Comput Geosci
DOI 10.1007/s10596-016-9599-7

ORIGINAL PAPER

Localization and regularization for iterative ensemble smoothers

Yan Chen[1] · **Dean S. Oliver**[2]

Received: 14 March 2016 / Accepted: 13 September 2016

Abstract Ensemble-based data assimilation methods have recently become popular for solving reservoir history matching problems, but because of the practical limitation on ensemble size, using localization is necessary to reduce the effect of sampling error and to increase the degrees of freedom for incorporating large amounts of data. Local analysis in the ensemble Kalman filter has been used extensively for very large models in numerical weather prediction. It scales well with the model size and the number of data and is easily parallelized. In the petroleum literature, however, iterative ensemble smoothers with localization of the Kalman gain matrix have become the state-of-the-art approach for ensemble-based history matching. By forming the Kalman gain matrix row-by-row, the analysis step can also be parallelized. Localization regularizes updates to model parameters and state variables using information on the distance between the these variables and the observations. The truncation of small singular values in truncated singular value decomposition (TSVD) at the analysis step provides another type of regularization by projecting updates to dominant directions spanned by the simulated data ensemble. Typically, the combined use of localization and TSVD is necessary for problems with large amounts of data. In this paper, we compare the performance of Kalman gain localization to two forms of local analysis for parameter estimation problems with nonlocal data. The effect of TSVD with different localization methods and with the use of iteration is also analyzed. With several examples, we show that good results can be obtained for all localization methods if the localization range is chosen appropriately, but the optimal localization range differs for the various methods. In general, for local analysis with observation taper, the optimal range is somewhat shorter than the optimal range for other localization methods. Although all methods gave equivalent results when used in an iterative ensemble smoother, the local analysis methods generally converged more quickly than Kalman gain localization when the amount of data is large compared to ensemble size.

Keywords Iterative ensemble smoother · Localization · Local analysis · Ensemble Kalman filter · Regularization · History matching

1 Introduction

When ensemble Kalman filter-like data assimilation methods are applied to large reservoir simulation models with many observations, it is necessary to apply some type of localization to reduce the effect of spurious correlations and to increase the number of degrees of freedom. Some of the standard approaches to localization work quite well on small- and medium-size problems, but the methods that are commonly used for subsurface parameter estimation problems are generally not efficient when the number of data and the number of model variables become large. This limits the ability to apply these methods to problems of assimilation of 4D seismic data and to methods in which all data are assimilated simultaneously. Since iterative ensemble smoothers have become fairly standard methods for history matching large reservoir models, we investigate the performance of

✉ Yan Chen
yanchen.ou@gmail.com

Dean S. Oliver
dean.oliver@uni.no

[1] Geoscience Research Centre, Total E&P, Aberdeen, Scotland

[2] Uni Research CIPR, Bergen, Norway

Published online: 03 November 2016

four localization approaches to data assimilation using an iterative ensemble Kalman smoother: (1) localization of the model covariance matrix, (2) localization of the Kalman gain matrix (computed one row at a time), (3) local analysis with observation taper, and (4) local analysis with Kalman gain taper.

Localization methods for ensemble-based data assimilation methods were developed for atmospheric and oceanography applications where, although the dynamics can be strongly nonlinear, data are often local and linear, i.e., observations are often of model states at spatially distributed locations. Using N_d to denote the dimension of the data space and N_m to denote the dimension of the model space, the observation operator $\mathbf{H}$ of size $N_\mathrm{d} \times N_\mathrm{m}$ for this type of local and linear data has only one nonzero element (equal to one) per row. If $\mathbf{H}$ is the observation operator, it is easy to directly apply tapering to the covariance of state variables ($\boldsymbol{\rho} \circ \mathbf{C}_\mathrm{M}$), in which $\mathbf{C}_\mathrm{M}$ is the covariance of the state vector, $\boldsymbol{\rho}$ is the tapering matrix of same size, and $\circ$ denotes element-wise multiplication (Schur product). After the tapering, matrices that involve $\mathbf{C}_\mathrm{M}$ in typical ensemble-based methods become

$$\mathbf{H}\left(\boldsymbol{\rho} \circ \mathbf{C}_\mathrm{M}\right)\mathbf{H}^\mathrm{T} = \left(\mathbf{H}\boldsymbol{\rho}\mathbf{H}^\mathrm{T}\right) \circ \left(\mathbf{H}\mathbf{C}_\mathrm{M}\mathbf{H}^\mathrm{T}\right) \tag{1}$$

$$\left(\boldsymbol{\rho} \circ \mathbf{C}_\mathrm{M}\right)\mathbf{H}^\mathrm{T} = \left(\boldsymbol{\rho}\mathbf{H}^\mathrm{T}\right) \circ \left(\mathbf{C}_\mathrm{M}\mathbf{H}^\mathrm{T}\right) . \tag{2}$$

Through the use of these identities, computation of the Schur product between matrices of size $N_\mathrm{m} \times N_\mathrm{m}$ is avoided, and the Schur product is only applied to matrices of size $N_\mathrm{d} \times N_\mathrm{d}$ (right-hand side of Eq. 1) and $N_\mathrm{m} \times N_\mathrm{d}$ (right-hand side of Eq. 2). The matrix $\mathbf{H}\boldsymbol{\rho}\mathbf{H}^\mathrm{T} \circ \mathbf{H}\mathbf{C}_\mathrm{M}\mathbf{H}^\mathrm{T}$, however, appears in a term that needs to be inverted at the analysis step. When the number of data is large, it is not possible to invert a matrix of size $N_\mathrm{d} \times N_\mathrm{d}$, so applications that apply localization to the covariance matrix $\mathbf{C}_\mathrm{M}$ typically assimilate data sequentially at the analysis step, for example Hamill et al. [17] assimilate data one by one and Houtekamer and Mitchell [18] assimilate data in batches. This sequential update takes place at a single analysis step, and should not be confused with sequential updating using data at different times. Assimilating data sequentially is only feasible when the observation operator $\mathbf{H}$ is inexpensive to evaluate, because simulated data need to be computed for each sequential update.

Elements of the taper matrix $\boldsymbol{\rho}$ are typically formed by evaluating a predefined correlation function ρ with appropriate distance as input. Some correlation functions are completely defined by a single parameter that controls the critical range, e.g., Gaspari and Cohn [15]. When the observation operator is local, this critical range depends primarily on the correlation scale of model variables (which is also the expected range of $\mathbf{C}_\mathrm{M}$). The optimal range of the tapering function is a function of ensemble size, with larger ensemble requires less localizaiton. The tapering function ρ in Furrer and Bengtsson [14] is optimized to minimize the difference between the tapered ensemble estimate $\boldsymbol{\rho} \circ \mathbf{C}_\mathrm{M}$ and the true covariance, and it directly incorporates ensemble size as an input parameter.

When ensemble-based data assimilation methods are applied to history matching problems, it has become common to use iterative ensemble smoothers [6, 10] instead of ensemble Kalman filters. In this case, only model parameters (e.g., gridblock permeability and porosity, etc.) are estimated, and the observation operator for typical production and seismic data is neither local or linear. We use $\mathbf{g}$ to denote a general observation operator and $\mathbf{G}$ to denote its linearized approximation (referred to data sensitivity), to clearly differentiate it from the special type of observation operate $\mathbf{H}$ mentioned earlier. The sensitivity $\mathbf{G}$ of production data like water cut and gas-oil-ratio to model parameters like permeability and porosity is typically not available, although their general behavior is often well understood [25]. Unfortunately, Eqs. 1 and 2 which allow efficient localization do not hold if $\mathbf{H}$ is replaced by the general observation operator $\mathbf{G}$, so the tapering matrices $\boldsymbol{\rho}_\mathrm{dd}$ and $\boldsymbol{\rho}_\mathrm{md}$ for correlation between different simulated data $\mathbf{G}\mathbf{C}_\mathrm{M}\mathbf{G}^\mathrm{T}$ and correlation between model parameters and data $\mathbf{C}_\mathrm{M}\mathbf{G}^\mathrm{T}$ cannot be easily derived from a tapering function that is appropriate for $\mathbf{C}_\mathrm{M}$. Because of the nonlocal nature of data sensitivity $\mathbf{G}$, the choice of the critical range for tapering matrices $\boldsymbol{\rho}_\mathrm{dd}$ and $\boldsymbol{\rho}_\mathrm{md}$ cannot be based on correlation scale of model variables alone as is the case for local data.

Chen and Oliver [3] studied the behavior of data sensitivity for different data types and suggested the choice of localization range should be a joint consideration of correlation scale of model parameters, data sensitivity, and ensemble size. They suggested that the range of the tapering function should be large enough to include the next line of wells outside a pattern for pattern flooding cases, and large enough to include important parameters in the direction of pressure support (e.g., aquifers). Since distances need to be computed between data and model parameters for tapering, a physical location needs to be assigned to each nonlocal observation. For data measured at well locations, the center location of open completions can often be used as the location for data; however, for certain types of data, e.g., water production rate, the location for data could be assigned at an intermediate location between the producer and the water source (i.e., injector or aquifers) as suggested in Chen and Oliver [2]. Seismic data that are used for history matching, typically are seismic attributes on either the seismic or the reservoir grid, and the location of the cell is a natural choice as the data location.

Emerick and Reynolds [7] suggested setting the critical range for localization to be the sum of the length scale of the prior covariance of model parameters and the length

scale of the well drainage area (from streamline simulation) and used Gaspari-Cohn function to compute elements of $\boldsymbol{\rho}_{\mathrm{md}}$ and $\boldsymbol{\rho}_{\mathrm{dd}}$, with distance between model parameters and data and between different data. Chen and Oliver [3] used one correlation function from Furrer and Bengtsson [14] to compute elements of $\boldsymbol{\rho}_{\mathrm{md}}$ then used an empirical formula to compute $\boldsymbol{\rho}_{\mathrm{dd}}$ from $\boldsymbol{\rho}_{\mathrm{md}}$. Similar to the nonlocal data in history matching problems, satellite radiances in atmospheric data assimilation applications also reflect spatially weighted average (mostly vertically) of state variables. The critical distance for localization of this type of nonlocal data is generally increased to reflect the length scale of vertical averaging [13].

When it is difficult to localize the product $\mathbf{G}\mathbf{C}_{\mathrm{M}}\mathbf{G}^{\mathrm{T}}$, or when it is infeasible to invert an $N_{\mathrm{d}} \times N_{\mathrm{d}}$ matrix, then an alternative is to ignore the localization of $\mathbf{G}\mathbf{C}_{\mathrm{M}}\mathbf{G}^{\mathrm{T}}$ and instead only apply localization to the entire Kalman gain matrix [2, 6, 10]. Localizing the Kalman gain matrix does not directly solve the problem of limited degrees of freedom, but it does alter the subspace in which updates are found and it does reduce spurious updates to model parameters. When iterative ensemble methods are used, the problem of insufficient degrees of freedom is reduced because a different subspace is used to construct updates at each iteration. Although the dimension of the model N_{m} is typically large, the Kalman gain matrix can be formed in segments and in parallel [8], so the Kalman gain localization is computationally feasible for large-scale problems. No difference in terms of the choice of tapering function for localizing covariance and Kalman gain has been mentioned in the literature, and typically the same criteria for $\boldsymbol{\rho}_{\mathrm{md}}$ have been used for tapering of the Kalman gain matrix. Chen and Oliver [3] observed artifacts in the updated model parameters in a 5-spot water flooding example when the Kalman gain matrix was localized with a localization region smaller than a single well pattern or only slightly bigger then a single pattern, but the artifacts disappeared when the localization region was made large enough (see examples in [2, 6, 10]).

Local analysis (also referred to as observation localization) is another commonly used localization method to reduce the impact of spurious correlations and to increase the degrees of freedom in updating. It is commonly used for data assimilation in atmospheric and oceanographic applications [1, 19, 26], but although several authors in petroleum engineering have used "regional analysis" [12, 30] and one paper has used local analysis with an observation cutoff [32], it appears that local analysis with tapering has not been used in petroleum applications. The main concept of local analysis is to decompose the entire model into local regions. Model variables within each local region are updated using only data that are within a critical distance to the local region. This local region can be a single grid point as shown by the gray cell in Fig. 1, or more commonly a single vertical column of a 3D model [19, 26]. Because of the data selection for update at each local region, the contribution from distant data that are likely to be dominated by noise is removed. In addition to data selection, the selected data for each update can be given weights by altering the elements of data noise covariance $\mathbf{C}_{\mathrm{D}}$ to gradually reduce the contribution from data as their distance to the current local region increases [13, 20]. Data selection and weighting of the data can be done consistently, for example, Hunt et al. [20] used the Gaspari-Cohn function with a chosen range to compute the weights (multiply to $\mathbf{C}_{\mathrm{D}}{}^{-1}$) and data with zero weight are not used for the local update.

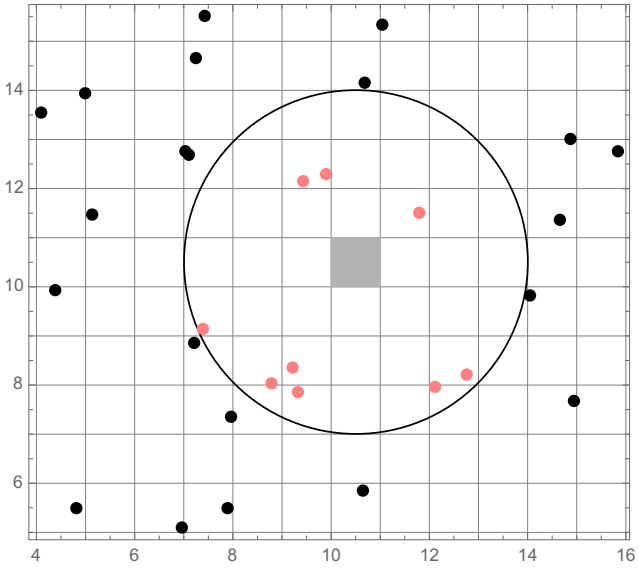


Fig. 1 Local analysis for updating of parameters in a single cell (*gray cell*). Only observations within critical radius (indicated by the *black circle*) are used to compute the update. The weights on the selected observations (*red dots*) are typically based on distance from the cell to be updated (*gray cell*)

Several authors [16, 22, 23] have noted that the use of the same distance taper in local analysis results in less localization than in covariance localization. Nerger et al. [23] showed that this difference in the effective localization range is especially large when the variance of the data noise is much smaller than the variance of the corresponding simulated data from the ensemble, and proposed a method to modify the weights in local analysis so that its behavior is more consistent with covariance localization. Sakov and Bertino [29] compared the behavior between covariance localization and local analysis for linear local observations with non-correlated errors using a simple toy problem. By analyzing the ensemble transform matrix, the results suggested that for strong data assimilation problem (accurate data), covariance localization might be better than local analysis. The same localization range, however, was used for covariance localization and local analysis in Sakov and Bertino [29].

This paper studies the performance of local analysis for regularizing updates from nonlocal data with comparison to the commonly used Kalman gain localization. Two types of tapering are tested for local analysis: one is tapering of $\mathbf{C_D}^{-1}$ as in Hunt et al. [20], the other is tapering of the local Kalman gain. The paper also aims at understanding the appropriate choice of localization range in local analysis for typical production data if it should be much different from what has been used for Kalman gain localization. The iterative ensemble smoother, Levenberg-Marquardt ensemble randomized maximum likelihood method (LM-EnRML) in Chen and Oliver [5], is used for history matching; however, findings in this paper are applicable to other similar iterative ensemble smoothers. Localization for iterative ensemble smoothers is simpler than localization for the ensemble Kalman filter since it is often more difficult to localize dynamic states (e.g., saturation and pressure) than model parameters [3]. The convergence behavior of LM-EnRML resulting from the combined use of truncated singular value decomposition (TSVD) and localization is also studied.

In the remainder of this paper, we first briefly review the methodology for LM-EnRML, the use of TSVD to reduce the number of effective data, then present different localization methods. The performance of various localization methods is compared using two simple toy problems with local and nonlocal data and a synthetic history matching example with production data from a large number of wells. We limit discussion to data with uncorrelated data noise (diagonal $\mathbf{C_D}$) in this paper. The results could be easily extended to non-diagonal $\mathbf{C_D}$ whose square root is easy to compute.

2 The iterative ensemble smoother

Here, we briefly describe an iterative ensemble smoother so that the implementation of local analysis can be understood. First, we define a few scalar variables: N_e is the ensemble size, N_m is the number of model parameters, N_d is the number of data, and ℓ and j are indices for iteration and model realization, respectively. We denote the normalized simulated data deviation ensemble as

$$\Delta\mathbf{D}^\ell = \frac{1}{\sqrt{N_\mathrm{e}-1}}\mathbf{C_D}^{-1/2}\,\mathbf{D}^\ell\left(\mathbf{I}_{N_\mathrm{e}} - \frac{1}{N_\mathrm{e}}\mathbf{1}_{N_\mathrm{e}}\mathbf{1}_{N_\mathrm{e}}^\mathrm{T}\right), \tag{3}$$

where $\mathbf{D}^\ell$ is a matrix of size $N_\mathrm{d} \times N_\mathrm{e}$ with each column being simulated data $\mathbf{g}(\mathbf{m}_j^\ell)$ from one model realization $\mathbf{m}_j^\ell$. Here, and elsewhere in the paper, $\mathbf{g}(\cdot)$ represents the function that maps model parameters to predicted data. The matrix $\mathbf{I}_{N_\mathrm{e}}$ is an identity matrix of size N_e and $\mathbf{1}_{N_\mathrm{e}}$ is a vector of length N_e with all elements equal to one. Data noise is assumed to follow a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{C_D})$, where $\mathbf{C_D}$ is a diagonal matrix of size N_d.

Similarly, the deviation ensemble of model parameters is

$$\Delta\mathbf{M}^\ell = \frac{1}{\sqrt{N_\mathrm{e}-1}}\mathbf{M}^\ell\left(\mathbf{I}_{N_\mathrm{e}} - \frac{1}{N_\mathrm{e}}\mathbf{1}_{N_\mathrm{e}}\mathbf{1}_{N_\mathrm{e}}^\mathrm{T}\right), \tag{4}$$

where $\mathbf{M}^\ell$ is a matrix of size $N_\mathrm{m} \times N_\mathrm{e}$ with each column being one realization of model parameters $\mathbf{m}_j^\ell$.

The updating formula at the ℓth iteration of the approximate form of LM-EnRML (Levenberg-Marquardt ensemble randomized maximum likelihood method, Chen and Oliver [5]) can be written as

$$\begin{aligned}\delta\mathbf{m}_j^\ell \;=\; & \Delta\mathbf{M}^\ell\Delta\mathbf{D}^{\ell\mathrm{T}}\left((1+\lambda_\ell)\mathbf{I}_{N_\mathrm{d}} + \Delta\mathbf{D}^\ell\Delta\mathbf{D}^{\ell\mathrm{T}}\right)^{-1} \\ & \mathbf{C_D}^{-1/2}\left(\mathbf{d}_{\mathrm{obs},j} - \mathbf{g}(\mathbf{m}_j^\ell)\right).\end{aligned} \tag{5}$$

In Eq. 5, $\delta\mathbf{m}_j^\ell$ of length N_m is the change to the jth realization of model parameters at iteration ℓ. The vector $\mathbf{d}_{\mathrm{obs},j}$ of length N_d is the perturbed data realization

$$\mathbf{d}_{\mathrm{obs},j} = \mathbf{d}_\mathrm{obs} + \mathbf{C_D}^{1/2}\mathbf{z}_j\,, \tag{6}$$

where $\mathbf{d}_\mathrm{obs}$ is the vector of observed data, and $\mathbf{z}_j$ of length N_d is white noise sampled from standard Gaussian distribution. The scalar λ in Eq. 5 is the Levenberg-Marquardt tuning parameter for iterations. The tuning of the λ in LM-EnRML follows the typical choice for Levenberg-Marquardt type regularization for optimization. The value of λ at the first iteration is typically large (similar order of magnitude as the data mismatch, for example 10^5). The value of λ is reduced by a factor (for example 10) with each successful iteration. The typical stopping criteria used for terminating iteration for the LM-EnRML include reaching maximum number of allowed iterations, reaching minimum reduction of data mismatch and reaching minimum changes to model parameters. Note that the approximate form of LM-EnRML (LM-EnRML approx) in Eq. 5 is very similar to ES-MDA (ensemble smoother with multiple data assimilation, Emerick and Reynolds [9]), except for the choice of the tuning parameter and the generation of the perturbed observations.

3 Truncated SVD for regularization

Typically, inversion of the matrix $(1+\lambda_\ell)\,\mathbf{I}_{N_\mathrm{d}} + \Delta\mathbf{D}^\ell\Delta\mathbf{D}^{\ell\mathrm{T}}$ in Eq. 5 is carried out in the ensemble space [11]. To simplify notation, the iteration index ℓ and the term $1+\lambda$ are suppressed in this section. We use $\mathbf{K}$ to represent the coefficient of the data misfit in Eq. 5, that is

$$\mathbf{K} = \Delta\mathbf{M}\Delta\mathbf{D}^\mathrm{T}(\mathbf{I}_{N_\mathrm{d}} + \Delta\mathbf{D}\Delta\mathbf{D}^\mathrm{T})^{-1}\,. \tag{7}$$

Note that $\mathbf{K}$ in Eq. 7 is very similar to the standard Kalman gain matrix in the ensemble Kalman filter except for a factor of $\mathbf{C_D}^{-1/2}$, and we will refer to $\mathbf{K}$ as Kalman gain in the following sections.

To provide regularization of the updates, we write the truncated singular value decomposition (TSVD) of $\Delta\mathbf{D}$ as

$$\Delta\mathbf{D} = \mathbf{U}_p\mathbf{W}_p\mathbf{V}_p^{\mathrm{T}}, \tag{8}$$

where p is the number of singular values retained in TSVD and the dimension of matrices $\mathbf{U}_p$, $\mathbf{W}_p$, and $\mathbf{V}_p$ are $N_\mathrm{d} \times p$, $p \times p$, and $N_\mathrm{e} \times p$, respectively. Using Eq. 8, the $\mathbf{K}$ matrix becomes

$$\begin{aligned}\mathbf{K} &\approx \Delta\mathbf{M}(\mathbf{U}_p^{\mathrm{T}}\Delta\mathbf{D})^{\mathrm{T}}(\mathbf{I}_p + \mathbf{W}_p^2)^{-1}\mathbf{U}_p^{\mathrm{T}} \\ &= \Delta\mathbf{M}\mathbf{V}_p\mathbf{W}_p(\mathbf{I}_p + \mathbf{W}_p^2)^{-1}\mathbf{U}_p^{\mathrm{T}} \\ &= \mathbf{K}_{\mathrm{eff}}\mathbf{U}_p^{\mathrm{T}},\end{aligned} \tag{9}$$

where $\mathbf{K}_{\mathrm{eff}}$ is Kalman gain for the effective data (data projected to the space spanned by columns of $\mathbf{U}_p^{\mathrm{T}}$).

Utilizing the coefficient matrix in Eq. 9, the updating step of the iterative ensemble smoother Eq. 5 becomes

$$\delta\mathbf{m}_j = \underbrace{\Delta\mathbf{M}(\overbrace{\mathbf{U}_p^{\mathrm{T}}\Delta\mathbf{D}}^{\Delta\mathbf{D}_{\mathrm{eff}}})^{\mathrm{T}}(\mathbf{I}_p + \mathbf{W}_p^2)^{-1}}_{\mathbf{K}_{\mathrm{eff}}\,(N_\mathrm{m}\times p)} \overbrace{\mathbf{U}_p^{\mathrm{T}}\mathbf{C}_\mathrm{D}{}^{-1/2}\left(\mathbf{d}_{\mathrm{obs},j} - \mathbf{g}(\mathbf{m}_j^\ell)\right)}^{\text{effective innovation } (p\times 1)}. \tag{10}$$

The need to form a matrix of dimension $N_\mathrm{m} \times N_\mathrm{d}$ is avoided in Eq. 10 by working in the effective data space. The product of $\Delta\mathbf{M}\Delta\mathbf{D}_{\mathrm{eff}}$ can be thought of as the cross covariance between the model parameters and the effective data.

The use of TSVD regularizes ensemble updates by projecting updates to dominant directions spanned by $\Delta\mathbf{D}$. Chen and Oliver [4] showed the benefit of truncating small singular values using the Brugge benchmark case [27]. The changes made to the model were much smaller and the final data match was better when small singular values were truncated. Examples (not in the context of reservoir fluid flow) showing similar effects can also be found in Figs. 14.1 and 14.2 of Evensen [11]. The truncation level in the TSVD is usually based on percentage of energy retained. Values between 90 to 99.9 % are typically used in history matching problems. Sætrom et al. [28] discussed the use of cross-validation to select the optimal truncation level. Note that the TSVD (8) is computed at each iteration so that the subspace in which updates are computed changes at each iteration. This effectively increases the degrees of freedom; however, the updates are still restricted to ensemble space unless localization is used, so TSVD is typically not sufficient to allow assimilation of large amounts of data. In addition, the use of iteration compensates for strong truncation in TSVD, thus the truncation level is not as important in iterative methods as it is for non-iterative methods (shown with the example in Section 5.1).

4 Localization methods

If localization is not used as in Eq. 10, updates at one iteration are a linear combination of columns of the covariance between model parameters and effective data. The degrees of freedom for model update are then limited by the number of singular values p retained in the decomposition, with $p < N_\mathrm{e}$. When the ensemble size is small, localization is necessary so that different linear combinations can be used for updating different model parameters in order to assimilate large amounts of independent data. In addition, the ensemble size is typically much smaller than the effective model dimension, in this case using localization is also essential for reducing the effect of spurious correlation. Note that the effective dimension of the model space is usually less than the number of model parameters N_m, because model parameters are usually spatially correlated.

4.1 Taper functions

The localization function that is used for tapering is typically a correlation function with compact support. A common choice is the Gaspari-Cohn function [15], whose only parameter is range R (critical length) of the function. Figure 2a show the Gaspari-Cohn function with range equal to 10. The function value reduces to 0.2 at distance equal to R and is zero for distances greater than $2R$. Furrer and Bengtsson [14] introduced a localization function that depends on covariance type, range R of the covariance function, and ensemble size. Figure 2b, c shows the Furrer-Bengtsson function for spherical and Gaussian covariance

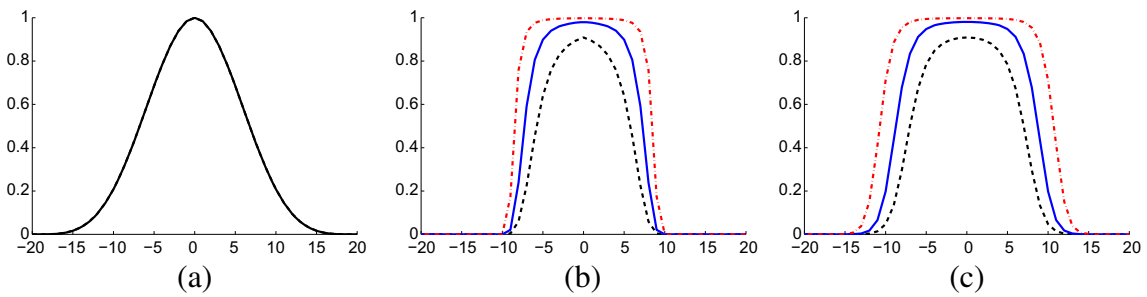


Fig. 2 **a** GC taper function [15], **b** and **c** taper function in Furrer and Bengtsson [14] with Spherical and Gaussian covariance function with ensemble size of 20 (*black dashed*), 100 (*blue solid*), and 1000 (*red dash-dotted*)

with range equal to 10 for different ensemble sizes. Unnormalized taper functions are shown in Fig. 2b, c. The normalized taper function is scaled so that the maximum value of the function is equal to one. In the remainder of the paper, we use ρ to denote the localization function with any optional parameters fixed so that ρ is only a function of distance $\rho(r)$. The bold $\boldsymbol{\rho}$ denotes a matrix or a vector with its elements evaluated at the appropriate distance using ρ.

4.2 Covariance localization and Kalman gain localization

For covariance localization, the covariance between model parameters and simulated data $\Delta\mathbf{M}^{\ell}\Delta\mathbf{D}^{\ell\mathrm{T}}$ and the covariance of simulated data $\Delta\mathbf{D}^{\ell}\Delta\mathbf{D}^{\ell\mathrm{T}}$ must both be localized, in which case the update at the analysis step becomes

$$\delta\mathbf{m}_j^{\ell} = \left[\boldsymbol{\rho}_{\mathrm{md}} \circ \left(\Delta\mathbf{M}^{\ell}\Delta\mathbf{D}^{\ell\mathrm{T}}\right)\right]\Big((1+\lambda_{\ell})\mathbf{I}_{N_{\mathrm{d}}} + \left[\boldsymbol{\rho}_{\mathrm{dd}} \circ \left(\Delta\mathbf{D}^{\ell}\Delta\mathbf{D}^{\ell\mathrm{T}}\right)\right]\Big)^{-1}\delta\mathbf{d}^{\ell} \tag{11}$$

where

$$\delta\mathbf{d}^{\ell} = {\mathbf{C}_{\mathrm{D}}}^{-1/2}\left(\mathbf{d}_{\mathrm{obs},j} - \mathbf{g}(\mathbf{m}_j^{\ell})\right). \tag{12}$$

In the remainder of this section, we omit the index for iteration ℓ and for realization j and the term $1+\lambda$ for simplification of notation.

In analogy to Eqs. 1 and 2, $\boldsymbol{\rho}_{\mathrm{dd}}$ and $\boldsymbol{\rho}_{\mathrm{md}}$ in Eq. 11 are usually chosen to be similar in effect to $\mathbf{G}\boldsymbol{\rho}_{\mathrm{mm}}\mathbf{G}^{\mathrm{T}}$ and $\boldsymbol{\rho}_{\mathrm{mm}}\mathbf{G}^{\mathrm{T}}$. The selection of the localization function ρ for computing elements of $\boldsymbol{\rho}_{\mathrm{md}}$ is typically a joint consideration of the correlation length of the covariance matrix for $\mathbf{m}_{\mathrm{pr}}$, the data sensitivity $\mathbf{G}$, and the ensemble size N_{e} [3, 7]. Because each row of $\mathbf{G}$ corresponds to an observation and each column of $\mathbf{G}$ corresponds to a model parameter, different ρ can be used for forming different parts of $\boldsymbol{\rho}_{\mathrm{md}}$ if the data sensitivity or the prior correlation length is very different for different types of data and different types of model parameters. Elements of $\boldsymbol{\rho}_{\mathrm{dd}}$ can be computed using the same localization function ρ as that for $\boldsymbol{\rho}_{\mathrm{md}}$ [7] or can be computed directly from $\boldsymbol{\rho}_{\mathrm{md}}$ [3].

In Eq. 11, because localization is applied to matrix $\Delta\mathbf{D}\Delta\mathbf{D}^{\mathrm{T}}$, the Kalman gain matrix cannot be written in terms of effective data as in Eq. 9 and it is necessary to invert the matrix of size $N_{\mathrm{d}} \times N_{\mathrm{d}}$, which can be computationally prohibitive for problems with large amounts of data. On the other hand, although N_{m} can be very large, it is possible to compute $\boldsymbol{\rho}_{\mathrm{md}} \circ \left(\Delta\mathbf{M}\Delta\mathbf{D}^{\mathrm{T}}\right)$ in segments.

In Kalman gain localization, tapering is directly applied to the $\mathbf{K}$ matrix in Eq. 9

$$\delta\mathbf{m} = \underbrace{\left[\boldsymbol{\rho}_{\mathrm{md}} \circ \left(\Delta\mathbf{M}(\mathbf{U}_p^{\mathrm{T}}\Delta\mathbf{D})^{\mathrm{T}}(\mathbf{I}_p + \mathbf{W}_p^2)^{-1}\mathbf{U}_p^{\mathrm{T}}\right)\right]}_{\mathbf{K}^{\mathrm{loc}}}\delta\mathbf{d}. \tag{13}$$

Because the effective data do not have physical locations associated to them, in order to apply localization, the actual $\mathbf{K}$ of size $N_{\mathrm{m}} \times N_{\mathrm{d}}$ is needed instead of the gain matrix $\mathbf{K}_{\mathrm{eff}}$ of size $N_{\mathrm{m}} \times p$ in Eq. 9. Similar to tapering of $\Delta\mathbf{M}\Delta\mathbf{D}^{\mathrm{T}}$ in Eq. 11, tapering of Kalman gain in Eq. 13 can also be formed in segments, for example row by row, and in parallel, so using Kalman gain localization is feasible even for large-scale problems.

4.3 Local analysis/observation localization

In local analysis, the entire model is decomposed into many local domains (i.e., domain decomposition). For earth science problems, it is typical to decompose a 3D model into vertical columns [19, 26]. A local update is carried out for model parameters within each local domain using only data that are within a critical distance to the local region (i.e., data selection). In addition, the selected data for each local update can be given weights based on their distance to the current local region (i.e., tapering).

Although the number of model parameters updated at one local analysis step will generally be more than one, we describe the procedure assuming one local update is performed for each model parameter without losing generality. Similar to the Kalman gain matrix $\mathbf{K}$ in Eq. 7 for the global analysis, the Kalman gain matrix at the ith local update is

$$\mathbf{K}_{(i)} = \Delta\mathbf{M}_{(i)}\Delta\mathbf{D}_{(i)}^{\mathrm{T}}(\mathbf{I}_{N_{\mathrm{d}_i}} + \Delta\mathbf{D}_{(i)}\Delta\mathbf{D}_{(i)}^{\mathrm{T}})^{-1}, \tag{14}$$

where (i) indicate the local update for the ith model parameter. $\Delta\mathbf{M}_{(i)}$ represents the ith row of $\Delta\mathbf{M}$, and $\Delta\mathbf{D}_{(i)}$ consists of the rows of $\Delta\mathbf{D}$ that correspond to the subset of data used for the ith local update. The data used for updating the ith model parameter $\mathbf{m}_{(i)}$ are those whose distance r to $\mathbf{m}_{(i)}$ results in $\rho(r)$ greater than a threshold (10^{-3} is used in this study). The number of data assimilated at the ith local update is N_{d_i}.

Similar to Eq. 8, the truncated singular value decomposition (TSVD) of $\Delta\mathbf{D}_{(i)}$ is

$$\Delta\mathbf{D}_{(i)} = \mathbf{U}_{p_i}\mathbf{W}_{p_i}\mathbf{V}_{p_i}^{\mathrm{T}}. \tag{15}$$

The columns of $\mathbf{U}_{p_i}$ and the diagonal elements of $\mathbf{W}_{p_i}$ are the p_i leading left singular vectors and singular values of $\Delta\mathbf{D}_{(i)}$. Using Eqs. 12, 14, and 15 and applying tapering to the local Kalman gain matrix, the ith model parameter $\mathbf{m}_{(i)}$ is updated using

$$\delta\mathbf{m}_{(i)} = \underbrace{\left[\boldsymbol{\rho}_{\mathrm{md}\,(i)} \circ \left(\Delta\mathbf{M}_{(i)}(\mathbf{U}_{p_i}^{\mathrm{T}}\Delta\mathbf{D}_{(i)})^{\mathrm{T}}(\mathbf{I}_{p_i} + \mathbf{W}_{p_i}^2)^{-1}\mathbf{U}_{p_i}^{\mathrm{T}}\right)\right]}_{\mathbf{K}_{(i)}{}^{\mathrm{loc}}}\delta\mathbf{d}_{(i)}, \tag{16}$$

where $\delta\mathbf{d}_{(i)}$ and $\boldsymbol{\rho}_{\mathrm{md}\,(i)}$ indicate taking the corresponding elements of $\delta\mathbf{d}$ and $\boldsymbol{\rho}_{\mathrm{md}}$ for the ith local update.

Table 1 Summary of different localization methods investigated

Method	Main characteristics
Kalman gain localization (13)	Global analysis with tapering on Kalman gain
LA Kalman gain taper (16)	Local analysis with tapering on local Kalman gain
LA observation taper (19)	Local analysis with tapering on observation noise

The updating equation for local analysis with tapering on the local Kalman gain, Eq. 16, is similar to Eq. 13 except that the TSVD is performed with a subset of data $\Delta\mathbf{D}_{(i)}$ at each local update (while in Eq. 13 the TSVD is only performed once with the entire data set). The number of data at each local update of local analysis is generally small compared to the total number of data. The TSVD of $\Delta\mathbf{D}_{(i)}$ in Eq. 15 can retain much more information compared to TSVD of $\Delta\mathbf{D}$ in Eq. 8. This results in faster convergence rate of local analysis methods compared to localization of the Kalman gain when used with the iterative ensemble smoother (as shown later with the examples). To avoid repeated calculation of TSVD in Eq. 15 for the same data set, it is sensible to group model parameters that have the same data set into one local update, for example, different properties at the same model cell or different properties at an entire column of the 3D model, etc. The local updates in local analysis are independent of each other and can be carried out in parallel.

In atmospheric and oceanography literature, a more common way to localize at each local update is to scale the data noise, so that noise of data that are far from $\mathbf{m}_{(i)}$ is increased to reduce the effect of distant data on the update. As shown in Sakov and Bertino [29], scaling the data noise is equivalent to scaling the ensemble anomalies. In the following, local analysis with observation taper is written in terms of scaling the ensemble anomalies. Let $\boldsymbol{\rho}_{(i)}$ be a column vector of length N_{d_i} with its elements computed based on the distance between each datum and the model parameter $\mathbf{m}_{(i)}$. The scaled anomalies of the simulated data ensemble are represented using

$$\Delta\mathbf{D}^{\rho}_{(i)} = \left(\boldsymbol{\rho}^{1/2}_{(i)}\mathbf{1}^{\mathrm{T}}_{N_\mathrm{e}}\right) \circ \Delta\mathbf{D}_{(i)}\,, \tag{17}$$

where $\mathbf{1}_{N_\mathrm{e}}$ is a column vector of length N_e with all elements equal to one.

Using the scaled data $\Delta\mathbf{D}^{\rho}_{(i)}$ in place of $\Delta\mathbf{D}_{(i)}$ in Eq. 14, the updating for model parameter $\mathbf{m}_{(i)}$ becomes

$$\delta\mathbf{m}_{(i)} = \Delta\mathbf{M}_{(i)}\Delta\mathbf{D}^{\rho\,\mathrm{T}}_{(i)}\left(\mathbf{I}_{N_{\mathrm{d}_i}}+\Delta\mathbf{D}^{\rho}_{(i)}\Delta\mathbf{D}^{\rho\,\mathrm{T}}_{(i)}\right)^{-1}\left[\boldsymbol{\rho}^{1/2}_{(i)}\circ\delta\mathbf{d}_{(i)}\right]. \tag{18}$$

Because $\mathbf{C}_\mathrm{D}{}^{1/2}$ appears in the definition of $\delta\mathbf{d}$ (12), $\delta\mathbf{d}_{(i)}$ also needs to be scaled in Eq. 18. Applying TSVD to the scaled data $\Delta\mathbf{D}^{\rho}_{(i)}$ similar to Eq. 15, Eq. 18 becomes

$$\delta\mathbf{m}_{(i)} = \Delta\mathbf{M}_{(i)}\Delta\mathbf{D}^{\rho\,\mathrm{T}}_{(i)}\mathbf{U}^{\rho}_{p_i}(\mathbf{I}_{p_i}+\mathbf{W}^{\rho 2}_{p_i})^{-1}\mathbf{U}^{\rho\,\mathrm{T}}_{p_i}\left[\boldsymbol{\rho}^{1/2}_{(i)}\circ\delta\mathbf{d}_{(i)}\right], \tag{19}$$

where the columns of $\mathbf{U}^{\rho}_{p_i}$ and the diagonal elements of $\mathbf{W}^{\rho}_{p_i}$ are the p_i leading left singular vectors and singular values of $\Delta\mathbf{D}^{\rho}_{(i)}$.

To show the similarity between tapering data noise or ensemble anomalies (18) and localizing covariances at each local update, we write covariance localization (analogy to Eq. 11) at each local update as

$$\begin{aligned}\delta\mathbf{m}_{(i)} &= \left[\boldsymbol{\rho}_{\mathrm{md}\,(i)}\circ\left(\Delta\mathbf{M}_{(i)}\Delta\mathbf{D}^{\mathrm{T}}_{(i)}\right)\right]\\ &\quad\times\left(\mathbf{I}_{N_{\mathrm{d}_i}}+\left[\boldsymbol{\rho}_{\mathrm{dd}\,(i)}\circ\left(\Delta\mathbf{D}_{(i)}\Delta\mathbf{D}^{\mathrm{T}}_{(i)}\right)\right]\right)^{-1}\delta\mathbf{d}_{(i)}\,.\end{aligned} \tag{20}$$

Defining

$$\boldsymbol{\rho}_{\mathrm{md}\,(i)} = \boldsymbol{\rho}^{\mathrm{T}/2}_{(i)} \tag{21}$$

and

$$\boldsymbol{\rho}_{\mathrm{dd}\,(i)} = \boldsymbol{\rho}^{1/2}_{(i)}\boldsymbol{\rho}^{\mathrm{T}/2}_{(i)}\,, \tag{22}$$

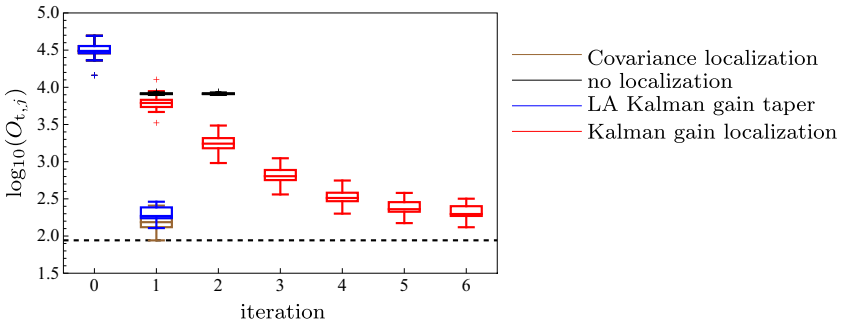


Fig. 3 Reduction of $\log_{10}$ transformed total objective function $O_{\mathrm{t},j}$ with iteration for three cases with localization and one case without localization. For all cases, the truncation level of TSVD is 100%. Each box shows distribution of 20 realizations. The *horizontal dashed line* shows the mean of $\log_{10}(O_\mathrm{t})$ from 20 posterior realizations computed using theoretical model covariance and true data sensitivity

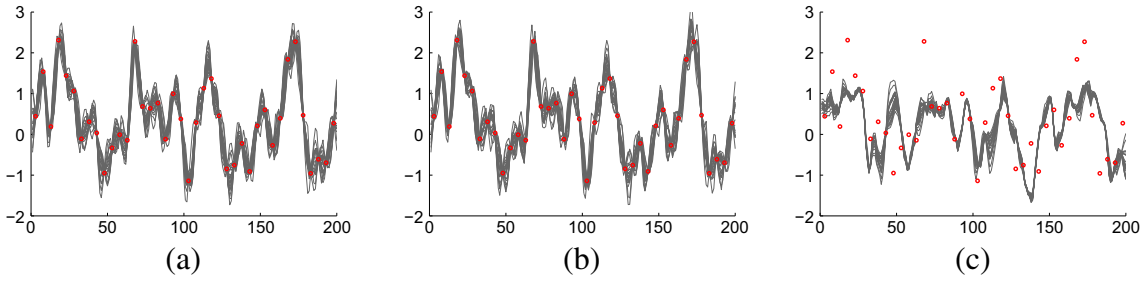


Fig. 4 The final ensemble for cases: **a** Kalman gain localization, **b** local analysis Kalman gain taper, and **c** no localization. The *black curves* show the ensemble of 20 realizations and the *red dots* are observed data $\mathbf{d}_{\text{obs}}$

we see that Eq. 20 is identical to Eq. 18 except for the scaling of $\delta\mathbf{d}_{(i)}$ with $\boldsymbol{\rho}_{(i)}^{1/2}$. Similar observations has been made by Sakov and Bertino [29]. The final updating equation for local analysis with observation taper (19) is, however, different from Eq. 20 because of the truncated singular value decomposition in Eq. 19.

5 Illustrative examples

Three data assimilation examples are used in this section to illustrate the behavior of the various localization methods on problems with different features. The first problem examines behavior of iterative assimilation of large amounts of linear local observations and the effect of truncation level of the TSVD. The second problem investigates the effect of the range of the taper function for various localization methods for assimilation of linear nonlocal observations. The third problem is a history matching problem with production data from 25 repeated 5-spot patterns from Chen and Oliver [3]. Because it is not computationally feasible to use covariance localization for history matching problems with large amounts of data, only the three localization methods shown in Table 1 are tested for most of the examples.

The performance of each method is evaluated based on the magnitude of the data mismatch, magnitude of the total objective function, and error in the posterior variance. The squared data mismatch O_{d} for each model realization j is defined as

$$O_{\text{d},j} = (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m}_j))^{\text{T}} \mathbf{C}_{\text{D}}{}^{-1} (\mathbf{d}_{\text{obs},j} - \mathbf{g}(\mathbf{m}_j)) \,. \tag{23}$$

This is the most common measure that is used in history matching problems. The changes from the initial/prior model realization $\mathbf{m}_{\text{pr},j}$ (squared model mismatch) is defined as

$$O_{\text{m},j} = \left(\mathbf{m}_{\text{pr},j} - \mathbf{m}_j\right)^{\text{T}} \mathbf{C}_{\text{M}}{}^{-1} \left(\mathbf{m}_{\text{pr},j} - \mathbf{m}_j\right) \,, \tag{24}$$

where $\mathbf{C}_{\text{M}}$ is the covariance of the prior model parameters. The total objective function O_{t} measures the quality of each model realization in terms of both data mismatch and model mismatch:

$$O_{\text{t},j} = \text{O}_{\text{d},j} + O_{\text{m},j} \,. \tag{25}$$

This total objective function is the cost function that LM-EnRML aims at minimizing [5]. Loosely, it requires that the updated model should stay as close as possible to the initial model unless changes are required to match data. The total objective function is only computed for the toy problems for which the theoretical $\mathbf{C}_{\text{M}}$ is used. For large-scale problems, computing an approximation of $\mathbf{C}_{\text{M}}{}^{-1}$ is not easy unless $\mathbf{C}_{\text{M}}$ takes certain special forms (e.g., [24, 31]). In most cases, however, when inverting $\mathbf{C}_{\text{M}}$ is necessary a pseudo inverse of the low rank ensemble approximation of $\mathbf{C}_{\text{M}}$ is used,

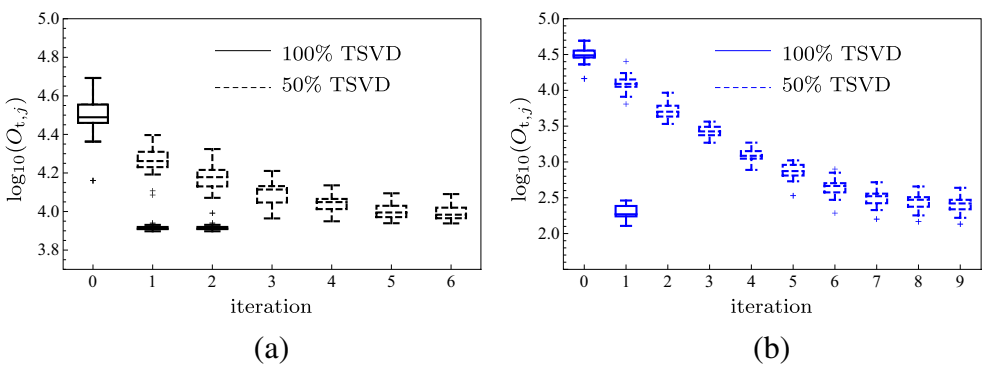


Fig. 5 Reduction of $\log_{10}$ transformed total objective function $O_{\text{t},j}$ with iteration for two cases: **a** no localization. **b** local analysis Kalman gain taper

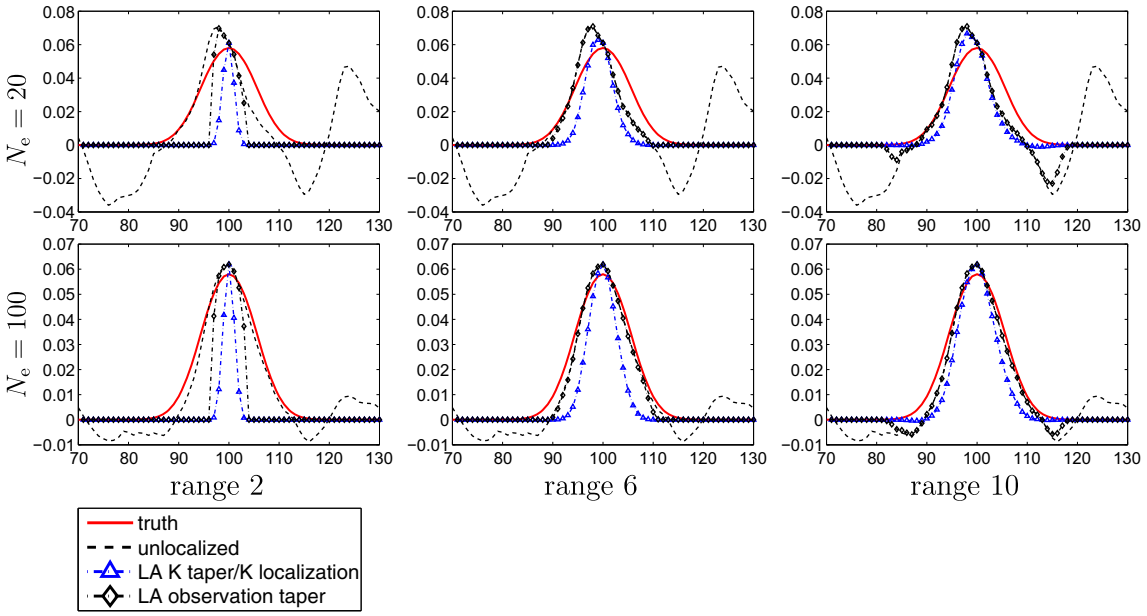


Fig. 6 Localized and unlocalized Kalman gain from an ensemble of size 20 (*top row*) and of size 100 (*bottom row*) for a 1D problem with one linear nonlocal data (27). The true Kalman gain is shown in *solid red*. Only model section from gridblock 70 to 130 is shown

despite the limitations of that approach. Symbols O_d, O_m, and O_t denote the ensemble mean of $O_{d,j}$, $O_{m,j}$, and $O_{t,j}$ respectively.

In addition, we use a separate measure O_c to quantify the goodness of the final variance after conditioning

$$O_c = (\mathbf{S}_t - \mathbf{S}_e)^T(\mathbf{S}_t - \mathbf{S}_e)\,, \tag{26}$$

where $\mathbf{S}_t$ is a vector containing the true standard deviation of the model parameters after conditioning to data and the $\mathbf{S}_e$ is the vector of standard deviations of the final ensemble. It is only possible to compute the analytical posterior covariance of model parameters for the toy examples. For the flow example, the final variability of the ensemble is benchmarked against a run with a large ensemble.

The termination criteria used in the examples for the approximate form of LM-EnRML are (1) maximum number of iterations equal to 20, (2) maximum number of inner iterations (tuning of λ) equal to 3, (3) minimum reduction of the ensemble mean of data mismatch (O_d) equal to 5%, and (4) minimum ensemble mean of data mismatch (O_d) equal to the number of data N_d. An iteration is accepted if the magnitude of O_d is reduced from the previous iteration. Although for the toy examples it would be possible to base the convergence criteria on the mean of the total objective function O_t, the acceptance criterion is based on O_d since that is what can be used in typical history matching problems.

Chen and Oliver [5] showed the behavior of O_d and O_m with iteration for a 1D multiphase flow problem. The model mismatch O_m is zero for the initial ensemble and increases with iteration. The data mismatch is typically large in magnitude for the initial ensemble, and decreases with iteration. When the approximate form of LM-EnRML is used and the reduction in data mismatch is used to terminate the iteration, there is a danger of over-fitting data. In this case, at later iterations, the data mismatch O_d may continue to reduce while the total objective function O_t increases. In the iteration

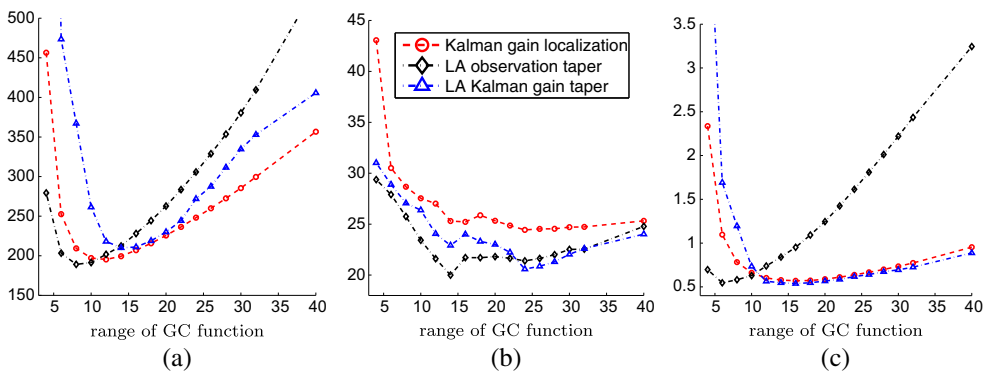


Fig. 7 Total objective function $\bar{O}_t$ (**a**), data mismatch $\bar{O}_d$ (**b**) and error in standard deviation $\bar{O}_c$ (**c**) of the final ensemble with different ranges used for the Gaspari-Cohn taper function. $\bar{O}_t$, $\bar{O}_d$, and $\bar{O}_c$ are averages of O_t, O_d, and O_c over 40 independent runs. (32 linear nonlocal data problem with $N_e = 20$)

Table 2 Localization range that gives the lowest $\bar{O}_t$ and $\bar{O}_c$ for cases shown in Fig. 7

	Optimal range	O_t
Kalman gain localization	12	195 ± 28
LA observation taper	8	189 ± 30
LA Kalman gain taper	14	210 ± 31

The mean and standard deviation of O_t for each localization range are shown before and after the $\pm$ sign. As a comparison, the theoretical mean and standard deviation of O_t is 66 and 9. (32 linear nonlocal data problem with $N_e = 20$)

termination criteria listed above, the minimum reduction of O_d is set relatively loosely and an additional termination criterion based on O_d less than the number of data N_d is added to reduce the danger of over-fitting.

5.1 Linear local data

The model parameters are spatially correlated Gaussian random variables at 200 discrete gridblocks. The prior distribution of the model parameters is multivariate Gaussian with zero mean and exponential covariance. The exponent, variance, and range for the covariance are equal to 1.9, 1, and 10 gridblocks, respectively. As a test for assimilating linear and local data, model parameters are observed at 40 evenly spaced locations with standard deviation of data noise being 0.05. The ensemble size is 20, which is less than the number of data (equal to 40). The Furrer-Bengtsson (FB) function with the model covariance and the ensemble size as input was used for tapering for all the localization methods. This choice of taper function is optimal for covariance localization with linear local data. The normalized FB function, for which the value of the function is equal to one at distance equal to zero, is used as suggested in Furrer and Bengtsson [14].

Figure 3 shows reduction of $\log_{10}$ transformed $O_{t,j}$ with iteration for different localization methods and for a case without using localization. The truncation level of TSVD is 100 % for all cases (meaning that all nonzero singular values are kept in SVD of Eqs. 8, 15, and 17). The horizontal dashed line in Fig. 3 shows the mean of $\log_{10}(O_{t,j})$ from 20 posterior realizations computed using the randomized maximum likelihood (RML) method [25] with the theoretical model covariance and true data sensitivity. The prior realizations and the observed data realizations are the same as those used for the ensemble method. For this linear problem, LM-EnRML approx with covariance localization converges in one iteration. Note that TSVD is not used with covariance localization (see Eq. 11). With Kalman gain localization, TSVD is applied to the data deviation ensemble consisting of all 40 data ($\Delta\mathbf{D}$ of size 40×20). Although all 19 nonzero singular values are kept, there are not enough degrees of freedom to match all 40 data at 1 iteration, and additional iterations are needed to obtain similar level of data match as covariance localization. For local analysis, an ensemble size of 20 provides enough degrees of freedom at each local update to match data, as a result the case with LA Kalman gain taper converges in a couple of iterations. The results with LA observation taper are very similar to those with LA Kalman gain taper and are not shown here. The final ensemble for the case with Kalman gain localization and LA Kalman gain taper are shown in Fig. 4a, b, in which data are matched well and the spatial continuity of the initial realizations are maintained.

When localization is not used, the ensemble variability collapsed quickly and the data mismatch stayed high at the final iteration (see Fig. 3). It is possible to reduce the rate of ensemble collapse by truncating small singular values, for example Fig. 5a compares the reduction of $O_{t,j}$ with iteration using 100 and 50% truncation for the TSVD for the case without localization. When the truncation level in TSVD is 50% and localization is not used, the quality of data match is poor and the ensemble variability is too small (Fig. 4c). When the number of independent data is large, regularization through TSVD alone is not sufficient to improve the quality of data match and localization is necessary.

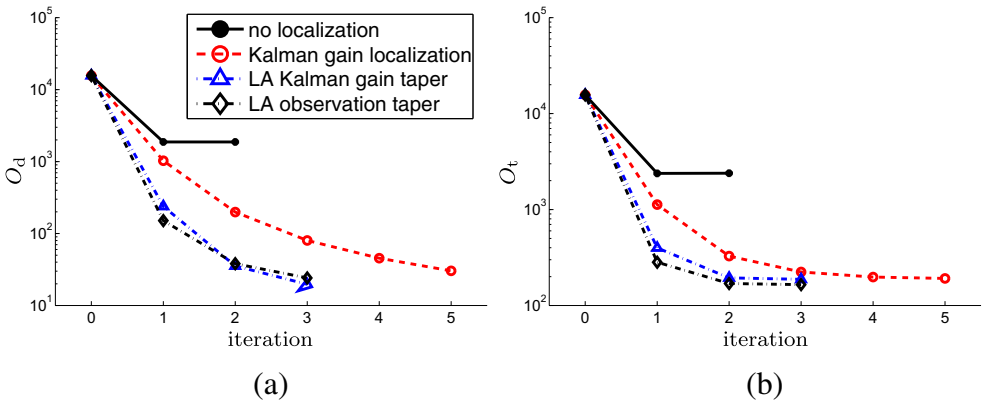


Fig. 8 The change of O_d and O_t with iteration for 1 of the 40 runs for the case with 32 linear nonlocal data with $N_e = 20$. The localization range that gave the best O_t in Table 2 for each localization method are chosen, i.e., 12 gridblocks for Kalman gain localization (*red circle*); 8 gridblocks for LA observation taper (*black diamond*); 14 gridblocks for LA Kalman gain taper (*blue triangle*)

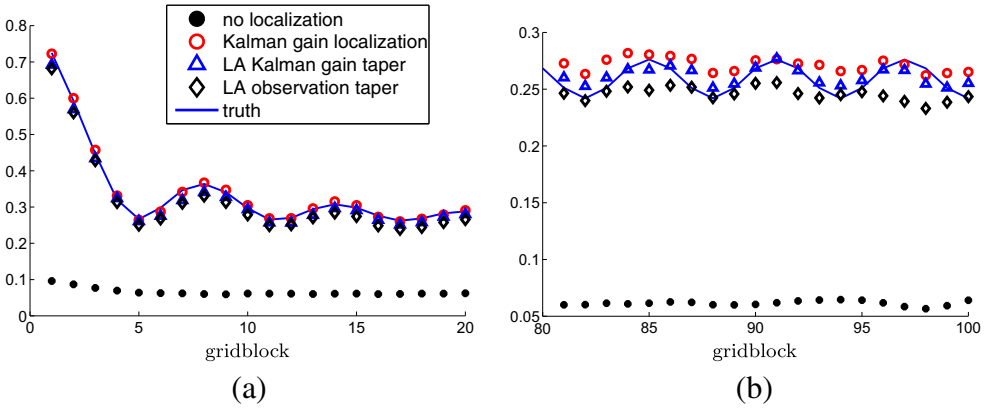


Fig. 9 The average of the standard deviation of the final ensemble ($\mathbf{S}_e$ in Eq. 26) over the 40 runs for the same 4 cases as shown in Fig. 8. The theoretical standard deviation ($\mathbf{S}_t$ in Eq. 26) is shown in *solid blue line* for comparison. Two representative model sections are shown, from gridblock 1 to 20 in (**a**) and from gridblock 81 to 100 in (**b**). (32 nonlocal linear data with $N_e = 20$)

The Levenberg-Marquardt tuning parameter λ is set to be zero at all iterations for results shown above. The effect of using nonzero λ to regularize the iterations is similar to the effect of truncating small singular values in the TSVD. Although not important for this simple linear problem, regularization through both λ and TSVD is necessary for inverse problems that are highly nonlinear. Li et al. [21] showed that at early iterations of typical history matching problems when data mismatch is large, it is beneficial to use a large value of λ to reduce erroneously large updates to the model and to improve the final data match. Evensen [11] Chen and Oliver [4] both showed benefit from using TSVD for nonlinear data assimilation problems. If 50 % TSVD is used with local analysis Kalman gain taper, a similar level of data match is achieved as the case with 100 % TSVD but with many more iterations as shown in Fig. 5b. This suggests that for typical history matching problems when regularization is important in addition to localization, even if a conservative truncation level is used for TSVD, satisfactory results can be achieved with additional iterations.

5.2 Linear nonlocal data

The model setup in this subsection is the same as in Section 5.1 except that the data are nonlocal. The behavior of the localized Kalman gain for nonlocal observations is first investigated using a case with only one nonlocal data measuring the average of model parameters from gridblock 95 to 105

$$g(\mathbf{m}) = \frac{1}{11}\sum_{i=95}^{105} m_i \,, \tag{27}$$

where m_i denotes the ith element of model parameters $\mathbf{m}$. The standard deviation of data noise is 0.05. The center of the region of nonzero sensitivity (at m_{100}) is used as the data location for localization. Gaspari-Cohn (GC) is used as the taper function for all localization methods.

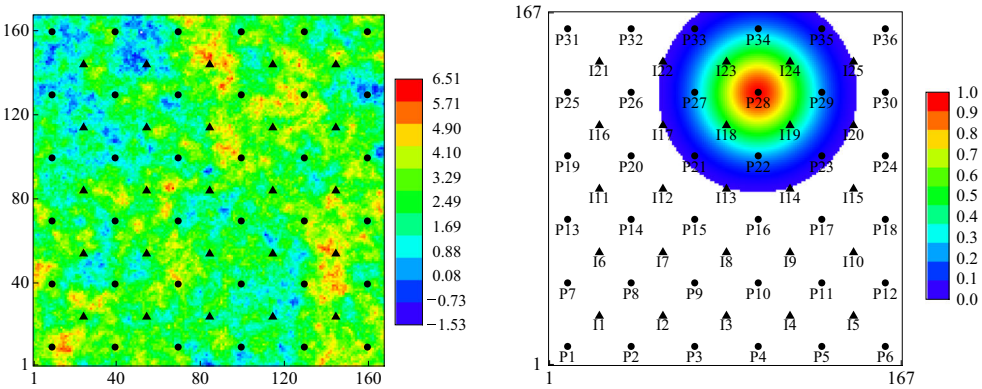
The iteration tuning parameter λ is set to be zero at all iterations and all nonzero singular values are kept (there is only one nonzero singular value for this one-data case). The unlocalized Kalman gain (black dashed), the localized Kalman gain from LA Kalman gain taper (blue triangle), and LA observation taper (black diamond) at the first iteration are compared with the theoretical Kalman gain (red solid) in Fig. 6. Each column of Fig. 6 shows results from using a different range for the GC function. The top row shows results from an ensemble of size 20, and the bottom row shows results from an ensemble size of 100.

The definition of the localized Kalman gain for Kalman gain localization and LA Kalman taper is defined in Eqs. 13

Table 3 Results for the cases shown in Figs. 8 and 9 in terms of the total number of iterations used, and the value of O_d, O_t, and O_c at the final iteration

	# of iteration	O_d	O_t	O_c
No localization	2 ± 0	1455 ± 723	2212 ± 820	10.4 ± 0.28
Kalman gain localization	5 ± 0.8	27 ± 3	195 ± 28	0.6 ± 0.15
LA observation taper	3 ± 0.7	26 ± 4	189 ± 30	0.6 ± 0.13
LA Kalman gain taper	3 ± 0.6	23 ± 5	210 ± 31	0.5 ± 0.13

The numbers shown before and after the ± sign are the mean and the standard deviation from the 40 runs, respectively. (32 nonlocal linear data with $N_e = 20$)



(a) (b)

Fig. 10 **a** The true log-permeability field from which synthetic historical data are generated. **b** GC taper function with range equal to 30 gridblocks evaluated with distance computed to the location of P28. Note that values less than 0.01 are truncated to zero. Locations of producers are shown as *circles* and injectors are shown as *triangles*

and 16, respectively. The localized Kalman gain for LA observation taper is defined as

$$\mathbf{K}_{(i)}^{\text{loc}} = \Delta\mathbf{M}_{(i)}\Delta\mathbf{D}_{(i)}^{\rho\,\mathrm{T}}\mathbf{U}_{p_i}^{\rho}(\mathbf{I}_{p_i} + \mathbf{W}_{p_i}^{\rho 2})^{-1}\mathbf{U}_{p_i}^{\rho\,\mathrm{T}}\text{diag}(\boldsymbol{\rho}_{(i)}^{1/2}), \tag{28}$$

where diag(**x**) represents a diagonal matrix with vector **x** on the diagonal. For problems with a single observation, Kalman gain localization and LA Kalman taper are equivalent. The localized Kalman gain shown in Fig. 6 for local analysis is a composite from all local updates. The localized Kalman gain from LA Kalman taper and Kalman gain localization is consistently narrower than that from LA observation taper with the same localization range, meaning that with the same localization range LA observation taper results in less localization. This is consistent with what has been observed with local data [16, 22].

The sensitivity of the quality of the updated ensemble, quantified by total objective function, data mismatch, and posterior variance (defined in Eqs. 25, 23, and 26), with respect to the range of the taper function is tested through a problem of assimilating 32 linear nonlocal data. Each data is an average of model parameters over 11 gridblock as in Eq. 27. The centers of the averaging are from gridblocks 7, 13, to 193, which gives 32 nonlocal data with overlapping sensitivity regions. This is designed to mimic some aspects of the behavior of data in typical history matching problems. Figure 7 plots total objective function $\bar{O}_\text{t}$ (first column), data mismatch $\bar{O}_\text{d}$ (second column), and error in standard deviation $\bar{O}_\text{c}$ (third column) of the final ensemble with different ranges used for the Gaspari-Cohn taper function. A wide range of choices for the localization range (from 4 gridblocks to 40 gridblocks) is tested in Fig. 7. The values $\bar{O}_\text{t}$, $\bar{O}_\text{d}$, and $\bar{O}_\text{c}$ in Fig. 7 are averages of O_t, O_d, and O_c from 40 independent runs with ensemble size of 20. O_t, O_d, and O_c are computed at the last iteration of each run. Both the initial ensemble and the synthetic truth are generated independently among the 40 runs, but they are the same for all the cases evaluated (different localization methods and no localization).

Since the total objective function O_t measures the quality of sampling of the posterior pdf using the LM-EnRML, it is a comprehensive measure for the quality of the final ensemble. The magnitude of data mismatch O_d and error in final variance O_c provide information on two specific aspects of the final ensemble. If a total objective function less than 250 is considered acceptable, we see that there is a relatively wide range of choices for the range of the GC taper for all three localization methods (Fig. 7a). Table 2 shows the range that results in the lowest average of O_t in Fig. 7 for each localization method, and the mean and standard deviation (before and after the plus/minus sign) of O_t from the 40 runs with the optimal range used for localization. Note that O_t is the mean of 20 realizations from each run.

For nonlocal observations, the range of localization is generally chosen based on the region of sensitivity and

Table 4 Description of five cases that are used to evaluate performance of various localization methods

Case	N_e	loc. method	Range of GC	TSVD	# of iter
1	100	None	–	95 %	7
2	100	Kalman localization	35	95 %	19
3	100	LA observation taper	20	95 %	11
4	100	LA Kalman gain taper	30	95 %	13
5	2000	None	–	99 %	9

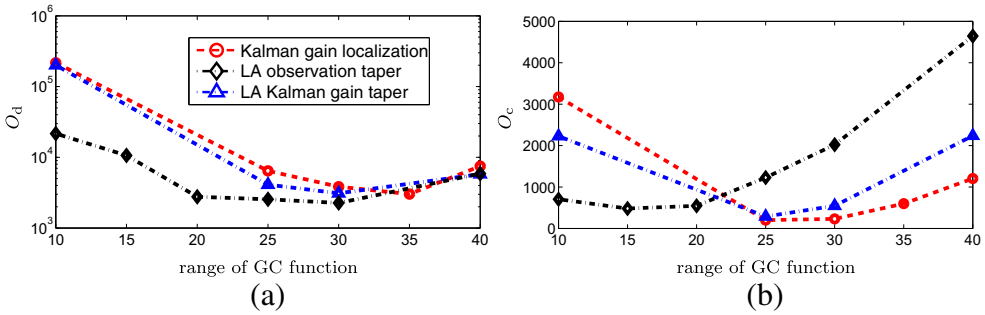


Fig. 11 The data mismatch O_d (**a**) and error in standard deviation O_c (**b**) of the final ensemble with different ranges used for the Gaspari-Cohn taper function for the flow example

on the range of the prior covariance [3]. In this example, the range of the prior covariance is 10 gridblocks and the distance to the edge of the region of averaging from the observation point is five gridblocks, so the cross covariance has a range of approximately 15 gridblocks, which would be the suggested localization range for Kalman gain localization [7]. Based on the optimal ranges shown in Table 2, the guidelines work well for Kalman gain localization and for local analysis with Kalman gain taper. It appears, however, that a somewhat shorter range should be used for local analysis with observation taper (consistent with what is shown in Fig. 6).

Just to benchmark the values of O_t shown in Table 2, we computed posterior realizations using the theoretical model covariance $\mathbf{C}_M$ and the true data sensitivity for 40 groups of 20 realizations. The synthetic truth is generated independently among the 40 groups, and simulated data from the true model is then perturbed to create 20 realizations of the observed data for conditioning the prior realizations (this is to mimic what is done in the ensemble based method). The resulting mean and standard deviation of O_t from the 40 groups are 66 and 9, respectively. Although the results with $N_e = 20$ with localization are still quite high compared to the theoretical values, the updated model realizations (not shown here) all appear quite satisfactory. If the truth is randomized in creating each of the posterior realization, the resulting $O_{t,j}$ from the posterior realizations follows a χ^2 distribution with mean equal to $2N_d$ [25].

Figure 8 shows the change of O_d and O_t with iteration for 1 of the 40 runs. Three cases with different localization methods and one case without using localization are compared. The localization range that gave the best O_t in Table 2 for each localization method are chosen, i.e., 12 gridblocks for Kalman gain localization (red circle); 8 gridblocks for LA observation taper (black diamond); 14 gridblocks for LA Kalman gain taper (blue triangle). Because of the fast reduction in ensemble variability and inability to further reduce O_d, the case without localization terminated after two iterations. Both local analysis methods converged faster than Kalman gain localization because TSVD of the local data ensemble in local analysis methods ($\Delta\mathbf{D}_{(i)}$ and $\Delta\mathbf{D}^{\rho}_{(i)}$ in Eqs. 15 and 17) removes less information than truncation of the singular values of the global data ensemble ($\Delta\mathbf{D}$ in Eq. 8) in Kalman gain localization.

Figure 9 shows the average of the standard deviation of the final ensemble ($\mathbf{S}_e$ in Eq. 26) over the 40 runs for the same 4 cases as shown in Fig. 8. The theoretical standard deviation ($\mathbf{S}_t$ in Eq. 26) is shown in Fig. 9 in solid blue line for comparison. Two representative model sections are shown, from gridblock 1 to 20 in (a) and from gridblock 81 to 100 in (b). The variability of the final ensemble without using localization (black dots) is clearly too low, while the variability of the final ensemble from the three localization methods are comparable to the theoretical values.

Table 3 summarizes results for the cases shown in Figs. 8 and 9 in terms of the total number of iterations used, and the value of O_d, O_t, and O_c at the final iteration. The numbers shown before and after the plus/minus sign are the mean and the standard deviation from the 40 runs, respectively. The rate of convergence is on average two iterations slower for Kalman gain localization compared to the two local analysis methods. All localization methods show acceptable results in terms of the three measures, while the case without localization clearly fail to obtain reasonable results.

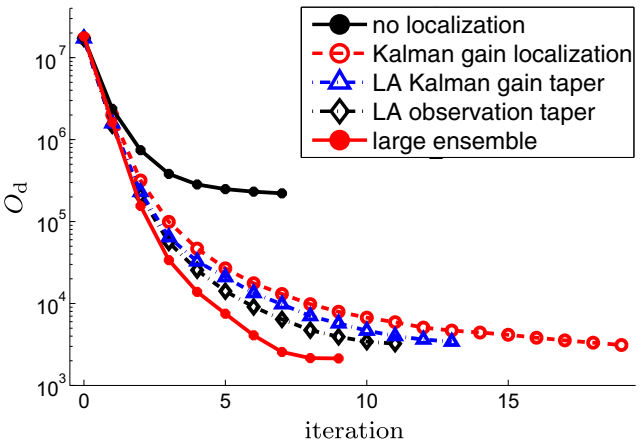


Fig. 12 The change of ensemble mean of the data mismatch O_d with iteration for the five cases shown in Table 4

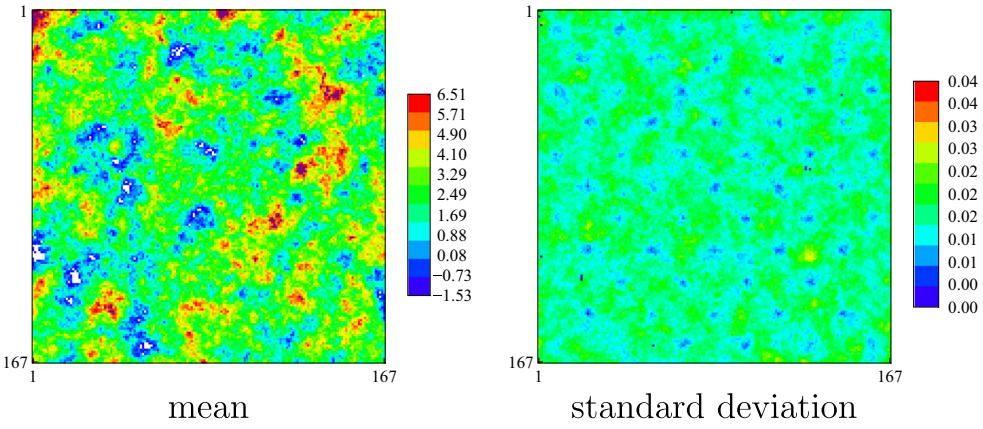


Fig. 13 Mean and standard deviation of the log-permeability ensemble at the final iteration (iteration 7) when no localization was used with $N_e = 100$ (case 1). The initial mean and std of log-permeability are 2.5 and 1.2, respectively

5.3 Flow example with 25 well patterns

The same synthetic 2D water flooding example used in Chen and Oliver [3] is used here to test the various localization methods. The simulation model has 167 × 167 gridblocks. The natural log transformed permeability (log-permeability) is assumed to follow a multivariate Gaussian distribution with mean 2.5 and standard deviation 1.2. The covariance function is exponential with exponent equal to 1 and range equal to 26 gridblocks. The true log-permeability field, which is used to generate synthetic data, is shown in Fig. 10a. There are 25 injectors (locations shown as triangles) and 36 producers (shown as circles). The injectors are controlled by constant water injection rate and the producers are controlled by constant bottom-hole pressure (BHP). Historical data consist of water and oil production rates and water injection rate and BHP at nine different times. In total, the number of data points is 1098 ($N_d = 1098$). The standard deviation of data noise is 8 bbl/day for both oil and water production rate, 5 bbl/day for water injection rate, and 10 psi for BHP. For this case, the only uncertain parameters are permeability at each model cell, meaning that the ensemble of reservoir models are identical to the true model except for the permeability map.

The Gaspari-Cohn (GC) taper function was used for localization. The data location for localization of all data types is the well location of the measurement, and the same localization range is used for all data. Figure 10b shows the GC taper function with range equal to 30 gridblocks evaluated with distance computed to the location of producer 28. It might be beneficial to use different tapering matrix for data obtained at different times at the same location (same well) since data sensitivity changes with time. It is, however, generally sufficient to disregard this time dependency if the localization region is chosen large enough for the largest region of data sensitivity (i.e., after water breakthrough).

Table 4 gives the description of the five cases that were run for comparison. The initial permeability ensemble is the same for cases 1 to 4. Case 5 with a large ensemble size without any localization is used as a benchmark. The localization range was chosen based on a joint consideration of the range of the prior covariance and the well pattern, but a few trials were also made to check the sensitivity and the limiting effect if the range is chosen to be too small

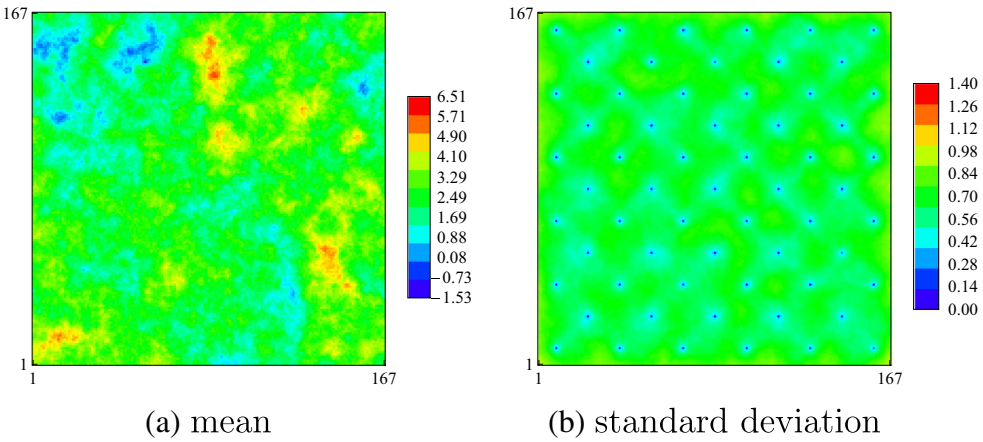


Fig. 14 Mean and standard deviation of the log-permeability ensemble at the final iteration (iteration 9) with $N_e = 2000$. No localization is used

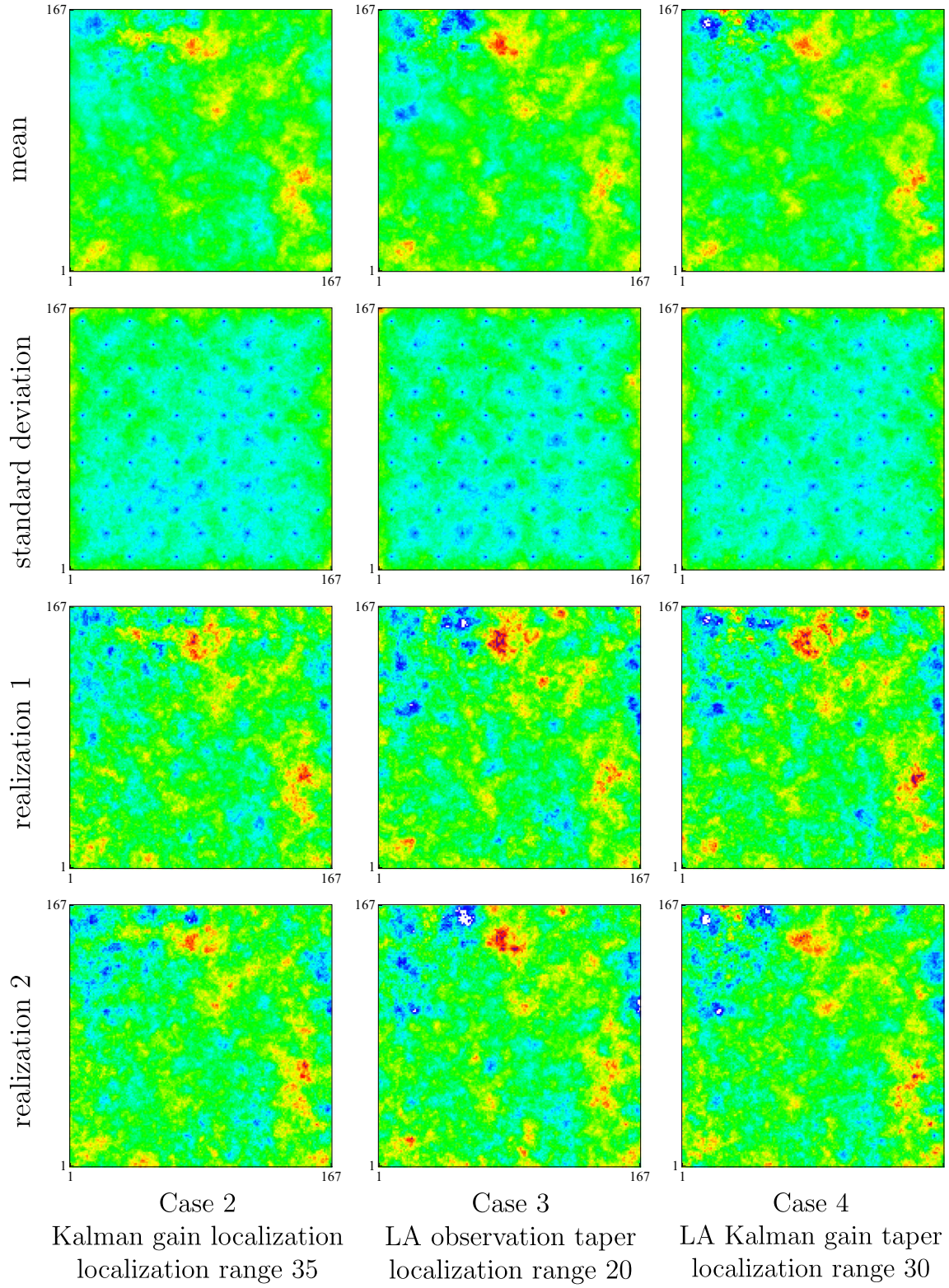


Fig. 15 Mean (*first row*) and standard deviation (*second row*) of the final log-permeability ensemble and two final realizations (*last two rows*). The same scale as in Fig. 14a is used for mean and realizations. The same scale as in Fig. 14b is used for standard deviation.

or too large. Figure 11 shows the data mismatch O_d and error in final variance O_c for various localization ranges. In this example, the theoretical variance is not possible to obtain so that the final standard deviation from case 5 (large ensemble size) is used as $\mathbf{S}_t$ in Eq. 26 to compute O_c. For all localization methods, if the range for localization

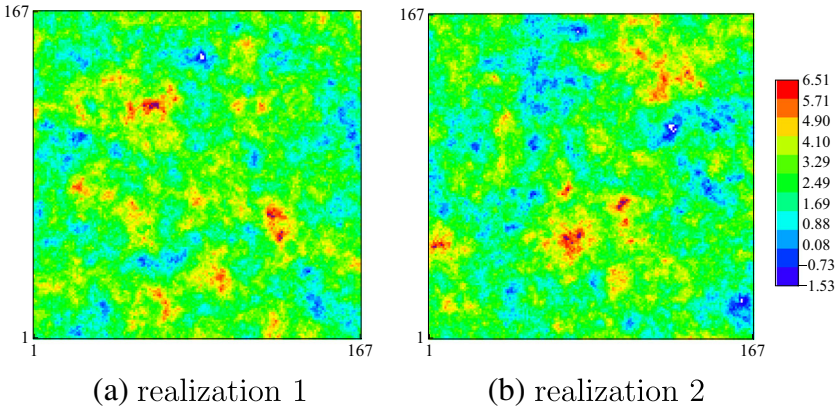


Fig. 16 The initial log-permeability realizations corresponding to the two final ones shown in the *bottom two rows* of Fig. 15

is chosen to be too small, the ensemble of updated property fields will often have artificially large variability and the data match will often be poor. If the range for localization is chosen to be too large, the ensemble may collapse and again the data match may be poor. The range of tapers for which results are acceptable in this example, i.e., 25 to 35 gridblocks for Kalman gain localization and local analysis Kalman gain taper, and 20 to 25 gridblocks for local analysis observation taper, is a substantial fraction of the interwell spacing (approximately 20 gridblocks). Similar to what was observed in Section 5.2, a considerably smaller range is needed for LA observation taper compared to the other two localization methods in order to obtain a similar level of final ensemble variance. Based on the empirical formula shown in Chen and Oliver [5], the initial value of the Levenberg-Marquardt parameter λ is 10,000 for all cases. The iteration termination criteria are the same as stated in the beginning of Section 5.

Figure 12 shows the ensemble mean of the data mismatch O_d with iteration for the five cases. When no localization is used with a typical ensemble size of 100, the iterative ensemble smoother failed to reduce data mismatch to an acceptable level. The mean and standard deviation of the final ensemble at iteration 7 for case 1 without using localization are shown in Fig. 13. Typical signs of sampling errors and insufficient degrees of freedom such as strong updates to model parameters and large reduction in ensemble variability can be observed for the final ensemble of case 1. The mean and standard deviation of the final ensemble for the case with a large ensemble size (case 5) are shown in Fig. 14, in which the mean is smooth and the reduction of standard deviation is mainly at locations of high data sensitivity.

As shown in Fig. 12, the data match at the final iteration from an ensemble of size 100 with different localization methods (cases 2, 3, and 4) are much better than the case without localization (case 1) and are not far from the case with a large ensemble size (case 5). The mean and standard deviation of the final ensemble for cases 2, 3, and 4 (shown in Fig. 15) are also comparable to the case with a large ensemble (case 5). Two final realizations from each of the localization methods are also shown in the bottom two rows of Fig. 15 with their corresponding initial realizations shown in Fig. 16 for comparison. The final history matched realizations show similar spatial continuity as the initial ones. Similar to what have been observed with the toy examples, local analysis methods (cases 3 and 4) converged faster than the case with Kalman gain localization (case 2).

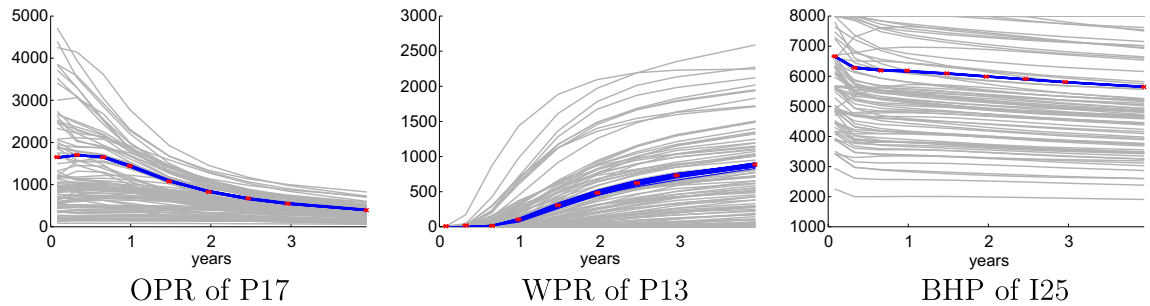


Fig. 17 Simulated data at three wells from the initial ensemble and the final ensemble for the case using LA Kalman gain taper (Case 4). Simulated data from the initial models are shown in *gray*, and simulated data from the final models (at iteration 13) are shown in *blue*. Observed data are shown as *red dots* with the error bar indicating one standard deviation of the data noise. The unit for rates is bbl/day, and the unit for pressure is psi

Figure 17 shows simulated data at three wells from the initial ensemble (gray) and the final ensemble (blue) at iteration 13 for the case using LA Kalman gain taper (case 4). The data match from other localization cases are very similar and are not shown here. As an aside, much better data matches were achieved at the final iteration with all the localization methods using the iterative ensemble smoother than those obtained using the ensemble Kalman filter (EnKF) with covariance localization shown in Chen and Oliver [3]. Although the taper function is slightly different in the two studies, the main reason for the improved data match with the iterative ensemble smoother is that EnKF without iteration did not handle nonlinearity correctly.

6 Conclusions

The ensemble-based methods with a small ensemble size require regularization and localization in order to obtain satisfactory results for problems with large amounts of model parameters and large amounts of independent data. The state-of-the-art approach in ensemble-based history matching typically uses truncated singular value decomposition (TSVD) for regularization and uses localization of the Kalman gain matrix to alter the subspace for updating different model parameters and to reduce spurious updates.

In this paper, we investigated the use of local analysis with the iterative ensemble smoother for history matching problems and compared its performance with Kalman gain localization. Two local analysis methods were investigated: one with tapering of the simulated data ensemble (or equivalently the observation noise) as typically used in numerical weather prediction community, and the other with tapering of the local Kalman gain matrix. The conclusions are the following:

1. Truncation of the SVD before inversion provides useful regularization, but TSVD alone is typically not sufficient to allow enough degrees of freedom to assimilate large amounts of data even if iteration is used.
2. When Kalman gain localization is used without local analysis, the subspace spanned by the singular vectors of TSVD may be too small to represent the data, but good assimilation results can be obtained when an iterative ensemble smoother is used.
3. The use of iteration compensates for strong truncation of the singular values in TSVD, thus the truncation level is not as important in iterative methods as it is for non-iterative methods
4. Good results can be obtained for all localization methods, if the localization range is chosen well. In general, Kalman gain localization and local analysis with Kalman gain taper have similar optimal localization ranges, while the optimal range is somewhat shorter for local analysis with observation taper. In most examples that we investigated, the range of appropriate values is relatively large so selection of a range to use is not always difficult.
5. Although all localization methods can give equivalent results if used iteratively, the local analysis methods generally converge more quickly than Kalman gain localization when the amount of data is large compared to ensemble size.

Acknowledgments The authors would like to thank Total for the permission to publish this work. Primary support for the second author has been 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).

References

1. Annan, J.D., Hargreaves, J.C., Edwards, N.R., Marsh, R.: Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. Ocean Model. **8**(1–2), 135–154 (2005). doi:10.1016/j.ocemod.2003.12.004. ISSN 1463–5003
2. Chen, Y., Oliver, D.S.: Ensemble-based closed-loop optimization applied to Brugge Field. SPE Reserv. Eval. Eng. **13**(1), 56–71 (2010a). doi:10.2118/118926-PA
3. Chen, Y., Oliver, D.S.: Cross-covariances and localization for EnKF in multiphase flow data assimilation. Comput. Geosci. **14**, 579–601 (2010b). doi:10.1007/s10596-009-9174-6. ISSN 1420-0597
4. 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. ISSN 1874-8961
5. 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. ISSN 1420–0597
6. Chen, Y., Oliver, D.S.: History matching of the Norne full-field model with an iterative ensemble smoother. SPE Reserv. Eval. Eng. **17**(2), 244–256 (2014). doi:10.2118/164902-PA
7. Emerick, A., Reynolds, A.: Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications. Comput. Geosci. **15**(2), 251–269 (2011). doi:10.1007/s10596-010-9198-y. ISSN 1420-0597
8. Emerick, A.A.: Analysis of the performance of ensemble-based assimilation of production and seismic data. J. Petrol. Sci. Eng. Online first (2016)
9. Emerick, A.A., Reynolds, A.C.: Ensemble smoother with multiple data assimilation. Comput. Geosci. **55**, 3–15 (2013a). doi:10.1016/j.cageo.2012.03.011. ISSN 0098-3004
10. Emerick, A.A., Reynolds, A.C.: History-matching production and seismic data in a real field case using the ensemble smoother with multiple data assimilation, SPE-164902. In: Proc of SPE RSS. The Woodlands (2013b)
11. Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, 2n edn.. Springer Verlag (2009)

12. Fahimuddin, A., Aanonsen, S.I., Skjervheim, J.-A.: 4D seismic history matching of a real field case with EnKF: Use of local analysis for model updating. In: SPE Annual Technical Conference and Exhibition, 19-22 September 2010. Florence (2010)
13. Fertig, E.J., Hunt, B.R., Ott, E., Szunyogh, I.: Assimilating non-local observations with a local ensemble Kalman filter. Tellus A **59**(5), 719–730 (2007). doi:10.1111/j.1600-0870.2007.00260.x
14. 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. ISSN 0047-259X
15. Gaspari, G., Cohn, S.E.: Construction of correlation functions in two and three dimensions. Q. J. R. Meteorol. Soc. **125**(554), 723–757 (1999)
16. Greybush, S.J., Kalnay, E., Miyoshi, T., Ide, K., Hunt, B.R.: Balance and ensemble Kalman filter localization techniques. Mon. Weather Rev. **139**(2), 511–522 (2011). doi:10.1175/2010MWR3328.1. ISSN 0027-0644
17. Hamill, T.M., Whitaker, J.S., Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon. Weather Rev. **129**(11), 2776–2790 (2001)
18. Houtekamer, P.L., Mitchell, H.L.: A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. **129**(1), 123–137 (2001)
19. Houtekamer, P.L., Mitchell, H.L.: Data assimilation using an ensemble Kalman filter technique. Mon. Weather Rev. **126**(3), 796–811 (1998)
20. Hunt, B.R., Kostelich, E.J., Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. Physica D: Nonlin. Phenom. **230**(1–2), 112–126 (2007)
21. Li, R., Reynolds, A.C., Oliver, D.S.: History matching of three-phase flow production data. SPE J. **8**(4), 328–340 (2003). doi:10.2118/87336-PA
22. Miyoshi, T., Yamane, S.: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. Mon. Weather Rev. **135**(11), 3841–3861 (2007). doi:10.1175/2007MWR1873.1
23. Nerger, L., Janjić, T., Schröter, J., Hiller, W.: A regulated localization scheme for ensemble-based Kalman filters. Q. J. R. Meteorol. Soc. **138**(664), 802–812 (2012). doi:10.1002/qj.945
24. Oliver, D.S.: Calculation of the inverse of the covariance. Math. Geol. **30**(7), 911–933 (1998)
25. Oliver, D.S., Reynolds, A.C., Liu, N. Inverse Theory for Petroleum Reservoir Characterization and History Matching, 1st edn. Cambridge University Press, Cambridge (2008)
26. Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D.J., Yorke, J.A.: A local ensemble Kalman filter for atmospheric data assimilation. Tellus A **56**(5), 415–428 (2004)
27. Peters, L., Arts, R.J., Brouwer, G.K., Geel, C.R., Cullick, S., Lorentzen, R.J., Chen, Y., Dunlop, K.N.B., Vossepoel, F.C., Xu, R., Sarma, P., Alhutali, A.H., Reynolds, A.C.: Results of the Brugge benchmark study for flooding optimization and history matching. SPE Reserv. Evalu. Eng. **13**(3), 391–405 (2010). doi:10.2118/119094-PA
28. Sætrom, J., Hove, J., Skjervheim, J.-A., Vabø, J.G.: Improved uncertainty quantification in the ensemble Kalman filter using statistical model-selection techniques. SPE J. **17**, 152–162 (2012). doi:10.2118/145192-PA
29. Sakov, P., Bertino, L.: Relation between two common localisation methods for the EnKF. Comput. Geosci. **15**(2), 225–237 (2011). doi:10.1007/s10596-010-9202-6. ISSN 1420-0597
30. Skjervheim, J.-A., Evensen, G.: An ensemble smoother for assisted history matching. In: SPE Reservoir Simulation Symposium, 21–23 February. The Woodlands, doi:10.2118/141929-MS (2011)
31. Weaver, A.T., Mirouze, I.: On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. Q. J. R. Meteorol. Soc. **139**(670), 242–260 (2013). doi:10.1002/qj.1955. ISSN 1477-870X
32. Zhao, Y., Reynolds, A.C., Li, G.: Generating facies maps by assimilating production data and seismic data with the ensemble Kalman filter, SPE-113990. In: Proc of SPE IOR Symp. Tulsa, doi:10.2118/113990-MS (2008)