

# Coarse-scale data assimilation as a generic alternative to localization

Kristian Fossum<sup>1</sup>  · Trond Mannseth<sup>1</sup>

Received: 27 April 2016 / Accepted: 27 October 2016  
© Springer International Publishing Switzerland 2016

**Abstract** Coarse-scale data assimilation (DA) with large ensemble size is proposed as a robust alternative to standard DA with localization for reservoir history matching problems. With coarse-scale DA, the unknown property function associated with each ensemble member is upscaled to a grid significantly coarser than the original reservoir simulator grid. The grid coarsening is automatic, ensemble-specific and non-uniform. The selection of regions where the grid can be coarsened without introducing too large modelling errors is performed using a second-generation wavelet transform allowing for seamless handling of non-dyadic grids and inactive grid cells. An inexpensive local-local upscaling is performed on each ensemble member. A DA algorithm that restarts from initial time is utilized, which avoids the need for downscaling. Since the DA computational cost roughly equals the number of ensemble members times the cost of a single forward simulation, coarse-scale DA allows for a significant increase in the number of ensemble members at the same computational cost as standard DA with localization. Fixing the computational cost for both approaches, the quality of coarse-scale DA is compared to that of standard DA with localization (using state-of-the-art localization techniques) on examples spanning a large degree of variability. It is found that coarse-scale DA is more robust with respect to variation in example type than each of the localization techniques considered with standard DA. Although the paper is concerned with two spatial

dimensions, coarse-scale DA is easily extendible to three spatial dimensions, where it is expected that its advantage with respect to standard DA with localization will increase.

**Keywords** Coarse-scale data assimilation · Localization · Second-generation wavelets · Upscaling

## 1 Introduction

Sound decision making regarding efficient petroleum-reservoir exploitation requires that uncertain porous-media property functions are estimated and that the associated uncertainties are quantified. The Bayesian framework facilitates uncertainty quantification as an integral part of the history matching procedure. The posterior probability density function (PDF) constitutes the formal Bayesian solution to the problem, but for all other than Gauss-linear problems, the posterior PDF can only be characterized through sampling. Markov chain Monte Carlo (MCMC) methods sample correctly from the posterior PDF, but are known to be prohibitively computationally expensive for realistic reservoir problems. Ensemble-based data assimilation (DA) methods, like the ensemble Kalman filter (EnKF) [23] and the ensemble smoother (ES) [44], sample approximately from the posterior PDF. (They sample correctly in the Gauss-linear case). In recent years, several variants of ensemble-based DA methods [9, 10, 20, 21, 38, 39] have been applied and discussed in a large number of publications concerned with porous-media flow problems. A review of ensemble-based DA methods for reservoir history matching is found in [2]. Evaluation and comparison of different ensemble-based DA methods are found in [22, 26–28, 40, 47].

---

✉ Kristian Fossum  
kristian.fossum@uni.no

<sup>1</sup> Uni Research CIPR, Allegaten 41, 5020 Bergen, Norway

Common to ensemble-based DA methods is that both computational cost and statistical credibility are primarily controlled by the ensemble size,  $N_e$ . The overall computational cost,  $Q$ , roughly equals  $N_e$  times the cost of a forward-model run,  $F$ . With fixed computational resources available, one is therefore often forced to use a smaller  $N_e$  than what one would have preferred to do if statistical quality was the only concern.

Utilizing a small or moderately sized  $N_e$  will give rise to two concerns. Firstly, sampling errors resulting from ensemble approximation of the involved covariance matrices will lead to spurious correlations between quantities that should be uncorrelated. Secondly, since the number of degrees of freedom in the assimilation equals  $N_e$ , attempting to assimilate a large number of data simultaneously, such as with time-lapse seismic data, may lead to ensemble collapse, or at least to unwarranted reduction of the variance in the analysed ensemble.

Attempting to alleviate the negative effects of using a small/moderate ensemble size, localization [37] is often applied. With localization, a spatial region is associated with each observation. Variables associated with grid cells inside this region are allowed to be influenced by the observation, while other variables are not. The most common way to define the region is to associate the region boundary with a distance (that may vary with spatial direction) from the observation location. In [3, 15], sensitivities were applied to determine localization regions, while [8, 19] argued that also the prior model needs to be taken into account.

Ideally, the localization region should be defined such that all real correlations are retained while no false correlations are allowed, but in practice, this is, of course, extremely hard to achieve. For porous-media flow problems, it is particularly challenging due to heterogeneities of porous rocks, which make the shape of the region an important factor in addition to its size. For such problems, it is therefore almost impossible to design a good *generic* localization method, that is, a localization method whose performance does not vary significantly with problem type. Alternative ways to alleviate the negative effects of using a small/moderate ensemble size in a more stable manner are therefore highly desirable.

A conceptually simple way to avoid, or at least significantly reduce, spurious correlations and/or unwarranted reduction of the variance in the analysed ensemble, would be to increase the ensemble size substantially. In order to be able to do that with a fixed  $Q$ ,  $F$  has to decrease by the same proportion as  $N_e$  increases. A way to secure a smaller  $F$  is to replace the reservoir simulator by a computationally less expensive proxy model. Various types of proxy models, like proper orthogonal decompositions (POD) and/or trajectory

piecewise linearization, have been introduced in the context of reservoir production optimization, reservoir history matching, and reservoir DA.

In this paper, an alternative approach, where the proxy is an upscaled version of the original reservoir model, will be utilized to do coarse-scale DA. This type of approach has been used, for example, in multilevel history matching [1], and to improve the mixing of the chain of samples for MCMC methods [18]. An advantage of this type of proxy is a high degree of transparency between the proxy and the original simulation model. The standard approach for modelling multiphase flow in a porous media is through finite volume/finite difference discretization of the governing equations [5]. It is well known that the accuracy of this type of schemes is controlled through the refinement of the mesh. Hence, the accuracy of the proposed proxy model can be entirely controlled through the coarsening of the simulation grid. After discretization and linearization, solving the multiphase flow equations consists of, for each grid point in time, iterating until the non-linear equations have converged where for each iteration of the non-linear equation we must solve a system of linear equations. The computational cost of a forward simulation is, therefore, proportional to the cost of solving the system of linear equations, which scales approximately as  $N_g^\eta$ , where  $N_g$  denotes the number of grid cells and  $\eta \in (1.25, 1.5)$  [4].

Clearly, it will be important to carefully select which regions of the grid that can be coarsened without introducing too large errors in the forward simulations, and which regions that cannot. Following [17, 48, 50], we will apply a wavelet transform to perform this selection. To increase the flexibility and applicability of this approach to non-regular grids and geometries, a second-generation wavelet transform [55] will be utilized.

We will compare the use of coarse-scale DA to DA using localization. Three types of localization will be considered. The first, termed *GC-localization*, is the commonly applied localization procedure proposed by Gaspari and Cohn [31]; the second, termed *CO-localization*, will utilize localization procedures proposed by Chen and Oliver [8]; while the third, termed *ER-localization*, will utilize the localization procedure proposed by Emerick and Reynolds [19].

The rest of the paper is organized as follows. Section 2 discuss ensemble-based DA and how upscaled reservoir models fit into the DA framework, while Section 3 discuss localization. Section 4 is devoted to wavelet transforms and their use within upscaling of reservoir models. Section 5 presents the numerical experiments, while Section 6 analyses the numerical experiments where coarse-scale DA is compared to localization. Finally, in Section 7, we summarize and conclude our findings.

## 2 Ensemble-based data assimilation

Let  $\mathbf{y} \in \mathbb{R}^{N_y}$  denote a joint state vector containing a parameter vector,  $\mathbf{m} \in \mathbb{R}^{N_m}$ , and a dynamic state vector,  $\mathbf{z} \in \mathbb{R}^{N_z}$ , that is,

$$\mathbf{y} = \begin{pmatrix} \mathbf{z} \\ \mathbf{m} \end{pmatrix}. \tag{1}$$

Let  $\mathbf{w} \in \mathbb{R}^{N_z}$  denote the forward-model operator, and let  $t_k$  denote an arbitrary time instance. Define the operator,  $\mathbf{g}_k$ , mapping the joint state from  $t_{k-1}$  to  $t_k$ , by

$$\mathbf{y}_k = \mathbf{g}_k(\mathbf{y}_{k-1}) + \mathbf{r}_k \tag{2}$$

$$= \begin{pmatrix} \mathbf{w}_k(\mathbf{z}_{k-1}, \mathbf{m}_{k-1}) \\ \mathbf{m}_{k-1} \end{pmatrix} + \begin{pmatrix} \mathbf{q}_k \\ \mathbf{0} \end{pmatrix}, \tag{3}$$

where  $\mathbf{q}$  denotes modelling errors, and let  $\mathbf{d}_{1:N_a} = \{\mathbf{d}_1, \dots, \mathbf{d}_{N_a}\}$  denote the sequence of measurement vectors to be assimilated. The observation model for  $\mathbf{d}_k \in \mathbb{R}^{N_{d_k}}$  is given as

$$\mathbf{d}_k = H_k \mathbf{y}_k + \boldsymbol{\xi}_k, \tag{4}$$

where  $H_k$  denotes a matrix of zeros and ones reflecting measurement locations,  $\boldsymbol{\xi}_k \sim \mathcal{N}(\mathbf{0}, \Sigma_{d_k})$  denotes measurement errors,  $\mathcal{N}$  denotes the Gaussian probability distribution and  $\Sigma_{d_k}$  denotes the covariance of the measurement errors.

### 2.1 Filter solutions

The filter solution to the DA problem obtained via the Bayesian approach is defined as the PDF,  $p(\mathbf{y}_{N_a} | \mathbf{d}_{1:N_a}) = p(\mathbf{z}_{N_a}, \mathbf{m} | \mathbf{d}_{1:N_a})$ . With non-restrictive, standard assumptions on the error statistics [24], the filter solution can be obtained by assimilating data sequentially,

$$p(\mathbf{z}_k, \mathbf{m} | \mathbf{d}_{1:k}) \propto p(\mathbf{d}_k | \mathbf{z}_k, \mathbf{m}) p(\mathbf{z}_k, \mathbf{m} | \mathbf{d}_{1:k-1}), \tag{5}$$

hence, the posterior PDF at  $t_{k-1}$ ,  $p(\mathbf{z}_k, \mathbf{m} | \mathbf{d}_{1:k-1})$ , becomes the prior PDF at  $t_k$ .

When the prior model is Gaussian and  $\mathbf{w}_k$  is linear for all involved  $k$ , all the posterior PDF's will be Gaussian, and the solution to the DA problem may be written as the Kalman-filter equations for the mean,  $\boldsymbol{\mu}_y$ , and covariance,  $\Sigma_y$ , at time instance  $t_k$ ,

$$\boldsymbol{\mu}_{y_k}^f = \mathbf{g}_k \boldsymbol{\mu}_{y_{k-1}}^a, \tag{6}$$

$$\Sigma_{y_k}^f = \mathbf{g}_k \Sigma_{y_{k-1}}^a \mathbf{g}_k^T + \Sigma_{q_k}, \tag{7}$$

$$\boldsymbol{\mu}_{y_k}^a = \boldsymbol{\mu}_{y_k}^f + \Gamma_k (\mathbf{d}_k - H_k \boldsymbol{\mu}_{y_k}^f), \tag{8}$$

$$\Sigma_{y_k}^a = \Sigma_{y_k}^f - \Gamma_k H_k \Sigma_{y_k}^f, \tag{9}$$

$$\Gamma_k = \Sigma_{y_k}^f H_k^T (H_k \Sigma_{y_k}^f H_k + \Sigma_{d_k})^{-1}, \tag{10}$$

where the superscripts  $f$  and  $a$  denote forecast and analysis, respectively, and  $\Gamma$  denotes the Kalman gain.

When  $\mathbf{w}_k$  is non-linear, the Kalman filter is no longer applicable, and approximate methods, like the ensemble Kalman filter (EnKF), are used instead. With the EnKF, the involved PDF's are represented by ensembles, and each ensemble member is updated using approximations to the Kalman-filter equations where the true covariances are approximated by empirical covariances. In the description that follows,  $N_e$  ensemble members corresponding to the quantities  $\mathbf{d}$ ,  $\mathbf{m}$ ,  $\mathbf{w}$  and  $\mathbf{z}$ , will be stored as columns in the ensemble matrices  $D$ ,  $M$ ,  $W$  and  $Z$ , respectively. Functional dependencies between ensemble members, like with  $z_e = \mathbf{w}(\mathbf{m}_e)$ , will be denoted by  $Z = W(M)$  when using the ensemble notation. Empirical covariances will be denoted by  $C$  and the approximate Kalman gain obtained from inserting empirical covariances in the expression for  $\Gamma$  in Eq. 10 will be denoted by  $K$ .

If the involved forward models are non-linear, the updated dynamic states obtained with EnKF,  $Z_k^a$ , and the dynamic states at  $t_k$  that would have been obtained by running the forward-model sequence from time zero with the updated parameters,  $M_k^a$ , as input,  $W_k(Z_0, M_k^a)$ , will be statistically inconsistent [58]. The half-iteration EnKF (HIEnKF) [58] avoids this inconsistency by simply replacing  $Z_k^a$  by  $W_k(Z_0, M_k^a)$  before continuing towards the next assimilation time. Hence, the HIEnKF is computationally more costly than the EnKF. We will apply HIEnKF in this paper, and we are interested in estimating  $\mathbf{m}$ , only. The latter is reflected in the version of the HIEnKF algorithm summarized in Algorithm 1.

---

#### Algorithm 1 HIEnKF

---

*Initialize*

$$M_0^a = M_{\text{prior}}$$

$$Z_0 = Z_{\text{prior}}$$

**for**  $k = 1, \dots, N_a$  **do**

**Forecast**

*Forecast parameters and states at  $t_k$  by*

$$M_k^f = M_{k-1}^a$$

$$Z_k^f = W_k(Z_0, M_{k-1}^a)$$

**Analysis**

*Define the HIEnKF Kalman gain for the parameters as*

$$K_k = C_{m_k^f, z_k^f} (C_{z_k^f} + \Sigma_{d_k})^{-1},$$

*and update the ensemble of parameters*

$$M_k^a = M_k^f + K_k (D_k - Z_k^f).$$

**end for**

---

## 2.2 Sequential ensemble-based data assimilation with upscaled forward models

For sequential DA to work properly, it is required that the dimension of the joint-state vector is constant for all assimilation times, confer Eq. 5. Since we apply upscaled forward models in the forecast runs, it is useful to consider how this requirement can be accommodated. Let  $\widehat{W}$  denote the forward model (in ensemble notation) incorporating upscaling, let  $\widehat{Z}$  denote its output, and let  $\widehat{D}$  denote the data observed on the coarse grid. The coarse-scale DA algorithm—HIEnKF with upscaling—is summarized in Algorithm 2.

---

### Algorithm 2 HIEnKF with upscaling

---

*Initialize*

$$M_0^a = M_{\text{prior}}$$

$$Z_0 = Z_{\text{prior}}$$

**for**  $k = 1, \dots, N_a$  **do**

**Forecast**

*Forecast parameters and upscaled states at  $t_k$  by*

$$M_k^f = M_{k-1}^a$$

$$\widehat{Z}_k^f = \widehat{W}_k(Z_0, M_{k-1}^a)$$

**Analysis**

*Define the upscaled HIEnKF Kalman gain for the parameters as*

$$\widehat{K}_k = C_{m_k^f z_k^f} (C_{z_k^f} + \Sigma_{\widehat{a}_k})^{-1},$$

*and update the ensemble of parameters*

$$M_k^a = M_k^f + \widehat{K}_k (\widehat{D}_k - \widehat{Z}_k^f).$$

**end for**

---

Note that the upscaling is considered as an implicit part of the forward model, and the joint-state vector of HIEnKF at the beginning of forecast stage  $k$  is  $(z_0, m_{k-1}^a)^T$ . The requirement that the dimension of the joint state vector should be constant for all assimilation times is, therefore, accommodated. This also holds for other DA algorithms, such as ES, iterative ES [10] and ES with multiple DA [21]. We stress that the main goal of this paper is to investigate the coarse-scale DA approach as a generic alternative to localization. Hence, since both the upscaling approach and localization could be utilized with any of the above-mentioned algorithms, any one of them could potentially be utilized in this investigation. The upscaling approach could, however, not be applied to the EnKF, since the joint-state vector at the beginning of forecast stage  $k$  then would have become  $(\widehat{z}_{k-1}^a, m_{k-1}^a)^T$ . In that case, a downscaling

of  $\widehat{z}_{k-1}^a$  to the scale of  $z_0$  would be required before performing the forecast with the reservoir simulator, see, e.g. [45]. For practical applications of the ensemble-based methods, there is a strong motivation to use methods that restart from initial time, see, e.g. [10, 11, 53]. Hence, considering the HIEnKF method instead of the EnKF is in line with how the ensemble-based methods are utilized in realistic applications.

One extra feature of this approach is that since the joint state vector remains in the same dimension, it is possible to combine localization with the coarse-scale DA approach. This will not be investigated in this work, but would be practical for cases where one knew that some correlations were spurious, e.g. if some areas of the reservoir, and the parameters in this area, were completely isolated from the one or several data points.

In the case of infinite ensemble size, Algorithm 2 converge to the Kalman filter Eqs. 8–9 only with  $z_k$  substituted by the upscaled approximation  $\widehat{z}_k$  [46]. Hence, the accuracy of our solution will depend on the approximation error  $\|z_k - \widehat{z}_k\|$  in some suitable norm. However, since the upscaled model is introduced for cases where the ensemble size is small to moderate, this limit has little practical interest. We will therefore assume that some accuracy in the forward model can be sacrificed for higher accuracy in the Monte-Carlo estimation.

## 2.3 Discussion of alternative approaches

The approach introduced above only places minor restrictions on the selected DA algorithm and requires little modification of the original code. Hence, it is clear that other proxy models could also be utilized, such as POD, or methods based on a neural network. However, to generate such proxy models, snapshots of the original output is required. Hence, a representative set of the original fine-scale reservoir simulator must be run before the DA algorithm can start. The quality of this approximation depends on the snapshots, and whether these contain sufficient variation to capture the potential variation in the reservoir simulator. Due to the dependence on snapshots, there is a more complex relationship between quality and cost for these types of proxy models. Hence, to further compare different proxy models, a thorough numerical analysis is required. This is not in the scope of this work.

An other alternative are methods based on the multilevel Monte-Carlo technique [32, 35], especially the multilevel EnKF (MLEnKF) introduced in [36]. Here, a set of proxy models that represents different levels of approximation of the forward model are utilized to generate a more efficient Monte Carlo algorithm. The major difference between

multilevel approach and the approach introduced here is that, in the case of infinite ensemble size, the former approach converge to Kalman filter Eqs. 8–9, while in our approach the accuracy of the forward simulator remains constant and, as mentioned above, the algorithm converges to an approximate solution. In addition, the multilevel approach is required to balance the computational effort between the different levels, while in our approach, all computational effort is made on a single level. A further comparison of the MLEnKF and data assimilation using upscaled methods are currently in progress, and will be presented in a future paper.

### 3 Localization

The localization methods selected for comparison with coarse-scale DA will now be presented. From Algorithm 1, the expression for the Kalman gain for  $m$  is (the subscript,  $k$ , and superscript,  $f$ , will be omitted in this section, for notational convenience)

$$K = C_{mz} (C_z + \Sigma_d)^{-1}. \tag{11}$$

Only localization of the individual covariance matrices in the expression for the Kalman gain will be considered. (In [8], it was found that localization of the individual covariance matrices seems to be more robust than direct localization of the Kalman gain.)

To describe the localization methods considered in a unified framework, some notation is introduced. The description of the methods utilizes a particular matrix associated with two functions and a particular matrix product. Let  $C_{uv}$  denote the covariance matrix for  $u$  and  $v$ , where  $u \in \mathbb{R}^I$  and  $v \in \mathbb{R}^J$  denote arbitrary vectors whose values vary with spatial location. The element  $c_{uv,ij}$  of  $C_{uv}$  then gives the covariance between  $u_i = u(x_i)$  and  $v_j = v(x_j)$ , where  $x_i$  and  $x_j$  denote two arbitrary spatial locations. Furthermore, let  $\rho(h)$  denote an arbitrary spatial correlation function with compact support, let  $\tau(\rho)$  denote a function to be specified below, let  $h_{ij} = \|x_i - x_j\|$  and let  $L^{\tau\rho}$  denote the  $I \times J$  matrix where  $l_{ij}^{\tau\rho} = \tau(\rho(h_{ij}))$ . The final piece needed to formally describe localization is the Schur (element wise) matrix product ‘ $\circ$ ’. This product is such that  $A = B \circ D$  results in  $a_{ij} = b_{ij}d_{ij}$ , for arbitrary  $I \times J$  matrices  $B$  and  $D$ . Localization of  $C_{uv}$  corresponding to  $\tau$  and  $\rho$  is then performed as  $L_{uv}^{\tau\rho} \circ C_{uv}$ . From Eq. 11, the general expression for the Kalman gain with localized covariance matrices then becomes,

$$K = (L_{mz}^{\tau\rho} \circ C_{mz}) (L_z^{\tau\rho} \circ C_z + C_d)^{-1}. \tag{12}$$

### 3.1 GC-localization

In [31], the authors introduce the Gaspari-Cohn formula for  $\rho(h)$ ,

$$\rho(h) = \begin{cases} -\frac{1}{4} \left(\frac{h}{\Lambda_{GC}}\right)^5 + \frac{1}{2} \left(\frac{h}{\Lambda_{GC}}\right)^4 + \frac{5}{8} \left(\frac{h}{\Lambda_{GC}}\right)^3 - \frac{5}{3} \left(\frac{h}{\Lambda_{GC}}\right)^2 + 1 & 0 \leq h \leq \Lambda_{GC} \\ \frac{1}{12} \left(\frac{h}{\Lambda_{GC}}\right)^5 - \frac{1}{2} \left(\frac{h}{\Lambda_{GC}}\right)^4 + \frac{5}{8} \left(\frac{h}{\Lambda_{GC}}\right)^3 + \frac{5}{3} \left(\frac{h}{\Lambda_{GC}}\right)^2 - 5 \frac{h}{\Lambda_{GC}} + 4 - \frac{2}{3} \left(\frac{h}{\Lambda_{GC}}\right)^{-1} & \Lambda_{GC} \leq h \leq 2\Lambda_{GC} \\ 0 & h > 2\Lambda_{GC}, \end{cases} \tag{13}$$

while they use the identity function for  $\tau(\rho)$ . In [31], the specification of  $\Lambda_{GC}$  is left to the user. We will use the commonly applied choice  $\Lambda_{GC} = \Lambda_{prior}$ , where  $\Lambda_{prior}$  denotes the correlation length of the prior model.

### 3.2 CO-localization

In [8], the authors express  $\tau(\rho)$  by Eq. 23 in [30],

$$\tau(\rho) = \frac{1}{1 + (1 + \alpha^2/\rho^2)/N_e}, \tag{14}$$

and they use an exponential correlation function,

$$\rho(h) = \alpha \exp(-3h/\Lambda_{exp}). \tag{15}$$

The selection of the correlation length,  $\Lambda_{exp}$ , used to calculate  $L_{mz}^{\tau\rho}$  is based on rules of thumb gained from experience obtained through investigations in [8] using large ensembles and either 1-D examples or 2-D examples with regular well patterns. They argued that a localization region for a particular datum with respect to a variable should cover the region of significant cross-correlation between that datum and variable. For data located in repeated five-spot patterns, they found that the localization region should be centred at the data location with  $\Lambda_{exp}$  approximately equal to the distance between neighbouring wells. For a single five-spot pattern, they found that the location of the localization region depends on the data type. Localization regions for bottom-hole pressure data can be selected in about the same manner as for repeated patterns, while localization regions for fluid-production data should be centred between the data location and the nearest injector with  $\Lambda_{exp}$  large enough to ensure that the region covers both these wells. They also point to the fact that there are several complicating issues that might make use of these rules of thumb challenging: ideally, localization regions should vary with time, irregular well patterns is not discussed, etc. Furthermore, they argue that the localization matrix for  $C_z$  can be expressed as

$$L_z^{\tau\rho} = (L_{mz}^{\tau\rho})^T L_{mz}^{\tau\rho}, \tag{16}$$

followed by a scaling such that the diagonal elements of  $L_z^{\tau\rho}$  become equal to unity.

### 3.3 ER-localization

In [19], the authors use the identity function for  $\tau(\rho)$ , while  $\rho(h)$  is given by the Gaspari-Cohn formula, Eq. 13. They argue that the critical length,  $\Lambda_{GC}$ , should be selected as  $\Lambda_{GC} = \Lambda_{\text{prior}} + \Lambda_{\text{sens}}$ , where  $\Lambda_{\text{sens}}$  denotes the radius of a circle with area equal to the region of non-zero sensitivity for the datum in question. If the prior model is anisotropic with principal correlation lengths  $\Lambda_{\text{prior},1}$  and  $\Lambda_{\text{prior},2}$ , the formula for  $\Lambda_{GC}$  is replaced by  $\Lambda_{GC,1} = \Lambda_{\text{prior},1} + \Lambda_{\text{sens}}$  and  $\Lambda_{GC,2} = \Lambda_{\text{prior},2} + \Lambda_{\text{sens}}$ . (Note that while the authors of [19] use an approximate method equivalent to using streamlines to calculate  $\Lambda_{\text{sens}}$ , we calculate  $\Lambda_{\text{sens}}$  from the sensitivity matrix.)

## 4 Upscaling

In the reservoir engineering workflow, there are (at least) two models describing the subsurface, the geocellular model and the simulation model. The former contains a highly detailed description of the geological subsurface (typically in the order of  $10^7 - 10^8$  grid cells), while the latter is an upscaled, coarser, representation of the same subsurface (typically in the order of  $10^5 - 10^6$  grid cells). In this paper, we propose to upscale the simulation model further to facilitating a faster reservoir simulator. The upscaling is automatic, ensemble-specific and non-uniform. We utilize a simple local-local upscaling method, see, e.g. [16, 25], where the porosity and permeability for each coarse grid cell are found by calculating, respectively, the arithmetic and harmonic mean of the corresponding fine grid cell properties.

To control the approximation error, a good way of determining which fine grid cells that should be upscaled is required. Following the approach introduced in [17, 48, 50], a wavelet transform is applied to perform this selection. Contrary to the previous works, we utilize a second-generation wavelet transform, which makes this approach extremely flexible and computationally efficient.

### 4.1 The discrete wavelet transform

The discrete wavelet transform was developed as a linear method for computing compact representations of functions or data sets by exploiting structures in the data or underlying functions. The goal is to create a transform that will give an informative, efficient and useful description of the function or signal by using localized oscillatory basis functions called wavelets. The specific design of the wavelets

allows the signal to be localized both in space and in frequency. Several petrophysical properties, such as absolute permeability  $K(x)$  and porosity  $\phi(x)$ , are correlated on all the length scales in the geological model, and the wavelet transform is therefore well suited to analyse these quantities.

We can generate a standard wavelet basis for any equation  $f(x) \in L^2$ , by the scaling function,  $\Phi_k(x)$ , and the wavelet function,  $\psi_{j,k}(x)$ . Where  $k$  indicates translation, and  $j$  indicates dilatation. Hence, the same, ‘mother’, wavelet function can be dilated and, therefore, be used on multiple scales. If a suitable scaling and wavelet function is defined, we can expand the function  $f(x)$  in the following form

$$f(x) = \sum_k c(k) \Phi_k(x) + \sum_k \sum_{j=0}^{\infty} d_j(k) \psi_{j,k}(x), \quad (17)$$

where  $c(k)$  and  $d_j(k)$  are coefficients. The *discrete wavelet transform* of the function  $f(x)$  is defined as determining the coefficients in the expansion (17).

The wavelet basis functions are localized in space and frequency, and the standard approach is to let  $j = 0$  be the function with the largest domain. Hence, one obtain a coarser representation of  $f(x)$  by truncating the infinite sum in Eq. 17, and we will later see that this is a useful feature with regards to upscaling. For more info regarding the discrete wavelet transform, see, e.g. [6, 13, 42, 49, 54].

### 4.2 The lifting scheme and second-generation wavelet transform

Traditionally, the basic tool for generating wavelet functions is the Fourier transform. However, there are a number of problems where the Fourier transform cannot be applied, e.g. problems defined on curves, surfaces or volumes. In addition, there are cases that are not translation and dilation invariant; this may be along boundaries or for signals with irregular samples. To deal with these problems, second-generation wavelets were introduced in [55].

Second-generation wavelets are generated via the *lifting scheme*, consisting of three steps: *split*, *predict* and *update*. In the first step, the signal  $(s_{j,k})$  at scale  $j$ , for a finite number of integers  $k$ , is split into two equally large groups. Then, in the second step, the elements in the second group is predicted from the elements in the first group, typically by some polynomial interpolation. The quality of the prediction—the detail coefficients—are stored as the difference between the predictions and the original elements in the second group.

Finally, in the third step, all the elements in the first group are updated such that the mean, or some higher order moment, of the original signal is preserved. The updated group—the smooth coefficients—can now be considered as

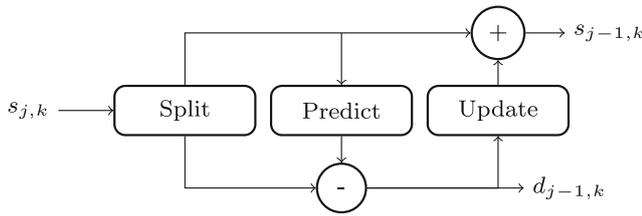


Fig. 1 The lifting scheme

a new and coarser representation of the original signal, and a new pass of the lifting scheme can be performed on the smooth coefficient. The smooth and detail coefficient, after transformation of  $s_{j,k}$  at scale  $j$ , are  $s_{j-1,k^*}$  and  $d_{j-1,k^*}$  respectively, for a new reduced set of integers  $k^*$ .

All the calculations in the lifting scheme can be performed in-place, i.e., the detail coefficients overwrites the values in the second group, and the smooth coefficients overwrites values in first group. Hence, the algorithm will not require more memory than is needed to store the original signal. Moreover, obtaining the inverse transform is trivial, one simply runs the algorithm backwards. The basic steps involved in the lifting scheme are illustrated in Fig. 1, and the basic steps of the inverse transform are illustrated in Fig. 2.

In the following, utilizing the lifting scheme, we will derive the simplest first generation wavelet transform, the Haar wavelet [34] and its second-generation counterpart, the unbalanced Haar transform. For more details regarding the lifting scheme, see, e.g. [56, 57].

4.3 The Haar-wavelet

The Haar wavelet splits the signal ( $s_{j,k}$ ) into its even ( $s_{j,2k}$ ) and odd ( $s_{j,2k-1}$ ) elements, and utilizes a prediction operator which assumes that the signal is a constant. Hence, the odd samples are predicted to have the same value as their even neighbours to the left, and the detail coefficients are

$$d_{j-1,k} = s_{j,2k+1} - s_{j,2k}. \tag{18}$$

It is clear that for constant signals, all detail coefficients are zero. We say that the order of the Haar wavelet predictor is one since it correctly eliminates zeroth order correlation in

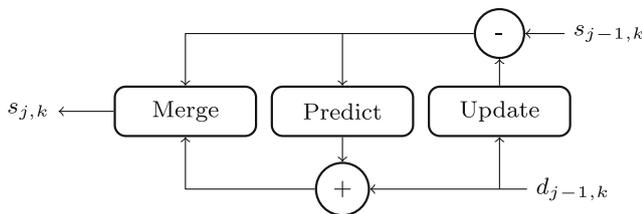


Fig. 2 The inverse transform

the signal. The update is designed to preserve the average of the signal, and the smooth coefficients are

$$s_{j-1,k} = s_{j,2k} + \frac{1}{2}d_{j-1,k} = \frac{s_{j,2k+1} + s_{j,2k}}{2}. \tag{19}$$

Since the Haar wavelet preserves the zeroth order moment, we say that the order of the update operator is one.

4.4 The unbalanced Haar transform

Based on the lifting scheme, it is possible to design a second-generation Haar wavelet transformation, the unbalanced Haar transform, originally introduced in [33]. The major difference between the Haar and the unbalanced Haar transform are the coefficients  $s_{j,k}$ . In the second-generation case, these do no longer represent the signal on the location  $x_{j,k}$ . Instead, the coefficients represent an average on the interval  $[x_{j,k}, x_{j,k+1}]$ , defined as

$$I_{j,k} = \int_{x_{j,k}}^{x_{j,k+1}} w(x) dx, \tag{20}$$

The unbalanced Haar wavelet transform is constructed by the lifting scheme as described above. The split, and predict step are identical to the original Haar wavelet transform but the update step is modified to preserve the weighted average

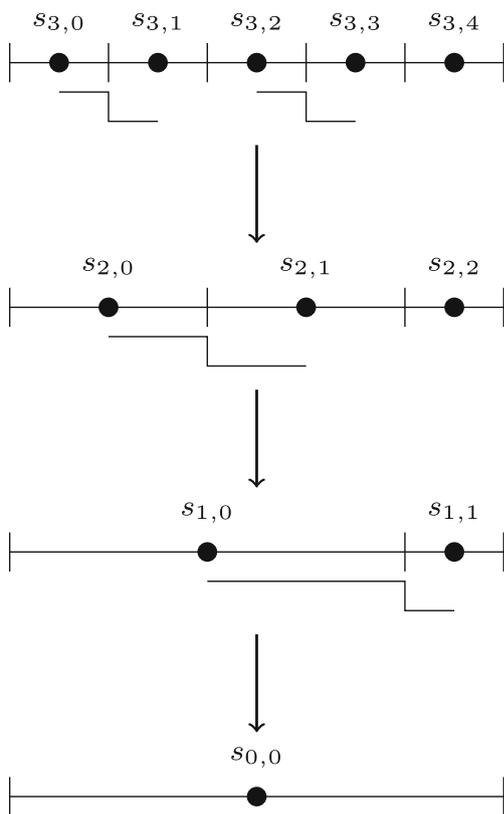
$$s_{j-1,k} = s_{j,2k} + \frac{I_{j,2k+1}}{I_{j-1,k}}d_{j-1,k}. \tag{21}$$

Since the unbalanced Haar transform allows different weight for different elements of the signal, it can be utilized for any realistic computational grid where the cell volumes vary. In addition, the transform can handle inactive or zero-volume cells by setting the weight of these cells equal to zero.

Because of the flexibility introduced by the unbalanced Haar transform, one can deal with non-dyadic grids in a number of ways, see, e.g. [41, ch. 4.4.5]. One approach, which is also utilized in our application, is based on the original idea of splitting the full signal into its even and odd points. For non-dyadic grids, this subdivision will, for some scale  $j$ , result in one more even than odd points, or, in other words, the last even point cannot be utilized to predict its odd neighbour. This is solved by postponing the prediction until we reach a scale where this is possible. This procedure is illustrated in Fig. 3.

4.5 2/3-dimensional wavelet transform

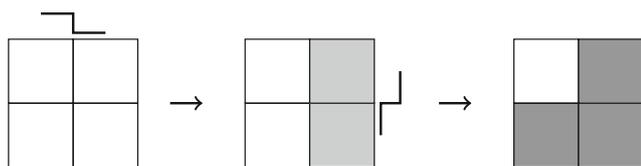
If we want to utilize the wavelet transform for realistic cases, we must be able to extend it to higher dimensions. This is, in principle, not problematic. There is, however, more ways of splitting the signal into two groups when we consider high-dimensional signals. Hence, there exists a wide range



**Fig. 3** Three passes of the unbalanced Haar transform. The vertical lines indicated the walls of the grid cell, and bullet denotes the cell-centres

of different approaches to this wavelet transform, including approaches that splits the signal adaptively [29].

The simplest approach in higher dimensional problems is, however, to decompose the original signal into several 1D signals and perform a 1D wavelet transform on each 1D signal. This repeated use of the 1D transform is not as flexible as wavelet transform with more complex splitting schemes, but it is still widely used in image compression because of its efficiency. One pass of the wavelet transform, performed on a two-dimensional image, is shown in Fig. 4. This approach is utilized in all numerical experiments.



**Fig. 4** 2-D wavelet transform. White tiles illustrate the signal, light grey tiles illustrates the detail coefficients after the transformation is performed for the first dimension, dark grey tiles illustrates the detail coefficients after the transformation is performed for the second dimension

### 4.6 Wavelet-based upscaling

The petrophysical parameter, discretized on the computational grid, can be considered as a piecewise constant function with discontinuities at the cell edges. During the upscaling procedure, suitable fine scale grid cells are merged into larger grid cells. The discretized parameters on the upscaled grid can, however, still be considered as piecewise constant functions that are constant over a larger domain. Recall that the detail coefficients in the Haar-wavelet are zero only for constant functions. Hence, the upscaling procedure is identical to performing an inverse Haar wavelet transform where selected detail coefficients have been set to zero. Because the Haar wavelet preserves the average value, the upscaled parameter automatically receives an averaged—upscaled—value.

The upscaling procedure we utilize takes advantage of the Haar wavelets similarity with a traditional upscaling procedure. However, since we seek an upscaling algorithm that should work on any grid, the method introduced here is based on the unbalanced Haar transform, where we assume that the weight function,  $w(x)$ , is a constant function over all active grid cells, and zero in the inactive grid cells. Hence, the value of the average on the interval  $I_{k,j}$  will only depend on the size of the active grid cells.

Before this procedure can be applied, we must determine some criterion that can be utilized for selecting which detail coefficients to set as zero prior to the inverse transform. This is handled in the manner proposed in [17, 48, 50–52]. This approach is based on critical-path analysis, which states that, for a heterogeneous media, most of the fluid flow takes place in the high permeable regions of the reservoir. To minimize the upscaling error, we want to retain a high resolution in the areas where there is a high degree of fluid flow, i.e. high permeability. To capture edges of areas with high degree of fluid flow, we also keep a high resolution in areas with rapidly changing permeability. We also want to ensure that cells that contains wells, or other predefined structures, are not upscaled. This is done by introducing two threshold values,  $\epsilon_s$  and  $\epsilon_d$ , defined as

$$\epsilon_s = \gamma_s \max_k s_{J,k}, \quad \epsilon_d = \gamma_d \max_k d_{J,k}, \tag{22}$$

where  $\gamma_s/\gamma_d$  are fractions and  $J$  is the resolution of the original signal. In addition, we introduce a matrix of booleans,  $B$ , where elements of  $B$  corresponding to cells we want to keep in the original grid is set to False while all others are set as True. After one step of the wavelet transform, one checks if  $s_{j-1,k} < \epsilon_s$  and  $d_{j-1,k} < \epsilon_d$  are satisfied. If these conditions are met, and if all affected elements in  $B$  are True, we set  $d_{j-1,k} = 0$ . This procedure is performed for all  $k$ , and is repeated on subsequently coarser scales. We stop the procedure when we reach a scale where the requirements are

no longer satisfied for any  $k$ . Hence, a new grid, containing averaged values in regions with low permeability and small variations, is generated by an inverse wavelet transform.

The efficiency of this method depends on the degree of heterogeneity in the permeability field. If there are few heterogeneities, i.e., the permeability distribution has low variance, the critical-path analysis states that the fluid flow takes place through many parts of the field. Hence, we must define the threshold for upscaling such that these parts are retained in high resolution, and thus few grid cells can be upscaled. However, if the permeability field is highly heterogeneous, i.e. the permeability distribution has a large variance, the critical-path analysis states that most of the fluid flow takes place in only a small part of the reservoir. Hence, the threshold values can be set such that only these are retained in high resolution, and large parts of the reservoir may be upscaled.

Note that for 2-D reservoirs, the wavelet transformation produces one  $s_{j-1,k}$  value and three detail coefficients (see Fig. 4). The shape of the final upscaled grid cell depends on which one of these three coefficients we set to zero. Different approaches has been tested numerically, and the best numerical behaviour is achieved when the coarse grid cells are square. To obtain square grid-blocks, we either truncate all, or none, of the three detail coefficient in the upscaling scheme. The only exception from this geometry happens when the grid is non-dyadic, and some part of the signal has been passed over. When such cells are upscaled, there will be a difference in the size of the fine grid cells, and the new merged cell is not necessarily square. An example is illustrated in Fig. 5, where the first pass of the transform produces non-square grid cells. Due to the way we handle non-dyadic grids, this will only happen on the edge of the reservoir.

Figure 6 illustrates how the porosity of layer 38 in the SPE-10 model [12] is upscaled using the wavelet approach (no wells are included in this demonstration).

There is an inherent cost related to the upscaling. Since we want to allow for a large number of ensemble members in the coarse-scale DA approach, it is important that this cost is low. Fortunately, the lifting scheme is a very efficient method for generating the second-generation wavelet transform. The wavelet representation of a signal can be achieved in  $O(N_g)$  operations, and, moreover, the inverse transform can be computed in  $O(N_g)$  operations [56]. Hence, this

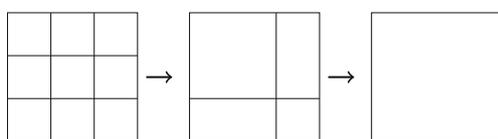


Fig. 5 Upscaling of a non-dyadic 2-D grid

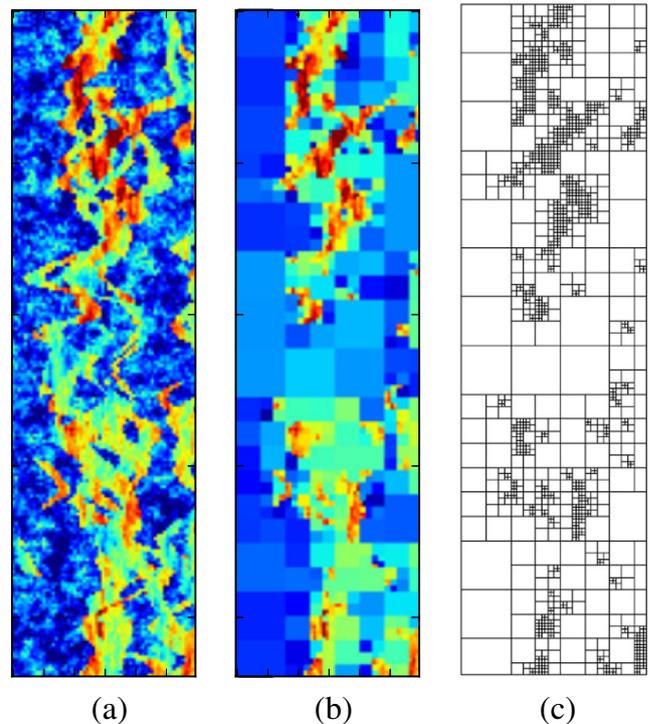


Fig. 6 2-D unbalanced Haar transform of porosity in layer 38 in SPE-10 model. **a** Original field. **b** Upscaled field. **c** Upscaled grid

upscaling method generates very little overhead, and compared to the reservoir model, where we solve a large number of linear equations each requiring  $O(N_g^\eta)$  ( $\eta \in (1.25, 1.5)$ ) operations, this cost is negligible.

### 5 Numerical experiments

The numerical experiments consists of DA problems that cannot be solved to reasonable accuracy when utilizing a small or moderately sized ensemble, and we will study various ways of improving the accuracy for such problems. More specifically, we compare the coarse-scale DA methods against the three different localization schemes discussed in Section 3. For all localization schemes, the localization matrix  $L_z^{\tau\rho}$  is calculated by Eq. 16. Throughout, we assume that the available computational resources will allow us to run 100 simulation runs of the original reservoir simulation model, hence,  $N_e = 100$  is selected for all experiments.

We assume that the absolute permeability is the only unknown petrophysical parameter, and we assume that the permeability is isotropic. Hence, the unknown parameter can be defined as  $\mathbf{m} \in \mathbb{R}^{N_g}$ , i.e. a vector with the same dimension as the number of grid cells.

Each ensemble member is upscaled with regards to its specific permeability field, and it is, therefore, not possible to define a unique value for the reduction of grid cells for

the whole ensemble. Instead, given truncation values  $\epsilon_s$  and  $\epsilon_d$ , we upscale the original ensemble containing  $N_e = 100$  members, and calculate the mean degree of upscaling as

$$\zeta = \frac{1}{N_e} \sum_{i=1}^{N_e} \frac{N_g}{(\widehat{N_g})_i}, \quad (23)$$

where  $(\widehat{N_g})_i$  is the number of grid cells after upscaling of ensemble member  $i$ . This procedure is inexpensive since it does not require a run of the reservoir simulator.

In the reservoir simulator, the flow equations have been discretized following a finite volume approach and the solution of the equations are found by some iterative scheme on the linearized equations, see, e.g. [5]. Since each non-linear iteration requires a solution of a linear system of equations, it is clear that the numerical cost of solving the flow equations in the reservoir simulator,  $F$ , is proportional to the cost of solving the linear system of equations. The cost of solving the linear system is a function of the number of grid cells. We, therefore, assume that  $F$  scales as  $N_g^\eta$ , where the value  $\eta \in (1.25, 1.5)$  depends on the choice of linear solver, see, e.g. [4]. Since the total overall computational cost of the DA method,  $Q$ , roughly equals  $N_e \times F$ , we can keep the total numerical cost constant by defining the number of ensemble members for the upscaled DA method,  $\widehat{N}_e$ , as

$$\widehat{N}_e = N_e \times \zeta^\eta. \quad (24)$$

It is clear that the number of upscaled ensemble members depends on the quality of the linear solver. To evaluate this, we will run experiments utilizing both  $\eta = 1.25$  and  $\eta = 1.5$ . To make the DA problem extra challenging for coarse-scale DA, we also test the unrealistic case where  $\eta = 1$ .

### 5.1 Example 1

In the first numerical experiment, we consider a  $60 \times 60$  2D reservoir model defined on a uniform Cartesian grid. This experiment is based on example 3 in [19]. The unknown model parameters are the log-permeability value in each grid cell. The prior ensemble is generated from an anisotropic spherical covariance function with major correlation length equal to 30 grid cells, minor correlation length equal to 10 grid cells and the direction of major correlation rotated  $45^\circ$  clockwise around the ordinate axis. The prior mean is set to 5.0 and the prior variance to 1.0. The true model, used to generate the true data, is selected as one realization drawn from the prior distribution and is shown in Fig. 7. For all models, the porosity is set to 0.2 throughout the field. The model contains six producing wells, P1–P6, and one injector, I1.

At the start of the simulation time, producers P1, P2 and P3 starts producing oil, while the injector, I1, starts injecting water. After 6000 days, the well P4 starts production,

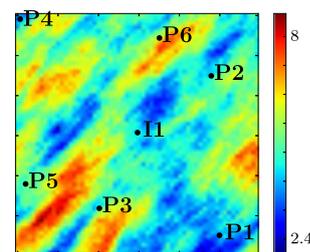


Fig. 7 True model

and after 9000 days the two final wells, P5 and P6, start production. The total production time is 12,000 days. All wells are controlled by bottom-hole pressure constraints,  $1.03 \times 10^4$  kPa for the producers and  $2.75 \times 10^4$  kPa for the injector. The observed data consists of oil and water production rates from the producing wells along with water injection rates from the injecting well. These quantities are observed each 150th day, resulting in 80 assimilation steps.

An ensemble of synthetic data was generated by adding uncorrelated Gaussian noise to the true data with a variance equal to 1% of the true value. To avoid unrealistically accurate data points, we omitted all data with value below 10, resulting in 594 data points.

For the upscaling, we set  $\epsilon_s$  and  $\epsilon_d$ , as defined in Eq. 22, by setting  $\gamma_s = 0.7$  and  $\gamma_d = 0.7$ . This results in  $\zeta = 3$ . An illustration of the upscaling for one realization of the prior model is given in Fig. 8, and oil production rate from well P2 for the original and the upscaled model is shown in Fig. 9.

For the three localization procedures, we must define the different localization ranges. For the GC-method and the ER-method, this does not introduce any extra work since the range is the prior range in the GC-method and prior plus drainage area in the ER-method. However, for the CO-approach, we must manually select the position and range of the localization, following the approach outlined in Section 3. The area of significant cross-correlation should be included in the localization region [8]. Therefore, to compare the localization functions, we have plotted the absolute value of the cross-correlation between the log-permeability

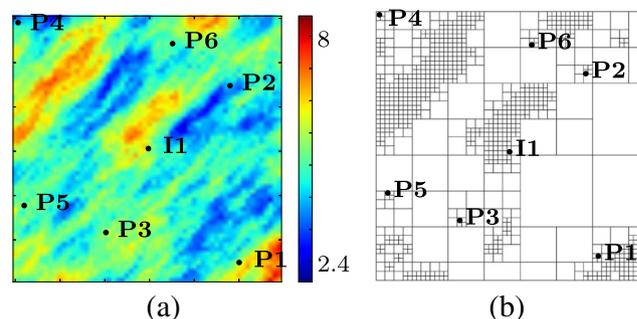
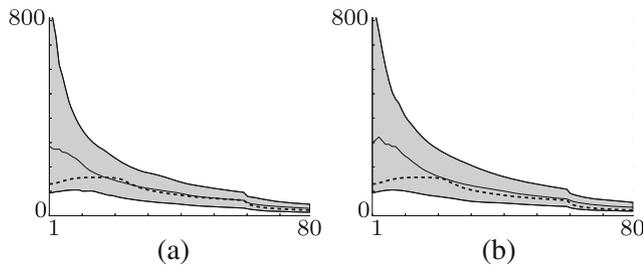


Fig. 8 Example 1, upscaling of one realization. a Original model. b Upscaled grid

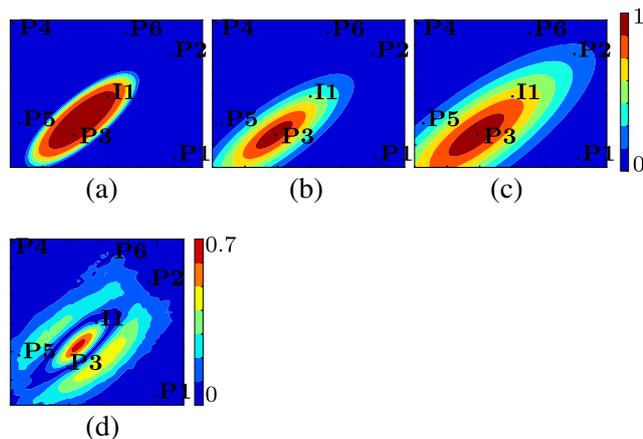


**Fig. 9** Oil production rate ( $\text{m}^3/\text{day}$ ) from well P2 for the different assimilation steps. The shaded area indicates 10 – 90% percentile, the solid line represents the median and the thick dashed line represents the true data. **a** Original. **b** Upscaled

and the selected datum in Fig. 10. The cross-correlation was calculated using a large ensemble containing  $1 \times 10^4$  ensemble members.

### 5.2 Example 2

In the second numerical experiment, we consider a  $167 \times 167$  2D reservoir model defined on a uniform Cartesian grid. This experiment is based on example 2 in [8]. The unknown model parameters are the log-permeability value in each grid cell. The prior ensemble is generated from an isotropic exponential covariance function with correlation length equal to 10 grid cells. The prior mean is set to 2.5 and the prior variance 1.44. The porosity is set to 0.2 throughout the field. For this case, the true model, used to generate the true data, is the same as utilized in [8], generously provided by [7]. The model contains 25 inverted five-spot patterns, giving a total of 36 producers and 25 injectors. Figure 11a shows the model utilized to generate the true data, with the well positions indicated.



**Fig. 10** Localization functions for water production rate at well P3 calculated, **a** CO, **b** GC, and **c** ER. **d** Absolute value of cross-correlation between water production rate, at well P3, and parameter at assimilation time 5, calculate using  $N_e = 1 \times 10^4$

All the producers are controlled by a fixed bottom-hole pressure constraint of  $6.89 \times 10^3 \text{ kPa}$ , while the injectors are controlled by fixed water injection rate constraint of  $238.5 \text{ m}^3/\text{day}$  and, in addition, a maximum bottom-hole pressure constraint of  $5.516 \times 10^4 \text{ kPa}$ . Data are available at nine different times, after 30, 120, 240, 360, 540, 780, 900, 1080 and 1440 days. At each assimilation time, the available data consists of oil production rates and water production rates from all producers, and bottom-hole pressure from the injectors. In addition, the water injection rate from the injectors are included since some injectors, in the true run, switches controls to fulfil the constraints. The total number of available data at each assimilation time is 122, hence, there is a total of 1098 data points.

An ensemble of synthetic data was generated by adding uncorrelated Gaussian noise to the true data with a variance equal to 1% of the true value for data with values above 10, and fixed variance equal to 0.1 for data with values below 10.

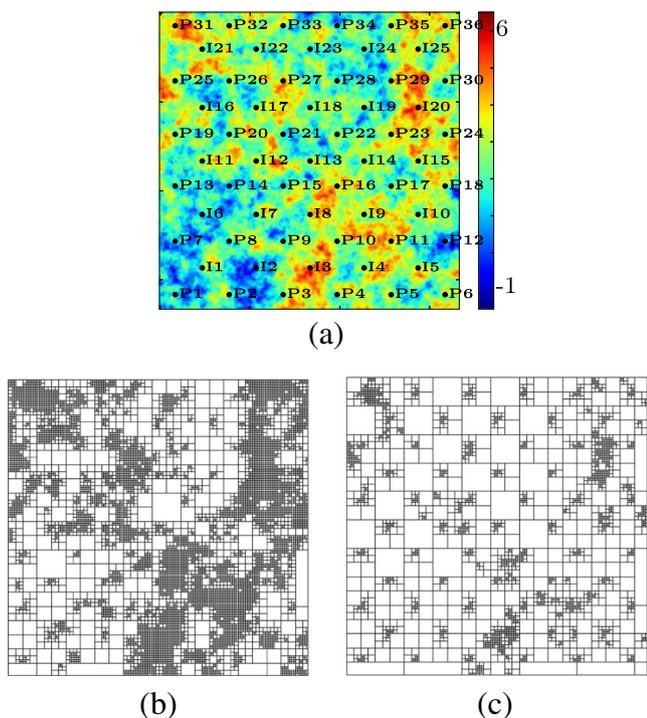
We note that this model is relatively densely populated with wells, and in this example, the flow-paths are determined, to a higher degree, by the well position rather than the permeability value. This makes the model difficult to upscale with our proposed technique, and we therefore test two different degrees of upscaling. The values of  $\epsilon_s$  and  $\epsilon_d$  (see Eq. 22), are determined by  $\gamma_{s/d} = 0.7$ , and  $\gamma_{s/d} = 0.5$ . These degrees of upscaling results in  $\zeta = 15$ , and  $\zeta = 7$ . Note that no grid cells containing wells can be part of an upscaled grid cell. Hence, relatively many grid cells cannot be upscaled, independent of their permeability value. The results from upscaling the true model, utilizing the different degrees of upscaling, are illustrated in Fig. 11. Oil production rate from well P15 for the original and the upscaled models are shown in Fig. 12.

For this experiment, the localization functions for CO, as utilized in [8], was provided by [7]. To illustrate the difference between the localization schemes, we have illustrated the localization functions, in addition to the cross-correlation (calculated using an ensemble size of  $1 \times 10^4$ ), for a selected datum in Fig. 13. Note that, for this case, all taper functions are centred at the well positions, but the localization range of the various methods varies greatly.

### 5.3 Example 3

In the third example, we consider a  $400 \times 400$  2D reservoir model defined on a uniform Cartesian grid, where the unknown model parameters are the log-permeability value in each grid cell, while the porosity is constant and equal 0.35.

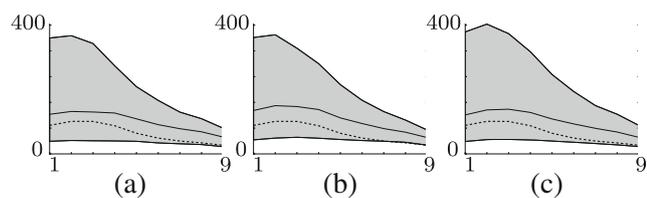
Contrary to examples 1 and 2, where the unknown parameter was described as a Gaussian field, we now consider an unknown parameter described as a categorical field



**Fig. 11** Upscaling of one realization. **a** True model. **b** Upscaled grid  $\zeta = 7$ . **c** Upscaled grid  $\zeta = 15$

with two distinctive regions and small Gaussian variation within each region. Both the true model and the prior ensemble are generated by a combination of the sequential indicator simulation (SIS) algorithm and the sequential Gaussian simulation (SGS) procedure, both implemented in the GSLIB software [14]. In the SIS, we simulate two threshold values of 0.1 and 1, with target proportions of 0.8 and 0.2, respectively. Both categories utilized the same isotropic spherical variogram with a variance of 1 and range of 500 grid cells. The Gaussian variations were simulated by the SGS with a mean value of 6 and a variance of 1 sampled from an isotropic spherical variogram model with a range of 10 grid cells. The final realizations were generated by multiplying the two fields.

The reservoir contains nine wells, five producers and four injectors, placed along the two diagonals going from



**Fig. 12** Oil production rate from well P15 ( $\text{m}^3/\text{day}$ ) for the different assimilation steps. The *shaded area* indicates 10 – 90% percentile, the *solid line* represents the median and the *thick dashed line* represents the true data. **a** Original model. **b** Upscaled prediction  $\zeta = 7$ . **c** Upscaled prediction  $\zeta = 15$

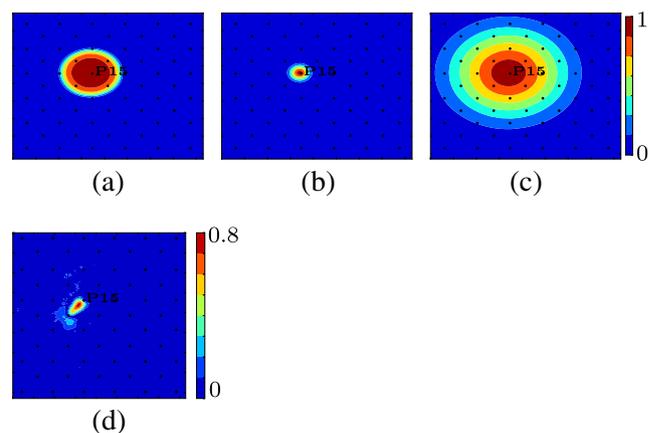
north-west to south-east and from south-west to north-east, and the flow pattern depends heavily on the spatial extent of the region with high permeability. To make the problem extra challenging, we condition the true field to have a high-permeable path, connecting the wells, going along both diagonals, while the prior model is only conditioned to be in the high permeable region at well positions. The true model, with well-positions indicated, is illustrated in Fig. 14, and an arbitrary realization of the prior model is illustrated in Fig. 15a.

All wells are controlled by fixed bottom-hole pressure constraints, the injectors at  $3 \times 10^4 \text{kPa}$  and the producers at  $1 \times 10^4 \text{kPa}$ . Oil and water production rate from the producers and water injection rate from the injectors are collected every 6 months for 8 years, giving a total of 224 data points.

An ensemble of synthetic data was generated by adding uncorrelated Gaussian noise to the true data with a variance equal to 1% of the true value for data with value above 10, and fixed variance equal to 0.1 for data with value below 10.

To perform the upscaling, we set  $\epsilon_s$  and  $\epsilon_d$  (see Eq. 22), by  $\gamma_s = 0.7$  and  $\gamma_d = 0.7$ . This degree of upscaling results in  $\zeta = 20$ . The upscaled grid, for the log-permeability realization shown in Fig. 15a, is illustrated in Fig. 15b, while the original and upscaled oil production curve for well P2 is shown in Fig. 16.

This example is so large that a simulation of the full model takes approximately 450 s. Hence, obtaining a solution utilizing  $1 \times 10^4$  ensemble members would take approximately 52 days, and we, therefore, cannot afford to run the case with a very large ensemble. The range and position of the CO-localization function must therefore be selected based on the well position. We select a localization range of 100 grid cells, and let the CO-localization regions, for the producers, be centred between the data location and the



**Fig. 13** **a–c** Localization functions for water production rate at well P15 calculated, from *left to right*, by CO, GC and ER. **d** Absolute value of cross-correlation between water production rate, at well P3, and parameter at assimilation time 5, calculate using  $N_e = 1 \times 10^4$

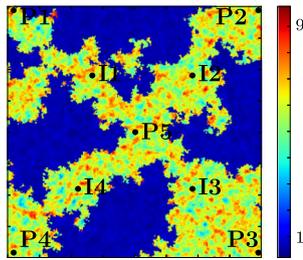


Fig. 14 Example 3, true model

nearest injector. For the injectors, we let the CO-localization regions be centred at the well locations. The GC-, and ER-localization regions are defined in their usual manner, but we let the range representing the prior model be 20 grid cells since we are not considering a Gaussian prior model. Figure 17 shows the various localization functions for the oil production rate at well P2.

### 6 Analysis of numerical results

In the following, we will present and analyse the results of the numerical investigation. The analysis will compare the proposed coarse-scale DA method to the methods that are based on localization. For each of the three examples, we will compare the methods based on their estimates of the posterior mean and standard deviation, in addition to the data match for some representative data points. Examples 1 and 2 are sufficiently small to allow DA utilizing a high number of ensemble members, and the result from this test is considered as the *reference* solution and will be presented for these examples. Note that since the forward model is non-linear, the reference solution does not represent samples from the Bayesian posterior distribution. However, since the EnKF also converges for non-linear forward models [43], the best result we are able to achieve in this framework is the EnKF in the limit of infinite ensemble size. Hence, it is reasonable to compare the coarse-scale DA and the localization approach to the reference solution.

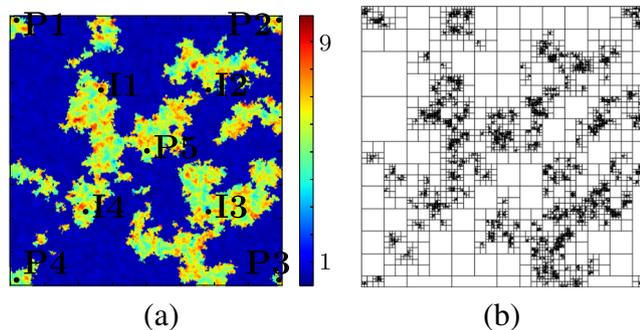


Fig. 15 Example 3. a Original model. b Upscaled grid

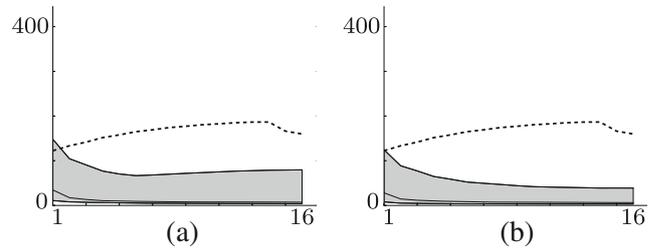


Fig. 16 Oil production rate ( $\text{m}^3/\text{day}$ ) from well P2 for the different assimilation steps. The shaded area indicates 10 – 90% percentile, the solid line represents the median and the thick dashed line represents the true data. a Original model. b Upscaled model

### 6.1 Results example 1

In addition to the high number of data (potentially 720 data points), which might cause the ensemble to collapse, this experiment is also sensitive to the effects of spurious correlations. Since the wells P4-P6 has a later production startup than the other wells, the grid cells surrounding P4-P6 can be affected by spurious correlation from earlier assimilation steps. This can make it especially difficult to assimilate the data from these wells [19].

In Fig. 18, we have plotted the posterior mean obtained when utilizing the various localization methods and upscaling. In addition, we have plotted the posterior mean obtained without localization and the reference posterior mean. The figure shows that all four methods obtain reasonably similar estimates of the mean that are all similar to the reference solution. The coarse-scale DA method, especially with  $\eta = 1.5$ , performs equally good as the distance based localization methods. From the result without localization, we clearly observe the necessity of implementing some method to handle the small ensemble size. The mean model without localization contains much roughness and does not resemble the reference solution.

In Fig. 19, the posterior standard deviation of the log-permeability, obtained by the four methods, are shown. The plot of the reference posterior standard deviation, and the plot of the posterior standard deviation obtained without localization is also shown in Fig. 19. From the figure, we observe that the three methods based on localization performs relatively similarly, and captures the reference solution quite well. Note that the same colorbar is utilized

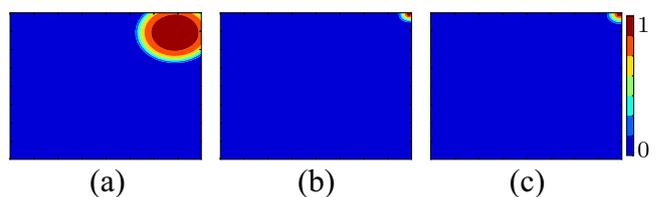
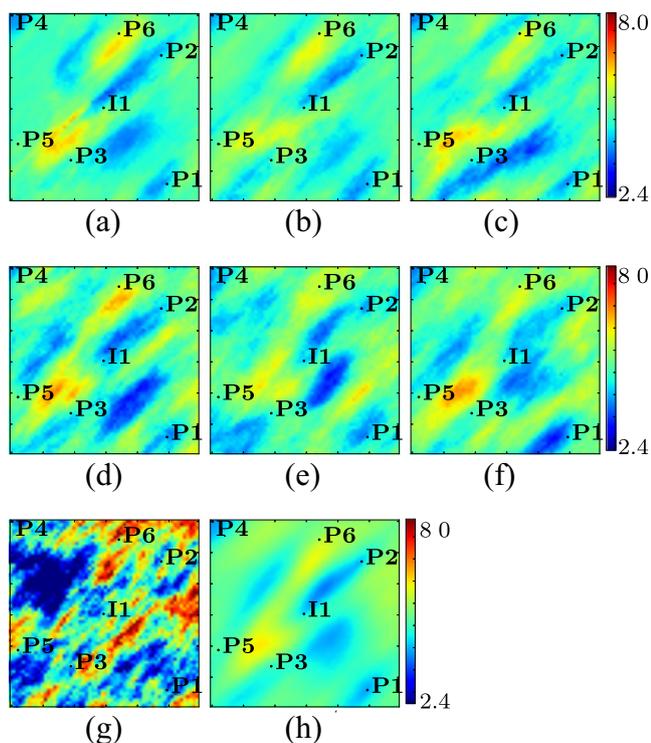


Fig. 17 a–c Localization functions for oil production rate at well P2. a CO. b GC. c ER



**Fig. 18** Mean log-permeability. **a** CO. **b** GC. **c** ER. **d** coarse-scale DA  $\eta = 1$ . **e** coarse-scale DA  $\eta = 1.25$ . **f** coarse-scale DA  $\eta = 1.5$ . **g** no localization. **h** high  $N_e$

for all plots in Fig. 19. For the cases with localization, some areas have a higher standard deviation than the reference solution, and there are areas where the values overshoot the range of the colorbar. The maximum standard deviation for the different localization methods are 1.14, 1.17 and 1.04, for the CO-method, the GC-method and the ER-method, respectively. The coarse-scale DA method, with  $\eta = 1$ , underestimates the standard deviation for this case, but, for  $\eta = 1.5$ , the standard deviation is, as expected, more similar to the reference solution. From the figure, it is also clear that if nothing is done with regards to the small ensemble size, the ensemble collapses and the standard deviation is significantly underestimated.

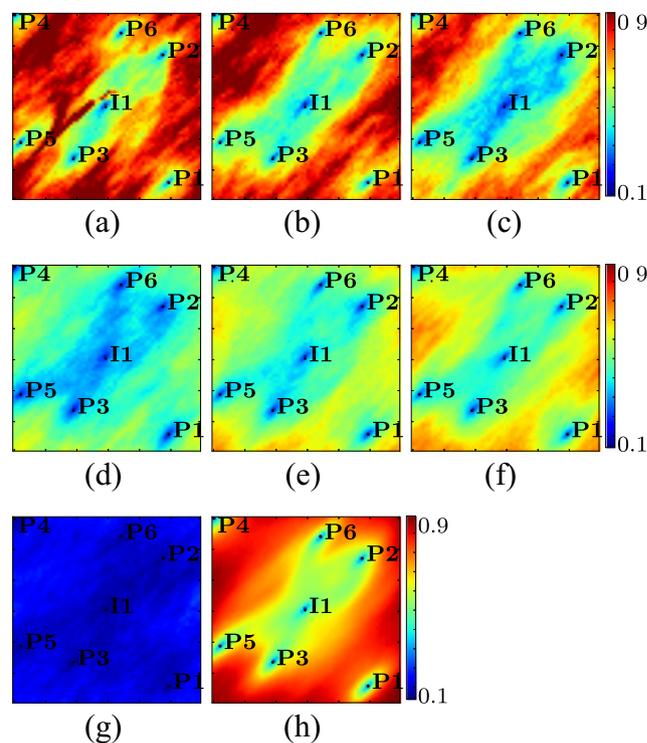
Figures 20, and 21 shows the water and oil production rates from well P2 and P4, respectively. These figures are obtained by rerunning the ensemble of models, obtained from the various methods, after all data have been assimilated. Again, we observe similar results from the methods that apply localization. For the water production rate, the median model for the coarse-scale DA method is closer to the true data, and, contrary to the localization methods, all data are captured within the shaded area. Hence, the coarse-scale DA method performs better than the localization methods for this data type. For the oil production rate, the coarse-scale DA slightly underestimates the spread in the prediction while all the localization methods slightly

overestimates the spread. However, even though that data from well P4 was expected to be affected by spurious correlations, all the coarse-scale DA cases, the GC method, and the ER method matches the data, while the CO method fails to capture all data within the shaded area. For the case where nothing is done to mitigate the effect of the ensemble size, we clearly observe that the ensemble collapses and is unable to match the data.

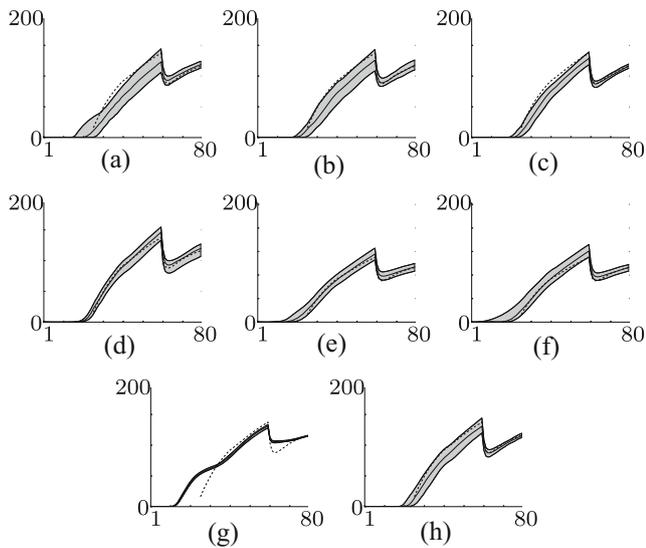
In summary, when combining the results from the mean, standard deviation, and the data match, it is clear that for this example the coarse-scale DA method with  $\eta = 1.5$  produces equally good results as the localization method with the best result, and better results than the localization method with the worst result. The coarse-scale DA cases with  $\eta = 1$ , and  $\eta = 1.25$  has better data match, equal estimates of the mean, but worse estimates of the standard deviation when compared with the localization methods.

### 6.2 Results example 2

Similar to example 1, this example was also designed to contain a high number of data (a total of 1098 data points). However, for this case, the data comes from a dense pattern of wells. Hence, the data from this experiment resembles seismic data with respect to spatial data density.



**Fig. 19** Standard deviation log-permeability. **a** CO. **b** GC. **c** ER. **d** coarse-scale DA  $\eta = 1$ . **e** coarse-scale DA  $\eta = 1.25$ . **f** coarse-scale DA  $\eta = 1.5$ . **g** No localization. **h** High  $N_e$



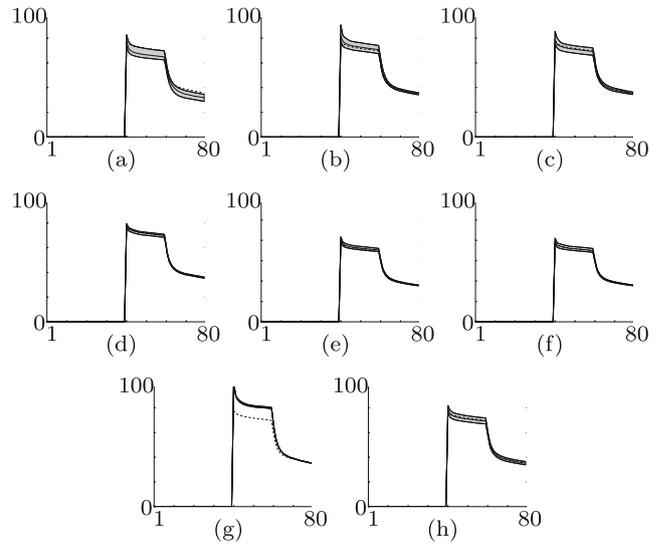
**Fig. 20** Water production rate ( $\text{m}^3/\text{day}$ ) from well P2 for the different assimilation steps. The *shaded area* indicates 10–90% percentile, the *solid line* represents the median and the *thick dashed line* represents the true data. **a** CO. **b** GC. **c** ER. **d** coarse-scale DA  $\eta = 1$ . **e** coarse-scale DA  $\eta = 1.25$ . **f** coarse-scale DA  $\eta = 1.5$ . **g** No localization. **h** High  $N_e$

Figure 22 shows the mean posterior log-permeability and Fig. 23 shows the standard deviation of the posterior log-permeability, obtained by the three localization methods and the various coarse-scale DA methods. In addition, Figs. 22 and 23 shows the mean and standard deviation of the posterior log-permeability without localization, and the reference mean and standard deviation of the posterior log-permeability obtained from a run with a very large ensemble.

Clearly, if nothing is done to handle the small ensemble size, the ensemble will lose all its variability and collapse into a mean model which has a high degree of roughness.

By comparing the results from the localization methods with the reference solution, we see the effect of the different localization ranges (illustrated in Fig. 13). The GC-method has the shortest range, and the mean model does not show significant updates between wells, especially in the northeast corner of the reservoir, and the standard deviation is still high between the wells. The CO-model has medium range and its mean and standard deviation is closest to the reference solution, while the mean for the ER-model, which has the longest range, is slightly rougher than the reference mean and the standard deviation for the ER-model is significantly lower than the reference standard deviation.

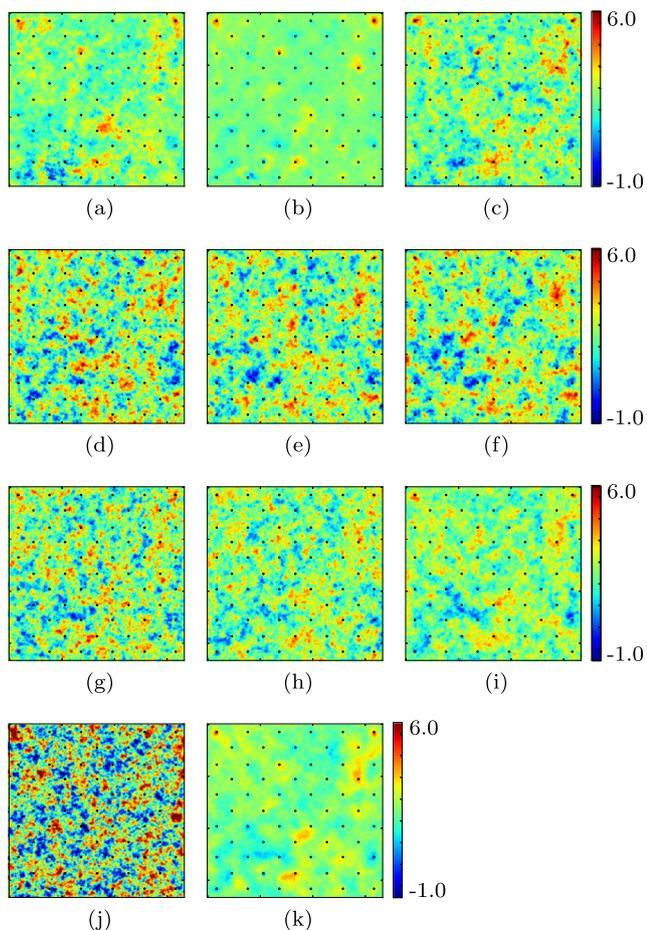
The results from the coarse-scale DA method shows a clear dependency on the degree of upscaling. For these methods, the best standard deviation is achieved by  $\zeta = 15$   $\eta = 1.5$  (Fig. 23f), which is the run with the highest number of ensemble members. The other coarse-scale DA cases



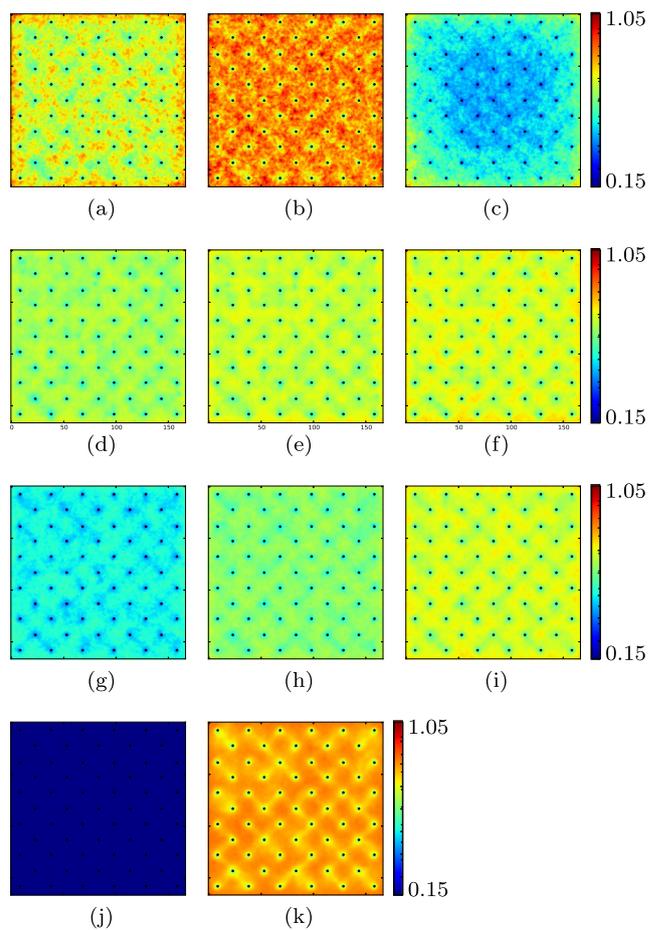
**Fig. 21** Oil production rate ( $\text{m}^3/\text{day}$ ) from well P4 for the different assimilation steps. The *shaded area* indicates 10–90% percentile, the *solid line* represents the median and the *thick dashed line* represents the true data. **a** CO. **b** GC. **c** ER. **d** coarse-scale DA  $\eta = 1$ . **e** coarse-scale DA  $\eta = 1.25$ . **f** coarse-scale DA  $\eta = 1.5$ . **g** no localization. **h** High  $N_e$

with  $\zeta = 15$  underestimates the standard deviation. However, the mean model obtained by  $\zeta = 15$   $\eta = 1.5$  is significantly rougher than the mean reference model, and there is little difference between the various cases with  $\zeta = 15$ . This effect is not seen in the cases with  $\zeta = 7$ . Here, for  $\eta = 1.5$ , we have a slightly lower standard deviation, and the standard deviation decreases with  $\eta$ , but we obtain a mean model which is smooth for  $\eta = 1.5$  and rougher as  $\eta$  decreases.

This effect can be explained by Fig. 11, which illustrates the upscaled grid for both degrees of upscaling. As mentioned above, due to pressure gradients between wells, the potential flow-paths in this experiment is determined more by the well position than the parameter field. For  $\zeta = 15$  only a small area, apart from grid cells containing wells, is kept in high resolution. Hence, areas between many wells that might contain important flow-paths have been upscaled. For  $\zeta = 7$ , a larger area, which naturally covers more flow-paths, is kept in high resolution. Due to the upscaling, all parameters in an upscaled grid cell becomes strongly correlated in the analysis step of the coarse-scale DA method. Since parameters that contribute to the flow get the largest updates and because these parameters are retained in high resolution, we usually do not observe correlations caused by the upscaling in the analysed models. However, when other factors than permeability are important for the flow-paths, we observe these correlations as roughness in the approximation of the posterior mean. By using a lower degree of upscaling, we can account for the fact that the well positions



**Fig. 22** Mean log-permeability. **a** CO. **b** GC. **c** ER. **d**  $\zeta = 15, \eta = 1$ . **e**  $\zeta = 15, \eta = 1.25$ . **f**  $\zeta = 15, \eta = 1.5$ . **g**  $\zeta = 7, \eta = 1$ . **h**  $\zeta = 7, \eta = 1.25$ . **i**  $\zeta = 7, \eta = 1.5$ . **j** No localization. **k** High  $N_e$



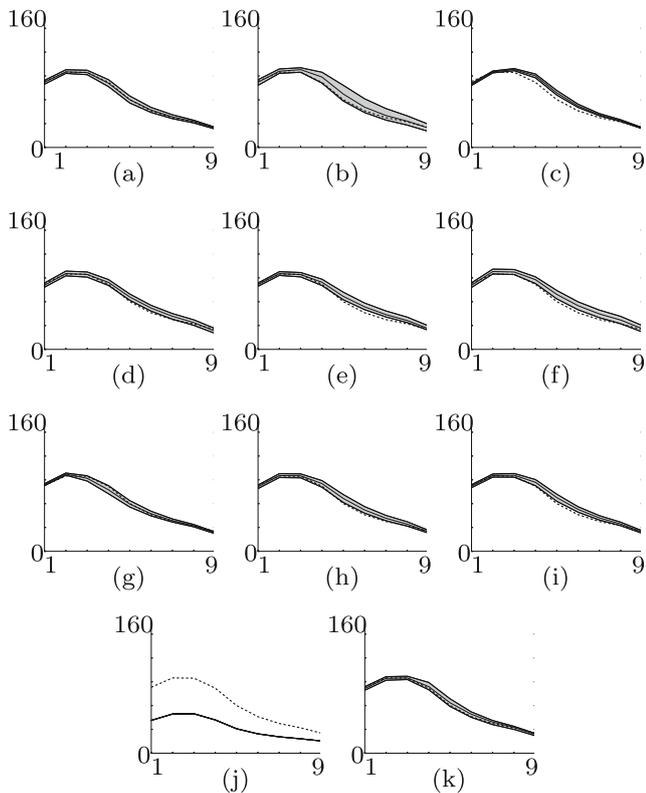
**Fig. 23** Standard deviation log-permeability from coarse-scale DA. **a** CO. **b** GC. **c** ER. **d**  $\zeta = 15, \eta = 1$ . **e**  $\zeta = 15, \eta = 1.25$ . **f**  $\zeta = 15, \eta = 1.5$ . **g**  $\zeta = 7, \eta = 1$ . **h**  $\zeta = 7, \eta = 1.25$ . **i**  $\zeta = 7, \eta = 1.5$ . **j** No localization. **k** High  $N_e$

are important for determining the flow-paths, and remove the unwanted effect of the correlations. Hence, since this example only permits a moderate degree of upscaling, it is a challenging problem for coarse-scale DA.

Figures 24 and 25 show the curves of the oil and water production rates for well P15 and P30, respectively, obtained by rerunning the ensembles after all data have been assimilated. This experiment contains a high number of wells, and we can only show two results here. However, the majority of the data matches are similar to the result shown in Fig. 24, which can be defined as a typical result. The results that are not typical will for the most part be similar to the data match shown in Fig. 25. For the oil production rate, all the methods perform similar. However, similar to the plot of standard deviation, we observe that the ER-method underestimates the spread while the GC-method overestimates the spread. In addition, we observe that some of the data-points for the ER-method are outside the 10–90% percentile region. For the water production rates, we observe

that neither the GC-method nor the coarse-scale DA method match the water breakthrough. However, the results for the two methods show no signs of collapse, such as in the case with no localization. It is reasonable to assume that for the GC-method, we cannot match the data due to the small localization radius. For the coarse-scale DA method, we observe that the match gets better as  $\zeta$  decreases. This shows the dependence on upscaling for this specific example and a better match could perhaps be achieved by further decreasing the degree of upscaling. However, this was not studied further. Again, we clearly observe the consequence of not doing anything to mitigate the small ensemble, the ensemble is fully collapsed away from the data.

In summary, the CO-method obtains the best result when comparing the mean, standard deviation and data match. However, the best coarse-scale DA method ( $\zeta = 7$  and  $\eta = 1.5$ ) obtains equally good, or slightly better, estimates of the mean and standard deviation compared to the ER and GC methods, and equally good data match as the GC-method.



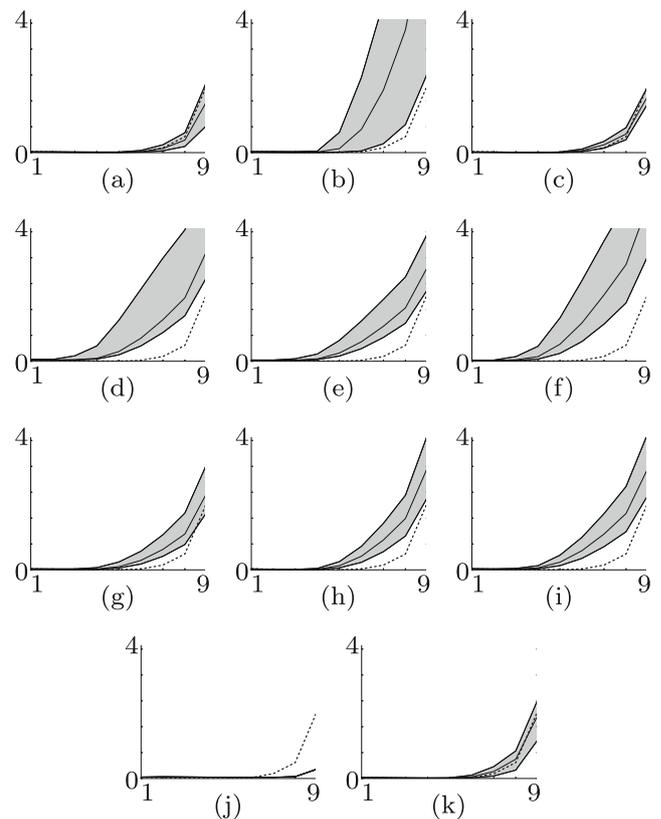
**Fig. 24** Oil production rate from well P15 in  $\text{m}^3/\text{day}$  for the different assimilation steps. The shaded area indicates 10–90% percentile, the solid line represents the median and the thick dashed line represents the true data. **a** CO. **b** GC. **c** ER. **d**  $\zeta = 15 \eta = 1$ . **e**  $\zeta = 15 \eta = 1.25$ . **f**  $\zeta = 15 \eta = 1.5$ . **g**  $\zeta = 7 \eta = 1$ . **h**  $\zeta = 7 \eta = 1.25$ . **i**  $\zeta = 7 \eta = 1.5$ . **j** No localization. **k** High  $N_e$

Hence, in summary, the coarse-scale DA method performs equally good, or better than the localization method with the worst result.

### 6.3 Results example 3

Compared to the two previous examples, example 3 is significantly larger and more complex. As mentioned above, this inhibits us from calculating a reference solution to the DA problem. In addition, due to the complexity of the problem, it is not possible to perform DA without localization. After the first analysis step, the updated ensemble is so rough that the flow equations will not converge, and it is therefore impossible to do any further predictions. Hence, DA on this model cannot be performed without some method to handle the low ensemble size.

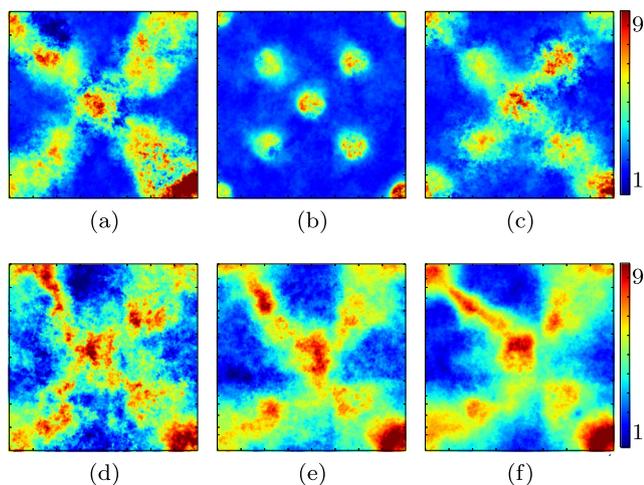
In Fig. 26, we have plotted the various approximations to the posterior mean. Since we do not have the reference posterior mean, we must compare the result to the true model (Fig. 14). It is clear that the localization range is essential for the result. The GC-method has the shortest range, and we clearly see that this method is unable to capture the channels



**Fig. 25** Water production rate from well P30 in  $\text{m}^3/\text{day}$  for the different assimilation steps. The shaded area indicates 10–90% percentile, the solid line represents the median and the thick dashed line represents the true data. **a** CO. **b** GC. **c** ER. **d**  $\zeta = 15 \eta = 1$ . **e**  $\zeta = 15 \eta = 1.25$ . **f**  $\zeta = 15 \eta = 1.5$ . **g**  $\zeta = 7 \eta = 1$ . **h**  $\zeta = 7 \eta = 1.25$ . **i**  $\zeta = 7 \eta = 1.5$ . **j** No localization. **k** High  $N_e$

connecting the wells. The ER-method, which has a larger range, also struggles to properly connect the channels. The CO-method, and all the coarse-scale DA methods are able to connect the channels better. However, we observe that the posterior mean from the CO-method, which has large localization range, is quite rough. The coarse-scale DA method with  $\eta = 1$  is also quite rough, and we observe that areas away from the channels are updated. As  $\eta$  increases, the result from the coarse-scale DA methods gets smoother, and there is less updates outside of the high permeable region. Hence, the coarse-scale DA methods does the best job of capturing the posterior mean, when compared to the true model.

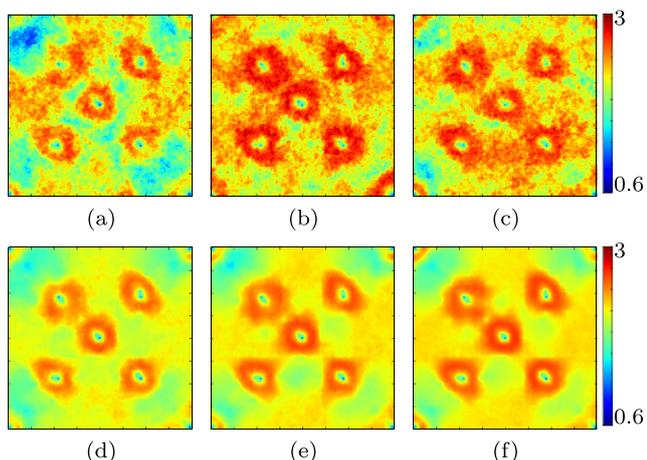
The same conclusions can be drawn from Fig. 27, showing the standard deviation. Here, since we do not have a reference solution, we can only compare the methods to each other. However, we clearly observe that the highest standard deviation is obtained by the GC-method, while the lowest is obtained by the CO-method. Hence, we can assume that the localization range of the GC-method is too short to reduce the standard deviation, while the range in the



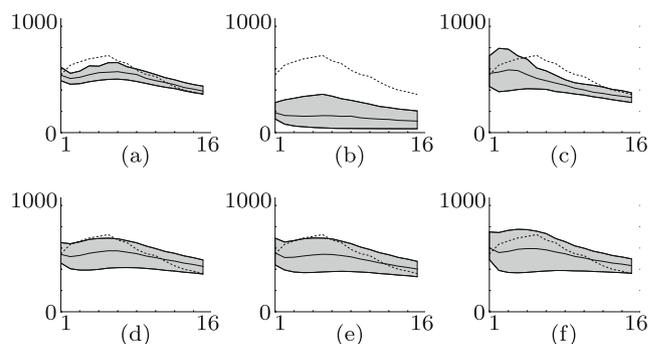
**Fig. 26** Mean log-permeability. **a** CO. **b** GC. **c** ER. **d** coarse-scale DA,  $\eta = 1$ . **e** coarse-scale DA,  $\eta = 1.25$ . **f** coarse-scale DA,  $\eta = 1.5$

CO-method is so long that the ensemble cannot retain sufficient spread. For the coarse-scale DA methods, the standard deviation gradually increase as  $\eta$  increases.

Figure 28 shows the posterior ensemble prediction of the oil production rate from well P5. We observe that none of the localization methods are able to match all the data. The CO-method has the best match of the methods applying localization, but only parts of the data are within the shaded area. The matches obtained by the coarse-scale DA methods are better than for all the localization methods. For  $\eta = 1$  and  $\eta = 1.25$ , almost all the data are within the shaded area, while for the case with  $\eta = 1.5$ , all data are within the shaded area. Hence, for this example, the coarse-scale DA methods obtains a better data match than any of the localization methods.



**Fig. 27** Standard deviation log-permeability. **a** CO. **b** GC. **c** ER. **d** coarse-scale DA,  $\eta = 1$ . **e** coarse-scale DA,  $\eta = 1.25$ . **f** coarse-scale DA,  $\eta = 1.5$



**Fig. 28** Oil production rate ( $\text{m}^3/\text{day}$ ) from well P5 for the different assimilation steps. The shaded area indicates 10 – 90% percentile, the solid line represents the median and the thick dashed line represents the true data. **a** CO. **b** GC. **c** ER. **d** coarse-scale DA,  $\eta = 1$ . **e** coarse-scale DA,  $\eta = 1.25$ . **f** coarse-scale DA,  $\eta = 1.5$

When comparing the mean, standard deviation and data match, it is clear that for this example the coarse-scale DA methods with  $\eta = 1.25$  and  $\eta = 1.5$  performs better than all of the localization methods. For the case with  $\eta = 1$ , the result is in summary equally good or slightly better than the best localization result.

### 7 Summary and conclusions

We have proposed and evaluated a new method for reservoir history matching problems, coarse-scale DA, that aims to remove, or at least significantly reduce, the need for localization. Coarse-scale DA utilize an automatic, ensemble-specific and non-uniform upscaling method to reduce the computational cost of running each ensemble member, allowing for a significant increase in the number of ensemble members at the same computational cost as standard DA with or without localization.

Three numerical experiments, with high degree of variability, show that the coarse-scale DA method is more robust with respect to variation in example types than each of the localization techniques considered with standard DA.

Experiment 1 was a relatively small history matching problem with a smoothly varying permeability field, and a high number of assimilation steps. We observed that, even with a moderate degree of upscaling, the coarse-scale DA method performed equally good or slightly better than the localization methods.

Experiment 2 also had a smoothly varying permeability field, which was densely populated with wells. This lead to a high number of possible flow-paths in the reservoir model, and the flow-paths depended more on the well positions than on the permeability field. Since the basis of our upscaling procedure is that the permeability field governs the flow-paths, this model was the most challenging to upscale.

However, for this example, we observed the importance of a flexible upscaling procedure. We could easily adjust the degree of upscaling by the truncation levels. The CO-method gave the best results for this case, but, we observed that the coarse-scale DA method performed equally good as the two other localization methods.

Experiment 3 had a different prior structure than the two previous examples. Here, the model consisted of a categorical field with small Gaussian variations. This was an extremely challenging field to history match, and it was impossible to perform DA without some method to handle the low ensemble size. For this example, we observed a strong dependency on the localization range, where a large region was required to capture the channel structure in the true model. However, as a consequence of a large localization region, the standard deviation was underestimated when compared to the results from other methods. For this example, the best result was obtained by the coarse-scale DA approach.

The results from the localization technique varied for the different examples, and for all examples the coarse-scale DA methods performed equally good or better than the localization methods with the worst result. Moreover, none of the DA methods with localization would consistently outperform the other methods. The best result from the coarse-scale DA methods was obtained when the cost of the linear solver scaled as  $N_g^{1.5}$ . For more efficient linear solvers, the coarse-scale DA method is most often equally good, and occasionally better than the localization approaches.

Examples 1–3 are synthetic 2D examples and realistic cases might be more complex, and/or have different features. However, if we consider upscaling in all three dimensions, there is a large potential for a greater increase in the ensemble size. Moreover, the method for retaining grid cells containing wells in high resolution can easily be extended to retain other structures, such as, faults in the highest resolution. Hence, we believe that the proposed method has high potential for realistic 3D cases.

We have seen that there is a larger variability in the results from the localization methods, and in general it is difficult to determine a generic localization method for history matching. Apart from the task of setting a suitable degree of upscaling, which does not involve any simulation runs, the coarse-scale DA algorithm does not need any other example specific input settings. The coarse-scale DA method is, therefore, a generic alternative to localization for history matching reservoir models.

**Acknowledgments** The authors are grateful for the financial support provided by the CIPR/IRIS cooperative research project “4D Seismic History Matching” which is funded by industry partners Eni, Petrobras, and Total, as well as the Research Council of Norway (PETROMAKS 2).

## References

1. Aanonsen, S.I.: Efficient history matching using a multiscale technique. *SPE Reser. Eval. Eng.* **11**(1) (2008). doi:[10.2118/92758-PA](https://doi.org/10.2118/92758-PA)
2. 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**, 393–412 (2009). doi:[10.2118/117274-PA](https://doi.org/10.2118/117274-PA)
3. Arroyo, E., Devegowda, D., Datta-Gupta, A., Choe, J.: Streamline-assisted ensemble Kalman filter for rapid and continuous reservoir model updating. *SPE Reser. Eval. Eng.* **11**(6) (2008). doi:[10.2118/104255-PA](https://doi.org/10.2118/104255-PA)
4. Axelsson, O.: *Iterative Solution Methods*. Cambridge University Press, Cambridge (1994)
5. Aziz, K., Settari, A.: *Petroleum Reservoir Simulation*. Elsevier Applied Science, New York (1979)
6. Burrus, S.C., Gopinath, R.A., Guo, H.: *Introduction to Wavelets and Wavelet Transforms: A Primer*. Prentice-Hall, New Jersey (1998)
7. Chen, Y.: Personal communication (2015)
8. Chen, Y., Oliver, D.S.: Cross-covariances and localization for EnKF in multiphase flow data assimilation. *Comput. Geosci.* **14**(4), 579–601 (2010). doi:[10.1007/s10596-009-9174-6](https://doi.org/10.1007/s10596-009-9174-6)
9. Chen, Y., Oliver, D.S.: Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math. Geosci.* **44**(1), 1–26 (2012). doi:[10.1007/s11004-011-9376-z](https://doi.org/10.1007/s11004-011-9376-z)
10. Chen, Y., Oliver, D.S.: Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput. Geosci.* **17**(4), 689–703 (2013). doi:[10.1007/s10596-013-9351-5](https://doi.org/10.1007/s10596-013-9351-5)
11. Chen, Y., Oliver, D.S.: History matching of the Norne full-field model with an iterative ensemble smoother. *SPE Reser. Eval. Eng.* **17**(02), 244–56 (2014). doi:[10.2118/164902-PA](https://doi.org/10.2118/164902-PA)
12. Christie, M.A., Blunt, M.J.: Tenth SPE comparative solution project: a comparison of upscaling techniques. *SPE Reser. Eval. Eng.* **4**(4), 308–317 (2001). doi:[10.2118/72469-PA](https://doi.org/10.2118/72469-PA)
13. Daubechies, I.: *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics. doi:[10.1137/1.9781611970104](https://doi.org/10.1137/1.9781611970104) (1992)
14. Deutsch, C.V., Journel, A.G.: *GSLIB: Geostatistical Software Library and User’s Guide*. In: *Applied Geostatistics Series*. 2nd edn. Oxford University Press, New York (1998)
15. Devegowda, D., Arroyo, E., Datta-Gupta, A., Douma, S.G.: Efficient and robust reservoir model updating using ensemble Kalman filter with sensitivity-based covariance localization. *SPE Reserv. Simul. Symp.* doi:[10.2118/106144-MS](https://doi.org/10.2118/106144-MS) (2007)
16. Durlofsky, L.: Upscaling of geocellular models for reservoir flow simulation: A review of recent progress. In: *7th International Forum on Reservoir Simulation* Bühl/Baden-Baden, pp. 23–27. Germany (2003)
17. Ebrahimi, F., Sahimi, M.: Multiresolution wavelet coarsening and analysis of transport in heterogeneous media. *Physica A: Stat. Mech. Appl.* **316**(1–4), 160–188 (2002). doi:[10.1016/S0378-4371\(02\)01199-8](https://doi.org/10.1016/S0378-4371(02)01199-8)
18. Efendiev, Y., Datta-Gupta, A., Ma, X., Mallick, B.: Efficient sampling techniques for uncertainty quantification in history matching using nonlinear error models and ensemble level upscaling techniques. *Water Resour. Res.* **45**(11) (2009). doi:[10.1029/2008WR007039](https://doi.org/10.1029/2008WR007039)
19. Emerick, A.A., Reynolds, A.C.: Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. *Comput. Geosci* **15**(2), 251–269 (2011). doi:[10.1007/s10596-010-9198-y](https://doi.org/10.1007/s10596-010-9198-y)
20. 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), 639–659 (2012). doi:[10.1007/s10596-012-9275-5](https://doi.org/10.1007/s10596-012-9275-5)

21. Emerick, A.A., Reynolds, A.C.: Ensemble smoother with multiple data assimilation. *Comput. Geosci.* **55**, 3–15 (2013). doi:[10.1016/j.cageo.2012.03.011](https://doi.org/10.1016/j.cageo.2012.03.011)
22. 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). doi:[10.1007/s10596-012-9333-z](https://doi.org/10.1007/s10596-012-9333-z)
23. Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5), 10,143–10,162 (1994)
24. Evensen, G., van Leeuwen, P.J.: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.* **128**, 1852–1867 (2000)
25. Farmer, C.L.: Upscaling: a review. *Int. J. Numer. Methods Fluids* **40**(1–2), 63–78 (2002). doi:[10.1002/flid.267](https://doi.org/10.1002/flid.267)
26. Fossum, K., Mannseth, T.: Parameter sampling capabilities of sequential and simultaneous data assimilation: I. Analytical comparison. *Inverse Prob.* **30**(11), 114,002 (2014). doi:[10.1088/0266-5611/30/11/114002](https://doi.org/10.1088/0266-5611/30/11/114002)
27. Fossum, K., Mannseth, T.: Parameter sampling capabilities of sequential and simultaneous data assimilation: II. Statistical analysis of numerical results. *Inverse Prob.* **30**(11), 114,003 (2014). doi:[10.1088/0266-5611/30/11/114003](https://doi.org/10.1088/0266-5611/30/11/114003)
28. Fossum, K., Mannseth, T.: Assessment of ordered sequential data assimilation. *Comput. Geosci.* **19**(4), 821–844 (2015). doi:[10.1007/s10596-015-9492-9](https://doi.org/10.1007/s10596-015-9492-9)
29. Fryzlewicz, P., Timmermans, C.: SHAH: SHape-Adaptive Haar wavelets for image processing. *J. Comput. Graph. Stat.* (August), 00–00 (2015). doi:[10.1080/10618600.2015.1048345](https://doi.org/10.1080/10618600.2015.1048345)
30. 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). doi:[10.1016/j.jmva.2006.08.003](https://doi.org/10.1016/j.jmva.2006.08.003)
31. Gaspari, G., Cohn, S.E.: Construction of correlation functions in two and three dimensions. *Quar. J. R. Meteorol. Soc.* **125**(554), 723–757 (1999). doi:[10.1002/qj.4971255417](https://doi.org/10.1002/qj.4971255417)
32. Giles, M.: Multi-level Monte Carlo path simulation. *Oper. Res.* **56**(3), 607–617 (2008). doi:[10.1287/opre.1070.0496](https://doi.org/10.1287/opre.1070.0496)
33. Girardi, M., Sweldens, W.: A new class of unbalanced haar wavelets that form an unconditional basis for  $L_p$  on general measure spaces. *J. Four. Anal. Appl.* **3**(4), 457–474 (1997). doi:[10.1007/BF02649107](https://doi.org/10.1007/BF02649107)
34. Haar, A.: Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen* **71**(1), 38–53 (1911). doi:[10.1007/BF01456927](https://doi.org/10.1007/BF01456927)
35. Heinrich, S.: Multilevel Monte Carlo methods. In: Margenov, S., Waśniewski, J., Yalamov, P. (eds.) *Springer Proceedings in Mathematics and Statistics, Lecture Notes in Computer Science*, vol. 65, pp. 58–67. Springer, Berlin / Heidelberg (2001)
36. Hoel, H., Law, K.J.H., Tempone, R.: Multilevel ensemble Kalman filtering. *SIAM J. Numer. Anal.* **54**(3), 1813–1839 (2016). doi:[10.1137/15M100955X](https://doi.org/10.1137/15M100955X)
37. Houtekamer, P.L., Mitchell, H.L.: Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **126**(3), 796–811 (1998). doi:[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)
38. Iglesias, M.A.: Iterative regularization for ensemble data assimilation in reservoir models, vol. 19. doi:[10.1007/s10596-014-9456-5](https://doi.org/10.1007/s10596-014-9456-5) (2015)
39. Iglesias, M.A., Law, K.J.H., Stuart, A.M.: Ensemble Kalman methods for inverse problems. *Inverse Prob.* **29**(4), 045,001 (2013). doi:[10.1088/0266-5611/29/4/045001](https://doi.org/10.1088/0266-5611/29/4/045001)
40. Iglesias, M.A., Law, K.J.H., Stuart, A.M.: Evaluation of Gaussian approximations for data assimilation in reservoir models. *Comput. Geosci.* **17**(5), 851–885 (2013). doi:[10.1007/s10596-013-9359-x](https://doi.org/10.1007/s10596-013-9359-x)
41. Jansen, M., Oonincx, P.: *Second Generation Wavelets and Applications*. Springer-Verlag, London (2005). doi:[10.1007/1-84628-140-7](https://doi.org/10.1007/1-84628-140-7)
42. Jensen, A., la Cour-Harbo, A.: *Ripples in Mathematics: The Discrete Wavelet Transform*. Springer-Verlag, Berlin Heidelberg (2001)
43. Le Gland, F., Monbet, V., Tran, V.D.: Large sample asymptotics for the ensemble Kalman filter. In: Crisan, D., Rozovskii, B. (eds.) *The Oxford Handbook of Nonlinear Filtering*, pp. 598–631. Oxford University Press (2011)
44. 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)
45. Lodoen, O., Omre, H.: Scale-corrected ensemble Kalman filtering applied to production-history conditioning in reservoir evaluation. *SPE J.* **13**(2), 1–34 (2008). doi:[10.2118/111374-PA](https://doi.org/10.2118/111374-PA)
46. Mandel, J., Cobb, L., Beezley, J.D.: On the convergence of the ensemble Kalman filter. *Appl. Math.* **56**(6), 533–41 (2011). doi:[10.1007/s10492-011-0031-2](https://doi.org/10.1007/s10492-011-0031-2)
47. Mannseth, T.: Comparison of five different ways to assimilate data for a simplistic weakly nonlinear parameter estimation problem. *Comput. Geosci.* **19**(4), 791–804 (2015). doi:[10.1007/s10596-015-9490-y](https://doi.org/10.1007/s10596-015-9490-y)
48. Mehrabi, A., Sahimi, M.: Coarsening of heterogeneous media: Application of wavelets. *Phys. Rev. Lett.* **79**(22), 4385–4388 (1997). doi:[10.1103/PhysRevLett.79.4385](https://doi.org/10.1103/PhysRevLett.79.4385)
49. Radunović, D.P.: *Wavelets from Math to Practice*. Springer Berlin Heidelberg, Berlin, Heidelberg (2009). doi:[10.1007/978-3-642-00614-2](https://doi.org/10.1007/978-3-642-00614-2)
50. Rasaei, M.R., Sahimi, M.: Upscaling of the permeability by multi-scale wavelet transformations and simulation of multiphase flows in heterogeneous porous media. *Comput. Geosci.* **13**(2), 187–214 (2009). doi:[10.1007/s10596-008-9111-0](https://doi.org/10.1007/s10596-008-9111-0)
51. Reza Rasaei, M., Sahimi, M.: Upscaling and simulation of waterflooding in heterogeneous reservoirs using wavelet transformations: application to the SPE-10 model. *Transp. Porous Media* **72**(3), 311–338 (2008). doi:[10.1007/s11242-007-9152-1](https://doi.org/10.1007/s11242-007-9152-1)
52. Sahimi, M., Heidarinasab, A., Dabir, B.: Computer simulation of conduction in heterogeneous materials: application of wavelet transformations. *Chem. Eng. Sci.* **59**(20), 4291–4303 (2004). doi:[10.1016/j.ces.2004.06.019](https://doi.org/10.1016/j.ces.2004.06.019)
53. Skjervheim, J.A., Evensen, G.: An ensemble smoother for assisted history matching. In: *Proceedings of SPE Reservoir Simulation Symposium*, 2003, pp. 1–15. Society of Petroleum Engineers. doi:[10.2118/141929-MS](https://doi.org/10.2118/141929-MS) (2011)
54. Strichartz, R.R.: Construction of orthonormal wavelets. In: Benedetto, J.J., Frazier, M.W. (eds.) *Wavelets: Mathematics and Applications*. CRC Press, Boca Raton (1994)
55. Sweldens, W.: The lifting scheme: a construction of second generation wavelets. *SIAM J. Math. Anal.* **29**(2), 511–46 (1998). doi:[10.1137/S0036141095289051](https://doi.org/10.1137/S0036141095289051)
56. Sweldens, W., Schröder, P.: Building you own wavelets at home. Part I: First generation wavelets. In: Klees, R., Haagmans, R. (eds.) *Wavelets in Geosciences, Lecture Notes in Earth Sciences*, vol. 90, pp. 72–107. Springer, Berlin Heidelberg (2000)
57. Sweldens, W., Schröder, P.: Building you own wavelets at home. Part II: Second generation wavelets. In: *Wavelets in Geosciences*, vol. 90, pp. 108–130. doi:[10.1007/BFb0011094](https://doi.org/10.1007/BFb0011094) (2000)
58. Wang, Y., Li, G., Reynolds, A.C.: Estimation of depths of fluid contacts and relative permeability curves by history matching using iterative ensemble-Kalman smoothers. *Spe J.* **15**(2), 509–525 (2010). doi:[10.2118/119056-PA](https://doi.org/10.2118/119056-PA)