Assessment of multilevel ensemble-based data assimilation for reservoir history matching

Kristian Fossum¹ · Trond Mannseth¹ · Andreas S. Stordal¹

Received: 3 July 2018 / Accepted: 17 October 2019 / Published online: 14 December 2019
 © The Author(s) 2019

Abstract

Multilevel ensemble-based data assimilation (DA) as an alternative to standard (single-level) ensemble-based DA for reservoir history matching problems is considered. Restricted computational resources currently limit the ensemble size to about 100 for field-scale cases, resulting in large sampling errors if no measures are taken to prevent it. With multilevel methods, the computational resources are spread over models with different accuracy and computational cost, enabling a substantially increased total ensemble size. Hence, reduced numerical accuracy is partially traded for increased statistical accuracy. A novel multilevel DA method, the multilevel hybrid ensemble Kalman filter (MLHENKF) is proposed. Both the expected and the true efficiency of a previously published multilevel method, the multilevel ensemble Kalman filter (MLEnKF), and the MLHENKF are assessed for a toy model and two reservoir models. A multilevel sequence of approximations is introduced for all models. This is achieved via spatial grid coarsening and simple upscaling for the reservoir models, and via a designed synthetic sequence for the toy model. For all models, the finest discretization level is assumed to correspond to the exact model. The results obtained show that, despite its good theoretical properties, MLEnKF does not perform well for the reservoir history matching problems considered. We also show that this is probably caused by the assumptions underlying its theoretical properties not being fulfilled for the multilevel reservoir models considered. The performance of MLHENKF, which is designed to handle restricted computational resources well, is quite good. Furthermore, the toy model is utilized to set up a case where the assumptions underlying the theoretical properties of MLEnKF are fulfilled. On that case, MLEnKF performs very well and clearly better than MLHENKF.

Keywords Ensemble-based history matching · Realistic computational resources · Multilevel methods

Nomenclature

$\mathbb{B}$	Bias
A	Empirical forecast cross covariance
C	Covariance
Q	Cost
N	Ensemble size
K	Kalman gain
L	Number of levels
w	Weights in MLEnKF
ρ	Mean squared error
$\mathbb{V}$	Variance
Ω	Scale matrix in Wishart distribution
C	Monte Carlo estimate of covariance

V	Empirical variance in toy model
E	Monte Carlo estimate of expectation
$\mathcal{E}^2$	Matrix of element-wise squared bias
x	Position in space
d	Data vector
H	Measurement operator
$\mathbb{E}$	Expectation
κ	Fixed cost
δ	Fixed MSE error
M	Arbitrary deterministic forward model
$\mathcal{N}$	Gaussian distribution
θ	Realization from $\mathcal{N}$
$\ \cdot\ $	Matrix norm
m	Parameter
$p()$	Probability density function
k	Permeability
ω	Arbitrary random variable
$\mathbb{X}$	Unspecified statistics
Y, Z	State vector

✉ Kristian Fossum
kristian.fossum@norce-research.no

¹ NORCE Energy, P.O. Box 22 Nygårdstangen, 5838, Bergen, Norway

- ϵ Level-specific error for toy models
- Ψ Grid-coarsening factor
- σ Average value over elements of Σ
- Σ Matrix of element-wise variance
- R Total available computational resources

1 Introduction

The parameter functions entering the reservoir model equations are uncertain, and sound decision-making therefore requires that available data are utilized to reduce and quantify the uncertainty in the models. Ensemble-based data assimilation (DA) methods, such as the ensemble Kalman filter (EnKF) [11], have therefore become popular for automated reservoir history matching [2, 39]. A range of ensemble-based methods have been introduced in the literature (see, e.g., [4, 5, 8, 9, 28, 29]), and several papers have evaluated and compared different methods [10, 14–16, 30, 36].

Ensemble-based methods rely on Monte Carlo (MC) estimates of the mean and covariance of the state vector. It is well known that the variance of the MC estimator can be large when the ensemble size is not sufficiently large. Moreover, with straightforward use of a standard ensemble-based method such as the EnKF, each member in the posterior ensemble is a linear combination of the members in the initial ensemble. Hence, the degrees of freedom are limited by the ensemble size. If a large amount of accurate data is available, there is typically insufficient degrees of freedom, resulting in problems such as spurious correlations and underestimation of variance, and in extreme cases, ensemble collapse.

To alleviate the problems arising from use of a too small ensemble size, localization [27] is often applied. With localization, a spatial region (the localization region) is defined for each observation, and only variables associated with a grid cell within the localization region are updated. Clearly, the definition of the localization region is essential for the success of this approach. Designing an appropriate localization region typically requires expert knowledge of the particular problem under investigation, and a significant amount of manual labor. More generic approaches to address problems arising from use of a too small ensemble size are therefore desirable. Even if such approaches are not expected to completely remove the need for localization for large cases, they may at least make localization more robust, in the sense that negative effects of not using an *optimal* localization region become less severe.

Such an alternative approach was recently proposed in [17]. Here, computationally less expensive reservoir simulations allowed for use of an increased ensemble size without increasing total computational cost. Computational savings were achieved by adaptive coarsening of the spatial

simulation grid and simplistic upscaling of reservoir properties. That approach performed at least equally well as a standard approach with localization (using state of the art localization methods with optimized localization regions) on small- to medium-sized problems with varying characteristics.

There is a wide range of options for reducing computational complexity by using proxy simulators [23, 25, 32, 41, 43]. As an example, while [17] utilized coarsening of the spatial grid, an alternative approach would be to coarsen the temporal grid as in [26]. With all proxy models used in this fashion, there is a tradeoff between numerical and MC errors. In a statistical framework, this is manifested in the mean squared error, ρ , which equals the sum of the variance and the squared bias. A computationally inexpensive proxy model allows for an increase in ensemble size, which in turn reduces the variance. The bias will, however, increase since a computationally inexpensive proxy model is also (generally speaking) inaccurate. There will always be an optimal proxy-model accuracy level where ρ is minimized. It is, however, difficult to select the optimal accuracy level in advance. Moreover, many proxy models require off-line computations as part of the training. The total computational gain and optimal accuracy level can therefore be hard to estimate.

In this paper, we consider a class of methods where no a priori selection of proxy-model accuracy level is required. With *multilevel methods*, one utilizes a sequence of proxy models with varying levels of computational complexity and numerical accuracy. Various flavors of multilevel schemes have been introduced in a wide range of applications. This spans from methods used to solve systems of equations arising from discretizations of differential equations (see, e.g., [42]) to methods used to solve parameter estimation problems [1, 6, 18, 21, 22, 26, 34, 35, 38]. We investigate a subclass of the latter type—multilevel ensemble-based methods for DA. In this context, the rationale for using multilevel methods is to utilize computationally less complex models to reduce variance while retaining computationally more complex models to reduce bias, hence, obtaining a small ρ .

The multilevel EnKF (MLEnKF) was introduced in [26] (see, also [6]). MLEnKF utilizes the main idea in multilevel MC (MLMC) [19] to construct an unbiased estimator. It has been shown that, under certain assumptions (discussed in Section 3), MLMC outperforms MC (see, [19]), while MLEnKF outperforms EnKF (see, [26]), in bringing ρ below a given positive number. We note, however, that the validity of these assumptions will depend on the particular model under investigation, on the sequence of multilevel models used to approximate it, and on the computational resources available. The challenge of assessing the validity of the assumptions is pointed out in [20].

In this paper, we want to assess if multilevel methods are suitable for reservoir history matching. Based on simple reasoning, we argue that the assumptions underlying MLMC and MLEnKF are difficult to satisfy for a typical reservoir history matching problem. A simple example illustrates this reasoning. We therefore propose a novel multilevel method, the *multilevel hybrid EnKF* (MLHENKF), with alternative underlying assumptions. We argue that these assumptions are more likely to be valid for reservoir history matching.

We assess the model-specific assumptions underlying MLMC and MLEnKF, as well as the assumptions underlying MLHENKF, for a toy model where we can easily control and design the structure of the associated multilevel sequence of approximate models, and for two selected reservoir models with associated multilevel sequences of approximate models. This leads to predictions of how MLEnKF and MLHENKF will perform in history matching. Finally, we perform history matching with EnKF, MLEnKF, and MLHENKF. When analyzing the history matching results, it is of interest both to assess which method that shows the better performance, but also to assess if the history matching results are in line with the above-mentioned predictions. For the latter assessment, the toy model is especially useful since different multilevel sequences can easily be designed.

The rest of this paper is organized as follows. Section 2 briefly discusses basic (i.e., non-multilevel) methods, like MC estimation, ensemble-based DA, and Bayesian model averaging, to ease the description of the multilevel methods which follows in Section 3. Section 4 describes the reservoir models and the toy model used to assess the validity of the different assumptions underlying MLEnKF and MLHENKF. Section 5 describes the setup of the numerical experiments, while Section 6 presents results from these experiments. Finally, in Section 7, we summarize the paper and conclude our findings.

2 Basic methods

2.1 Monte Carlo estimation

Let M denote an arbitrary deterministic forward model, ω an arbitrary random variable, $Y(\omega)$ a state vector associated with M , and $\{Y^n\}_{n=1}^N$ an ensemble of independent realizations of Y . The Monte Carlo (MC) estimators, $E(Y)$ and $C(Y)$ for the expectation, $\mathbb{E}[Y]$, and covariance, $\mathbb{C}[Y]$, are then

$$E(Y) = \frac{1}{N} \sum_{n=1}^N Y^n, \quad (1)$$

$$C(Y) = \frac{1}{N-1} \sum_{n=1}^N (Y^n - E(Y)) (Y^n - E(Y))^T. \quad (2)$$

The computational cost and variance of $E(Y)$ and $C(Y)$ are given by

$$Q(E(Y)) = Q(C(Y)) = N Q(Y^n), \quad (3)$$

$$\mathbb{V}(E(Y)) = \frac{1}{N} \mathbb{V}(Y^n), \quad (4)$$

$$\mathbb{V}(C(Y)_{i,j}) = \frac{1}{(N-1)} \left(\Omega_{i,i} \Omega_{j,j} + \Omega_{i,j}^2 \right), \quad (5)$$

where $Q(Y^n)$ and $\mathbb{V}(Y^n)$ denote the cost and variance of a single realization, respectively, and $\Omega_{i,j}$ denotes the element on the i th row and j th column of the scale matrix in the Wishart distribution for $C(Y)$ [37].

Later on, it will be useful to specify the size of the ensemble that the MC estimates are based on. In these cases, the notation $E^N(Y)$ and $C^N(Y)$ will be applied. We will also use the notation $X^N(Y)$ to denote the MC estimate of $\mathbb{X}[Y]$, where $\mathbb{X}$ is an unspecified statistics.

2.2 Ensemble-based data assimilation

Some ensemble-based methods assimilate all available data simultaneously while other methods separate data into groups and assimilate these sequentially. For our purpose, it will suffice to describe a single assimilation step of a basic ensemble-based method. In that case, the descriptions of the two options coincide.

Let Z_{pr}^n denote a prior-ensemble member, that is, a realization from the prior distribution of the state vector. The forecast ensemble, $\{Y^n\}_{n=1}^N$, is then obtained by running each ensemble member through the forward model,

$$Y^n = M(Z_{pr}^n). \quad (6)$$

Empirical estimates of the two first moments of the forecast distribution are obtained as $E(Y)$ and $C(Y)$, respectively.

Let d^n denote a realization of the data to be assimilated, let $\mathbb{C}_d$ denote the covariance of the data errors, and let H denote a matrix where each element is either zero or one, designed to select those elements in Y corresponding to data positions. An arbitrary ensemble member after assimilation of the data, Z^n , is then given by

$$Z^n = Y^n + K(d^n - H Y^n), \quad (7)$$

where the Kalman gain is given by

$$K = C(Y) H^T \left(H C(Y) H^T + \mathbb{C}_d \right)^{-1}. \quad (8)$$

Empirical estimates of the two first moments of the analysis distribution (i.e., the distribution after assimilation of the data) are obtained as $E(Z)$ and $C(Z)$, respectively.

Even though we limit ourselves to a single assimilation step of a basic ensemble-based method in this paper, we remark that with sequential assimilation of more than one data group, the next step can be performed by substituting Z^n for Z_{pr}^n in Eq. 6 and repeating the process described above.

In applied ensemble-based data assimilation, the user must alleviate problems arising from the use of a too small ensemble size. For reservoir history matching, distance-based localization is the most used technique. Here a localization matrix is multiplied element-wise with either the Kalman gain (8) matrix or the covariance matrix (2). In this paper, we will localize the Kalman gain, and we will use a Gaspari-Cohn correlation function, centered at the data position, as localization matrix.

2.3 Bayesian model averaging

Assume that multiple numerical models, all forecasting the same quantity, are available. With Bayesian model averaging (BMA), one does not need to select a single model among several competing models. Instead, inferences are made utilizing all models. In general, the BMA method assumes that all the models are unbiased and predict the same quantity, albeit with different model assumptions. With S models, the law of total probability gives

$$p(Y) = \sum_{s=1}^S p(M_s) p(Y|M_s), \quad (9)$$

where $p(Y|M_s)$ is the forecast PDF based on model M_s alone, and $p(M_s)$ is the probability of model M_s being the correct one. Hence, $p(Y)$ is a weighted average of the various forecast PDFs. The weights, $p(M_s)$, might potentially be estimated from a training data set.

The BMA expectation and covariance of Y can be obtained from the laws of total expectation and covariance, respectively, and are given in Appendix 1.

3 Multilevel methods

Let $\{M_l\}_{l=0}^L$ denote a sequence of deterministic models such that the models $M_0, \dots, M_{L-1}$ approximate M_L with increasing accuracy and computational cost. The models in the sequence will be denoted *multilevel models*, and we assume that M_L is exact. Furthermore, let $\{Y_l\}_{l=0}^L$ be a sequence of random variables such that $Y_l(\omega)$ denotes the state vector of M_l . Hence, the quantities $Y_0, \dots, Y_{L-1}$ approximate Y_L with increasing accuracy and computational cost.

3.1 Multilevel Monte Carlo

3.1.1 Description

The multilevel Monte Carlo (MLMC) method utilizes the identity

$$\mathbb{X}[Y_L] \equiv \mathbb{X}[Y_0] + \sum_{l=1}^L (\mathbb{X}[Y_l] - \mathbb{X}[Y_{l-1}]), \quad (10)$$

to construct a multilevel estimator for $\mathbb{X}[Y_L]$. To this end, let $\{N_l\}_{l=0}^L$ be a sequence of positive integers, let $\{\omega_l\}_{l=0}^L$ be independent realizations of ω , and let $\{Y_l^n(\omega_l)\}_{n=0}^{N_l}$ and $\{Y_{l-1}^n(\omega_l)\}_{n=0}^{N_l}$ be ensembles of realizations associated with $Y_l(\omega_l)$ and $Y_{l-1}(\omega_l)$, respectively. The multilevel estimator may then be expressed as

$$X_{ML}(Y_L) = X^{N_0}(Y_0(\omega_0)) + \sum_{l=1}^L \Delta_l X^{N_l}(Y(\omega_l)), \quad (11)$$

where $\Delta_l X^{N_l}(Y(\omega_l)) = X^{N_l}(Y_l(\omega_l)) - X^{N_l}(Y_{l-1}(\omega_l))$. Since X is an unbiased estimator for $\mathbb{X}$, it follows from Eq. 10 that $X_{ML}(Y_L)$ is an unbiased estimator for $\mathbb{X}[Y_L]$.

Inserting C for $\mathbb{X}$ leads to

$$C_{ML}(Y_L) = C^{N_0}(Y_0(\omega_0)) + \sum_{l=1}^L \Delta_l C^{N_l}(Y(\omega_l)), \quad (12)$$

while inserting $\mathbb{E}$ for $\mathbb{X}$, defining $Y_{\Delta l}^n(\omega_l) = Y_l^n(\omega_l) - Y_{l-1}^n(\omega_l)$, and utilizing the linearity of $\mathbb{E}(Y)$, gives

$$E_{ML}(Y_L) = E^{N_0}(Y_0(\omega_0)) + \sum_{l=1}^L E^{N_l}(Y_{\Delta l}(\omega_l)). \quad (13)$$

Throughout, the subscript Δl denotes the difference between level l and $l-1$.

3.1.2 Theoretical efficiency

The efficiency of the MLMC estimator, with optimally distributed computational resources (denoted by the superscript o), can be compared with the MC estimator by estimating the computational resources required to produce an estimate with a fixed MSE error, $\rho = \delta^2$. It is possible to present arguments for why

$$Q_\delta^o(E_{ML}(Y_L)) < Q_\delta(E(Y_L)), \quad (14)$$

where $Q_\delta^o(E_{ML}(Y_L))$ and $Q_\delta(E(Y_L))$ denote the computational resources required by the MLMC and MC estimators, respectively (see Appendix 2 for more details).

In addition, Theorem 2.1 in [20] states that, under certain conditions on a geometric sequence of levels, $\rho_{ML} < \delta^2$ with computational complexity bounds depending on these conditions. Two of the conditions concern the monotonic increase and decrease with l of the upper bounds on $Q_{\Delta l}^n$ and $V_{\Delta l}^n$, respectively

$$Q_{\Delta l}^n \leq c_1 2^{\gamma l}, \quad (15)$$

$$V_{\Delta l}^n \leq c_2 2^{-\beta l}, \quad (16)$$

where c_1 , c_2 , γ , and β are positive constants (we refer to [20] for further details).

3.1.3 Expected efficiency for real applications

We are interested in assessing whether or not multilevel methods (where the selected multilevel structure of $\{M_l\}_{l=0}^L$ will be defined in Section 4) are suitable for reservoir history matching. Giles [20] points out the challenge in assessing whether or not the assumptions of his Theorem 2.1 are valid for real applications. In addition, we remark that for many real applications, including reservoir history matching, $Q_\delta(E(Y_L))$ is so large that MC is currently impractical. For such applications, Eq. 14 does not ensure that MLMC would be practical either. In the following, we will therefore consider computationally less expensive alternatives to MLMC.

3.2 Multilevel EnKF

3.2.1 Description

MLEnKF can largely be described as EnKF with a multilevel Kalman gain (for full details of MLEnKF, we refer to [26]). A brief summary of the main steps is given in Appendix 3.

3.2.2 Expected efficiency for real applications

The main idea of MLEnKF is inherited from MLMC. From the discussion in Section 3.1.2 and Appendix 2, it seems reasonable to assess whether or not $\mathbb{V}_{\Delta l}^n$ is monotonically decreasing with l for a particular problem in order to assess if one can expect MLEnKF to represent an improvement over EnKF for that problem (for the reservoir models we consider, $Q_{\Delta l}^n$ will always increase monotonically). It is, however, evident that fulfilment of this condition is not *sufficient* to ensure that MLEnKF will perform well. Therefore, we state that models where this condition is fulfilled have a higher *MLMC potential* than models where it is not. We will perform numerical experiments to assess predictions obtained from evaluation of the MLMC potential.

The multilevel empirical covariance, $C_{ML}(Y_L)$, is central for the MLEnKF. It is unbiased, so its quality depends on the matrix of element-wise variances, $\mathbb{V}_e[C_{ML}(Y_L)]$. Since all elements in the sequence $\{\omega_l\}_{l=0}^L$ are independent, this matrix can be expressed as a sum of contributions from the individual terms on the right-hand side of Eq. 12

$$\mathbb{V}_e[C_{ML}(Y_L)] = \mathbb{V}_e[C^{N_0}(Y_0(\omega_0))] + \sum_{l=1}^L \mathbb{V}_e[\Delta_l C^{N_l}(Y(\omega_l))]. \quad (17)$$

The main line of arguments in favor of MLMC in [20] is for small values of l , the computational cost is small, so a low variance in $\Delta_l C^{N_l}(Y(\omega_l))$ can be obtained by selecting a large N_l . For large values of l , the variance in

Table 1 Motivating example, bias, and variance for different levels

Variable	Variance
$A_0^{N_0}(Y_0(\omega_0))$	$1.1N_0^{-1}$
$A_1^{N_1}(Y_1(\omega_1))$	$1.4N_1^{-1}$
$A_2^{N_2}(Y_2(\omega_2))$	$24.2N_2^{-1}$
$\Delta_1 A^{N_1}(Y(\omega_1))$	$0.04N_1^{-1}$
$\Delta_2 A^{N_2}(Y(\omega_2))$	$23.3N_2^{-1}$

$\Delta_l C^{N_l}(Y(\omega_l))$ will be small since M_l and M_{l-1} should approximate M_L well.

For the latter argument to hold, M_l and M_{l-1} must not differ too much. This requires that L is sufficiently large. Since some computational resources have to be distributed to each level, this in turn requires that R is sufficiently large. For many real applications, such as reservoir history matching, this may not be the case.

We will now provide a simple example with three different levels, and a realistic difference between M_l and M_{l-1} (details regarding the numerical setup of this example are found in Appendix 4.1).

Let the empirical forecast cross covariance between simulated output at data locations and parameters be denoted by A , that is, $A = CH^T$, where C denotes the empirical covariance of Y obtained by one of the methods considered. We assess the relative magnitudes of variances (level specific and level difference) by bootstrapping carried out using a sample of 2000 permeability realizations on each level, and with 1000 replicates. We assume that M_2 is the true model. Results obtained for the different levels when averaging the bootstrapped results over the matrix elements are summarized in Table 1. Note that variances are presented as functions of the level-specific sample sizes. Hence, the averaged bootstrapped results can be obtained by inserting $N_0 = N_1 = N_2 = 2000$ in the table.

Fig. 1 Mean squared error as a function of computational resources. Full line, MLEnKF; dashed line, EnKF

The level-specific results for variance in Table 1 can be utilized to calculate ρ for the two different covariance estimation strategies as functions of total computational resources, R , utilizing (4) for EnKF, and Eq. 39 for MLEnKF. These functions are shown in Fig. 1. To explain how to interpret the functions on the figure, let the available computational resources be expressed in terms of the number of simulations one can afford on the finest level. Since all calculations with EnKF are performed on the finest level, the numbers on the horizontal axis of Fig. 1 correspond directly to N_2 for EnKF. This is not the case for MLEnKF. Here, R must be distributed among the levels, and N_0 , N_1 , and N_2 correspond to optimal sample sizes (see Eq. 41).

Figure 1 shows that EnKF has lower ρ than MLEnKF for all values of R . This example model is clearly not suitable for MLEnKF. From Table 1, we see that the variance of $\Delta_l A^{N_l}(Y(\omega_l))$ is not monotonically decreasing, and the highest variance value is in fact from $l = L$.

3.3 Multilevel hybrid EnKF

Referring to Section 2.3 and Appendix 1, let $s = l + 1$ and let the $L + 1$ models be the multilevel models introduced in the first paragraph of Section 3. As with BMA, the involved models predict the same quantity (here, Y_L), but now with varying accuracy. Hence, contrary to BMA, all models but one are biased. The multilevel estimator to be described below will also be biased. The working hypothesis is that it may still be useful for problems where variance errors dominate the bias.

3.3.1 Description

Firstly calculate

$$\hat{Y}_l^n = M_l(Z_{pr}^n) \quad (18)$$

for $l \in [0, L]$ and $n \in [1, N_l]$. Then, let an arbitrary member in forecast ensemble number l be given by

$$Y_l^n = \hat{Y}_l^n + \{E^{N_L}(M_L(Z_{pr})) - E^{N_l}(M_l(Z_{pr}))\}, \quad (19)$$

where the terms inside the braces are included to ensure that $\mathbb{E}[Y_l] = \mathbb{E}[Y_L]$. Empirical estimates of the two first moments of the forecast distribution are obtained from

$$E_{BA}^{\sim}(Y_L) = \sum_{l=0}^L w_l E^{N_l}(Y_l | M_l), \quad (20)$$

$$C_{BA}^{\sim}(Y_L) = \sum_{l=0}^L w_l \times \{C^{N_l}(Y_l | M_l) + (E^{N_l}(Y_l | M_l) - E^{N_L}(Y_L)) \times (E^{N_l}(Y_l | M_l) - E^{N_L}(Y_L))^T\}. \quad (21)$$

Note that the expressions for $E_{BA}^{\sim}(Y_L)$ and $C_{BA}^{\sim}(Y_L)$ are very similar to those for the BMA estimators $E_{BA}(Y)$ and $C_{BA}(Y)$, derived in Appendix 1. The only difference (in addition to the change of indexing noted at the start of the current section) is that an unspecified weight, w_l , has been substituted for $p(M_l)$.

The general idea of utilizing an empirical forecast covariance based on a weighted average of different empirical forecast covariances is not new. It was introduced in [24], combining the EnKF with the 3DVAR method, and denoted a *hybrid* scheme. Since we utilize the same general idea (albeit in an entirely different manner), we denote the novel method the *multilevel hybrid EnKF* (MLHENKF). While the notation used in Eqs. 20 and 21 is convenient to illustrate the relation to BMA, these expressions can also be written as

$$E_{MH}(Y_L) = \sum_{l=0}^L w_l E^{N_l}(Y_l), \quad (22)$$

$$C_{MH}(Y_L) = \sum_{l=0}^L w_l \{C^{N_l}(Y_l) + (E^{N_l}(Y_l) - E^{N_L}(Y_L)) \times (E^{N_l}(Y_l) - E^{N_L}(Y_L))^T\}. \quad (23)$$

To specify w_l , let us first consider what would happen if we select $w_l = p(M_l)$. Since we have assumed that M_L is exact while M_l ; $l = 0, \dots, L - 1$, are approximations to M_L , this choice would have resulted in $w_L = 1$ and $w_l = 0$; $l = 0, \dots, L - 1$. Hence, only M_L would have contributed to $E_{MH}(Y_L)$ and $C_{MH}(Y_L)$, and in that case, MLHENKF would reduce to EnKF, which is clearly unwanted. Since we target situations where variance errors dominates bias, we aim for low values for $\mathbb{V}[E_{MH}(Y_L)]$ and $\|\mathbb{V}_e[C_{MH}(Y_L)]\|$ by selecting large weights for terms where $\mathbb{V}[E^{N_l}(Y_l)]$ and $\|\mathbb{V}_e[C^{N_l}(Y_l)]\|$ are small. Since $\mathbb{V}[E^{N_l}(Y_l)]$ and $\|\mathbb{V}_e[C^{N_l}(Y_l)]\|$ are proportional to $\frac{1}{N_l}$ and $\frac{1}{N_l-1}$, respectively, we simply select w_l proportional to N_l ,

$$w_l = \frac{N_l}{\sum_{l=0}^L N_l}. \quad (24)$$

An arbitrary ensemble member after assimilation of the data is given by

$$Z_l^n = Y_l^n + K_{MH}(d^n - H Y_l^n), \quad (25)$$

where K_{MH} is given by

$$K_{MH} = C_{MH}(Y_L) H^T (H C_{MH}(Y_L) H^T + \mathbb{C}_d)^{-1}, \quad (26)$$

Empirical estimates of the two first moments of the analysis distribution are given by $E_{MH}(Z_L)$ and $C_{MH}(Z_L)$. It is straightforward to show that when w_l is given by Eq. 24,

Fig. 2 Flowchart MLHENKF

$E_{MH} \equiv E^{N_S}$, that is E_{MH} is equivalent to the MC estimator based on all ensemble members on all involved levels. Moreover, due to Eq. 19

$$\mathbb{B}_e[C_{MH}(Y_L)] = \frac{1}{\sum_{l=0}^L N_l} \sum_{l=0}^L N_l \mathbb{B}_e[C^{N_l}(Y_l(\omega_l))], \quad (27)$$

and

$$\mathbb{V}_e[C_{MH}(Y_L)] = \frac{1}{\sum_{l=0}^L N_l} \sum_{l=0}^L N_l \mathbb{V}_e[C^{N_l}(Y_l(\omega_l))]. \quad (28)$$

Since this is a new method, we include a simple flowchart in Fig. 2 to illustrate the algorithmic details. The number of levels used in the method is problem dependent and we cannot provide a general recommendation for this.

3.3.2 Expected efficiency for real applications

It is important to assess if bias or variance errors dominate when calculating $C_{MH}(Y_L)$ for typical reservoir history matching problems. It has been shown [13] that the magnitudes of the true correlations between parameters and simulated data play an important role regarding this issue. We will illustrate this by a toy example involving one case where the true correlation is zero and one case where the true correlation is high.

Consider two multilevel model sequences, $\{M_l^1\}_{l=0}^3$ and $\{M_l^2\}_{l=0}^3$, with corresponding correlations at the different levels listed in Table 2, and assume that the correlations at level 3 are true. Furthermore, assume that the computational costs for both model sequences increase monotonically with l , such that it is natural to let N_l decrease monotonically with l , as in Table 2. We estimate the true correlations for M^1 and M^2 in two ways. Firstly, we use MC results obtained with M_3^1 and M_3^2 with sample size N_3 . Secondly,

Table 2 Correlations and ensemble sizes

l	0	1	2	3
M_l^1 correlations	0.2	0.1	0.05	0
M_l^2 correlations	0.7	0.8	0.85	0.9
N_l	250	150	50	8

we use weighted averages of MC results obtained with $\{M_l^1\}_{l=0}^3$ and $\{M_l^2\}_{l=0}^3$ with weights given by Eq. 24 (i.e., with $L = 3$) and $\{N_l\}_{l=0}^3$ listed in Table 2. The estimations are repeated a large number of times, allowing the kernel density estimates (KDEs) for the PDFs of the true correlations, shown in Fig. 3, to be formed.

Figure 3 a shows KDEs for M^1 . The PDF estimated with M_3^1 has a very high variance. The PDF estimated from weighted averages of results from $\{M_l^1\}_{l=0}^3$ has a much lower variance, but is biased. The bias in the latter PDF is, however, much smaller than the standard deviation in the former PDF.

Figure 3 b shows KDEs for M^2 . The PDF estimated with M_3^2 has a moderately large variance. The PDF estimated from weighted averages of results from $\{M_l^2\}_{l=0}^3$ has a lower variance, but it is biased. The bias in the latter PDF is somewhat larger than the standard deviation in the former PDF.

The results in Fig. 3 indicate that one can expect MLHENKF to perform better than EnKF on problems with weak true correlations, but not necessarily on problems with strong true correlations. Since real history matching problems with production data contains a large number of uncertain parameters and spatially sparsely distributed data, many true correlations between data and parameters will be weak, and hence, MLHENKF is expected to perform better than EnKF on such problems.

4 Test models

We are interested in the applicability of multilevel DA methods to reservoir history matching, and consider two model reservoirs for testing purposes. For all reservoirs, the original simulation model corresponds to M_L while $\{M_l\}_{l=0}^{L-1}$ represent approximations to M_L where accuracy and computational cost increase with l . (See Section 4.1 for a specific description of the multilevel structure of $\{M_l\}_{l=0}^L$.) In addition, to evaluate the limitations of the methods, we consider a bivariate toy model where we can easily adjust the forecast bias and variance for $\{M_l\}_{l=0}^{L-1}$ (see Appendix 4.2 for details regarding the toy model).

4.1 Reservoir models

We consider a multilevel structure where $\{M_l\}_{l=0}^L$ represent a sequence of reservoir-flow models such that M_L denotes the flow model discretized on the original simulation grid, while $\{M_l\}_{l=0}^{L-1}$ denotes flow models discretized on coarser spatial grids. The grid resolution increases with l , and the grid is coarsened by merging Ψ grid cells in each coordinate direction, as illustrated in Fig. 4 for the case where $L = 2$ and $\Psi = 2$.

Fig. 3 Sample correlations. Solid curves, based on results from M_3^1 and M_3^2 . Dashed curves, based on weighted averages of results from $\{M_l^1\}_{l=0}^3$ and $\{M_l^2\}_{l=0}^3$

Several of the reservoir-flow model input parameters may vary from grid cell to grid cell. Upscaling is required to represent these properties on the coarser grids. We utilize a simple local-local upscaling method (see, e.g., [7, 12]), where the upscaled properties are found by proper averaging of the fine-cell properties. This scheme is exact only for one-dimensional, single-phase flow. We consider two-dimensional, two-phase flows, so the output from M_l will be biased for $l < L$, and the bias will generally increase as l decreases. Note that this method for generating a sequence of reservoir-flow models is very fast and easily automated. Hence, we assume that no additional computational overhead is generated. More complex upscaling techniques, with potentially less bias, could be used to generate a multilevel sequence. Alas, increased accuracy comes with added computational on-line, or off-line, cost (see Appendices 4.3 and 4.4 for details on RM1 and RM2).

5 Numerical investigation

In Section 3.2.2, it was found that the MLMC potential of a model problem is expected to be important for the performance of MLEnKF, while Section 3.3.2 indicates that the relative magnitudes of bias and variance for a reservoir history matching problem are suitable for MLHENKF. The numerical investigation has been designed with this in mind.

Firstly, we investigate the MLMC potential for all models by studying the variances of the matrix elements in the sequence $\{\Delta_l A^{N_l}(Y(\omega_l))\}_{l=1}^L$. Secondly, for the reservoir models, we estimate the ρ for the Kalman gain calculated using EnKF, MLEnKF, and MLHENKF. Finally, we perform DA on all models using EnKF, MLEnKF, and MLHENKF. We want to assess to what degree the predictions obtained by looking at the MLMC potential carry over to DA, and to assess if MLEnKF and/or MLHENKF perform better than EnKF for reservoir history matching.

The MSE for the Kalman gain and the variance of the different terms in the MLMC estimator are unknown. To evaluate the MLMC potential and ρ , bootstrapping is therefore applied. For all models, $N_l = 2000$ for each level. For RM1 and RM2, bootstrapping is performed with 5000 replicates of the ensembles at each level, while for the toy model, the bootstrapping is performed with 2000 replicates.

The DA part of the investigation takes into account that the total computational resources available in a real application to reservoir history matching are quite restricted. For such large cases, a direct link between the cost and the number of grid cells for each level exists. The governing equations of the reservoir-flow model are solved by an iterative scheme, where each iteration requires the solution of a system of linear equations. For large cases, the solution of the linear system is the dominating computational cost. Then, the computational cost of running the reservoir-flow

Fig. 4 Illustration of multilevel family of grids with $L = 2$ and $\psi = 2$. **a** $l = 2$. **b** $l = 1$. **c** $l = 0$

model once, Q_l^n , is approximately proportional to the cost of solving the system of linear equations. This cost scales approximately as $(G_l)^\nu$, where G_l denotes the number of grid cells in the simulation grid at level l , and the exponent $\nu \in [1.25, 1.5]$ [3].

We will use this measure of computational cost for RM1 and RM2, and we set $\nu = 1.25$ in all these experiments. Since Ψ has been specified earlier, and since $\{G_l\}_{l=0}^5$ are easily calculated when Ψ is known, normalization with respect to Q_5^n results in

$$[\hat{Q}_0^n, \hat{Q}_1^n, \hat{Q}_2^n, \hat{Q}_3^n, \hat{Q}_4^n, \hat{Q}_5^n] \approx \left[\frac{1}{4929}, \frac{1}{871}, \frac{1}{135}, \frac{1}{32}, \frac{1}{5}, 1 \right]. \quad (29)$$

A description of the setup of the numerical experiments constituting the investigation follows, while results from the experiments are presented in Section 6.

To understand the behavior of the different multilevel DA methods, it is convenient to control the evolution of the level special-mean and variance. To this end, we design a multilevel family of bivariate toy models, $\{M_l\}_{l=0}^L$. Let $m^T = (m_1, m_2)$, where m_1 and m_2 are independent, and define the outcome from M_L by

$$y_L(m) = 10 + 0.2m_1 + m_2. \quad (30)$$

Hence, m_1 is weakly correlated to the data while m_2 is strongly correlated to the data. We define the outcomes from M_l for $l \in [0, L-1]$ by

$$y_l(m) = y_L(m) + \epsilon_l, \quad (31)$$

where ϵ_l is a level-specific error term. It is useful to define ϵ_l such that one can relate the level-specific mean and variance to the high-resolution mean and variance via two control variables. To achieve this, we let

$$\epsilon_l = \mu_l E(y_L(m)) + \theta_l, \quad (32)$$

where θ_l denotes a realization from $\mathcal{N}(0, \gamma_l V(y_L(m)))$, $\mathcal{N}$ denotes the Gaussian distribution and V denotes the empirical variance. The variables μ_l and γ_l are then defined as the fractions of the level-specific empirical mean and variance to the high-resolution empirical mean and variance, respectively;

$$\mu_l = \frac{E(y_l)}{E(y_L)}, \quad (33)$$

$$\gamma_l = \frac{V(y_l)}{V(y_L)}. \quad (34)$$

The prior model for m is $\mathcal{N}(\bar{m}, C_m)$, where

$$\bar{m} = [0.5, 1.1]^T,$$

and

$$C_m = \begin{pmatrix} 2, & 0 \\ 0, & 2 \end{pmatrix}.$$

Three different multilevel structures are defined; firstly, a case with low errors that decrease monotonically with l (low error), secondly, a case mimicking the structure of RM2 (reservoir error), and, finally, a case in between the low-error and the reservoir-error cases (medium error). The corresponding values of μ_l and γ_l are listed in Table 5.

The reservoir-error case is established by selecting Y as a single reservoir-simulation output, the well oil production rate from P after 45 timesteps. We estimate $\{E(y_l)\}_{l=0}^L$ and $\{V(y_l)\}_{l=0}^L$ from the level-specific forecasts of RM2 by bootstrapping, and calculate μ_l and γ_l using Eqs. 55 and 56, thereby ensuring that the level-specific errors in the toy model have the same relative sizes as those of RM2.

In Fig. 16a, we plot the kernel density estimates (KDEs), calculated from 2000 runs, of the level-specific forecasts of the selected simulation output from RM2. The KDEs of the level-specific forecasts from the toy model (Fig. 16b) shows that, with the estimated values for μ_l and γ_l , the multilevel structure of the toy model is similar to that of RM2. (In Fig. 16, we plot KDEs only for $l = 1, 3$ and 5. However, KDEs for $l = 0, 2$ and 4 show similar behavior.)

For reasons given in the first paragraph of Section 3.3.2, the toy model is designed such the correlation between m_1 and y is much smaller than between m_2 and y . Hence, variance dominates for m_1 while bias dominates for m_2 .

5.1 MLMC potential

Let $\Sigma_{\Delta_l} = \mathbb{V}_e[\Delta_l A^{N_l}(Y(\omega_l))]$. We calculate Σ_{Δ_l} by bootstrapping, and since all elements in Σ_{Δ_l} are nonnegative, we gauge the MLMC potential by taking the mean over its elements, σ_{Δ_l} , for $l \in [1, L]$.

5.2 MSE

Let $\mathcal{E}_l^2$ denote the matrix of element-wise squared biases, $\mathbb{B}_e^2[K^*]$, while $\Sigma = \mathbb{V}_e[K^*]$. Here $*$ denotes the method applied to calculate K . We calculate both $\mathcal{E}^2$ and Σ by bootstrapping.

Since all elements in $\mathcal{E}^2$ and Σ are nonnegative, we gauge their relative magnitudes by their respective means over the elements, ξ^2 and σ . The estimate of ρ is obtained by summing ξ^2 and σ .

5.3 Ensemble-based DA

We test EnKF, MLEnKF, and MLHENKF on RM1, RM2, and the toy model with three different error sequences,

Fig. 5 $\log_2(\sigma_0)$, and $\log_2(\sigma_{\Delta_l})$ versus l for $l \in [1, L]$. Solid curve, RM1; dashed curve, RM2

$\{\epsilon_l\}_{l=0}^4$, corresponding to low error, medium error, and reservoir error, respectively.

5.3.1 Toy model

We generate an ensemble of synthetic data for the DA by drawing from $\mathcal{N}(10, 0.1)$. All numerical experiments are run 50 times, and for each run, we store the mean of the analyzed state, facilitating calculation of the KDEs of the mean estimates.

To calculate $\{N_l^o\}_{l=0}^5$, knowledge of $\{Q_l^n\}_{l=0}^5$ is required. In Section 4.2, we selected $\{\epsilon_l\}_{l=0}^4$ for the reservoir-error case to mimic the behavior of RM2. We therefore choose to mimic also the cost structure of RM2 (even though the true Q_l^n obviously does not change with l for the toy model), that is, we assume that the normalized level-specific costs are given by Eq. 29 for the reservoir-error case. We also keep the same cost structure for the low-error and medium-error cases, to be able to assess the effect on the estimation results of changes in error type for a fixed cost structure.

5.3.2 Reservoir models

We generate an ensemble of synthetic data for the DA by running M_L with a reference permeability field drawn from the prior model, and adding random errors drawn from $\mathcal{N}(0, \mathbb{C}_d)$, where $\mathbb{C}_d$ is given as a diagonal matrix with elements given as the square of the standard deviation defined in Appendix 4.3.

We assume that the total computational resources available for each algorithm, EnKF, MLEnKF, and MLHENKF, are proportional to the cost of running RM1/RM2 on level 5, that is, $R = Q = N_5 Q_5^n$. With application of EnKF to reservoir history matching, an ensemble size of 100 is common practice. We will therefore select $N_5 = 100$ for both RM1 and RM2. We will, however, also run RM1 and RM2 with $N_5 = 10$, and the reason is as follows. For RM1/RM2, $G_5 = 3600$, while grid sizes are $\mathcal{O}(10^5 - 10^6)$ for a field case. Since sampling errors generally increase with grid size and decrease with ensemble size, it is of interest to select a low N_5 for the problem we consider in order to better mimic the sampling errors for a field case.

Both MLEnKF and MLHENKF will be run with the sequence of level-specific ensemble sizes, $\{N_l^o\}_{l=0}^5$, that minimize $\mathbb{V}[E_{ML}(Y_5)]$ when Q is fixed. We evaluate the quality of the results by comparing them to results from EnKF run on level 5 with an ensemble size, $N_5^* = 5000$, large enough to make sampling errors negligible, but which is unrealistic in real applications.

To test the effect of localization, we perform localized EnKF on RM1. To select the effective radius of the localization matrix, we perform the DA using a wide range of different radii. We then selected the radius that gave the best result. This technique is not possible for realistic cases and must be considered as an upper bound for the quality of localized EnKF. In addition, we perform localized MLHENKF on RM2 with $N_5 = 10$.

To provide a fair evaluation of the methods, it is necessary to provide plots in the same color range. However, with a unified color range, some details might be missing from the plots. For this reason, we provide two plots of almost all the results, one plot with the unified color range and one with an automatically selected color range.

6 Numerical results

6.1 MLMC potential

Results for the toy model are given in Appendix 5.1. In summary, MLEnKF performs very well for the low-error case and poorly for all others.

Figure 5 shows $\log_2(\sigma_0)$, and $\log_2(\sigma_{\Delta_l})$ for $l \in [1, L]$. Table 3 lists values for $\{N_l^o\}_{l=0}^L$ obtained by inserting the estimated values for σ_0 , $\{\sigma_{\Delta_l}\}_{l=1}^L$, and the

Table 3 Reservoir models. Estimates of $\{N_l^o\}_{l=0}^L$

l		0	1	2	3	4	5
RM1	$N_L \times$	123.6	43.6	14	2.2	0.97	0.56
RM2	$N_L \times$	55.6	22	8	1.1	0.5	0.77

Table 4 Reservoir models. Estimates of $\frac{\mathbb{V}_\kappa^o(E_{ML}(Y_L))}{\mathbb{V}_\kappa^o(E(Y_L))}$

RM1	RM2
0.27	0.81

corresponding costs, into Eq. 44. Table 4 lists values for $\mathbb{V}_\kappa^o(E_{ML}(Y_L))/\mathbb{V}_\kappa^o(E(Y_L))$ obtained by inserting the estimated values for σ_0 , $\{\sigma_{\Delta_l}\}_{l=1}^L$, and the corresponding costs, into Eq. 43, and the calculated value for σ_L into Eq. 4.

Figure 5 shows that $\log_2(\sigma_{\Delta_l})$ does not decrease monotonically when l increases for any of the reservoir models. Table 4 shows that the variance of MLEnKF is lower than that of EnKF for RM1 and comparable for RM2. In summary, MLEnKF cannot be expected to perform significantly better than EnKF for the reservoir models.

6.2 MSE

Figure 6 shows the $\log_2(\rho)$ of the Kalman gain for both reservoir models. MLHENKF provides the most accurate (in terms of ρ) estimate of Kalman gain for the range of computational resources considered. Note that both MLEnKF and EnKF will asymptotically be more accurate than MLHENKF since they are unbiased estimators. However, for standard computational resources, MLHENKF will give the best result.

6.3 Ensemble-based DA

6.3.1 Toy model

Details regarding this experiment can be found in Appendix 5.2. As can be expected from Section 3.2.2 and Appendix 5.1, MLEnKF performs very well in the low-error case and poorly in the medium-error and reservoir-error cases.

As can be expected from Section 3.3.2, MLHENKF performs better in estimation of the parameter that is weakly

correlated to the data than in estimation of the parameter that is strongly correlated to the data. In a real reservoir history matching application with production data, most parameters will be weakly correlated to the data.

The toy model experiments indicate that MLMC potential is a good predictor of MLEnKF, and that MLHENKF depends on correlation strengths. In summary, it is expected that MLHENKF will outperform both EnKF and MLEnKF in a real reservoir history matching application with production data.

6.3.2 Reservoir model 1

Normal ensemble size Figure 7 a shows the reference $\log k(x)$ mean field, that is, the field obtained with EnKF and $N_5 = 5000$. Results in this paragraph other than the reference solution are obtained with computational resources corresponding to a normal ensemble size, $N_5 = 100$. Hence, the different methods use the same computational resources. Figure 7 b–e show the $\log k(x)$ mean fields estimated with MLEnKF, EnKF, localized EnKF, and MLHENKF, respectively. Results obtained with MLHENKF are good; however, it fails to capture the highest values. The results from the EnKF are too rough and there are signs of spurious correlations. Localization does significantly improve the quality of the EnKF result. Results obtained with MLEnKF are poor, with severe overshoot and undershoot. Figure 8 a shows the reference $\log k(x)$ variance field, while Fig. 8 b–e show the $\log k(x)$ variance fields estimated with MLEnKF, EnKF, localized EnKF, and MLHENKF, respectively.

Results obtained with MLHENKF are similar to the reference solution in large parts of the reservoir, but the variance close to the wells is too high, and the region with reduced variance is too big. Results obtained with EnKF have too low variance in large parts of the reservoir, but captures the variance in the vicinity of the wells reasonably well. Localization increases the variance, but the results obtained with localized EnKF contain large overshoots and the structure is not similar to the reference solution. Results

Fig. 6 Mean squared error of Kalman gain as a function of computational resources. Full line, MLEnKF; dashed line, EnKF; dotted line, MLHENKF. **a** RM1. **b** RM2

Fig. 7 RM1. Estimated $\log k(x)$ mean. Computational resources corresponding to $N_5 = 100$. **a–e** Left plot is given with the reference solution color range, right plot is given with automatically generated color range

obtained with MLEnKF are poor, with severe overshoot and undershoot.

Small ensemble size Figure 9a shows the reference $\log k(x)$ mean field. Results in this paragraph other than the reference solution are obtained with computational resources corresponding to a small ensemble size, $N_5 = 10$. Figure 9 b–e show the $\log k(x)$ mean fields estimated with MLEnKF, EnKF, localized EnKF, and MLHENKF, respectively.

Fig. 8 RM1. Estimated $\log k(x)$ variance. Computational resources corresponding to $N_5 = 100$. **a–e** Left plot is given with the reference solution color range, right plot is given with automatically generated color range

Results obtained with EnKF, localized EnKF, and MLEnKF are poor, while MLHENKF performs quite well also for this tough test. Figure 10 a shows the reference $\log k(x)$ variance field, while Fig. 10 b–e show the $\log k(x)$ variance fields estimated with MLEnKF, EnKF, localized EnKF, and MLHENKF, respectively.

Results obtained with EnKF have far too low variance almost everywhere, and localized EnKF has higher variance at several locations, but does not capture the structure of the reference solution, while results obtained with

Fig. 9 RM1. Estimated $\log k(x)$ mean. Computational resources corresponding to $N_5 = 10$. **a–e** Left plot is given with the reference solution color range, right plot is given with automatically generated color range

MLEnKF overshoot and undershoot at some locations. Results obtained with MLHENKF are similar to those obtained in the previous paragraph with $N_5 = 100$, but with somewhat rougher structure.

Summary of results for RM1 Results obtained with MLHENKF are quite good, and outperforms both EnKF and the localized EnKF, both when computational resources

Fig. 10 RM1. Estimated $\log k(x)$ variance. Computational resources corresponding to $N_5 = 10$. **c–e** Left plot is given with the reference solution color range, right plot is given with automatically generated color range

correspond to normal and small ensemble sizes. In line with our expectations based on the discussion in the third paragraph in Section 6.3.1, MLEnKF performs poorly for both ensemble sizes.

6.3.3 Reservoir model 2

Normal ensemble size Figure 11a shows the reference $\log k(x)$ mean field. Figure 11 b–d show the $\log k(x)$ mean fields estimated with MLEnKF, EnKF, and MLHENKF,

Fig. 11 RM2. Estimated $\log k(x)$ mean. Computational resources corresponding to $N_5 = 100$. **a–d** Left plot is given with the reference solution color range, right plot is given with automatically generated color range

respectively. Figure 12a shows the reference $\log k(x)$ variance field, while Fig. 12 b–d show the $\log k(x)$ variance fields estimated with MLEnKF, EnKF, and MLHENKF, respectively. In broad terms, similar comments regarding the performances of the different methods as for RM1 with normal ensemble size apply also here. However, the MLHENKF mean contain less structure for this example.

Small ensemble size Figure 13a shows the reference $\log k(x)$ mean field. Figure 13b–e show the $\log k(x)$ mean fields estimated with MLEnKF, EnKF, MLHENKF, and localized MLHENKF, respectively. Figure 14a shows the reference $\log k(x)$ variance field, while Fig. 14b–e show the $\log k(x)$ variance fields estimated with MLEnKF, EnKF, MLHENKF, and localized MLHENKF, respectively. Estimated means for EnKF and MLEnKF are significantly out of range in many locations, while estimated variances are significantly (EnKF) and severely (MLEnKF) out of range almost everywhere. Results for MLHENKF are within

Fig. 12 RM2. Estimated $\log k(x)$ variance. Computational resources corresponding to $N_5 = 100$. **c–d** Left plot is given with the reference solution color range, right plot is given with automatically generated color range

the range of the reference solution and reasonably good. By adding localization to the MLHENKF, the result for the mean is improved, while the result for the variance is worsened by the localization.

Summary of results for RM2 MLHENKF outperforms both EnKF and MLEnKF. MLHENKF results are, however, not quite as good as for RM1. This probably reflects that bias is larger for RM2 than for RM1, due to errors induced by grid coarsening and upscaling of the fault present in RM2 in models M_4 – M_0 . In the small ensemble case, it was shown that adding localization to MLHENKF affects the results.

7 Summary and conclusions

Multilevel ensemble-based DA methods were considered as an alternative to standard (single-level) ensemble-based DA for reservoir history matching problems. While standard ensemble-based DA methods become unreliable

Fig. 13 RM2. Estimated $\log k(x)$ mean. Computational resources corresponding to $N_5 = 10$. **c–e** Left plot is given with the reference solution color range, right plot is given with automatically generated color range

for applications with restricted computational resources, the multilevel approach aims to retain both high statistical accuracy and numerical accuracy by combining the different levels in a multilevel numerical model. In this paper, we investigated two multilevel ensemble-based DA methods, one previously published, the MLEnKF, and one novel method, the MLHENKF.

The expected efficiency of the two multilevel methods relied on the sequence of multilevel approximations for the

Fig. 14 RM2. Estimated $\log k(x)$ variance. Computational resources corresponding to $N_5 = 10$. **c–e** Left plot is given with the reference solution color range, right plot is given with automatically generated color range

different multilevel models. Based on this, we formulated predictors of the DA results from the two multilevel methods. To test these predictors, we investigated a toy model with a controllable multilevel sequence. However, our main focus was reservoir history matching, and we also investigated the multilevel sequence of two multilevel reservoir models. Following the investigation of expected efficiency, we performed history matching for the toy model and the reservoir models.

The numerical investigation showed that MLHENKF consistently outperformed the standard (single-level) ensemble-based DA for reservoir history matching. The discussion of expected efficiency explained the reasons for this result. For problems with low true correlation between the parameter and data, where the bias effect was smaller than the variance effect, MLHENKF gave good results, while for models where the multilevel correction variance, σ_{Δ_l} , was monotonically decreasing, the MLENKF gave superior result. Hence, depending on the multilevel model, different multilevel methods may be optimal. The numerical investigation showed that MLHENKF was suitable for the multilevel reservoir model used in this paper.

Other multilevel reservoir models, more suitable for MLENKF, can be designed. To achieve a satisfactory evolution of σ_{Δ_l} , there can only be a small numerical difference between the model approximations at level l and at level $l - 1$. However, to ensure that first levels of the multilevel model are cheap, while restricting the size of the total amount of levels (L), the difference in computational cost of two consecutive levels must be sufficiently large. For the type of multilevel reservoir models considered in this paper, a significant difference in computational cost also gives a large difference in numerical accuracy. A satisfactory evolution of σ_{Δ_l} can therefore only be achieved if L is sufficiently large. This will, in turn, require that the total available resources is sufficiently large, which is not possible for real applications such as reservoir history matching where the computational resources are restricted.

Since it is difficult to imagine multilevel reservoir models with small L and satisfactory evolution of σ_{Δ_l} , the MLHENKF seems to be more suitable than MLENKF for reservoir history matching. Most of the parameters in the reservoir model are weakly correlated to the data, and we have demonstrated that for a relatively simple multilevel model, the MLHENKF gives the most accurate estimate of the Kalman gain, with reasonable computational resources available.

In this paper, we have focused on the original algorithms without localization. However, depending on the complexity of the problem and the available computational resources, all methods investigated could benefit from a correctly implemented localization. This was demonstrated for the EnKF for RM1, and for MLHENKF for RM2. Note, since the statistical accuracy is higher for the multilevel methods, localization is expected to become more robust, i.e., negative effects of not using an *optimal* localization region becomes less severe. Combining the multilevel DA with localization on large cases is a possible topic for future work.

Funding information The authors are grateful for the financial support provided by the NORCE research project Assimilating 4D Seismic Data: Big Data Into Big Models (4DSEIS) which is funded by industry partners Aker BP, Equinor, Lundin, Repsol, and Total EP NORGE, as well as the Research Council of Norway (PETROMAKS 2)

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix 1. Bayesian model averaging

The expectation and covariance of Y is obtained from the laws of total expectation and total covariance, resulting in (see, e.g., [33])

$$\mathbb{E}[Y] = \sum_{s=1}^S p(M_s) \mathbb{E}[Y|M_s], \quad (35)$$

$$\begin{aligned} \mathbb{C}[Y] = \sum_{s=1}^S p(M_s) \{ & \mathbb{C}[Y|M_s] \\ & + (\mathbb{E}[Y|M_s] - \mathbb{E}^S(\mathbb{E}[Y|M_s])) \\ & \times (\mathbb{E}[Y|M_s] - \mathbb{E}^S(\mathbb{E}[Y|M_s]))^T \}. \end{aligned} \quad (36)$$

Empirical estimates for $\mathbb{E}[Y]$ and $\mathbb{C}[Y]$ are obtained by replacing expectations and covariances on the right-hand sides of Eqs. 35 and 36 by their MC counterparts

$$E_{BA}(Y) = \sum_{s=1}^S p(M_s) E^{N_s}(Y|M_s), \quad (37)$$

$$\begin{aligned} C_{BA}(Y) = \sum_{s=1}^S p(M_s) \{ & C^{N_s}(Y|M_s) \\ & + (E^{N_s}(Y|M_s) - E^{N_S}(Y)) \\ & \times (E^{N_s}(Y|M_s) - E^{N_S}(Y))^T \}, \end{aligned} \quad (38)$$

where $N_S = \sum_{s=1}^S N_s$.

Appendix 2. Multilevel Monte Carlo - theoretical efficiency

Since all elements in the sequence $\{\omega_l\}_{l=0}^L$ are independent, $\mathbb{V}[E_{ML}(Y_L)]$ can be expressed as a sum of contributions from the individual terms on the right-hand side of Eq. 13

$$\mathbb{V}[E_{ML}(Y_L)] = N_0^{-1} \mathbb{V}[Y_0^n(\omega_0)] + \sum_{l=1}^L N_l^{-1} \mathbb{V}[Y_{\Delta_l}^n(\omega_l)]. \quad (39)$$

The computational cost involved in evaluating $E_{ML}(Y_L)$ is

$$Q(E_{ML}(Y_L)) = N_0 Q(Y_0^n(\omega_0)) + \sum_{l=1}^L N_l Q(Y_{\Delta_l}^n(\omega_l)). \quad (40)$$

Let δ denote a real number, and let Q_δ denote the computational cost required to achieve $\mathbb{V}[E_*(Y_L)] = \delta^2$, where the subscript “*” can denote either MC or MLMC.

Furthermore, let $\{N_l^o\}_{l=0}^L$ be the optimal sample sizes for MLMC, that is, those that minimize $\mathbb{V}[E_{ML}(Y_L)]$ under the constraint that $Q_\delta(E_{ML}(Y_L)) = \kappa$ is fixed. One may then show (see, e.g., [20]) that the optimal ensemble sizes are given by

$$N_l^o = \delta^{-2} \sqrt{\frac{\mathbb{V}_{\Delta l}^n}{Q_{\Delta l}^n}} \left(\sqrt{Q_0^n \mathbb{V}_0^n} + \sum_{l=1}^L \sqrt{Q_{\Delta l}^n \mathbb{V}_{\Delta l}^n} \right), \quad (41)$$

where the shorthand notation Q_0^n has been introduced to denote $Q(Y_0^n(\omega_0))$, and so on. The corresponding cost and minimized variance may be expressed as

$$Q_\delta^o(E_{ML}(Y_L)) = \delta^{-2} \left(\sqrt{Q_0^n \mathbb{V}_0^n} + \sum_{l=1}^L \sqrt{Q_{\Delta l}^n \mathbb{V}_{\Delta l}^n} \right)^2, \quad (42)$$

$$\mathbb{V}_\kappa^o(E_{ML}(Y_L)) = \kappa^{-1} \left(\sqrt{Q_0^n \mathbb{V}_0^n} + \sum_{l=1}^L \sqrt{Q_{\Delta l}^n \mathbb{V}_{\Delta l}^n} \right)^2, \quad (43)$$

respectively. Let R denote the total available computational resources. Inserting Eqs. 42 into 41 and assuming that $\kappa = R$, the optimal ensemble size for level l can be given as

$$N_l^o = R \sqrt{\frac{\mathbb{V}_{\Delta l}^n}{Q_{\Delta l}^n}} \left(\sqrt{Q_0^n \mathbb{V}_0^n} + \sum_{l=1}^L \sqrt{Q_{\Delta l}^n \mathbb{V}_{\Delta l}^n} \right)^{-1}. \quad (44)$$

Giles [20] presents arguments for why

$$Q_\delta^o(E_{ML}(Y_L)) < Q_\delta(E(Y_L)) \quad (45)$$

in the two extreme cases where either $\sqrt{Q_0^n \mathbb{V}_0^n}$ or $\sqrt{Q_{\Delta L}^n \mathbb{V}_{\Delta L}^n}$ dominates the other terms, such that either $Q_\delta^o(E_{ML}(Y_L)) \approx \delta^{-2} Q_0^n \mathbb{V}_0^n$ or $Q_\delta^o(E_{ML}(Y_L)) \approx \delta^{-2} Q_{\Delta L}^n \mathbb{V}_{\Delta L}^n$.

Appendix 3. Multilevel EnKF

With MLEnKF there will be L forecast ensembles, and two corresponding (but otherwise arbitrary) ensemble members in forecast ensembles number $l-1$ and l are given by

$$Y_{l-1}^n(\omega_l) = M_{l-1}(Z_{pr}^n(\omega_l)), \quad (46)$$

$$Y_l^n(\omega_l) = M_l(Z_{pr}^n(\omega_l)). \quad (47)$$

Empirical estimates of the two first moments of the forecast distribution are obtained as $E_{ML}(Y_L)$ and $C_{ML}(Y_L)$, see, Eqs. 13 and 12. The corresponding ensemble members after assimilation of the data are given by

$$Z_{l-1}^n(\omega_l) = Y_{l-1}^n(\omega_l) + K_{ME}(d^n - H Y_{l-1}^n(\omega_l)), \quad (48)$$

$$Z_l^n(\omega_l) = Y_l^n(\omega_l) + K_{ME}(d^n - H Y_l^n(\omega_l)), \quad (49)$$

where the multilevel Kalman gain is given by

$$K_{ME} = C_{ML}(Y_L) H^T (H C_{ML}(Y_L) H^T + \mathbb{C}_d^{-1}). \quad (50)$$

Empirical estimates of the two first moments of the analysis distribution are obtained as $E_{ML}(Z_L)$ and $C_{ML}(Z_L)$, respectively. It has been shown [26] that $E_{ML}(Z_L)$ and $C_{ML}(Z_L)$ converge to the KF analysis mean and covariance in the case where $N_l \rightarrow \infty$ for all involved values of l . Hence, the MLEnKF is an unbiased estimator for the KF.

Appendix 4. Numerical models

4.1 Motivating example

A square 2D reservoir with slightly compressible two-phase flow of oil and water is modelled. M_L is discretized on a 90×90 grid, while the approximate models are generated with $L = 2$ and $\Psi = 3$. Here Ψ denotes the number of grid cells that are merged in each coordinate direction during coarsening. The permeability field, $k(x)$, on the 90×90 grid is considered as the unknown parameter function, while porosity is modelled as homogeneous and equal to 0.4.

The prior model for $k(x)$ is Gaussian with mean $3.5 \times 10^{-13} \text{m}^2$ and covariance generated from a spherical variogram model with standard deviation $3.7 \times 10^{-14} \text{m}^2$ and range 10 grid cells. Furthermore, the anisotropy ratio is 0.5, and the main principal anisotropy axis is along the horizontal axis. A realization from the prior model is shown in Fig. 15a, while its representations on level 1 and level 0 are shown in Fig. 15 b and c, respectively.

The reservoir contains a five-spot well pattern with an injector in the center and producers in each corner. The water injection rate is set to $0.003 \text{m}^3/\text{s}$ and the production wells operate at a constant pressure of $25 \times 10^6 \text{Pa}$.

The flow equations are solved by the SimSim [31] reservoir simulator. The four production rates and the BHP in the injector are measured every 30 days for 240 days, starting at day 10. Hence, there are 40 measurements available altogether. The level-specific computational cost, Q_l^n , is gauged by the wall-clock time for runs on the different levels. Normalization with respect to Q_2^n results in

$$[\hat{Q}_0^n, \hat{Q}_1^n, \hat{Q}_2^n] \approx \left[\frac{1}{15}, \frac{1}{4}, 1 \right]. \quad (51)$$

4.2 Toy model

To understand the behavior of the different multilevel DA methods, it is convenient to control the evolution of the level special-mean and variance. To this end, we design a multilevel family of bivariate toy models, $\{M_l\}_{l=0}^L$. Let

Fig. 15 **a** A realization from the prior model for $k(x)$. **b** Its representation on level 1. **c** Its representation on level 0

Table 5 Toy model error-level coefficients

	l	0	1	2	3	4
Low error	μ_l	1.5	1.05	1.005	1.0005	1.00005
	γ_l	1.5	1.05	1.005	1.0005	1.00005
Medium error	μ_l	2	1.8	1.4	1.4	1.1
	γ_l	2	1.8	1.4	1.5	1.1
Reservoir error	μ_l	1.97	1.78	1.61	1.61	1.5
	γ_l	1.21	2.16	2.85	2.86	2.56

Fig. 16 Kernel density estimates of forecasts. Level 1: dotted curve, level 3: dashed curve, level 5: solid curve. **a** RM2. **b** Toy model.

Fig. 17 Log- $k(x)$ realizations with well locations. **a** RM1. **b** RM2 (with fault position indicated)

$m^T = (m_1, m_2)$, where m_1 and m_2 are independent, and define the outcome from M_L by

$$y_L(m) = 10 + 0.2m_1 + m_2. \quad (52)$$

Hence, m_1 is weakly correlated to the data while m_2 is strongly correlated to the data. We define the outcomes from M_l for $l \in [0, L - 1]$ by

$$y_l(m) = y_L(m) + \epsilon_l, \quad (53)$$

where ϵ_l is a level-specific error term. It is useful to define ϵ_l such that one can relate the level-specific mean and variance to the high-resolution mean and variance via two control variables. To achieve this we let

$$\epsilon_l = \mu_l E(y_L(m)) + \theta_l, \quad (54)$$

where θ_l denotes a realization from $\mathcal{N}(0, \gamma_l V(y_L(m)))$, $\mathcal{N}$ denotes the Gaussian distribution and V denotes the empirical variance. The variables μ_l and γ_l are then defined as the fractions of the level-specific empirical mean and variance to the high-resolution empirical mean and variance, respectively;

$$\mu_l = \frac{E(y_l)}{E(y_L)}, \quad (55)$$

$$\gamma_l = \frac{V(y_l)}{V(y_L)}. \quad (56)$$

The prior model for m is $\mathcal{N}(\bar{m}, C_m)$, where

$$\bar{m} = [0.5, 1.1]^T,$$

and

$$C_m = \begin{pmatrix} 2, & 0 \\ 0, & 2 \end{pmatrix}.$$

Three different multilevel structures are defined; firstly, a case with low errors that decrease monotonically with l (low error), secondly, a case mimicking the structure of RM2 (reservoir error), and, finally, a case in between the low-error and the reservoir-error cases (medium error). The corresponding values of μ_l and γ_l are listed in Table 5.

The reservoir-error case is established by selecting Y as a single reservoir-simulation output, the well oil production

rate from P after 45 timesteps. We estimate $\{E(y_l)\}_{l=0}^L$ and $\{V(y_l)\}_{l=0}^L$ from the level-specific forecasts of RM2 by bootstrapping, and calculate μ_l and γ_l using Eqs. 55 and 56, thereby ensuring that the level-specific errors in the toy model have the same relative sizes as those of RM2.

In Fig. 16a, we plot the kernel density estimates (KDEs), calculated from 2000 runs, of the level-specific forecasts of the selected simulation output from RM2. The KDEs of the level-specific forecasts from the toy model (Fig. 16) shows that, with the estimated values for μ_l and γ_l , the multilevel structure of the toy model is similar to that of RM2. (In Fig. 16, we plot KDEs only for $l = 1, 3$ and 5. However, KDEs for $l = 0, 2$ and 4 show similar behavior.)

For reasons given in the first paragraph of Section 3.3.2, the toy model is designed such the correlation between m_1 and y is much smaller than between m_2 and y . Hence, variance dominates for the m_1 while bias dominate for m_2 .

4.3 Reservoir model 1

A square 2D reservoir with compressible two-phase flow of oil and water is modelled. M_L is discretized on a 60×60 grid, while the approximate models are generated with $L = 5$ and $\Psi = 2$. The log permeability field, $\log k(x)$, on the 60×60 grid is considered as the unknown parameter function, while porosity is modelled as homogeneous and equal to 0.2.

The prior model for $\log k(x)$ is Gaussian with mean 5 and covariance generated from a spherical variogram model with standard deviation 1 and range 30 grid cells. Furthermore, the anisotropy ratio is 0.33, and the main principal anisotropy axis is rotated 45 degrees counterclockwise with respect to the positive horizontal axis.

The reservoir contains a producer (P) in the lower left corner, and an injector (I) in the upper right corner. Both wells are controlled by the bottom hole pressure (BHP), with a target of 27.5×10^6 Pa for the injector and 10.3×10^6 Pa for the producer. Figure 17 a shows a realization from the prior log permeability model, along with the positions of the wells.

The flow equations are solved by the commercial reservoir simulator ECLIPSE [40]. There are 80 report steps, and at each step, the oil production rate from P and the water injection rate from I are reported. Hence, there are 160 measurements available altogether. The standard deviation is defined as 20% of the observed values.

4.4 Reservoir model 2

Reservoir model 2 (RM2) (Fig. 17b) is an exact copy of Reservoir model 1 (RM1), except that in RM2 an impermeable fault from the lower right corner to the upper left corner, ending five grid cells from the boundaries, is present in the reservoir. The fault makes the upscaling procedure more challenging, and is added to introduce more bias in $\{M_l\}_{l=0}^{L-1}$. The flow equations are solved by ECLIPSE and the data setup is as for RM1.

Appendix 5. Numerical results—toy model

5.1 MLMC potential

Figure 18 shows $\log_2(\sigma_0)$, and $\log_2(\sigma_{\Delta_l})$ for $l \in [1, 5]$. Table 6 lists values for $\{N_l^o\}_{l=0}^5$ obtained by inserting the

Fig. 18 $\log_2(\sigma_0)$, and $\log_2(\sigma_{\Delta_l})$ versus l for $l \in [1, L]$ for Toy model. Solid curve: low error, dashed curve: medium error, dotted curve: reservoir error

Table 6 Toy model. Estimates of $\{N_l^o\}_{l=0}^5$

l		0	1	2	3	4	5
Low error	$N_L \times$	493.9	39.07	7.66	2.63	1.8	0.37
Medium error	$N_L \times$	185.1	5.76	5.48	0.73	1.62	0.57
Reservoir error	$N_L \times$	62.65	12.41	2.26	0.015	0.17	0.92

estimated values for σ_0 , $\{\sigma_{\Delta_l}\}_{l=1}^5$, and the costs in Eqs. 29, into 44. Table 7 lists values for $\mathbb{V}_k^o(E_{ML}(Y_L))/\mathbb{V}_k^o(E(Y_L))$ obtained by inserting the estimated values for σ_0 , $\{\sigma_{\Delta_l}\}_{l=1}^5$, and the costs in Eqs. 29, into 43, and the calculated value for σ_5 into Eq. 4. Figure 18 shows that $\log_2(\sigma_{\Delta_l})$ decreases monotonically when l increases in the low-error case, while σ_{Δ_l} does not show systematic behavior in the two other cases. Table 6 shows that considerably more computational resources can be distributed to the lower levels in the low-error case than in the two other cases. Table 7 shows that the variance of MLEnKF is orders of magnitude lower than that of EnKF in the low-error case, while it is comparable to that of EnKF in the two other cases.

5.2 Ensemble-based DA

Figure 19 shows KDEs for m_1 obtained with 50 mean values from the analyzed ensemble.

Figure 19 a shows results obtained in the low-error case. MLEnKF approximates the true posterior very well, while MLHENKF approximates the true posterior clearly better than EnKF, but clearly not as well as MLEnKF. Figure 19 b and c show results obtained in the medium-error and reservoir-error cases. Here, MLEnKF and EnKF both perform poorly. MLHENKF is biased, particularly in the reservoir-error case, but the variance is small and closer to the true posterior when comparing with the other methods.

Figure 20 shows KDEs for m_2 obtained with 50 mean values from the analyzed ensemble.

Figure 20 a shows results obtained in the low-error case. MLEnKF approximates the true posterior clearly better than EnKF, which does not perform well, while MLHENKF performs worse than EnKF. Figure 20 b and c show results obtained in the medium-error and reservoir-error cases. EnKF approximates the true posterior slightly better than MLEnKF and MLHENKF, which both perform poorly.

Fig. 19 KDEs of m_1 from the analyzed ensemble. Solid curve: True posterior. Dashed curve: MLHENKF. Dash-dotted curve: EnKF. Dotted curve: MLEnKF. **a** Low error. **b** Medium error. **c** Reservoir error.

Table 7 Toy model. Estimates of $\frac{\mathbb{V}_k^o(\mathbb{E}_{ML}(Y_L))}{\mathbb{V}_k^o(\mathbb{E}(Y_L))}$

Low error	Medium error	Reservoir error
0.026	0.28	1.72

Fig. 20 KDEs of m_2 from the analyzed ensemble. Solid curve: True posterior. Dashed curve: MLHENKF. Dash-dotted curve: EnKF. Dotted curve: MLEnKF. **a** Low error. **b** Medium error. **c** Reservoir error.

References

- Aanonsen, S.I., Eydinov, D.: A multiscale method for distributed parameter estimation with application to reservoir history matching. *Comput. Geosci.* **10**(1), 97–117 (2006). <https://doi.org/10.1007/s10596-005-9012-4>
- Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C., Vallès, B.: The ensemble Kalman filter in reservoir engineering—a review. *SPE J.* **14**(3), 393–412 (2009). <https://doi.org/10.2118/117274-PA>
- Axelsson, O.: *Iterative Solution Methods*. Cambridge University Press, Cambridge (1994)
- Chen, Y., Oliver, D.S.: Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math. Geosci.* **44**(1), 1–26 (2012). <https://doi.org/10.1007/s11004-011-9376-z>
- Chen, Y., Oliver, D.S.: Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput. Geosci.* <https://doi.org/10.1007/s10596-013-9351-5> (2013)
- Chernov, A., Hoel, H., Law, K.J.H., Nobile, F., Tempone, R.: Multilevel ensemble Kalman filtering for spatially extended models. *arXiv:1608.08558* (2016)
- Durlafsky, L.: Upscaling of geocellular models for reservoir flow simulation: a review of recent progress. 7th International Forum on Reservoir Simulation Bülh/Baden-Baden, Germany, 23–27 (2003)
- Emerick, A.A., Reynolds, A.C.: History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Comput. Geosci.* **16**(3), 639–59 (2012). <https://doi.org/10.1007/s10596-012-9275-5>
- Emerick, A.A., Reynolds, A.C.: Ensemble smoother with multiple data assimilation. *Comput. Geosci.* **55**, 3–15 (2013). <https://doi.org/10.1016/j.cageo.2012.03.011>
- Emerick, A.A., Reynolds, A.C.: Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. *Comput. Geosci.* **17**(2), 325–350 (2013). <https://doi.org/10.1007/s10596-012-9333-z>
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5), 10,143 (1994). <https://doi.org/10.1029/94JC00572>
- Farmer, C.L.: Upscaling: a review. *Int. J. Numer. Methods Fluids* **40**(1–2), 63–78 (2002). <https://doi.org/10.1002/fld.267>
- Flowerdew, J.: Towards a theory of optimal localisation. *Tellus Series A: Dynamic Meteorology and Oceanography* **67**(1). <https://doi.org/10.3402/tellusa.v67.25257> (2015)
- Fossum, K., Mannseth, T.: Parameter sampling capabilities of sequential and simultaneous data assimilation: I. Analytical comparison. *Inverse Problems* **114**(11), 002 (2014). <https://doi.org/10.1088/0266-5611/30/11/114002>
- Fossum, K., Mannseth, T.: Parameter sampling capabilities of sequential and simultaneous data assimilation: II. Statistical analysis of numerical results. *Inverse Problems* **114**(11), 003 (2014). <https://doi.org/10.1088/0266-5611/30/11/114003>
- Fossum, K., Mannseth, T.: Assessment of ordered sequential data assimilation. *Comput. Geosci.* (1). <https://doi.org/10.1007/s10596-015-9492-9> (2015)
- Fossum, K., Mannseth, T.: Coarse-scale data assimilation as a generic alternative to localization. *Comput. Geosci.* **21**(1), 167–186 (2017). <https://doi.org/10.1007/s10596-016-9602-3>
- Gentilhomme, T., Oliver, D.S., Mannseth, T., Caumon, G., Moya, R., Doyen, P.: Ensemble-based multi-scale history-matching using second-generation wavelet transform. *Comput. Geosci.* **19**(5), 999–1025 (2015). <https://doi.org/10.1007/s10596-015-9517-4>
- Giles, M.B.: Multi-level Monte Carlo path simulation. *Oper. Res.* **56**(3), 607–617 (2008). <https://doi.org/10.1287/opre.1070.0496>
- Giles, M.B.: Multilevel Monte Carlo methods. *Acta Numerica* **24**(2015), 259–328 (2015). <https://doi.org/10.1017/S096249291500001X>
- Grimstad, A.A., Mannseth, T.: Nonlinearity, scale, and sensitivity for parameter estimation problems. *SIAM J. Sci. Comput.* **21**(6), 2096–2113 (2000). <https://doi.org/10.1137/S1064827598339104>
- Grimstad, A.A., Mannseth, T., Nævdal, G., Urkedal, H.: Adaptive multiscale permeability estimation. *Comput. Geosci.* **7**(1), 1–25 (2003)
- Guo, Z., Reynolds, A.C.: INSIM-FT in three-dimensions with gravity. *J. Comput. Phys.* **380**, 143–169 (2019). <https://doi.org/10.1016/j.jcp.2018.12.016>
- Hamill, T.M., Snyder, C.: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Weather. Rev.* **128**(8), 2905–2919 (2000). [https://doi.org/10.1175/1520-0493\(2000\)128<2905:AHEKFS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFS>2.0.CO;2)
- He, J., Sarma, P., Durlafsky, L.J.: Reduced-order flow modeling and geological parameterization for ensemble-based data assimilation. *Comput. Geosci.* **55**, 54–69 (2013). <https://doi.org/10.1016/j.cageo.2012.03.027>. <http://linkinghub.elsevier.com/retrieve/pii/S0098300412001227>
- Hoel, H., Law, K.J.H., Tempone, R.: Multilevel ensemble Kalman filtering. *SIAM J. Numer. Anal.* **54**(3), 1813–1839 (2016). <https://doi.org/10.1137/15M100955X>
- Houtekamer, P.L., Mitchell, H.L.: Data assimilation using an ensemble Kalman filter technique. *Mon. Weather. Rev.* **126**(1969), 796–811 (1998). [https://doi.org/10.1175/1520-0493\(1998\)126<0796:DAUAEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2)
- Iglesias, M.A.: Iterative regularization for ensemble data assimilation in reservoir models, vol. 19. <https://doi.org/10.1007/s10596-014-9456-5> (2015)
- Iglesias, M.A., Law, K.J.H., Stuart, A.M.: Ensemble Kalman methods for inverse problems. *Inverse Problems* **045**(4), 001 (2013). <https://doi.org/10.1088/0266-5611/29/4/045001>
- Iglesias, M.A., Law, K.J.H., Stuart, A.M.: Evaluation of Gaussian approximations for data assimilation in reservoir models. *Comput. Geosci.* **17**(5), 851–85 (2013). <https://doi.org/10.1007/s10596-013-9359-x>
- Jansen, J.D.: SimSim: a simple reservoir simulator (2011)
- Lerlertpakdee, P., Jafarpour, B., Gildin, E.: Efficient production optimization with flow-network models. *SPE J.* **19**(06), 1083–1095 (2014). <https://doi.org/10.2118/170241-pa>
- Li, X., Tsai, F.T.: Bayesian model averaging for groundwater head prediction and uncertainty analysis using multimodel and multimethod. *Water Resour. Res.* **45**(9), 1–14 (2009). <https://doi.org/10.1029/2008WR007488>
- Lødøen, O.P., Omre, H.: Scale-corrected ensemble Kalman filtering applied to production-history conditioning in reservoir evaluation. *SPE J.* **13**(02), 177–194 (2008). <https://doi.org/10.2118/111374-PA>
- Mannseth, T.: Permeability identification from pressure observations: some foundations for multiscale regularization. *Multiscale Mode Simul.* **5**(1), 21–44 (2006). <https://doi.org/10.1137/050630167>
- Mannseth, T.: Comparison of five different ways to assimilate data for a simplistic weakly nonlinear parameter estimation problem. *Comput. Geosci.* (1). <https://doi.org/10.1007/s10596-015-9490-y> (2015)
- Muirhead, R.J.: *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Statistics. Wiley, Hoboken (2005). <https://doi.org/10.1002/9780470316559>. <https://www.jstor.org/stable/2987858?origin=crossrefhttp://doi.wiley.com/10.1002/9780470316559>
- Nævdal, G., Mannseth, T., Brusdal, K., Nordtvedt, J.E.: Multiscale estimation with spline wavelets, with application to two-phase

- porous-media flow. *Inverse Problems* **16**(2), 315–332 (2000). <https://doi.org/10.1088/0266-5611/16/2/304>
39. Oliver, D.S., Chen, Y.: Recent progress on reservoir history matching: a review. *Comput. Geosci.* **15**(1), 185–221 (2010). [10.1007/s10596-010-9194-2](https://doi.org/10.1007/s10596-010-9194-2)
  40. Schlumberger Ltd: ECLIPSE Reservoir simulation software: technical description. Schlumberger Software, London, UK (2009)
  41. Tarrahi, M., Elahi, S.H., Jafarpour, B.: Fast linearized forecasts for subsurface flow data assimilation with ensemble Kalman filter. *Comput. Geosci.*, 929–952. <https://doi.org/10.1007/s10596-016-9570-7>. <http://link.springer.com/10.1007/s10596-016-9570-7> (2016)
  42. Trottenberg, U., Oosterlee, C.: Schuller a.: Multigrid (2001)
  43. Wanderley de Holanda, R., Gildin, E., Jensen, J.L.: A generalized framework for capacitance resistance models and a comparison with streamline allocation factors. *J Petrol Sci Eng* **162**(October 2017), 260–282 (2018). <https://doi.org/10.1016/j.petrol.2017.10.020>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.