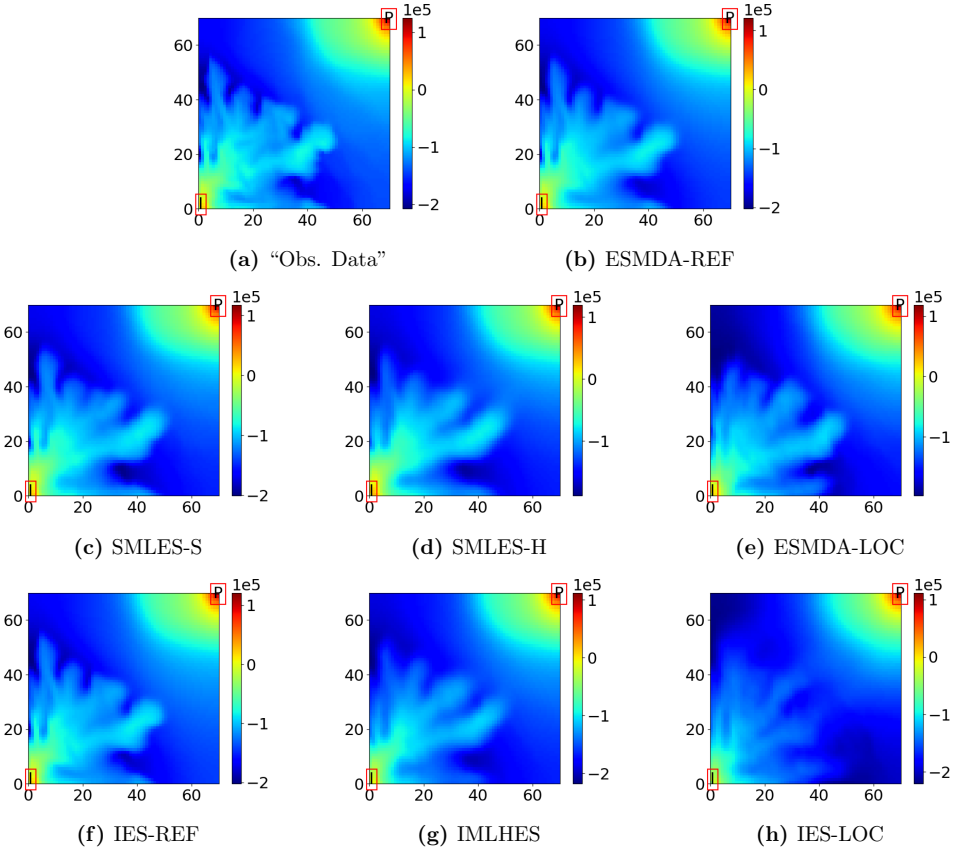
[8] Mario Trani, Rob Arts, Olwijn Leeuwenburgh, Jan Brouwer, and Sippe Douma. The importance of localization in the assimilation of 4D seismic data in the data assimilation process using the EnKF. In *SEG Technical Program Expanded Abstracts 2009*, pages 3835–3839. Society of Exploration Geophysicists, 2009.

[9] Mingliang Liu and Dario Grana. Time-lapse seismic history matching with an iterative ensemble smoother and deep convolutional autoencoder. *Geophysics*, 85(1):M15–M31, 2020.

[10] RV Soares, Xiaodong Luo, Geir Evensen, and Tuhin Bhakta. 4D seismic history matching: Assessing the use of a dictionary learning based sparse representation method. *Journal of Petroleum Science and Engineering*, 195:107763, 2020.

[11] Zhen Yin, Tao Feng, and Colin MacBeth. Fast assimilation of frequently acquired 4D seismic data for reservoir history matching. *Computers & Geosciences*, 128:30–40, 2019.

[12] Xiaodong Luo, Tuhin Bhakta, Morten Jakobsen, and Geir Nævdal. An ensemble 4D-seismic history-matching framework with sparse representation based on wavelet multiresolution analysis. *SPE Journal*, 22(03):985–1010, 2017.



(a) "Obs. Data" **(b)** ESMDA-REF

(c) SMLES-S **(d)** SMLES-H **(e)** ESMDA-LOC

(f) IES-REF **(g)** IMLHES **(h)** IES-LOC

Figure 17: Experiment III–Mean of posterior time-lapse bulk impedance field ($\frac{m}{s}\frac{kg}{m^3}$) in comparison with observation data in the second vintage

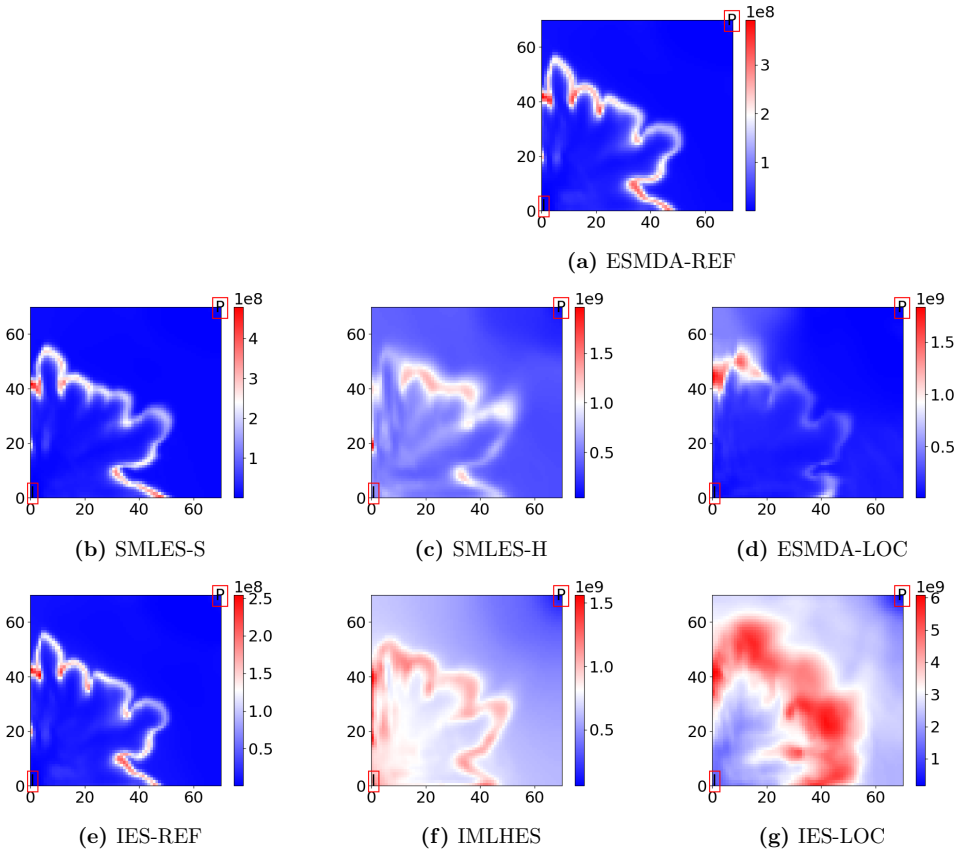
[13] JK Przybysz-Jarnut, JD Jansen, and A Gisolf. Joint assimilation of production data and time-lapse density changes from seismics using the representer method. In *ECMOR XI-11th European Conference on the Mathematics of Oil Recovery*, pages cp–62. European Association of Geoscientists & Engineers, 2008.

[14] Mario Trani, Rob Arts, and Olwijn Leeuwenburgh. Seismic history matching of fluid fronts using the ensemble Kalman filter. *SPE Journal*, 18(01):159–171, 2013.

[15] Yanhui Zhang and Olwijn Leeuwenburgh. Image-oriented distance parameterization for ensemble-based seismic history matching. *Computational Geosciences*, 21(4):713–731, 2017.

[16] Trond Mannseth and Kristian Fossum. Assimilating spatially dense data for subsurface applications–balancing information and degrees of freedom. *Computational Geosciences*, 22(5):1323–1349, 2018.

[17] Gilson Moura Silva Neto, Ricardo Vasconcellos Soares, Geir Evensen, Alessandra Davolio, and Denis José Schiozer. Subspace ensemble randomized maximum likelihood with local analysis for time-lapse-seismic-data assimilation. *SPE Journal*, 26(02):1011–1031, 2021.



(a) ESMDA-REF

(b) SMLES-S **(c)** SMLES-H **(d)** ESMDA-LOC

(e) IES-REF **(f)** IMLHES **(g)** IES-LOC

Figure 18: Experiment III–Variance of posterior time-lapse bulk impedance field ($(\frac{m}{s}\frac{kg}{m^3})^2$), in the second vintage. Note that the color bars have very different scales.

[18] Peter L Houtekamer and Herschel L Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123–137, 2001.

[19] Gregory Gaspari and Stephen E Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, 1999.

[20] Yan Chen and Dean S Oliver. Cross-covariances and localization for EnKF in multiphase flow data assimilation. *Computational Geosciences*, 14(4):579–601, 2010.

[21] Alexandre Emerick and Albert Reynolds. Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. *Computational Geosciences*, 15(2):251–269, 2011.

[22] Olwijn Leeuwenburgh, Jan Brouwer, and Mario Trani. Ensemble-based conditioning of reservoir models to seismic data. *Computational Geosciences*, 15(2):359–378, 2011.

[23] Alexandre A Emerick. Analysis of the performance of ensemble-based assimilation of production and seismic data. *Journal of Petroleum Science and Engineering*, 139:219–239, 2016.

Algorithm 1 SMLES–Original Formulation

Define $\{\mathcal{M}_l\}_{l=1}^{L}$
Define $\{A_l\}_{l=1}^{L}$
for $j = 1, ..., N_1$ **do**
  $z_{1,j}^{\text{pri}} \leftarrow \mathcal{N}\left(\mathrm{E}(Z^{\text{pri}}), \mathrm{C}(Z^{\text{pri}})\right)$
end for
for $l = 1, ..., L$ **do**
  for $j = 1, ..., N_l$ **do**
    $y_{l,j} \leftarrow \mathcal{M}_l(z_{l,j}^{\text{pri}})$
  end for
  $\mathrm{C}(\widetilde{D}_l) \leftarrow A_l U_L^l \mathbb{C}(D) {U_L^l}^T {A_l}^T$
  $K_l^S \leftarrow \mathrm{C}(Z_l^{\text{pri}}, Y_l)\left(\mathrm{C}(Y_l) + \mathrm{C}(\widetilde{D}_l)\right)^{-1}$
  for $j = 1, ..., N_l$ **do**
    $\widetilde{d}_{l,j} \leftarrow \mathcal{N}(U_L^l \mathbb{E}(D), \mathrm{C}(\widetilde{D}_l))$
    $z_{l,j}^{\text{pos}} \leftarrow z_{l,j}^{\text{pri}} + K_l^S(\widetilde{d}_{l,j} - y_{l,j})$
  end for
  for $j = 1, ..., N_{l+1}$ **do**
    $z_{l+1,j}^{\text{pri}} \leftarrow z_{l,j}^{\text{pos}}$
  end for
end for

[24] Miguel Alfonzo and Dean S Oliver. Seismic data assimilation with an imperfect model. *Computational Geosciences*, 24(2):889–905, 2020.

[25] Jincong He, Pallav Sarma, and Louis J Durlofsky. Reduced-order flow modeling and geological parameterization for ensemble-based data assimilation. *Computers & Geosciences*, 55:54–69, 2013.

[26] Mohammadali Tarrahi, Siavash Hakim Elahi, and Behnam Jafarpour. Fast linearized forecasts for subsurface flow data assimilation with ensemble Kalman filter. *Computational Geosciences*, 20(5):929–952, 2016.

[27] Kristian Fossum and Trond Mannseth. Coarse-scale data assimilation as a generic alternative to localization. *Computational Geosciences*, 21(1):167–186, 2017.

[28] Sam Hatfield, Aneesh Subramanian, Tim Palmer, and Peter Düben. Improving weather forecast skill through reduced-precision data assimilation. *Monthly Weather Review*, 146 (1):49–62, 2018.

[29] Kristian Fossum, Trond Mannseth, and Andreas S Stordal. Assessment of multilevel ensemble-based data assimilation for reservoir history matching. *Computational Geosciences*, 24(1):217–239, 2020.

[30] Håkon Hoel, Kody JH Law, and Raúl Tempone. Multilevel ensemble Kalman filtering. *SIAM Journal on Numerical Analysis*, 54(3):1813–1839, 2016.

[31] Rafael J de Moraes, Hadi Hajibeygi, and Jan Dirk Jansen. A multiscale method for data assimilation. *Computational Geosciences*, 24(2):425–442, 2020.

[32] K Fossum and T Mannseth. A novel multilevel method for assimilating spatially dense data. In *ECMOR XVI-16th European Conference on the Mathematics of Oil Recovery*, volume 2018, pages 1–12. European Association of Geoscientists & Engineers, 2018.

Algorithm 2 SMLES–Hybrid Formulation

Define $\{\mathcal{M}_l\}_{l=1}^{L}$
Define $\{A_l\}_{l=1}^{L}$
Define $\{w_l\}_{l=1}^{L}$
for $j = 1, ..., N_1$ **do**
  $z_{1,j}^{\text{pri}} \leftarrow \mathcal{N}\left(\mathrm{E}(Z^{\text{pri}}), \mathrm{C}(Z^{\text{pri}})\right)$
end for
for $\mathcal{L} = 1, ..., L$ **do**
  for $j = 1, ..., N_{\mathcal{L}}$ **do**
    $y_{\mathcal{L},j} \leftarrow \mathcal{M}_{\mathcal{L}}(z_{\mathcal{L},j}^{\text{pri}})$
  end for
  $\mathrm{E}(Z_{\mathcal{L}}) \leftarrow \frac{1}{N_{\mathcal{L}}-1} \sum_{j=1}^{N_{\mathcal{L}}} z_{\mathcal{L},j}^{\text{pri}}$
  $\mathrm{E}_{\mathcal{L}}^{\mathrm{H}}(\mathrm{Z}) \leftarrow \left(\sum_{l=1}^{\mathcal{L}} w_l \,\mathrm{E}(Z_l)\right) / \sum_{l=1}^{\mathcal{L}} w_l$
  $\mathrm{E}(Y_{\mathcal{L}}) \leftarrow \frac{1}{N_{\mathcal{L}}-1} \sum_{j=1}^{N_{\mathcal{L}}} y_{\mathcal{L},j}$
  $\mathrm{C}(Y_{\mathcal{L}}) \leftarrow \frac{1}{N_{\mathcal{L}}-1} \sum_{j=1}^{N_{\mathcal{L}}} (y_{\mathcal{L},j} - \mathrm{E}(Y_{\mathcal{L}}))(y_{\mathcal{L},j} - \mathrm{E}(Y_{\mathcal{L}}))^T$
  $\mathrm{C}(Z, Y_{\mathcal{L}}) \leftarrow \frac{1}{N_{\mathcal{L}}-1} \sum_{j=1}^{N_{\mathcal{L}}} (z_{\mathcal{L},j}^{\text{pri}} - \mathrm{E}(Z_{\mathcal{L}}))(y_{\mathcal{L},j} - \mathrm{E}(Y_{\mathcal{L}}))^T$
  $\mathrm{E}_{\mathcal{L}}^{\mathrm{H}}(Y) \leftarrow \left(\sum_{l=1}^{\mathcal{L}} w_l U_l^{\mathcal{L}} \,\mathrm{E}(Y_l)\right) / \sum_{l=1}^{\mathcal{L}} w_l$
  $\mathrm{C}_{\mathcal{L}}^{\mathrm{H}}(Z, Y) \leftarrow \sum_{l=1}^{\mathcal{L}} w_l \{\mathrm{C}(Z, Y_l) U_l^{\mathcal{L}^T} + \left(\mathrm{E}(Z_l) - \mathrm{E}_{\mathcal{L}}^{\mathrm{H}}(\mathrm{Z})\right) \left(U_l^{\mathcal{L}} \,\mathrm{E}(Y_l) - \mathrm{E}_{\mathcal{L}}^{\mathrm{H}}(Y)\right)^T\} / \sum_{l=1}^{\mathcal{L}} w_l$
  $\mathrm{C}_{\mathcal{L}}^{\mathrm{H}}(Y) \leftarrow \sum_{l=1}^{\mathcal{L}} w_l \{U_l^{\mathcal{L}} \,\mathrm{C}(Y_l) U_l^{\mathcal{L}^T} + \left(U_l^{\mathcal{L}} \,\mathrm{E}(Y_l) - \mathrm{E}_{\mathcal{L}}^{\mathrm{H}}(Y)\right) \left(U_l^{\mathcal{L}} \,\mathrm{E}(Y_l) - \mathrm{E}_{\mathcal{L}}^{\mathrm{H}}(Y)\right)^T\} / \sum_{l=1}^{\mathcal{L}} w_l$
  for $j = N_1, ..., N_{\mathcal{L}}$ **do**
    if $j \leq N_{\mathcal{L}+1}$ **then**
      $\mathrm{C}(\widetilde{D}_{\mathcal{L}}) \leftarrow A_{\mathcal{L}} U_L^{\mathcal{L}} \mathbb{C}(D) U_L^{\mathcal{L}^T} A_{\mathcal{L}}{}^T$
    else
      $\mathrm{C}(\widetilde{D}_{\mathcal{L}}) \leftarrow \left(\left(U_L^{\mathcal{L}} \mathbb{C}(D) U_L^{\mathcal{L}^T}\right)^{-1} - \sum_{l=1}^{\mathcal{L}-1} U_{\mathcal{L}}^{l\,T} \,\mathrm{C}(\widetilde{D}_l)^{-1} U_{\mathcal{L}}^{l}\right)^{-1}$
    end if
    $K_{\mathcal{L}}^{H} \leftarrow \left[\mathrm{C}_{\mathcal{L}}^{\mathrm{H}}(Z, Y)^T, \mathrm{C}_{\mathcal{L}}^{\mathrm{H}}(Y)\right]^T \left(\mathrm{C}_{\mathcal{L}}^{\mathrm{H}}(Y) + \mathrm{C}(\widetilde{D}_{\mathcal{L}})\right)^{-1}$
    $\widetilde{d}_{\mathcal{L},j} \leftarrow \mathcal{N}(\mathrm{E}(\widetilde{D}_{\mathcal{L}}), \mathrm{C}(\widetilde{D}_{\mathcal{L}}))$
    $\left[z_{\mathcal{L},j}^{\text{pos}T}, y_{\mathcal{L},j}^{\text{pos}T}\right]^T \leftarrow \left[z_{\mathcal{L},j}^{\text{pri}\,T}, y_{\mathcal{L},j}{}^T\right]^T + K_{\mathcal{L}}^{H}(\widetilde{d}_{l,j} - y_{l,j})$
  end for
  for $j = 1, ..., N_{l+1}$ **do**
    $z_{l+1,j}^{\text{pri}} \leftarrow z_{l,j}^{\text{pos}}$
  end for
  $\mathrm{E}(Z_{\mathcal{L}}) \leftarrow \frac{1}{N_{\mathcal{L}} - N_{\mathcal{L}+1} - 1} \sum_{j=N_{\mathcal{L}+1}+1}^{N_{\mathcal{L}}} z_{\mathcal{L},j}^{\text{pos}}$
  $\mathrm{E}(Y_{\mathcal{L}}) \leftarrow \frac{1}{N_{\mathcal{L}} - N_{\mathcal{L}+1} - 1} \sum_{j=N_{\mathcal{L}+1}+1}^{N_{\mathcal{L}}} y_{\mathcal{L},j}^{\text{pos}}$
  $\mathrm{C}(Y_{\mathcal{L}}) \leftarrow \frac{1}{N_{\mathcal{L}} - N_{\mathcal{L}+1} - 1} \sum_{j=N_{\mathcal{L}+1}+1}^{N_{\mathcal{L}}} (y_{\mathcal{L},j}^{\text{pos}} - \mathrm{E}(Y_{\mathcal{L}}))(y_{\mathcal{L},j}^{\text{pos}} - \mathrm{E}(Y_{\mathcal{L}}))^T$
  $\mathrm{C}(Z, Y_{\mathcal{L}}) \leftarrow \frac{1}{N_{\mathcal{L}} - N_{\mathcal{L}+1} - 1} \sum_{j=N_{\mathcal{L}+1}+1}^{N_{\mathcal{L}}} (z_{\mathcal{L},j}^{\text{pos}} - \mathrm{E}(Z_{\mathcal{L}}))(y_{\mathcal{L},j}^{\text{pos}} - \mathrm{E}(Y_{\mathcal{L}}))^T$
end for

[33] M Nezhadali, T Bhakta, K Fossum, and T Mannseth. A novel approach to multilevel data assimilation. In *ECMOR XVII*, volume 2020, pages 1–13. European Association of Geoscientists & Engineers, 2020.

[34] Håkon Hoel, Gaukhar Shaimerdenova, and Raúl Tempone. Multilevel ensemble Kalman filtering based on a sample average of independent EnKF estimators. *Foundations of Data Science*, 2(4):351, 2020.

[35] Andrey A Popov, Changhong Mou, Adrian Sandu, and Traian Iliescu. A multifidelity ensemble Kalman filter with reduced order control variates. *SIAM Journal on Scientific Computing*, 43(2):A1134–A1162, 2021.

[36] Gabriel Moldovan, Guillame Lehnasch, Laurent Cordier, and Marcello Meldi. A multigrid/ensemble Kalman filter strategy for assimilation of unsteady flows. *Journal of Computational Physics*, page 110481, 2021.

[37] Mohammad Nezhadali, Tuhin Bhakta, Kristian Fossum, and Trond Mannseth. Iterative multilevel assimilation of inverted seismic data. *Computational Geosciences*, 26(2):241–262, 2022.

[38] Peter Jan Van Leeuwen and Geir Evensen. Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review*, 124(12):2898–2913, 1996.

[39] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.

[40] Alexandre A Emerick and Albert C Reynolds. History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Computational Geosciences*, 16(3):639–659, 2012.

[41] Mohammad Nezhadali, Tuhin Bhakta, Kristian Fossum, and Trond Mannseth. Multilevel assimilation of inverted seismic data with correction for multilevel modeling error. *Frontiers in Applied Mathematics and Statistics*, 7, https://doi.org/10.3389/fams.2021.673077, 2021. doi: 10.3389/fams.2021.673077. URL `https://doi.org/10.3389/fams.2021.673077`.

[42] Kristian Fossum and Trond Mannseth. Parameter sampling capabilities of sequential and simultaneous data assimilation: I. analytical comparison. *Inverse Problems*, 30(11):114002, 2014.

[43] Kristian Fossum and Trond Mannseth. Parameter sampling capabilities of sequential and simultaneous data assimilation: II. statistical analysis of numerical results. *Inverse Problems*, 30(11):114003, 2014.

[44] Guy Chavent and Jun Liu. Multiscale parametrization for the estimation of a diffusion coefficient in elliptic and parabolic problems. *IFAC Proceedings Volumes*, 22(4):193–202, 1989.

[45] Alv Arne Grimstad and Trond Mannseth. Nonlinearity, scale, and sensitivity for parameter estimation problems. *SIAM Journal on Scientific Computing*, 21(6):2096–2113, 2000.

[46] Trond Mannseth. Assimilation of multiple linearly dependent data vectors. *Computational Geosciences*, 24(1):349–354, 2020.

[47] Khalid Aziz and Antonin Settari. *Petroleum reservoir simulation*. Applied Science Publ. Ltd., London, UK, 1979.

[48] Ltd Schlumberger. *Eclipse reservoir simulation software v2016. Technical Description Manual*, 2016.

[49] Michael Batzle and Zhijing Wang. Seismic properties of pore fluids. *Geophysics*, 57(11): 1396–1408, 1992.

[50] Abul Fahimuddin. *4D seismic history matching using the ensemble Kalman filter (EnKF): possibilities and challenges*. PhD thesis, The University of Bergen, 2010.

[51] Owe Axelsson. *Iterative solution methods*. Cambridge university press, 1996.

[52] Reinhard Furrer and Thomas Bengtsson. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis*, 98 (2):227–255, 2007.

[53] Geir Evensen. Sampling strategies and square root analysis schemes for the EnKF. *Ocean dynamics*, 54(6):539–560, 2004.

[54] Marco A Iglesias, Kody JH Law, and Andrew M Stuart. Evaluation of Gaussian approximations for data assimilation in reservoir models. *Computational Geosciences*, 17(5): 851–885, 2013.

Chapter 13

Paper E

Towards application of multilevel data assimilation in realistic reservoir history-matching problems

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230855935 (print)
9788230859636 (PDF)