



Data-space inversion with ensemble smoother

Mateus M. Lima¹ · Alexandre A. Emerick¹ · Carlos E. P. Ortiz²

Received: 22 March 2019 / Accepted: 4 January 2020 / Published online: 5 March 2020
© Springer Nature Switzerland AG 2020

Abstract

Reservoir engineers use large-scale numerical models to predict the production performance in oil and gas fields. However, these models are constructed based on scarce and often inaccurate data, making their predictions highly uncertain. On the other hand, measurements of pressure and flow rates are constantly collected during the operation of the field. The assimilation of these data into the reservoir models (history matching) helps to mitigate uncertainty and improve their predictive capacity. History matching is a nonlinear inverse problem, which is typically handled using optimization and Monte Carlo methods. In practice, however, generating a set of properly history-matched models that preserve the geological realism is very challenging, especially in cases with intricate prior description, such as models with fractures and complex facies distributions. Recently, a new data-space inversion (DSI) approach was introduced in the literature as an alternative to the model-space inversion used in history matching. The essential idea is to update directly the predictions from a prior ensemble of models to account for the observed production history without updating the corresponding models. The present paper introduces a DSI implementation based on the use of an iterative ensemble smoother and demonstrates with examples that the new implementation is computationally faster and more robust than the earlier method based on principal component analysis and gradient-driven optimization. The new DSI is also applied to estimate the production forecast in a real field with long production history and a large number of wells. For this field problem, the new DSI obtained forecasts comparable with a more traditional ensemble-based history matching.

Keywords Data-space inversion · Uncertainty quantification · Ensemble smoother · History matching

1 Introduction

Reservoir characterization from static and dynamic data allows to create numerical models that can be used to simulate the performance of petroleum reservoirs under different operating conditions. These models are essential for efficient exploitation and management of oil and gas fields. The incorporation of static data is typically done

using geostatistics while the incorporation of dynamic data is done using assisted history-matching methods.

History matching is usually a very difficult task, which involves the integration of interdisciplinary teams and an intensive use of computational resources. A complete study may require months of work and the results are not always satisfactory. The task is even harder if one needs to provide uncertainty estimates, in which case several alternative history-matched models must be generated. In the last few decades, the advances in assisted (or semi-automatic) history-matching techniques were notorious. Yet, history matching remains one of the most time-consuming steps of a field study because the size and complexity of the models have also increased significantly in the same period. [39] present a review of the main history-matching methods proposed in the literature. Among these methods, the ones based on the ensemble Kalman filter (EnKF) [17, 19] have become quite popular, especially because of their ease of implementation and integration with commercial reservoir simulators and the ability to generate multiple models with large number of uncertainty parameters at an affordable computational cost. Despite the relative success

✉ Alexandre A. Emerick
aemerick@gmail.com

Mateus M. Lima
mateusmartins@yahoo.com.br

Carlos E. P. Ortiz
capico.LENEP@gmail.com

¹ Petrobras, Av. Horacio de Macedo 950, Cidade Universitaria, Rio de Janeiro, RJ, 21941-915, Brazil

² Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, Brazil

in a number of recent field cases reported in the literature; see, for example, [1, 7, 9, 16, 20, 30, 36], generating a set of models properly conditioned to all historical data and still preserving the geological realism is very challenging, especially in cases with intricate prior description, such as models with fractures and complex facies distributions.

A new approach known as data-space inversion (DSI) [51] has drawn attention in the literature as an alternative to the model-space inversion approach used in history matching. The basic idea behind DSI is to update directly the predictions from a prior ensemble of models to account for the observed production history without updating the corresponding models. The upside of this approach is to be able to provide an ensemble of forecasts without going through the time-consuming history-matching step. Because there are no model inversions, there are no concerns about losing the geological realism. The downside is that DSI provides only the forecast estimates for a fixed production strategy. In practice, however, reservoir engineers may also be interested in having the corresponding models to study different drainage strategies. For this reason, data- and model-space inversions should be considered complementary rather than alternative (or even competing) approaches.

The DSI method introduced by [51] uses principal component analysis (PCA) to reparameterize the predicted data from the prior ensemble into a lower dimensional space and the randomized maximum likelihood (RML) method [40] to generate samples of the posterior distribution of predicted data given the observations. The authors used a data transformation before PCA to improve linearity. [52] extended the original DSI method introducing a more general data transformation procedure. They tested the method in a model of a complex fractured reservoir and obtained reasonable uncertainty estimates of the production forecast. [26] modified the DSI approach to allow changes in the well controls during the forecast period so that the method could be used for life-cycle optimization. They noted that the method required a larger prior ensemble to better represent a wider range of possibilities. They showed that the proposed method combined to a direct search optimization algorithm [2] was able to improve the expected net-present value of a reservoir model.

Similar ideas of DSI have appeared before in the literature. For example, [27] and [41] applied linear regression to combine the ozone forecast of ensembles of models. In the atmospheric literature, these methods are known as aggregation methods or ensemble forecast [34]. [8] also used a DSI-type of approach to invert 4D seismic impedance data directly to pressure and water saturation. [46] proposed a method named prediction-focused analysis (PFA) based on projecting the prior predictions into a low-dimensional space and using kernel

smoothing to estimate the joint distribution of historical and forecasted data. Using the joint distribution, they could predict the uncertainty estimates in a tracer transport problem. However, PFA method seems applicable only to problems with few data points because it requires the projection to very few dimensions (two or three dimension in the examples presented in the paper). [44] modified the PFA method using canonical functional component analysis to improve the linearity in the projected data. [45] used the same approach in a reservoir problem and concluded that the method provided uncertainty estimates of production forecast in reasonable agreement with rejection sampling. More recently, [23] used similar ideas from DSI to estimate the uncertainty reduction in a study to compute the value of information of data-acquisition plans. [25] applied machine learning techniques (neural networks and support vector regression) to DSI. They concluded that the method can be a more efficient alternative to history matching; however, the method fails to provide satisfactory forecast if the predictions from the prior ensemble (training set) are too far from the expected true response.

In the present paper, we introduce a new DSI implementation based on the use of an iterative ensemble smoother and demonstrate with examples that the new DSI is computationally faster and more robust than the procedure proposed in [51, 52]. Moreover, we apply the new DSI to a real field case with long production history and large number of wells and show that the method provides forecasts comparable with a more traditional ensemble-based history-matching process. The rest of the paper is organized as follows: Section 2 reviews the DSI method proposed in [51, 52] and the new DSI method. Section 3 presents three reservoir test problems. The first problem is a small synthetic case used to demonstrate that the proposed method provides results similar to the original DSI with a lower computational cost. The second problem is a benchmark history-matching case [3, 35] constructed with data from real reservoir in Campos Basis. This problem is used to compare the methods in a more realistic situation with a large number of data points. The last problem corresponds to a real brown-field case where the proposed method is compared against an ensemble-based history matching. Section 5 summarizes the conclusions of the paper.

2 Methodology

2.1 Preliminaries

Let $\mathbf{m} \in \mathbb{R}^{N_m}$ denote the vector of uncertain parameters of a reservoir model with a historical production period t_h . Our goal is to predict the production performance for a period t_f after t_h . Let $\mathbf{d} \in \mathbb{R}^{N_d}$ denote the vector of predicted

production data, which is a nonlinear function of \mathbf{m} , that is, $\mathbf{d} = \mathbf{d}(\mathbf{m})$. In the applications of interest of this paper, \mathbf{d} is the result of a reservoir simulation. This vector contains predicted data from both, history, \mathbf{d}_h , and forecast periods, \mathbf{d}_f , that is,

$$\mathbf{d} = \begin{bmatrix} \mathbf{d}_h \\ \mathbf{d}_f \end{bmatrix}. \tag{1}$$

Let $\mathbf{d}_{\text{obs}} \in \mathbb{R}^{N_{d,h}}$ denote the vector of field observations, which is corrupted with an additive random noise, \mathbf{e}_d , due to measurement errors such that

$$\mathbf{d}_{\text{obs}} = \mathbf{d}_{\text{true}} + \mathbf{e}_d, \tag{2}$$

where \mathbf{d}_{true} is the true (noiseless) data. Moreover, assume that model errors are also additive and

$$\mathbf{d}_{\text{true}} = \mathbf{d}_h(\mathbf{m}_{\text{true}}) + \mathbf{e}_m, \tag{3}$$

where \mathbf{m}_{true} is the vector containing the “true” values for the model parameters and \mathbf{e}_m is a vector of model errors. Under these conditions, it is straightforward to show [54] that the likelihood of \mathbf{m} is

$$\mathcal{L}(\mathbf{m}|\mathbf{d}_{\text{obs}}) = \text{const} \times \exp\left(-\frac{1}{2}(\mathbf{d}_{\text{obs}} - \mathbf{d}_h(\mathbf{m}))^\top \mathbf{C}_e^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{d}_h(\mathbf{m}))\right), \tag{4}$$

where

$$\mathbf{C}_e = \text{cov}[\mathbf{e}_d] + \text{cov}[\mathbf{e}_m] \tag{5}$$

is the total data-error covariance matrix. Note that we use the term “model error” to refer to any imperfection in the model used to represent the real reservoir. Examples of sources of model errors include numerical and discretization errors, simplifications of the physics, and insufficient parameterization. The assumption in Eq. 3 is that all sources of model errors can be aggregated into a random vector with $E[\mathbf{e}_m] = \mathbf{0}$ and known covariance. Even though these are strong assumptions which are unlikely to hold in reality, it is important to note that including $\text{cov}[\mathbf{e}_m]$ in the matrix \mathbf{C}_e reduces the weights attributed to data helping to partially compensate for deficiencies in the models; see, for example, [38, 53].

If the prior model follows a multivariate Gaussian distribution, $\mathbf{m} \sim \mathcal{N}(\mathbf{m}_{\text{pr}}, \mathbf{C}_m)$, then the posterior probability density function (PDF) of \mathbf{m} given \mathbf{d}_{obs} has the form

$$\begin{aligned} p(\mathbf{m}|\mathbf{d}_{\text{obs}}) &= \text{const} \times \mathcal{L}(\mathbf{m}|\mathbf{d}_{\text{obs}})p(\mathbf{m}) \\ &= \text{const} \times \exp(-\mathcal{O}(\mathbf{m})), \end{aligned} \tag{6}$$

where

$$\begin{aligned} \mathcal{O}(\mathbf{m}) &= \frac{1}{2}(\mathbf{d}_{\text{obs}} - \mathbf{d}_h(\mathbf{m}))^\top \mathbf{C}_e^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{d}_h(\mathbf{m})) \\ &\quad + \frac{1}{2}(\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_{\text{pr}}). \end{aligned} \tag{7}$$

The model \mathbf{m} that minimizes $\mathcal{O}(\mathbf{m})$ corresponds to the maximum a posteriori [54]. In practice, however, we are interested in sampling the posterior PDF to quantify uncertainty. In this case, one alternative is the RML method [40], which provides an approximate sampling of $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$. Each RML sample is obtained by minimizing a modified version of Eq. 7 given by

$$\begin{aligned} \mathcal{O}_{\text{rml}}(\mathbf{m}) &= \frac{1}{2}(\mathbf{d}_{\text{obs}}^* - \mathbf{d}_h(\mathbf{m}))^\top \mathbf{C}_e^{-1}(\mathbf{d}_{\text{obs}}^* - \mathbf{d}_h(\mathbf{m})) \\ &\quad + \frac{1}{2}(\mathbf{m} - \mathbf{m}^*)^\top \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}^*), \end{aligned} \tag{8}$$

where $\mathbf{d}_{\text{obs}}^* \sim \mathcal{N}(\mathbf{d}_{\text{obs}}, \mathbf{C}_e)$ and $\mathbf{m}^* \sim \mathcal{N}(\mathbf{m}_{\text{pr}}, \mathbf{C}_m)$.

2.2 Data-space inversion

In this section, we review the data-space inversion (DSI) procedure as proposed in [51] and later improved in [52]. The main idea behind the method is to use PCA to write the vector of predicted data as

$$\mathbf{d}_{\text{pca}} = \bar{\mathbf{d}} + \mathbf{C}_d^{1/2}\mathbf{x}, \tag{9}$$

where $\bar{\mathbf{d}}$ and \mathbf{C}_d are the mean and covariance of \mathbf{d} , respectively. Both are computed using a prior ensemble, that is,

$$\bar{\mathbf{d}} = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{d}_j \tag{10}$$

and

$$\begin{aligned} \mathbf{C}_d &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{d}_j - \bar{\mathbf{d}})(\mathbf{d}_j - \bar{\mathbf{d}})^\top \\ &= \Delta\mathbf{D}\Delta\mathbf{D}^\top, \end{aligned} \tag{11}$$

where

$$\Delta\mathbf{D} = \frac{1}{\sqrt{N_e - 1}} [\mathbf{d}_1 - \bar{\mathbf{d}}, \dots, \mathbf{d}_{N_e} - \bar{\mathbf{d}}]. \tag{12}$$

The square root of \mathbf{C}_d in Eq. 9 is computed using the singular value decomposition (SVD) of $\Delta\mathbf{D}$

$$\Delta\mathbf{D} = \mathbf{U}\Sigma\mathbf{V}^\top, \tag{13}$$

where \mathbf{U} is a $N_d \times N_d$ orthogonal matrix containing the left singular values of $\Delta\mathbf{D}$, which are equivalent to the left eigenvectors of \mathbf{C}_d ; Σ is a $N_d \times N_e$ matrix containing as non-zero elements the singular values of $\Delta\mathbf{D}$, or, equivalently, the square root of the eigenvalues of \mathbf{C}_d . The matrix \mathbf{V}

contains the right singular values of $\Delta \mathbf{D}$. The square-root of \mathbf{C}_d becomes

$$\mathbf{C}_d^{1/2} = \mathbf{U}\Sigma. \tag{14}$$

The vector \mathbf{x} in Eq. 9 is a sample from a standard normal distribution, that is, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. In practice, we truncate small singular values using an energy criterium, which means that we consider only the $N_r \leq \max\{N_d, N_e - 1\}$ largest singular values such that

$$\frac{\sum_{i=1}^{N_r} \sigma_i}{\sum_{i=1}^{\min\{N_d, N_e-1\}} \sigma_i} \geq \xi \tag{15}$$

where $\xi \leq 1$ is the energy threshold, typically selected between 0.9 and 0.99 and σ_i is the i th singular value of $\Delta \mathbf{D}$. Truncating small singular values has two positive side effects: it reduces the dimension of \mathbf{x} and introduces regularization in the inversion. Both are important because the DSI method uses RML for sampling, which requires solving several minimization problems.

Sun and Durlofsky [51] noted that the direct application of Eq. 9 may result in nonphysical values for the predicted data, for example negative production or pressure. According to the authors, this problem occurs mainly before water breakthrough time. Therefore, they proposed to apply a data transformation to the prior realizations of \mathbf{d} before PCA. The transformation is based on shifting and compressing/stretching the time series. They claim and illustrate in a example that the transformed vectors, $\hat{\mathbf{d}}$, have a more Gaussian prior distribution. However, this procedure is difficult to apply in cases with frequent changes in well controls. In [52], the authors proposed to use an inverse Gaussian anamorphosis procedure using the empirical cumulative density function (CDF) computed using the N_e prior realizations of \mathbf{d} . Figure 1 illustrates the process, where each component of the transformed vector $\hat{\mathbf{d}}$ is computed as

$$\hat{d}_i = \text{cdf}_1^{-1}(\text{cdf}_2(d_{\text{pca},i})), \tag{16}$$

where $\text{cdf}_1(\cdot)$ and $\text{cdf}_2(\cdot)$ are the CDF's of \mathbf{d} and \mathbf{d}_{pca} , respectively.

The final step of DSI is to use RML to generate a posterior ensemble of predicted data. The RML objective function can be written in terms of the vector of PCA coefficients as

$$\begin{aligned} \mathcal{O}_{\text{rml}}(\mathbf{x}) = & \frac{1}{2} (\mathbf{d}_{\text{obs}}^* - \hat{\mathbf{d}}_h(\mathbf{x}))^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}}^* - \hat{\mathbf{d}}_h(\mathbf{x})) \\ & + \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*), \end{aligned} \tag{17}$$

where $\mathbf{x}^* \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The minimization of Eq. 17 can be done with any optimization method. [51] used the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method [37], in which case it is necessary to compute the gradient of $\mathcal{O}_{\text{rml}}(\mathbf{x})$ with respect to \mathbf{x} . However, $\hat{\mathbf{d}}_h(\mathbf{x})$

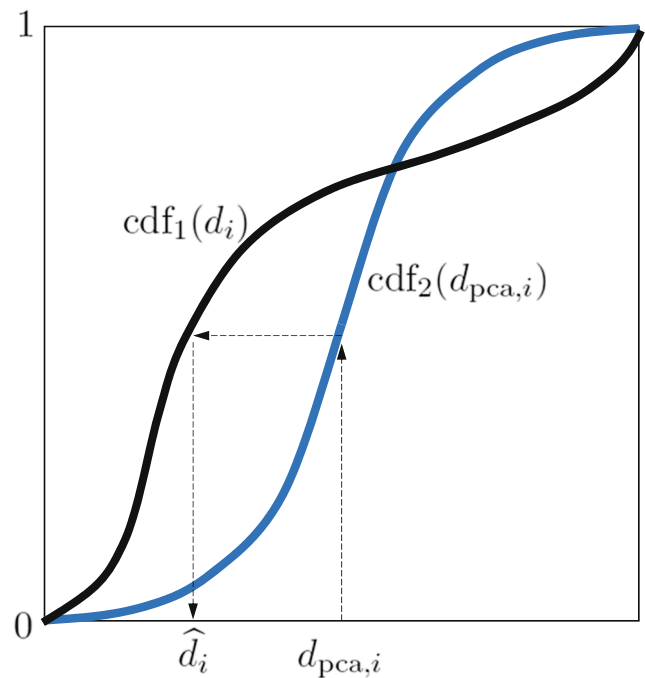


Fig. 1 Histogram transformation used in DSI

is a nonlinear function of \mathbf{x} with no general analytical form because of the transformation of Eq. 16. One alternative is to compute numerical gradients, which may be computationally expensive if we have a large number of data points. Jiang [26] proposed to ignore the transformation of Eq. 16 and use the analytical gradients computed with \mathbf{d}_{pca} instead of $\hat{\mathbf{d}}$. Our limited set of tests indicated that this procedure in fact improved the computational performance of our DSI implementation, which uses the limited-memory BFGS [37] implementation available in the C# library Accord.NET [48].

2.3 Data-space inversion with ensemble smoother

The ensemble smoother (ES) was introduced by [55] as an alternative to the sequential data assimilation scheme of EnKF. The first application of ES for history matching was presented by [47], which concluded that the method is faster than EnKF with similar results. Despite the good results presented in [47], some authors [5, 6, 14, 15] observed that ES tends to result in unreasonable data matches when applied to more complex history-matching problems. The reason for the poor performance is because ES is similar to applying a single Gauss-Newton iteration to minimize a RML-type of objective function with a sensitivity matrix estimated based on the prior ensemble [43]. This fact lead to the development of several iterative forms of ES; see, for example, [5, 6, 15, 31, 49, 50]. Among the iterative forms of ES, the ensemble smoother with multiple data assimilation (ES-MDA) [15] is a popular choice. The

popularity of ES-MDA can be attributed mainly to its good performance in history-matching problems [9, 16, 20, 36] and its simplicity of implementation. In fact, ES-MDA is essentially equivalent to repeat ES a few times with the data-error covariance matrix, \mathbf{C}_e , multiplied by coefficients α_k 's to avoid overweighing the measurements. The choice of the coefficients and the number of repetitions of ES, N_a , must obey the condition $\sum_{k=1}^{N_a} \alpha_k^{-1} = 1$ to ensure that ES-MDA samples the correct posterior PDF in the linear-Gaussian case [15]. For nonlinear problems, the choice of the α_k 's and N_a has a major impact in the performance of the method. There are recent works proposing methods to select α_k 's and N_a [9, 10, 29, 33, 42]. However, here we use simplest choice, which consists of selecting N_a in advance and setting $\alpha_k = N_a$, for $k = 1, \dots, N_a$.

The application of ES-MDA to DSI, that is, to generate samples of the predicted data vector, $\mathbf{d} = [\mathbf{d}_h^T, \mathbf{d}_f^T]^T$, given the vector of observations, \mathbf{d}_{obs} , is straightforward. We can write the resulting DSI-ESMDA update equation as

$$\mathbf{d}_j^{k+1} = \mathbf{d}_j^k + \mathbf{R} \circ \mathbf{K}^k (\mathbf{d}_{obs} + \sqrt{\alpha_k} \mathbf{e}_j^k - \mathbf{d}_h^k), \tag{18}$$

for $k = 1, \dots, N_a$ and $j = 1, \dots, N_e$, where \mathbf{K} is a modified version of the Kalman gain given by

$$\mathbf{K}^k = \Delta \mathbf{D}^k (\Delta \mathbf{D}_h^k)^T \left((\Delta \mathbf{D}_h^k (\Delta \mathbf{D}_h^k)^T + \alpha_k \mathbf{C}_e) \right)^{-1}. \tag{19}$$

$$r\left(\frac{h}{L}\right) = \begin{cases} -\frac{1}{4} \left(\frac{h}{L}\right)^5 + \frac{1}{2} \left(\frac{h}{L}\right)^4 + \frac{5}{8} \left(\frac{h}{L}\right)^3 - \frac{5}{3} \left(\frac{h}{L}\right)^2 + 1, & \text{if } 0 \leq \frac{h}{L} \leq 1 \\ \frac{1}{12} \left(\frac{h}{L}\right)^5 - \frac{1}{2} \left(\frac{h}{L}\right)^4 + \frac{5}{8} \left(\frac{h}{L}\right)^3 + \frac{5}{3} \left(\frac{h}{L}\right)^2 - 5 \left(\frac{h}{L}\right) + 4 - \frac{2}{3} \left(\frac{h}{L}\right)^{-1}, & \text{if } 1 \leq \frac{h}{L} \leq 2 \\ 0 & \text{if } \frac{h}{L} > 2 \end{cases}, \tag{21}$$

where h is a ‘‘distance’’ and L is a parameter known as ‘‘critical length,’’ which corresponds to the distance where the correlation function decays to approximately 0.21. Note that the matrix \mathbf{R} has the same shape of the Kalman gain and that each row of the Kalman gain corresponds to the correction term applied to each variable updated with Eq. 18, that is, entries of the vector \mathbf{d} . Each column of the Kalman gain corresponds to an observation used in the conditioning process. Therefore, h in Eq. 21 corresponds to the distance between entries of \mathbf{d} and \mathbf{d}_{obs} . In the applications of interest of this paper, both \mathbf{d} and \mathbf{d}_{obs} contain data from wells in the field. Therefore, the ratio h/L is computed based on the spatial distance between wells. Besides the spatial distance, we also introduced the time difference between data points in the calculation of h/L to account for the fact that early data have lower correlations

In the above equations, $\Delta \mathbf{D}$ was defined before (12) and $\Delta \mathbf{D}_h$ includes only the predicted data corresponding to the historical period, that is,

$$\Delta \mathbf{D}_h = \frac{1}{\sqrt{N_e - 1}} [\mathbf{d}_{h,1} - \bar{\mathbf{d}}_h, \dots, \mathbf{d}_{h,N_e} - \bar{\mathbf{d}}_h]. \tag{20}$$

The vector \mathbf{e} is a sample from $\mathcal{N}(\mathbf{0}, \mathbf{C}_e)$ and \mathbf{R} is the localization matrix with ‘‘o’’ denoting the Schur (element-wise) product. The matrix inversion required in Eq. 19 is done using the subspace inversion method [18] as described in the Appendix section of [13]. This procedure involves the truncated SVD of a rescaled matrix $\mathbf{C}_e^{-1/2} \Delta \mathbf{D}_h$ to avoid loss of relevant information when removing small singular values. The number of singular values retained is computed using the same energy criterium of Eq. 15.

The reason for introducing the Schur product between the localization matrix and the Kalman gain in Eq. 18 is twofold: it regularizes the estimates of the Kalman gain removing long-distance spurious correlations and increases the degrees of freedom to assimilate data [24]. In fact, localization proved to improve substantially the results of ensemble methods applied to history matching; see, for example, [4, 11, 12].

The regularization introduced by the Schur product is obtained by constructing the localization matrix using a correlation function with compact support. A common choice is the fifth-order compact correlation function from [22], in which case each entry of the matrix \mathbf{R} is computed using

with the predictions than the late data. The final expression for computing h/L is given by

$$\frac{h}{L} = \sqrt{\left(\frac{\Delta x'}{L_{x'}}\right)^2 + \left(\frac{\Delta y'}{L_{y'}}\right)^2 + \left(\frac{\Delta t}{T}\right)^2}, \tag{22}$$

where

$$\begin{bmatrix} \Delta x' \\ \Delta y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}. \tag{23}$$

In the above equations, Δx and Δy correspond to the spatial distance between the wells in the orthogonal directions x and y , while $\Delta x'$ and $\Delta y'$ are the corresponding distances rotated by an angle θ . This allows to consider anisotropy in the localization function as discussed in [12]. Δt is the time difference between data points. $L_{x'}$, $L_{y'}$ and T are the critical lengths in the ‘‘directions’’ x' , y' and t , respectively.

Besides removing long-distance spurious correlations, localization also increases the degrees of freedom to assimilate data. To show that, first consider the case without localization. Using Eq. 19 in Eq. 18, we can define the vector \mathbf{w}_j^k as

$$\mathbf{w}_j^k \equiv \left(\Delta \mathbf{D}_h^k \left(\Delta \mathbf{D}_h^k \right)^\top + \alpha_k \mathbf{C}_e \right)^{-1} \left(\mathbf{d}_{\text{obs}} + \sqrt{\alpha_k} \mathbf{e}_j^k - \mathbf{d}_h^k \right). \tag{24}$$

Hence

$$\begin{aligned} \mathbf{d}_j^{k+1} &= \mathbf{d}_j^k + \Delta \mathbf{D}^k \left(\Delta \mathbf{D}_h^k \right)^\top \mathbf{w}_j^k \\ &= \mathbf{d}_j^k + \frac{1}{N_e - 1} \sum_{i=1}^{N_e} \left(\mathbf{d}_i^k - \bar{\mathbf{d}}^k \right) \left(\mathbf{d}_{h,i}^k - \bar{\mathbf{d}}_h^k \right)^\top \mathbf{w}_j^k \tag{25} \\ &= \mathbf{d}_j^k + \frac{1}{N_e - 1} \sum_{i=1}^{N_e} \beta_{i,j}^k \left(\mathbf{d}_i^k - \bar{\mathbf{d}}^k \right), \end{aligned}$$

where $\beta_{i,j}^k \equiv \left(\mathbf{d}_{h,i}^k - \bar{\mathbf{d}}_h^k \right)^\top \mathbf{w}_j^k$ is a scalar. The vector $\bar{\mathbf{d}}^k$ is a linear combination of the vectors \mathbf{d}_i^k 's. Hence, it is possible to write \mathbf{d}_j^{k+1} as

$$\mathbf{d}_j^{k+1} = \sum_{i=1}^{N_e} \gamma_{i,j}^k \mathbf{d}_i^k, \tag{26}$$

where $\gamma_{i,j}^k$'s are scalars. Equation 26 means that \mathbf{d}_j^{k+1} is a linear combination of the vectors \mathbf{d}_i^k 's. Applying this result recursively to all MDA iterations, we conclude that the final predictions from DSI-ESMDA without localization are simply linear combinations of the prior ones. This effectively means that we have at most N_e coefficients (degrees of freedom) available to assimilate data. Note that the number of degrees of freedom of the standard DSI is also limited by the number of PCA coefficients which is $N_r \leq \min\{N_d, N_e - 1\}$.

Considering now the case with localization, we can write

$$\begin{aligned} \mathbf{d}_j^{k+1} &= \mathbf{d}_j^k + \left(\mathbf{R} \circ \mathbf{K}^k \right) \delta \mathbf{d}_j^k \\ &= \mathbf{d}_j^k + \sum_{i=1}^{N_{d,h}} \left(\mathbf{r}_i \circ \kappa_i^k \right) \delta d_{i,j}^k, \end{aligned} \tag{27}$$

where $\delta \mathbf{d}_j^k \equiv \mathbf{d}_{\text{obs}} + \sqrt{\alpha_k} \mathbf{e}_j^k - \mathbf{d}_h^k$ and $\delta d_{i,j}^k$ is its i th entry. \mathbf{r}_i and κ_i^k correspond to the i th columns of the localization

Table 1 Geostatistical parameters used to create the reference model and the prior ensemble

Parameter	Porosity	Log-permeability
Mean	0.22	7.2 ln-mD
Standard deviation	0.05	0.60 ln-mD
Variogram type	Spherical	Spherical
Variogram maximum range	1124.0 m	1124.0 m
Variogram minimum range	281.0 m	281.0 m
Azimuth	45°	45°

Test case 1

and Kalman gain matrices, respectively. Writing the update equation for the n th entry of the vector \mathbf{d}_j^k , we have

$$\begin{aligned} d_{n,j}^{k+1} &= d_{n,j}^k + \sum_{i=1}^{N_{d,h}} \kappa_{n,i}^k r_{n,i} \delta d_{i,j}^k \\ &= d_{n,j}^k + \sum_{i=1}^{N_{d,h}} \kappa_{n,i}^k \eta_{n,i,j}. \end{aligned} \tag{28}$$

For each n , we have a different coefficient $\eta_{n,i,j}$. Therefore, Eq. 28 means that each component of \mathbf{d}_j^{k+1} may be computed with a different linear combination of the $N_{d,h}$ columns of \mathbf{K}^k . Thus, localization expands the degrees of freedom to assimilate data.

The application of DSI-ESMDA is similar to the standard DSI, both methods require to run reservoir simulations only for the prior ensemble of models to generate the prior ensemble of predicted data. After that, DSI-ESMDA applies Eq. 18 N_a times to generate the posterior ensemble. Note that the N_a iterations are necessary because the prior ensemble is not Gaussian (if the prior is Gaussian, MDA is equivalent to a single ES update). Moreover, note that using Gaussian anamorphosis to transform variables only ensures that the marginal distributions of individual d_i 's are Gaussian, the joint distribution which is updated with

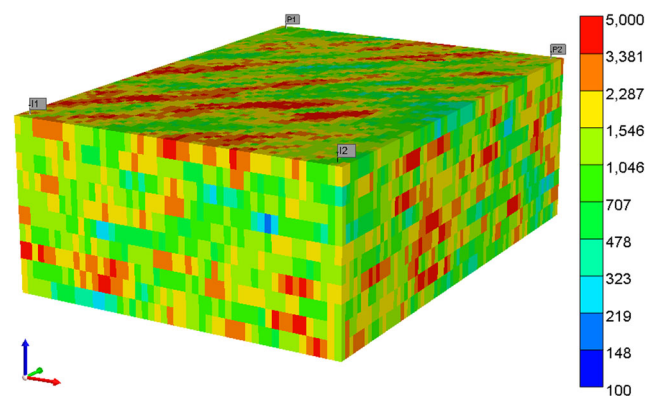


Fig. 2 Permeability field in mD. Test case 1

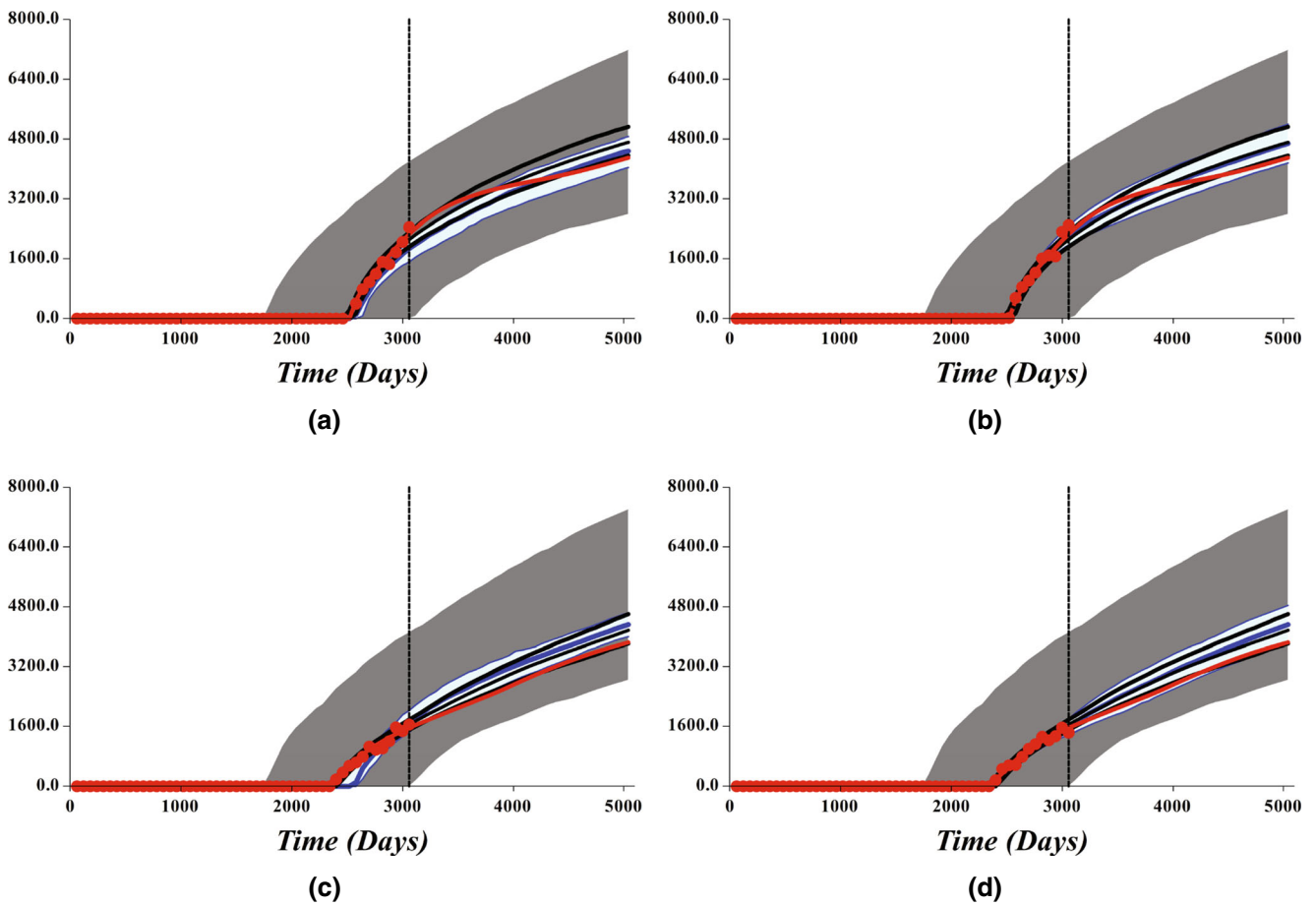


Fig. 3 Water production rate in m^3/days for wells P1 (first row) and P2 (second row). Test case 1. **a, c** DSI. **b, d** DSI-ESMDA. The red dots are the observed data points and the red line is the prediction from the reference model. The grey region corresponds to the predictions within the percentiles P10–P90 obtained with the prior ensemble. The light blue region corresponds to the predictions within the percentiles

P10–P90 obtained with DSI or DSI-ESMDA. The blue line corresponds to the percentile P50 obtained with DSI or DSI-ESMDA. The black lines correspond to the percentiles P10, P50, and P90 obtained by the history-matched models using ES-MDA. The vertical dashed line indicates the end of the history and beginning of the forecast period

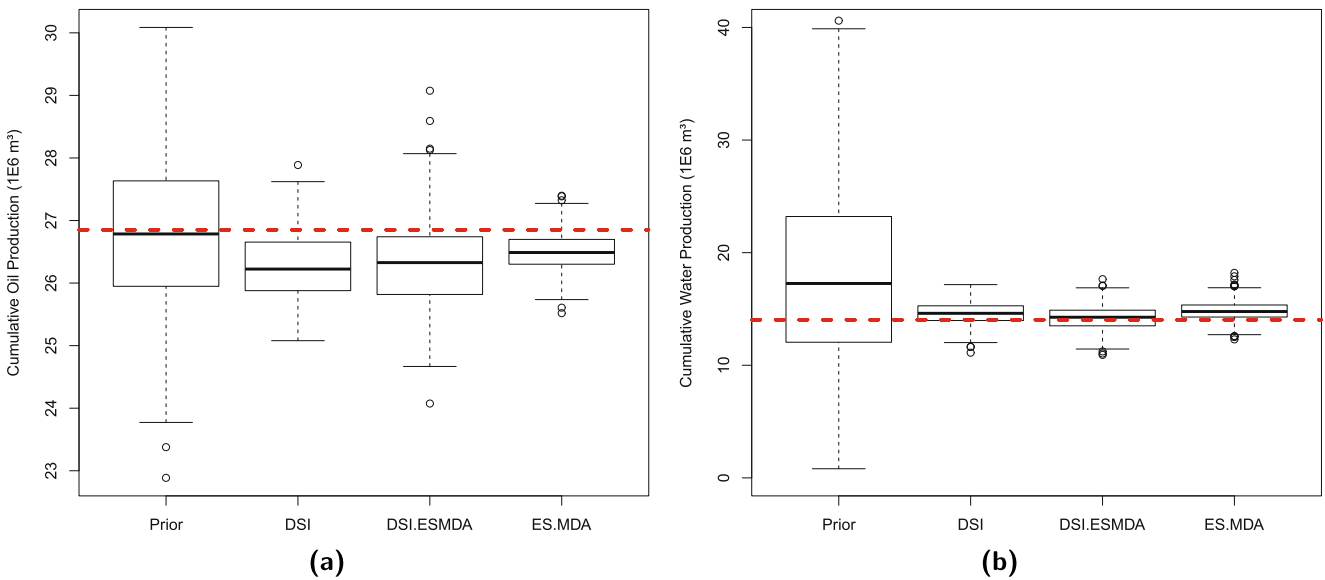


Fig. 4 Field cumulative production in 10^6 m^3 . **a** Oil and **b** water. Test case 1. The dashed red line indicates the cumulative production of the reference case

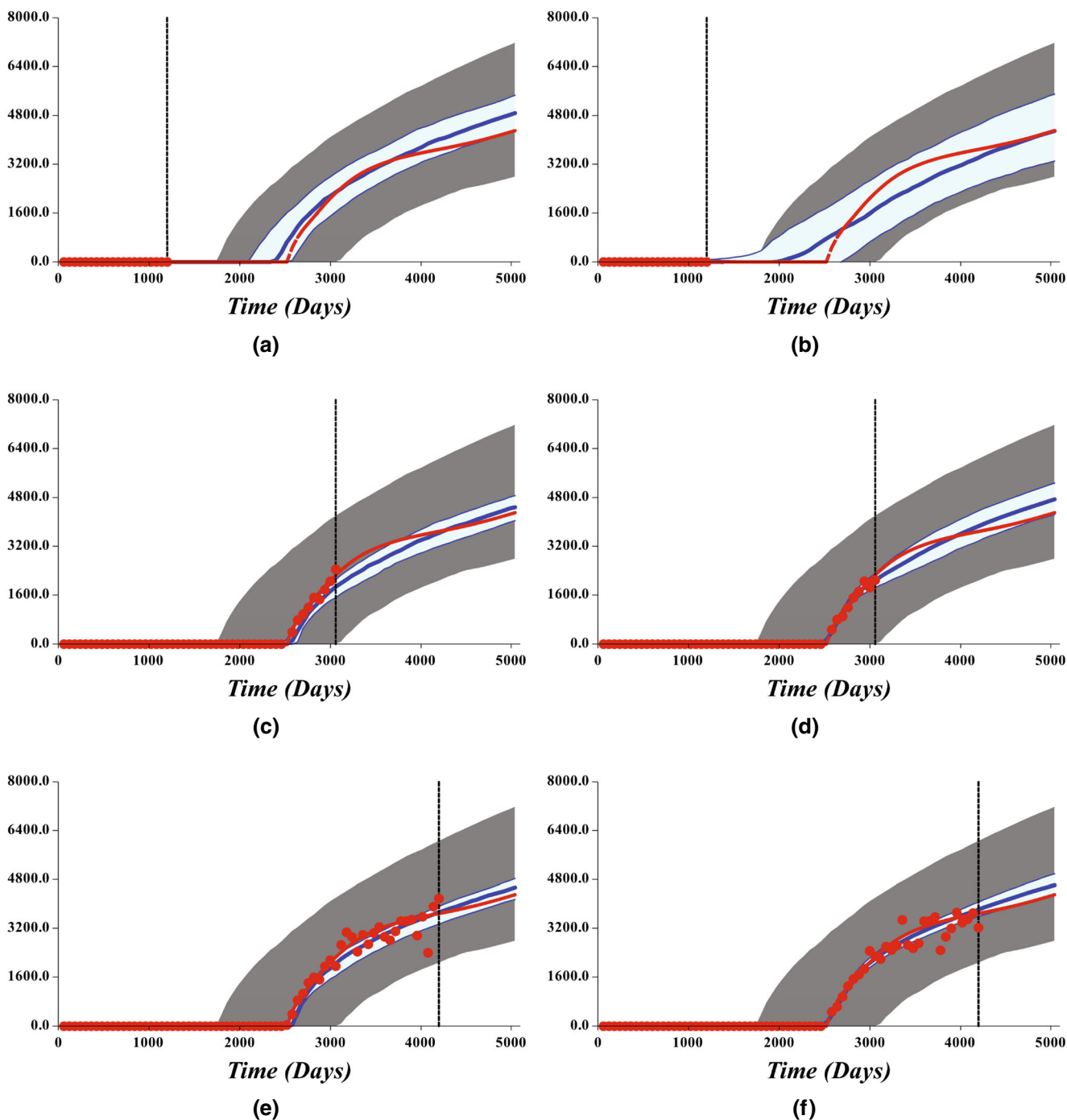


Fig. 5 Water production rate in m^3/days for well P1 considering different number of observed data points 20 (first row), 51 (second row), and 70 data points (third row). Test case 1. **a, c, e** DSI. **b, d, f** DSI-ESMDA. The red dots are the observed data points and the red line is the prediction from the reference model. The grey region corresponds to the predictions within the percentiles P10–P90 obtained with the

prior ensemble. The light blue region corresponds to the predictions within the percentiles P10–P90 obtained with DSI or DSI-ESMDA. The blue line corresponds to the percentile P50 obtained with DSI or DSI-ESMDA. The vertical dashed line indicates the end of the history and beginning of the forecast period

DSI-ESMDA may not be Gaussian. In our tests, we noticed that only a few MDA iterations are required. In all cases presented in the next section, we use $N_a = 4$. DSI-ESMDA can also be applied to the same parameterization

of DSI, that is, update the PCA coefficients with the data transformation of Eq. 16. However, our initial tests showed that this procedure did not improve the results, actually the results were slightly worse. Moreover, using the PCA

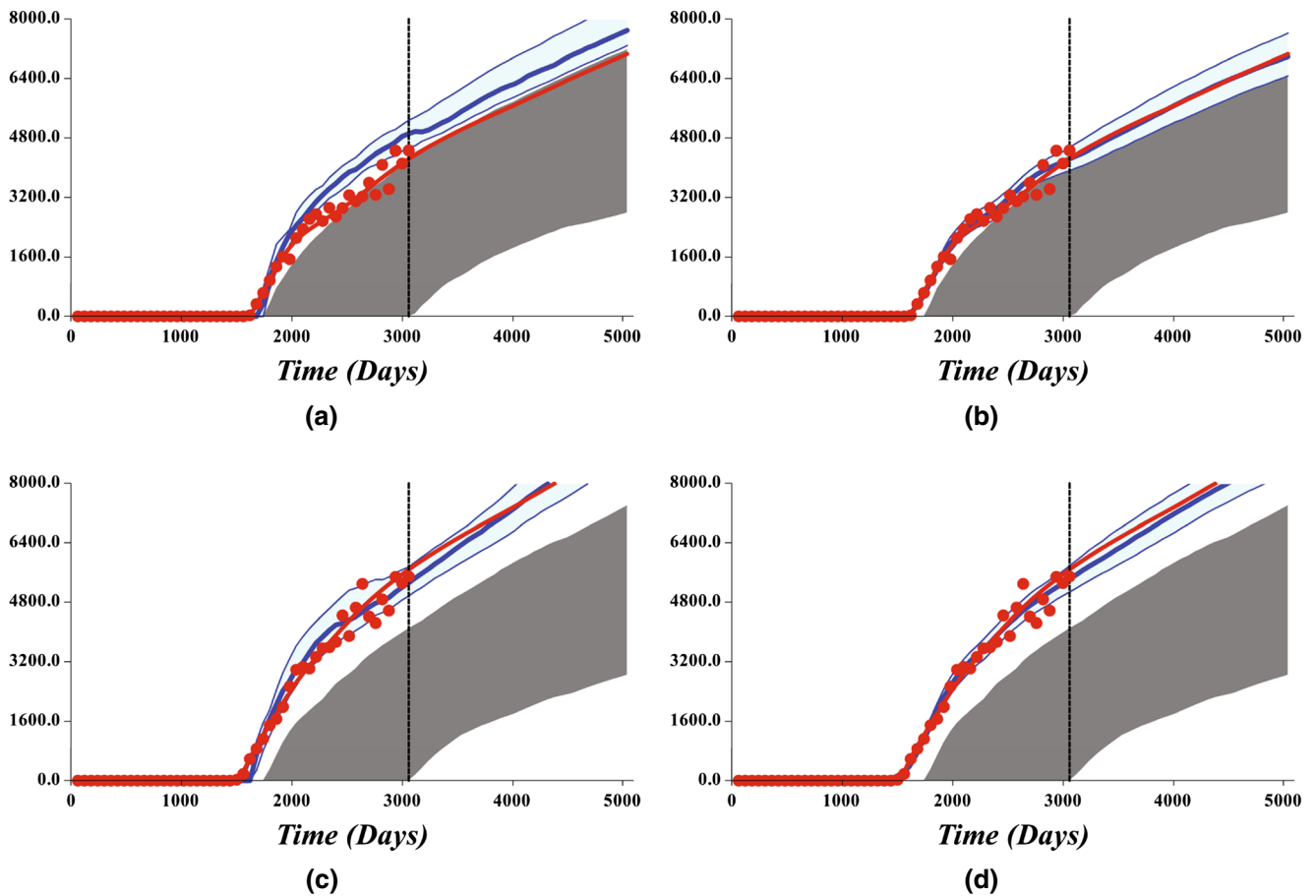


Fig. 6 Water production rate in m³/days for wells P1 (first row) and P2 (second row) for a case with poor coverage of predictions from the prior ensemble. Test case 1. **a, c** DSI. **b, d** DSI-ESMDA. The red dots are the observed data points and the red line is the prediction from the reference model. The grey region corresponds to the predictions within the percentiles P10–P90 obtained with the prior ensemble. The light

blue region corresponds to the predictions within the percentiles P10–P90 obtained with DSI or DSI-ESMDA. The blue line corresponds to the percentile P50 obtained with DSI or DSI-ESMDA. The vertical dashed line indicates the end of the history and beginning of the forecast period

parameterization of DSI would prevent to use the Kalman gain localization procedure described in this section. The only correction we applied is to truncate in zero if the final predicted data is negative, which in our tests occurred with water production data before breakthrough. Note that we apply this truncation only to the final estimates, not during the DSI-ESMDA iterations.

3 Test cases

3.1 Test case 1

The first test problem is a synthetic reservoir case created with rock and fluid properties typically found in the Campos Basin, Brazil. The reservoir model contains 50 × 70 × 10 gridblocks with uniform size of 50 m × 50 m × 5 m. A reference (true) case was generated using sequential

Gaussian simulation for modeling porosity and sequential Gaussian co-simulation for permeability using the porosity as secondary variable and a correlation coefficient of 0.95. Table 1 summarizes the geostatistical parameters used to generate the reference model. We assumed no vertical correlation between the layers of the model to make the problem more challenging for data assimilation. The model has two vertical oil producing and two vertical water injection wells placed on the borders of the reservoir as illustrated in Fig. 2. The producers operated under as specified bottom-hole pressure (BHP) of 25,000 kPa while the injectors operate with a BHP of 35,000 kPa.

We generated an ensemble of 500 realizations for testing DSI and DSI-ESMDA using the same geostatistical parameters of the reference model. The synthetic measurements correspond to oil and water rate at both production wells and water injection rate at both injectors. These observations were corrupted by adding random Gaussian noise with

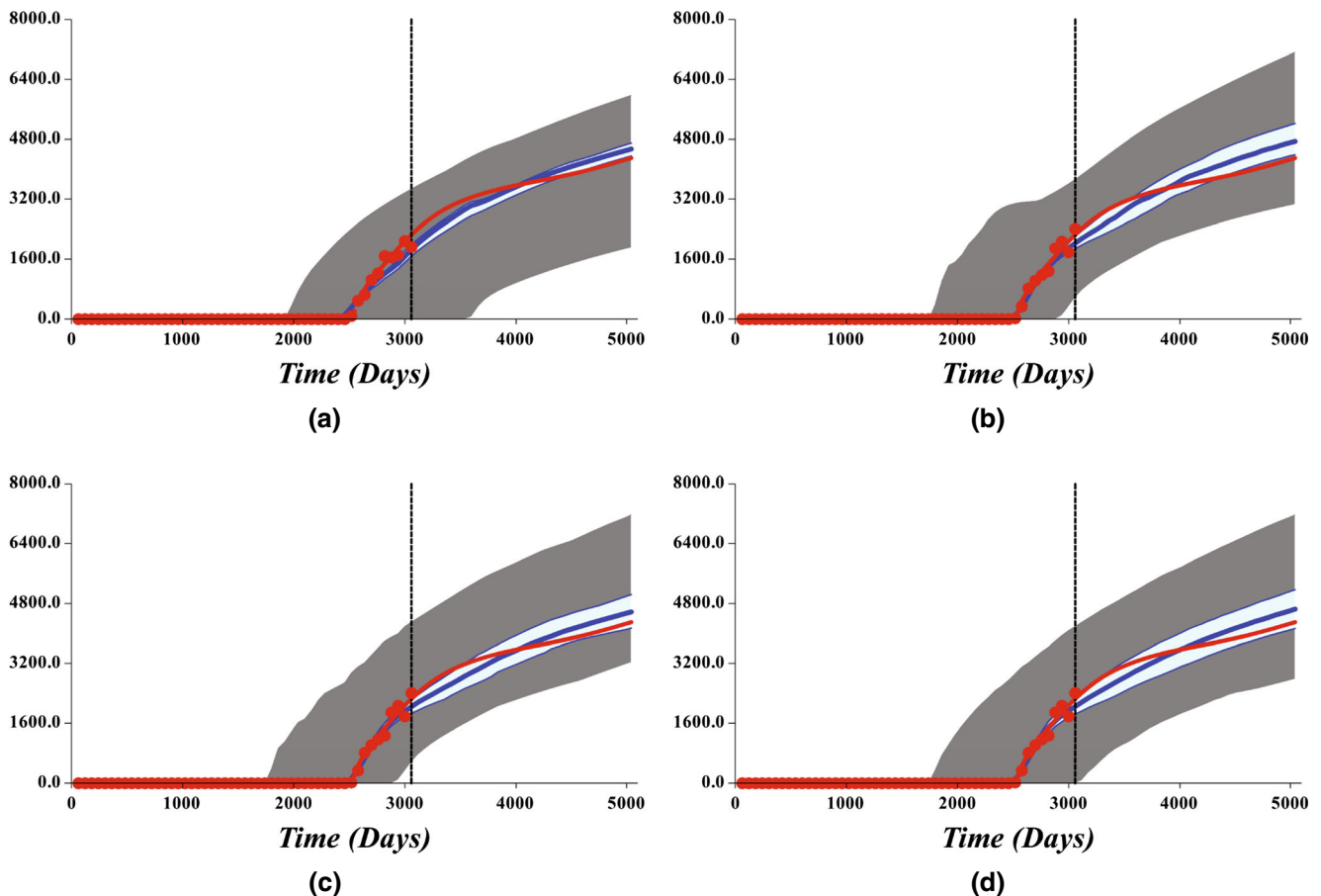


Fig. 7 Water production rate in m^3/days for well P1 obtained with DSI-ESMDA with different ensemble sizes. Test case 1. **a** $N_e = 10$, **b** $N_e = 50$, **c** $N_e = 100$, and **d** $N_e = 500$. The red dots are the observed data points and the red line is the prediction from the reference model. The grey region corresponds to the predictions within the percentiles

P10–P90 obtained with the prior ensemble. The light blue region corresponds to the predictions within the percentiles P10–P90 obtained with DSI-ESMDA. The blue line corresponds to the percentile P50 obtained with DSI-ESMDA. The vertical dashed line indicates the end of the history and beginning of the forecast period

zero mean and standard deviation corresponding to 10% of the data predicted by the reference model. The total number of observed data points is $N_{d,h} = 51$ and the number of predicted points is $N_{d,f} = 33$. For DSI, we kept 99% of the singular value energy (15) which corresponded to 98 singular values. For DSI-ESMDA, we also kept 99% of the singular value energy in the subspace inversion. We used $N_a = 4$ MDA iterations. Neither spatial nor temporal localizations were applied in this case. For comparisons, we also applied a standard history matching (that is, model-space inversion) using ES-MDA to update the same prior ensemble of 500 realizations with $N_a = 4$ without localization. Even though we use the history-matching results of ES-MDA as reference for comparisons, it is important to note that ES-MDA provides only an approximation of the uncertainty in the production forecast because this method is not guaranteed to converge to the correct sampling of the posterior PDF in nonlinear problems. The objective is to compare the DSI results against a history-matching procedure that is used in practice. There are rigorous sampling

methods, such as Markov chain Monte Carlo and rejection sampling [40] that could be used to generate the reference solution. Unfortunately, these methods are computationally prohibitive, even for a simple problem such as the one describe in this section. One alternative would be to simplify the problem by reducing the size of the model and the number of data points. However, this has already been done in [51], where it is shown that DSI obtained reasonable results compared with rejection sampling for a problem with few data points.

Figure 3 shows the predicted water production rate for both wells of the model obtained with DSI, DSI-ESMDA. For comparisons, we included in each plot the water rate predicted by the history-matched models with ES-MDA. The results in this figure indicate that both DSI methods obtained similar predictions, which are in reasonable agreement with the standard ES-MDA. Figure 4 shows the field cumulative production of oil and water predicted by the prior ensemble and by each method. The cumulative production from the reference case is also included in

Table 2 CPU time for inversion (70 data points)

Method	CPU time (s)
DSI	1506
DSI-ESMDA	2

Test case 1

this figure for comparisons. This figure shows that the uncertainty ranges predicted by the methods are similar. Figure 5 shows the predicted water production rate for the first well considering three different sizes of the historical period. The main difference is observed for the case with no water breakthrough in the production history (cases with 20 data points in Fig. 3). For this case, DSI-ESMDA predicts a larger uncertainty range in the water production rate. In order to test the robustness of the methods, we selected another reference model with predictions outside the P10–P90 range from the prior ensemble and the results are presented in Fig. 6. Despite the poor coverage of the prior ensemble, both methods obtained forecasts in reasonable agreement with the reference. However, we note that DSI-ESMDA obtained a slightly better data match for the first well (Fig. 6 a and b) and a forecast range centered around the reference. Finally, we also investigated the effect of the ensemble size in DSI-ESMDA. Figure 7 shows the predicted water rate obtained with ensembles sizes of 10, 50, 100, and 500. The results in this figure show indication of ensemble collapse for $N_e = 10$. On the other hand, an ensemble 100 prior realizations converged to a forecast very close to the case with 500 realizations.

Both methods require to run reservoir simulations of the prior ensemble, which is clearly the dominating computational time. DSI also requires to solve the RML optimizations with the BFGS method, while DSI-ESMDA requires to apply the analysis (18) $N_a = 4$ times for each ensemble member. We measured the CPU running time of the inversion part of each method and the results are reported in Table 2. All computations were performed in the same computer (Intel Core i5-4690 CPU 3.5 GHz and 16 GB RAM). The results in this table show that DSI-ESMDA is notably faster than DSI, requiring only two seconds against approximately 25 min for DSI.

3.2 Test case 2

The second test case is a more realistic history-matching problem known as UNISIM-I-H [3]. This problem is based on actual data from Namorado Field (Campos Basis, Brazil). The UNISIM-I-H model has $81 \times 58 \times 20$ gridblocks, but only 37,000 are active. All gridblocks have a

uniform size of $100 \text{ m} \times 100 \text{ m} \times 8 \text{ m}$. The original dataset is available for download at [35]. The dataset consists of 500 realizations of petrophysical properties (porosity, permeability in the three orthogonal directions and net-to-gross ratio). Besides the 500 petrophysical realizations, the UNISIM-I-H case also includes five global parameters, whose prior uncertainties were modeled as independent triangle distributions with values shown in Table 3. In this field, there are 25 long horizontal wells (14 producers and 11 water injectors). Figure 8 shows the position of the wells projected in the first layer of the model. The oil-producing wells are perforated close to the top and the water injection wells are perforated near to the bottom of the reservoir. The observed data were generated by adding random Gaussian data-error to the data predicted by a reference fine-scale model (UNISIM-I-R). The UNISIM-I-R was constructed with a higher level of geological details in a grid with 3.5 million active gridblocks [3]. The observed data correspond to monthly “measurements” of oil and water rate and the noise level was assumed as 10% of the data predicted by the model UNISIM-I-R. All wells are controlled by specified BHP during the historical and forecast periods.

For this test problem, we applied DSI, DSI-ESMDA, and DSI-ESMDA with localization. We also applied a model-based inversion (history matching) with ES-MDA for comparisons. The dimension of the matrix $\Delta \mathbf{D}$ for the UNISIM-I-H case is 12825×500 and the SVD retaining 99% of the singular values energy resulted in a vector of PCA coefficients with 395 elements for DSI. For DSI-ESMDA, we used $N_a = 4$ data assimilations and kept 99% of the singular values in the subspace inversion. DSI-ESMDA with localization considered a critical length computed with $L_{x'} = L_{y'} = 2000 \text{ m}$ and $T = 6000$ days. These values were selected based on our previous experience testing ES-MDA for history matching this benchmark problem. Later, we present a small sensitivity analysis on the choice of the critical length. The history matching with ES-MDA also used $N_a = 4$ and a spatial localization with critical length of 2000 m. Note that in the history matching case, the Kalman gain is applied to update model parameters (porosity, permeability, etc.). Therefore, it does not seem appropriate to introduce the “time distance” to compute the localization matrix in this case.

In our tests, the DSI implementation failed to achieve reasonable data matches for the UNISIM-I-H case. This result came as a surprise because the same problem was not observed with DSI-ESMDA using the same prior ensemble. We conducted a series of experiments trying to solve this problem, but the reason for the poor performance of the optimizations is still unclear. Among our experiments, we tried to use both numerical and the approximate analytical

Table 3 Prior distribution of global parameters

Parameter	Mode	Min.	Max.
Rock compressibility in (cm ² /kgf)	5.3×10^{-5}	1×10^{-5}	9.6×10^{-5}
Depth of the oil-water contact at the East block (Fig. 8)	3174	3169	3179
Maximum water relative permeability	0.35	0.15	0.55
Corey exponent of water relative permeability	2.4	2.0	3.0
Multiplier for vertical permeability	1.5	0.0	3.0

UNISIM-I-H

gradients during the minimizations; however, the large majority of the optimization did not converge properly. We also introduced a rescaling procedure before the truncated SVD. Basically, instead of computing the SVD of the matrix $\Delta\mathbf{D}$ (as in Eq. 13), we applied SVD to a scaled version

$$\widehat{\Delta\mathbf{D}} \equiv \mathbf{C}_e^{-1/2} \Delta\mathbf{D} = \widehat{\mathbf{U}} \widehat{\Sigma} \widehat{\mathbf{V}}^T, \quad (29)$$

and computed the square root of \mathbf{C}_d for PCA as

$$\mathbf{C}_d^{1/2} = \mathbf{C}_e^{1/2} \widehat{\mathbf{U}} \widehat{\Sigma}. \quad (30)$$

The rationale for this procedure is to avoid losing relevant information during the truncation of small singular values because $\Delta\mathbf{D}$ may be poorly scaled. However, this procedure did not result in significant improvements. We also tried the optimizations keeping all 499 singular values to preserve

the maximum number of degrees freedom provided by the prior ensemble to match data. In this case, we observed an even worse performance. Conversely, we also tried to reduce the number of singular values by keeping only 95% and 90% of the energy, which resulted in 237 and 154 singular values, respectively. The idea was to check if using a small number of PCA coefficients would help to regularize the optimizations. Again, no noticeable improvements were observed. Finally, following the recommendation of one of the reviewers, we tried the pre-selection procedure proposed in [51], in which case only the 20% best prior simulated data are used. In this case, we observed an improvement in the data matches, however at the cost of a large reduction in the spread of predictions.

In order to evaluate the quality of the final data matches, we computed the data-mismatch objective function nor-

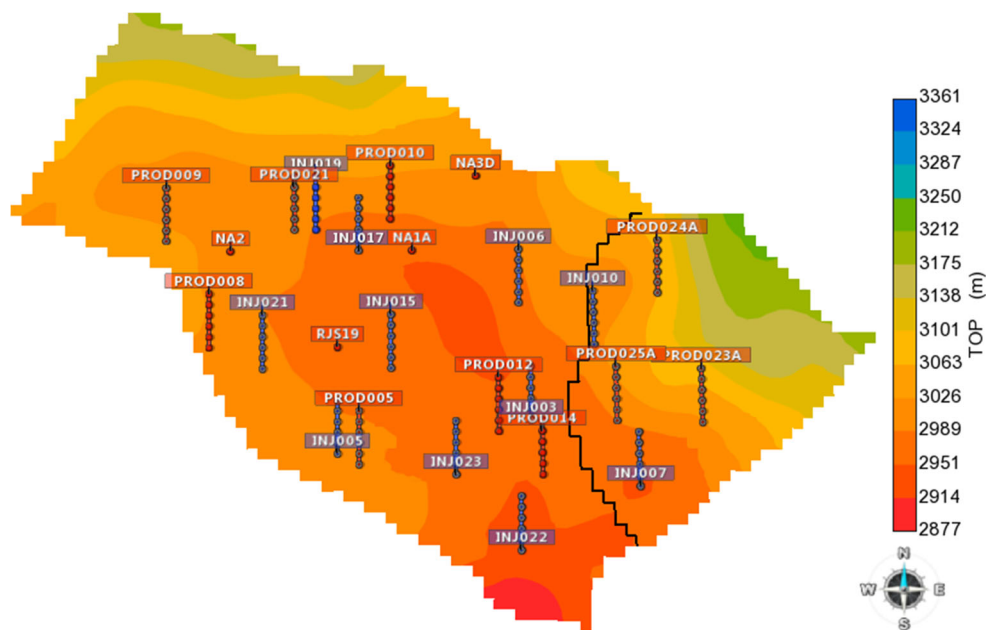
**Fig. 8** Position of the wells. UNISIM-I-H

Table 4 Normalized data-mismatch objective function

Case	Mean	Standard deviation
Prior	456.7	333.462
DSI	26.6	7.379
DSI with pre-selection	13.8	0.370
DSI-ESMDA	2.7	0.003
DSI-ESMDA with localization	1.9	0.008
ES-MDA	3.0	0.109

UNISIM-I-H

malized by the number of observed data points ($N_{d,h} = 9900$) for 500 posterior realizations obtained with DSI, DSI with pre-selection, DSI-ESMDA, DSI-ESMDA with localization, and the history matching with ES-MDA. The

normalized data mismatch objective function was computed as

$$\begin{aligned} \mathcal{O}_{N,d} &= \frac{1}{2N_{d,h}} (\mathbf{d}_{\text{obs}} - \mathbf{d}_h)^\top \mathbf{C}_e^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{d}_h) \\ &= \frac{1}{2N_{d,h}} \sum_{i=1}^{N_d} \left(\frac{d_{\text{obs},i} - d_{h,i}}{\sigma_{e,i}} \right)^2, \end{aligned} \quad (31)$$

where $\sigma_{e,i}$ is the data-error standard deviation of the i th datum. The last equality in Eq. 31 holds only for independent data errors. In the Appendix section of [38], it is shown that the expectation of $\mathcal{O}_{N,d}$ for a set of RML samples in a linear problem should be one half. Even though this value is valid only for linear problems, it serves as a reference. For example, if we have a predicted curve where the difference with all observed data points is exactly one standard deviation of the data error, then the normalized

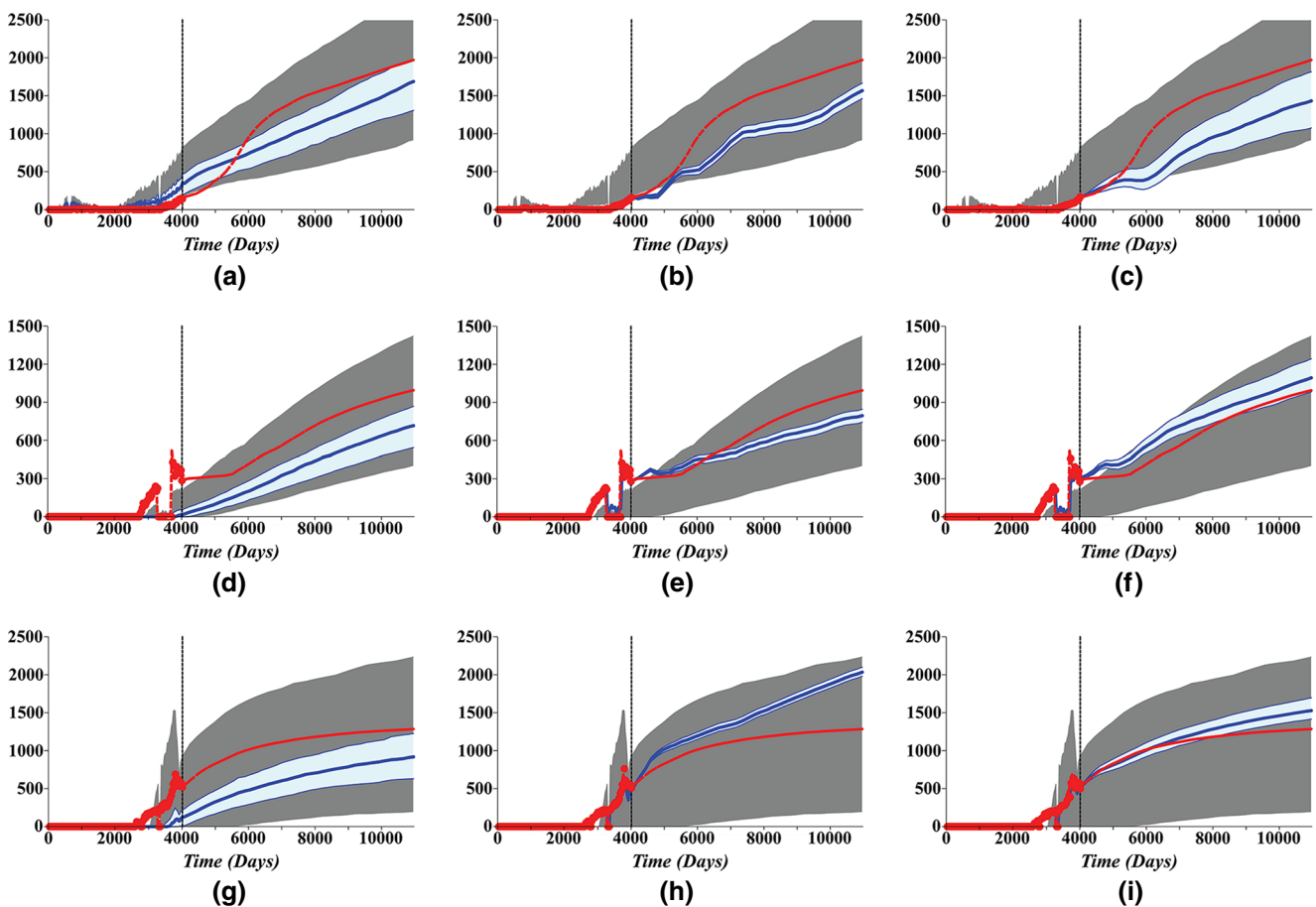


Fig. 9 Water production rate in m^3/days for three wells with good coverage of predictions from the prior ensemble. Well NA3D (first row), RJS19 (second row), and PROD24A (third row). **a, d, g** DSI. **b, e, h** DSI-ESMDA. **c, f, i** DSI-ESMDA with localization. UNISIM-I-H. The red dots are the observed data points and the red line is the prediction from the reference model. The grey region corresponds to

the predictions within the percentiles P10–P90 obtained with the prior ensemble. The light blue region corresponds to the predictions within the percentiles P10–P90 obtained with DSI-ESMDA. The blue line corresponds to the percentile P50 obtained with DSI-ESMDA. The vertical dashed line indicates the end of the history and beginning of the forecast period

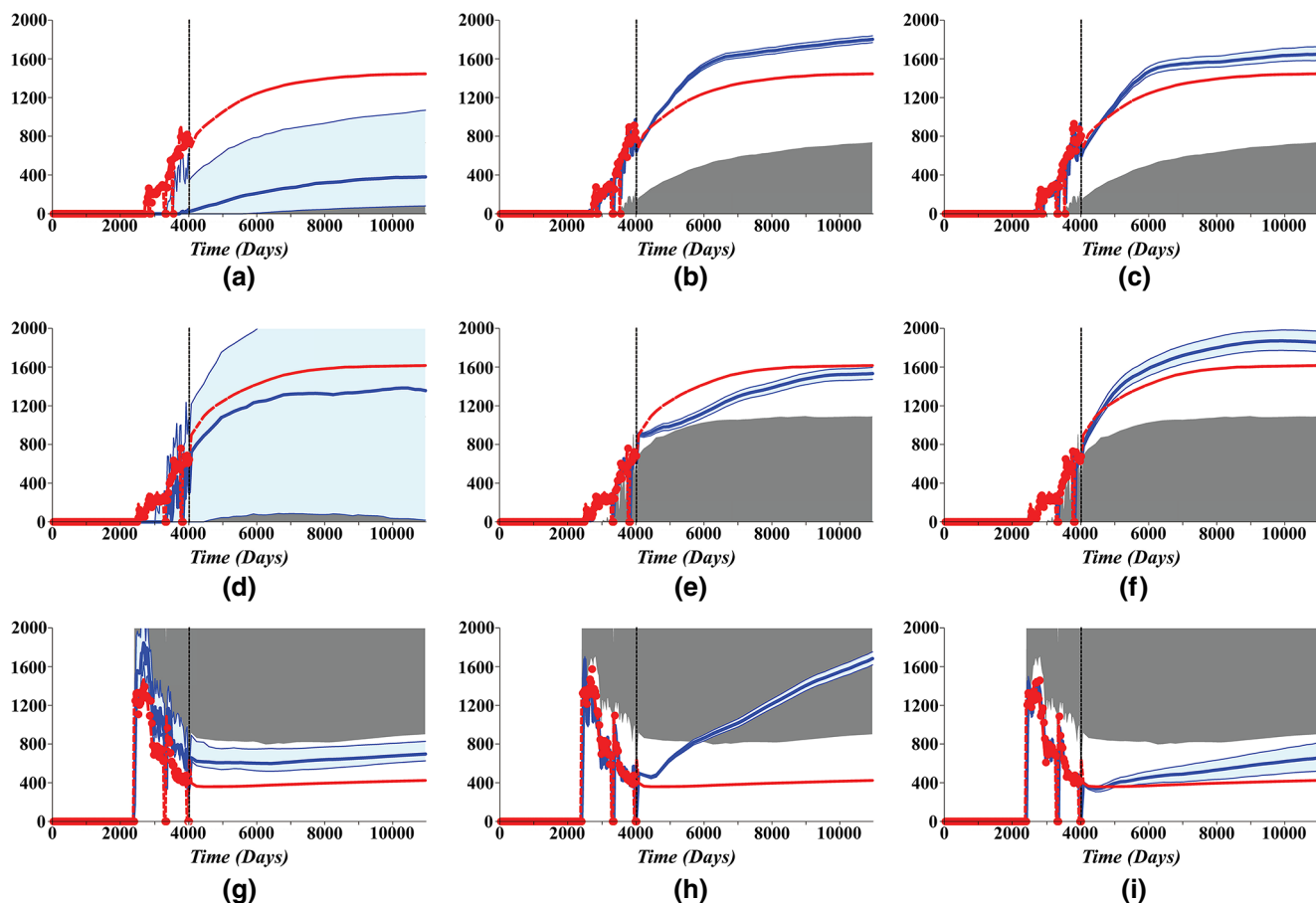


Fig. 10 Water rate in m^3/days for three wells with poor coverage of predictions from the prior ensemble. Well PROD021 (first row), PROD025A (second row), and INJ023 (third row). **a, d, g** DSI. **b, e, h** DSI-ESMDA. **c, f, i** DSI-ESMDA with localization. UNISIM-I-H. The red dots are the observed data points and the red line is the prediction from the reference model. The grey region corresponds to

the predictions within the percentiles P10–P90 obtained with the prior ensemble. The light blue region corresponds to the predictions within the percentiles P10–P90 obtained with DSI-ESMDA. The blue line corresponds to the percentile P50 obtained with DSI-ESMDA. The vertical dashed line indicates the end of the history and beginning of the forecast period

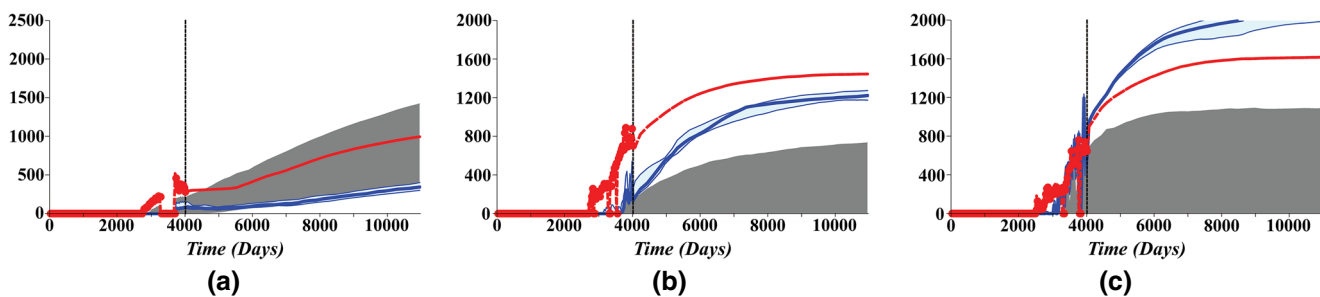


Fig. 11 Water production rate in m^3/days for three wells obtained with DSI with pre-selection. **a** RJS19, **b** PROD021 (first row), and **c** PROD025A. UNISIM-I-H. The red dots are the observed data points and the red line is the prediction from the reference model. The grey region corresponds to the predictions within the percentiles P10–P90

obtained with the prior ensemble. The light blue region corresponds to the predictions within the percentiles P10–P90 obtained with DSI. The blue line corresponds to the percentile P50 obtained with DSI. The vertical dashed line indicates the end of the history and beginning of the forecast period

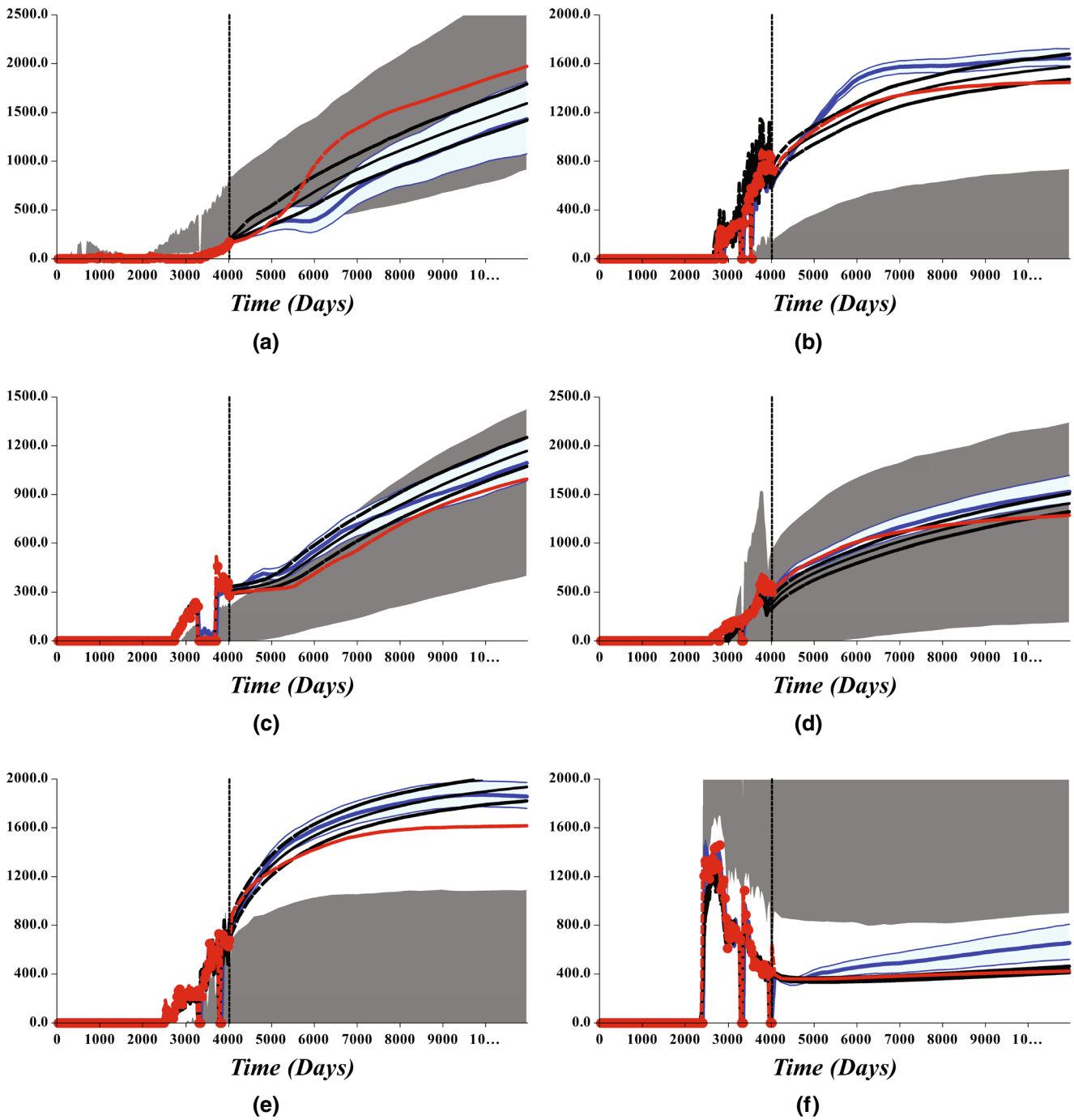


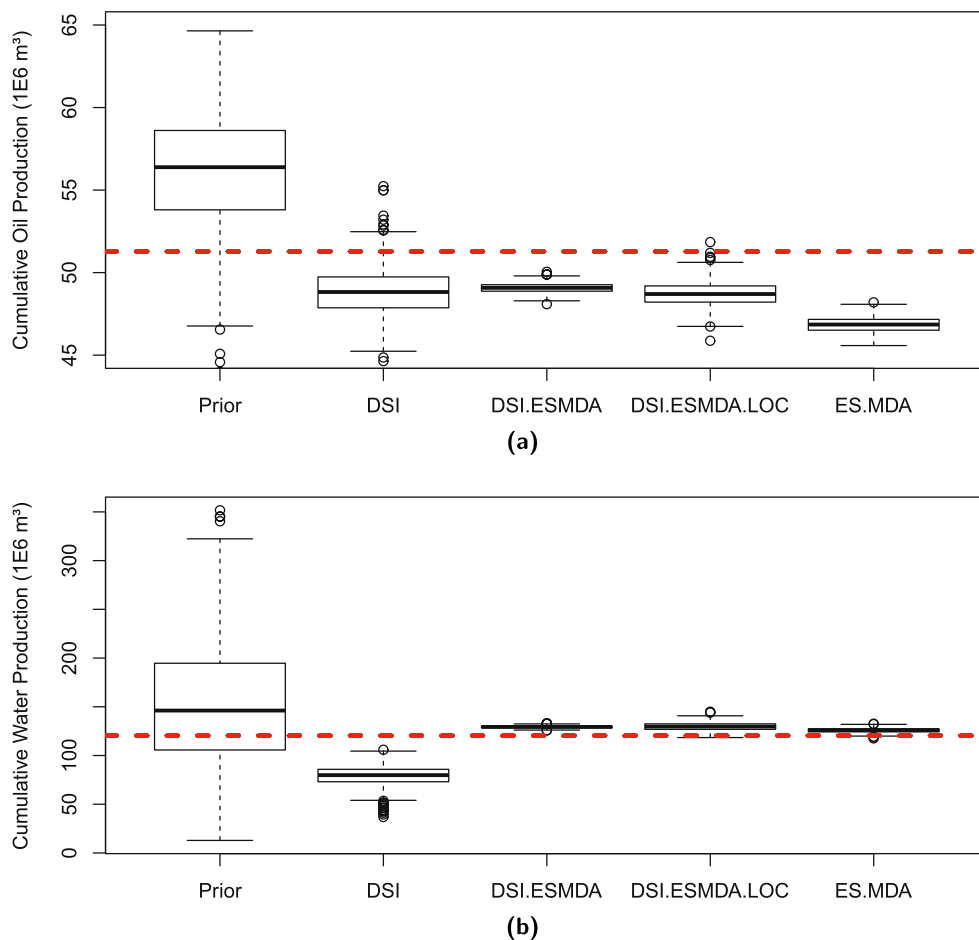
Fig. 12 Water rate in m³/days for six wells (a NA3D, b PROD021, c RJS19, d PROD024A, e PROD025A, f INJ023) obtained with DSI-ESMDA with localization and history matching with ES-MDA. UNISIM-I-H. The red dots are the observed data points and the red line is the prediction from the reference model. The grey region corresponds to the predictions within the percentiles P10–P90 obtained with the prior ensemble. The light blue region corresponds

to the predictions within the percentiles P10–P90 obtained with DSI-ESMDA. The blue line corresponds to the percentile P50 obtained with DSI-ESMDA. The black lines correspond to the percentiles P10, P50, and P90 obtained by the history-matched models using ES-MDA. The vertical dashed line indicates the end of the history and beginning of the forecast period

objective function of this curve is exactly 0.5. Analogously, we have $\mathcal{O}_{N,d} = 2$ and $\mathcal{O}_{N,d} = 4.5$ for two and three standard deviations, respectively. This effectively means that it would be difficult to justify in practice the acceptance

of a data match with $\mathcal{O}_{N,d}$ significantly larger than five. Table 4 shows the mean and standard deviation of $\mathcal{O}_{N,d}$ for each case. Clearly, DSI resulted too large values for $\mathcal{O}_{N,d}$, indicating the lack of convergence of the minimizations.

Fig. 13 Field cumulative production in 10^6 m^3 . **a** Oil and **b** water. UNISIM-I-H. The dashed red line indicates the cumulative production of the reference case



The pre-selection improved the data matches, but the mean $\mathcal{O}_{N,d}$ is still large. The objective functions from DSI-ESMDA and DSI-ESMDA with localization have mean roughly ten times smaller. The history matching case with ES-MDA also resulted in a mean objective function significantly smaller than DSI. In terms of computational cost, the DSI method required approximately 28 h to complete, while the DSI-ESMDA (with and without localization) require around 45 min each.

The prior ensemble for the UNISIM-I-H case was provided by the authors of the benchmark and it is noteworthy that even though this is a synthetic problem, the predictions from the prior ensemble do not span the data from the reference case for several wells. This is clearly not an ideal situation for applying DSI or even any history-matching method. Ideally, the prior ensemble should provide a reasonable estimate of the prior uncertainty, at the very least, it should be able to span the observations. In practice, we should revise the prior ensemble before using for data assimilation. However, here for the purposes of the investigation, we decided to test the methods with this deficient prior ensemble. Figures 9 and 10 show the results obtained by DSI, DSI-ESMDA, and DSI-ESMDA with

localization for three wells with good and poor coverage of the predictions from the prior ensemble, respectively. We selected these wells because they are representative of the results observed in this problem. Figure 9 indicates that DSI failed to match data even for the wells with good prior coverage. As a result, the prediction from the reference model lies outside the predicted P10–P90 range. For the wells with poor coverage (Fig. 10), DSI seems to fail to reduce the uncertainty range properly, for example, the posterior predictions for wells PROD021 and PROD025A have almost the same uncertainty range of the prior ones. Figure 11 shows the results for three selected wells obtained using DSI with pre-selection. The results in this figures illustrate the fact that pre-selection improves the data matches but the ranges of predictions are very narrow. DSI-ESMDA obtained reasonable data matches for all wells; however, the predictions are clearly too restricted as they do not span the reference. DSI-ESMDA with localization improved the results significantly, although it was not able to span the reference predictions, especially for the wells with poor prior coverage. Figure 12 compares the results of DSI-ESMDA with localization with the history matching with ES-MDA. It is interesting to note that

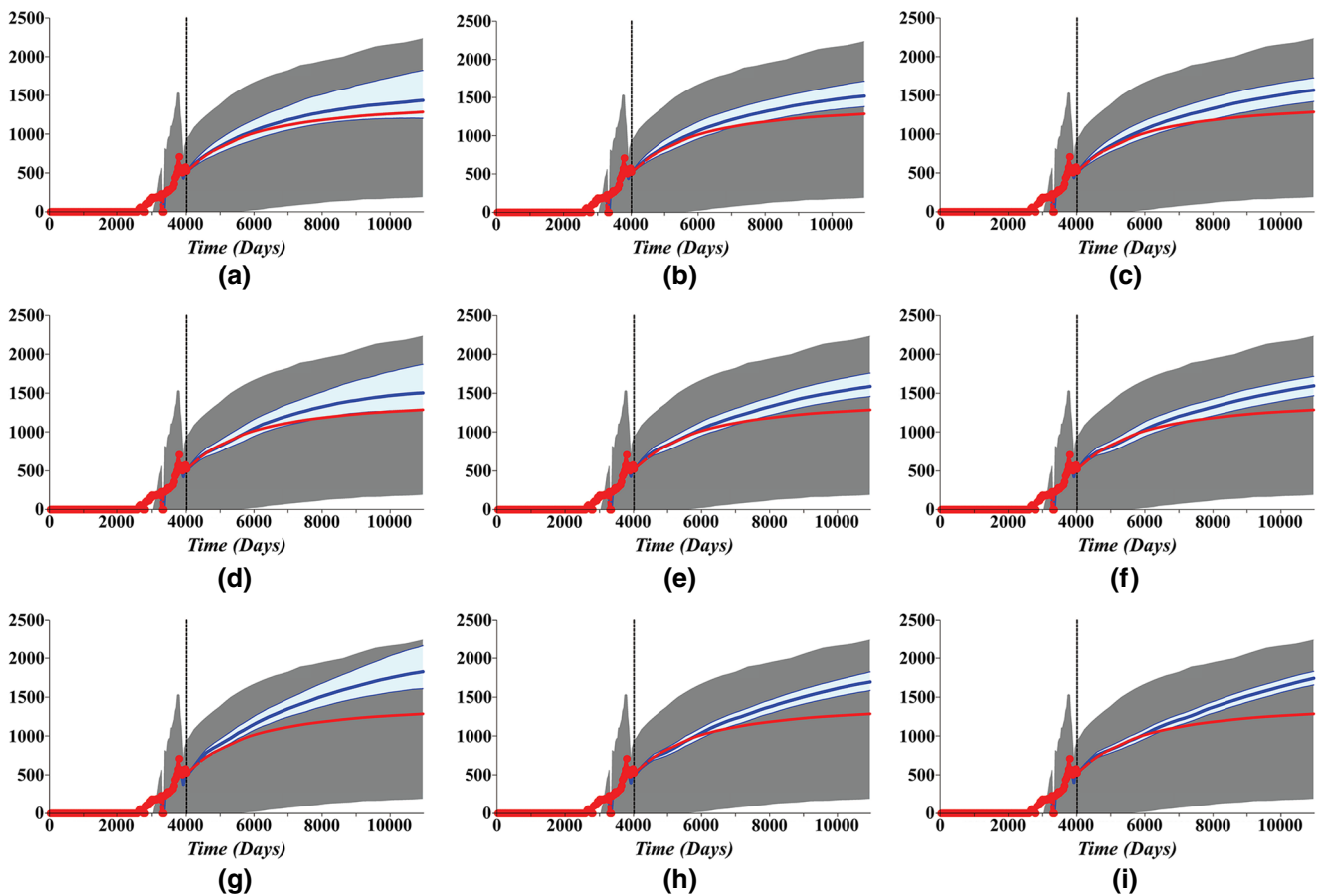


Fig. 14 Water production rate in $m^3/days$ for the PROD24A obtained by DSI-ESMDA with different localization configurations. **a** $L_{x'} = L_{y'} = 1000$ m and $T = 3000$ days, **b** $L_{x'} = L_{y'} = 1000$ m and $T = 6000$ days, **c** $L_{x'} = L_{y'} = 1000$ m and $T = 12000$ days, **d** $L_{x'} = L_{y'} = 2000$ m and $T = 3000$ days, **e** $L_{x'} = L_{y'} = 2000$ m and $T = 6000$ days, **f** $L_{x'} = L_{y'} = 2000$ m and $T = 12000$ days, **g** $L_{x'} = L_{y'} = 4000$ m and $T = 3000$ days, **h** $L_{x'} = L_{y'} = 4000$ m and $T = 6000$ days, and **i** $L_{x'} = L_{y'} = 4000$ m and $T = 12000$ days.

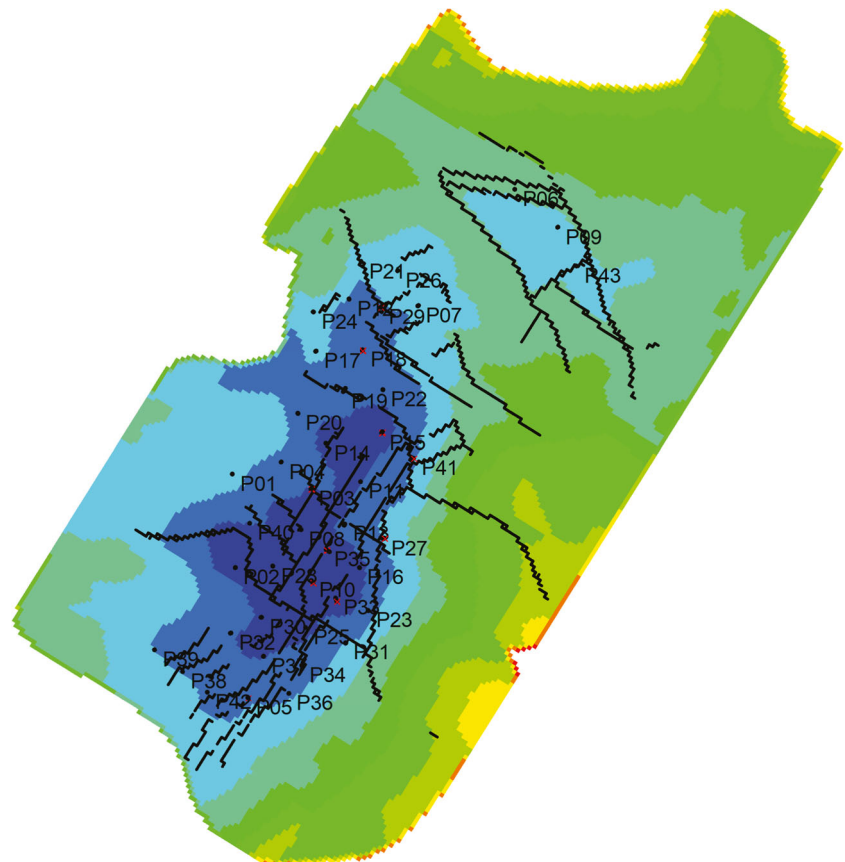
The red dots are the observed data points and the red line is the prediction from the reference model. The grey region corresponds to the predictions within the percentiles P10–P90 obtained with the prior ensemble. The light blue region corresponds to the predictions within the percentiles P10–P90 obtained with DSI-ESMDA. The blue line corresponds to the percentile P50 obtained with DSI-ESMDA. The vertical dashed line indicates the end of the history and beginning of the forecast period

the history matched models also suffer from the lack of representativeness of the prior ensemble; see, for example, the plots (a) and (d) in Fig. 12. Figure 13 shows boxplots of field cumulative oil and water production predicted by the prior ensemble and by each method. The cumulative production from the reference case is also included in this figure for comparisons. Overall, the predictions are biased compared with the reference production, which is probably explained by the problems with the prior ensemble.

The previous results clearly show the importance of using localization with DSI-ESMDA, which requires the selection of appropriate critical lengths. The same is true for standard history-matching applications of ES-MDA, except for the fact that we introduced a time localization in DSI-ESMDA. However, the “optimal” selection of the localization scheme is an open problem, despite the large number of works in this direction; see, for example, [4, 12, 21, 28, 32,

56] and references therein. The result presented for the UNISIN-I-H case assumed a localization with $L_{x'} = L_{y'} = 2000$ m and $T = 6000$ days. Here, we investigate the effect of this choice by running DSI-ESMDA with different configurations of the critical length. Figure 14 summarizes the results in terms of the predicted water rate for the well PROD24A. We selected this well because it has a good prior coverage, but the prediction range obtained by DSI-ESMDA with our original choice of critical length does not span the entire reference forecast, as illustrated in Fig. 9i. The results in Fig. 14 indicate that smaller critical lengths tend to improve the predicted forecast range for this well. In particular, selecting $L_{x'} = L_{y'} = 1000$ m (first row of Fig. 14) provides a noticeable improvement in the result. The cases with larger critical length (third row of Fig. 14), on the other hand, show a deterioration of the results with the predictions deviating from the reference forecast.

Fig. 15 Top view of the field case model showing the large number of wells and faults



4 Field case

The field case corresponds to a large offshore turbidite reservoir in Campos Basin. Currently there are 24 wells producing oil in the field from a total of 43 vertical wells that have been drilled during 18 years of operation. The main recovery mechanism is pressure maintenance by a large aquifer. A total of 500 prior realizations of the reservoir were built by the asset team of the field using the best practices available to date in the Company. These realizations include facies, porosity, permeability, and net-to-gross ratio for each one of 11 producing zones. The engineers of the field also selected the oil-water relative permeability curves as parameters with relevant uncertainty. The relative permeability curves in six different regions

of the model were parameterized using Corey functions with the exponents and the maximum relative permeability of water selected as uncertainty parameters with Gaussian prior distributions. The model has approximately 500,000 active gridblocks (total of $114 \times 238 \times 60$ with an average size of $75 \text{ m} \times 75 \text{ m} \times 2 \text{ m}$). This field has a large number of sealing faults as illustrated in Fig. 15. All wells are controlled by specified total liquid production rate during the historical period and the forecast. The observed data corresponds to monthly measurements of oil and water rate at the well with a total of 26703 data points.

Unlike the UNISIM-I-H case, the predictions from the prior ensemble cover reasonably well the observed data for the large majority of the wells in the field, which creates a more favorable situation for the application of the method. Once again, the optimizations in our DSI implementation failed to converge, providing unreasonable results for this field problem. DSI-ESMDA without localization showed severe reduction in the posterior variance, showing indications of ensemble collapse. Both results were not included in this paper. The only acceptable results obtained in this field are the ones from DSI-ESMDA with localization. In this case, we use the same configuration of the case UNISIM-I-H to define the critical length for localization, that is, $L_{x'} = L_{y'} = 2000 \text{ m}$ and

Table 5 Normalized data-mismatch objective function

Case	Mean	Standard deviation
Prior	32.19	11.309
DSI-ESMDA (with localization)	0.55	0.004
ES-MDA	4.11	0.581

Field case

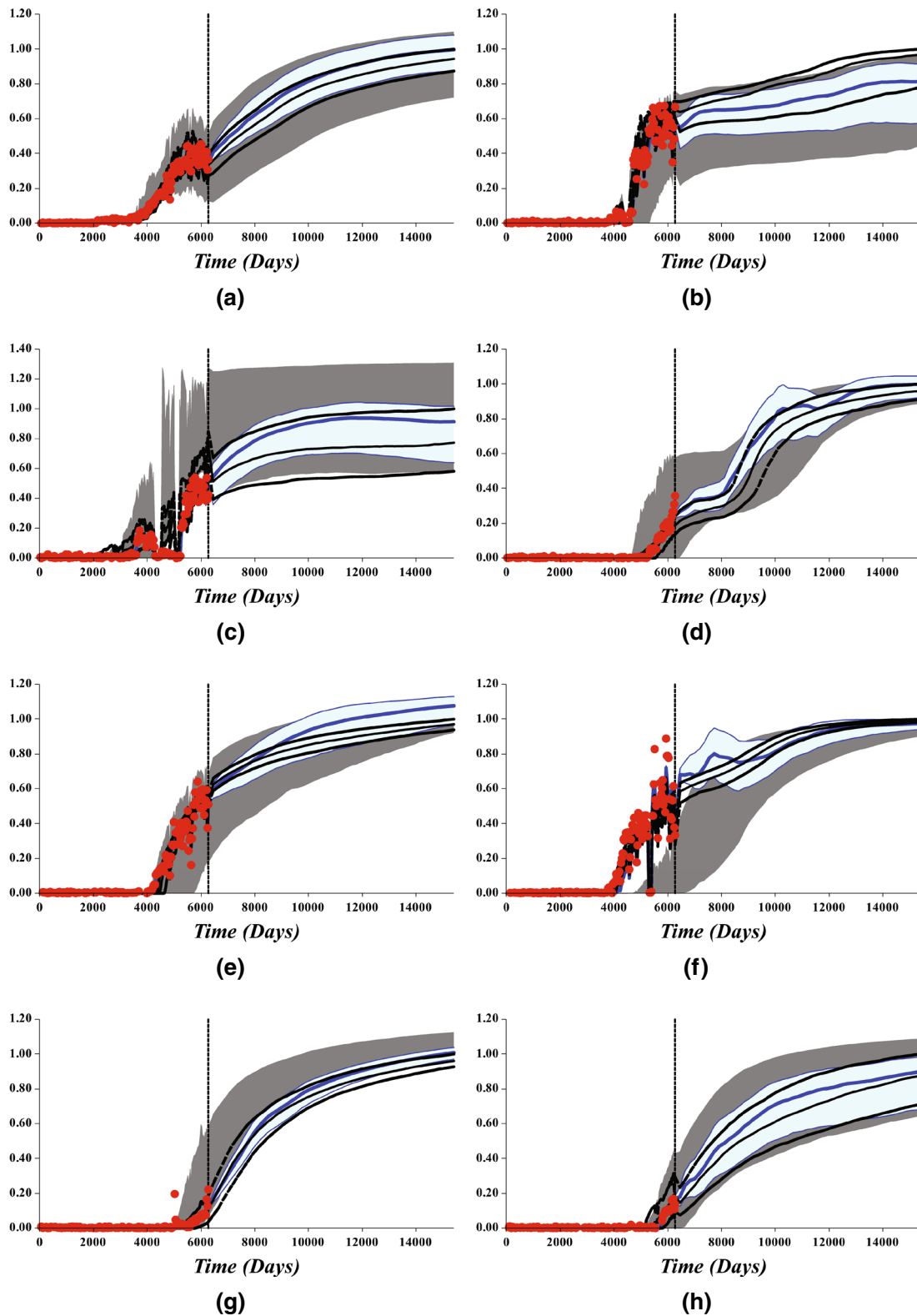


Fig. 16 Normalized water production rate obtained with DSI-ESMDA with localization. **a** Total field production. **b–h** seven selected wells. Field case. The red dots are the observed data points and the red line is the prediction from the reference model. The grey region corresponds to the predictions within the percentiles P10–P90 obtained with the prior ensemble. The light blue region corresponds to the predictions

within the percentiles P10–P90 obtained with DSI-ESMDA. The blue line corresponds to the percentile P50 obtained with DSI-ESMDA. The black lines correspond to the percentiles P10, P50, and P90 obtained by the history-matched models using ES-MDA. The vertical dashed line indicates the end of the history and beginning of the forecast period

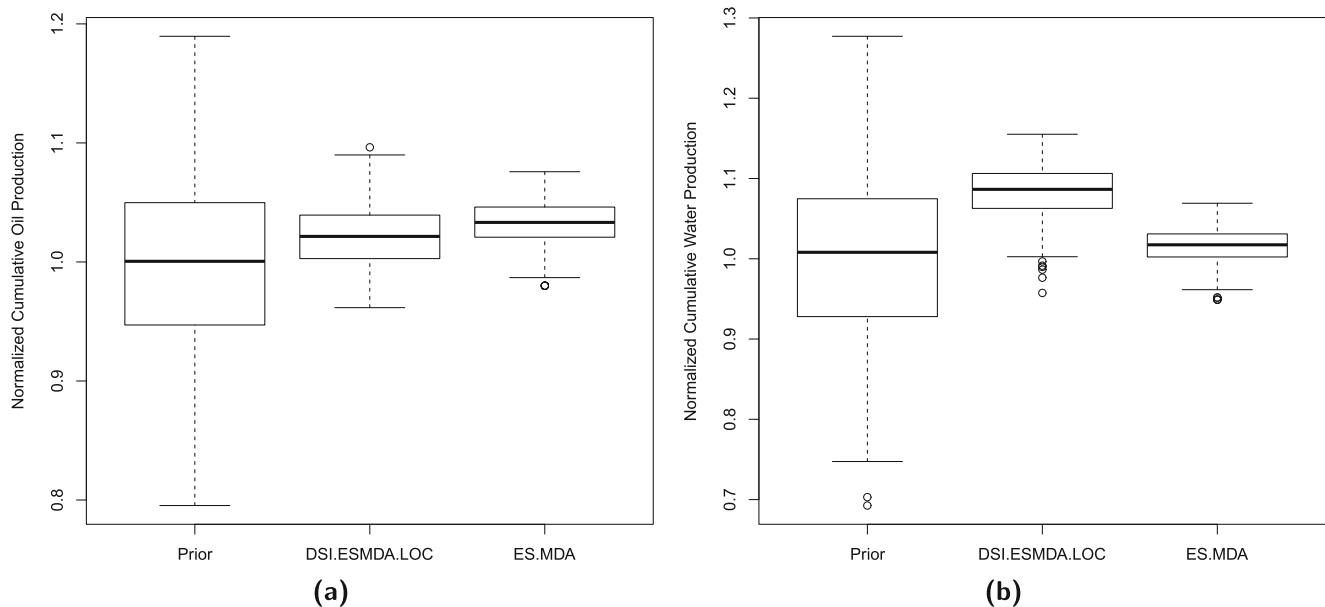


Fig. 17 Normalized field cumulative production. **a** Oil and **b** water. Field case

$T = 6000$ days. For comparisons, we history matched the same ensemble of models using ES-MDA with $N_d = 4$ and localization with a critical length of 2000 m. Table 5 presents the mean and standard deviation of the normalized data mismatch objective function. The results in this table indicate that DSI-ESMDA with localization obtained an excellent data match, with mean objective function of 0.55. The history matching with ES-MDA also improved significantly the data match, but with higher values of objective function. Figure 16 shows the water production rate for eight representative wells of the field. The values in this figure were normalized to preserve the confidentiality of the information. Overall, the forecasts after DSI-ESMDA with localization are comparable with the forecasts from the history-matched models, despite some clear differences observed in some wells. Figure 17 shows boxplots of field cumulative oil and water production. The main difference between DSI-ESMDA and ES-MDA is for the cumulative water production, where DSI-ESMDA predicts a higher water production in the field.

5 Conclusions

This paper introduced a new DSI implementation based on the data assimilation method ES-MDA. The new DSI-ESMDA was compared with the original DSI proposed in [51, 52], which is based on PCA to reparameterize the predicted data from a prior ensemble combined to RML for sampling. The new implementation preserves the main advantage of the original DSI, namely, it is able to provide an ensemble of production forecasts requiring

reservoir simulations only for a prior ensemble of models. We compared the DSI-ESMDA with the original DSI in two synthetic reservoir problems. We also applied DSI-ESMDA to a real field case with long production history and large number of wells. Based on the results for these test problems, we can summarize the following conclusions:

- The proposed DSI-ESMDA is computationally faster than the original DSI. Even though the time to execute the reservoir simulations for the prior models tends to be dominant in both methods, the difference in the computational time for the inversion can be relevant for large problems. For example, for the UNISIM-I-H case, the original DSI required approximately 28 h in a stand-alone computer while the DSI-ESMDA required only 45 min.
- The performance of both DSI implementations is highly dependent on the ability of the prior ensemble to provide reasonable estimates of the prior uncertainty. The same is also true for the more traditional model inversion with ES-MDA.
- The results indicate that the proposed DSI-ESMDA is more robust than the original DSI. In our tests, the optimizations required by DSI have not converged properly for the UNISIM-I-H case. As a result, DSI failed to obtain acceptable data matches and reasonable production forecasts. DSI-ESMDA, on the other hand, was able to match the observed data for all wells.
- The use of spatial and temporal localization improved significantly the results of the DSI-ESMDA method when the number of data point is large. However,

localization requires an appropriate choice of the critical length, which is problem-dependent.

- The proposed DSI-ESMDA method with localization obtained forecasts of production comparable with the forecast provided by the posterior models generated with history matching using ES-MDA for a real field problem.
- The proposed method is very simple to implement. It does not require data transformations and it is straightforward integrate with different types of data and models.

Acknowledgments The authors would like to thank Petrobras for supporting this research and for the permission to publish this paper.

References

1. Abadpour, A., Adejare, M., Chugunova, T., Mathieu, H., Haller, N.: Integrated geo-modeling and ensemble history matching of complex fractured carbonate and deep offshore turbidite fields, generation of several geologically coherent solutions using ensemble methods. In: Proceedings of the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, number SPE-193028-MS. <https://doi.org/10.2118/193028-MS> (2018)
2. Audet, C., Jr., J.E.D.: Mesh adaptive direct search algorithms for constrained optimization. *SIAM J. Optim.* **17**(1), 188–217 (2006). <https://doi.org/10.1137/040603371>
3. Avansi, G.D., Schiozer, D.J.: UNISIM-I: synthetic model for reservoir development and management applications. *Int. J. Model. Simul. Petrol. Indust.* **9**(1):21–30. <http://www.ijmspi.org/ojs/index.php/ijmspi/article/view/152> (2015)
4. Chen, Y., Oliver, D.S.: Cross-covariance and localization for enKF in multiphase flow data assimilation. *Comput. Geosci.* **14**(4), 579–601 (2010). <https://doi.org/10.1007/s10596-009-9174-6>
5. Chen, Y., Oliver, D.S.: Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math. Geosci.* **44**(1), 1–26 (2012). <https://doi.org/10.1007/s11004-011-9376-z>
6. Chen, Y., Oliver, D.S.: Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput. Geosci.* **17**, 689–703 (2013). <https://doi.org/10.1007/s10596-013-9351-5>
7. Chen, Y., Oliver, D.S.: History matching of the Norne full-field model with an iterative ensemble smoother. *SPE Reserv. Eval. Eng.* **17**(2). <https://doi.org/10.2118/164902-PA> (2014)
8. Emerick, A.A.: Estimation of pressure and saturation fields from time-lapse impedance data using the ensemble smoother. *J. Geophys. Eng.* **11**(3):035007, 2014. <https://doi.org/10.1088/1742-2132/11/3/035007>
9. Emerick, A.A.: Analysis of the performance of ensemble-based assimilation of production and seismic data. *J. Pet. Sci. Eng.* **139**, 219–239 (2016). <https://doi.org/10.1016/j.petrol.2016.01.029>
10. Emerick, A.A.: Analysis of geometric selection of the data-error covariance inflation for ES-MDA. arXiv:1812.00924v1 [math.NA]. arXiv:1812.00924 (2018)
11. Emerick, A.A., Reynolds, A.C.: History matching a field case using the ensemble Kalman filter with covariance localization. *SPE Reserv. Eval. Eng.* **14**(4), 423–432 (2011a). <https://doi.org/10.2118/141216-PA>
12. 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 (2011b). <https://doi.org/10.1007/s10596-010-9198-y>
13. 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). <https://doi.org/10.1007/s10596-012-9275-5>
14. Emerick, A.A., Reynolds, A.C.: Investigation on the sampling performance of ensemble-based methods with a simple reservoir model. *Comput. Geosci.* **17**(2), 325–350 (2013a). <https://doi.org/10.1007/s10596-012-9333-z>
15. Emerick, A.A., Reynolds, A.C.: Ensemble smoother with multiple data assimilation. *Comput. Geosci.* **55**, 3–15 (2013b). <https://doi.org/10.1016/j.cageo.2012.03.011>
16. Emerick, A.A., Reynolds, A.C.: History matching of production and seismic data for a real field case using the ensemble smoother with multiple data assimilation. In: Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, number SPE-163675-MS. <https://doi.org/10.2118/163675-MS> (2013c)
17. Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5), 10143–10162 (1994). <https://doi.org/10.1029/94JC00572>
18. Evensen, G.: Sampling strategies and square root analysis schemes for the enKF. *Ocean Dyn.* **54**(6), 539–560 (2004). <https://doi.org/10.1007/s10236-004-0099-2>
19. Evensen, G.: Data Assimilation the Ensemble Kalman Filter. Springer, Berlin (2007)
20. Evensen, G.: Analysis of iterative ensemble smoothers for solving inverse problems *Computational Geosciences.* <https://doi.org/10.1007/s10596-018-9731-y> (2018)
21. 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). <https://doi.org/10.1016/j.jmva.2006.08.003>
22. Gaspari, G., Cohn, S.E.: Construction of correlation functions in two and three dimensions. *Q. J. Roy. Meteorol. Soc.* **125**(554), 723–757 (1999). <https://doi.org/10.1002/qj.49712555417>
23. He, J., Sarma, P., Bhark, E., Tanaka, S., Chen, B., Wen, X.-H., Kamath, J.: Quantifying expected uncertainty reduction and value of information using ensemble-variance analysis. *SPE J.* **23**(2). <https://doi.org/10.2118/182609-PA> (2018)
24. Houtekamer, P.L., Mitchell, H.L.: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**(1):123–137. [10.1175/1520-0493\(2001\)129<0123:ASEKFFj>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFFj>2.0.CO;2) (2001)
25. Jeong, H., Sun, A.Y., Lee, J., Min, B.: A learning-based data-driven forecast approach for predicting future reservoir performance. *Adv. Water Resour.* **118**, 95–109 (2018). <https://doi.org/10.1016/j.advwatres.2018.05.015>
26. Jiang, S.: Data-Space Inversion with Variable Well Controls in the Prediction Period. Master's thesis, Stanford University (2018)
27. Krishnamurti, T.N., Kishtawal, C.M., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E.: Multimodel ensemble forecasts for weather and seasonal climate. *J. Clim.* **13**(23):4196–4216. [https://doi.org/10.1175/1520-0442\(2000\)013<4196:MEFFWA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2) (2000)
28. Lacerda, J.M., Emerick, A.A., Pires, A.P.: Methods to mitigate loss of variance due to sampling errors in ensemble data assimilation with non-local model parameters. *J. Pet. Sci. Eng.* **172**, 690–706 (2019). <https://doi.org/10.1016/j.petrol.2018.08.056>

29. Le, D.H., Emerick, A.A., Reynolds, A.C.: An adaptive ensemble smoother with multiple data assimilation for assisted history matching. *SPE J.* **21**(6), 2195–2207 (2016). <https://doi.org/10.2118/173214-PA>
30. Lorentzen, R.J., Luo, X., Bhakta, T., Valestrand, R.: History Matching the Full Norne Field Model Using Seismic and Production Data. *SPE Journal*, Preprint. <https://doi.org/10.2118/194205-PA> (2019)
31. Luo, X., Stordal, A.S., Lorentzen, R.J., Nævdal, G.: Iterative ensemble smoother as an approximate solution to a regularized minimum-average-cost problem: Theory and applications. *SPE J.* **20**(5). <https://doi.org/10.2118/176023-PA> (2015)
32. Luo, X., Bhakta, T., Nævdal, G.: Correlation-based adaptive localization with applications to ensemble-based 4D-seismic history matching. *SPE J.* **23**(2), 396–427 (2018). <https://doi.org/10.2118/185936-PA>
33. Ma, X., Hetz, G., Wang, X., Bi, L., Stern, D., Hoda, N.: A robust iterative ensemble smoother method for efficient history matching and uncertainty quantification. In: Proceedings of the SPE Reservoir Simulation Conference, Montgomery, Texas, USA, number SPE-182693-MS. <https://doi.org/10.2118/182693-MS> (2017)
34. Mallet, V., Stoltz, G., Mauricette, B.: Ozone ensemble forecast with machine learning algorithms. *J. Geophys. Res. Atmos.* **114**(D5). <https://doi.org/10.1029/2008JD009978> (2009)
35. Maschio, C., Avansi, G.D., Santos, A.A., Schiozer, D.J.: UNISIM-I-H: Case Study for History Matching. Dataset. www.unisim.cepetro.unicamp.br/benchmarks/br/unisim-i/unisim-i-h (2013)
36. Maucec, M., Ravanelli, F.M.D.M., Lyngra, S., Zhang, S.J., Alramadhan, A.A., Abdelhamid, O.A., Al-Garni, S.A.: Ensemble-based assisted history matching with rigorous uncertainty quantification applied to a naturally fractured carbonate reservoir. In: Proceedings of the SPE Annual Technical Conference and Exhibition, Dubai, UAE, number SPE-181325-MS. <https://doi.org/10.2118/181325-MS> (2016)
37. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (2006)
38. Oliver, D.S., Alfonzo, M.: Calibration of imperfect models to biased observations. *Comput. Geosci.* **22**, 145–161 (2018). <https://doi.org/10.1007/s10596-017-9678-4>
39. Oliver, D.S., Chen, Y.: Recent progress on reservoir history matching: a review. *Comput. Geosci.* **15**(1), 185–221 (2011). <https://doi.org/10.1007/s10596-010-9194-2>
40. Oliver, D.S., Reynolds, A.C., Liu, N.: Inverse Theory for Petroleum Reservoir Characterization and History Matching. Cambridge University Press, Cambridge (2008)
41. Pagowski, M., Grell, G.A., McKeen, S.A., Dévényi, W.J.M., Bouchet, V.S., Gong, W.F., Mchenry, J.N., Peckham, S., Mcqueen, J.T., Moffet, R., Tang, Y.: A simple method to improve ensemble-based ozone forecasts. *Geophys. Res. Lett.* **320**(7). <https://doi.org/10.1029/2004GL022305> (2005)
42. Rafiee, J., Reynolds, A.C.: Theoretical and efficient practical procedures for the generation of inflation factors for ES-MDA. *Inverse Probl.* **33**(11):115003. <https://doi.org/10.1088/1361-6420/aa8cb2> (2017)
43. Reynolds, A.C., Zafari, M., Li, G.: Iterative forms of the ensemble Kalman filter. In: Proceedings of 10th European Conference on the Mathematics of Oil Recovery, Amsterdam. <https://doi.org/10.3997/2214-4609.201402496> (2006)
44. Satija, A., Caers, J.: Direct forecasting of subsurface flow response from non-linear dynamic data by linear least-squares in canonical functional principal component space. *Adv. Water Resour.* **77**, 69–81 (2015). <https://doi.org/10.1016/j.advwatres.2015.01.002>
45. Satija, A., Caers, J.: Direct forecasting of reservoir performance using production data without history matching. *Comput. Geosci.* **21**(2), 315–333 (2017). <https://doi.org/10.1007/s10596-017-9614-7>
46. Scheidt, C., Renard, P., Caers, J.: Prediction-focused subsurface modeling Investigating the need for accuracy in flow-based inverse modeling. *Math. Geosci.* **47**(2), 173–191 (2015). <https://doi.org/10.1007/s11004-014-9521-6>
47. Skjervheim, J.-A., Evensen, G., Hove, J., Vabø, J.G.: An ensemble smoother for assisted history matching. In: Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands, Texas, number SPE-141929-MS. <https://doi.org/10.2118/141929-MS> (2011)
48. Souza, C.R.: Accord.NET Framework Broyden-Fletcher-Goldfarb-Shanno class. http://accord-framework.net/docs/html/T_Accord_Math_Optimization_BroydenFletcherGoldfarbShanno.htm (2017)
49. Stordal, A.S.: Iterative Bayesian inversion with Gaussian mixtures: finite sample implementation and large sample asymptotics. *Comput. Geosci.* **19**(1), 1–15 (2014). <https://doi.org/10.1007/s10596-014-9444-9>
50. Stordal, A.S., Elsheikh, A.H.: Iterative ensemble smoothers in the annealed importance sampling framework. *Adv. Water Resour.* **86**, 231–239 (2015). <https://doi.org/10.1016/j.advwatres.2015.09.030>
51. Sun, W., Durlafsky, L.J.: A new data-space inversion procedure for efficient uncertainty quantification in subsurface flow problems. *Math. Geosci.* **49**, 679–715 (2017). <https://doi.org/10.1007/s11004-016-9672-8>
52. Sun, W., Hui, M.-H., Durlafsky, L.J.: Production forecasting and uncertainty quantification for naturally fractured reservoirs using a new data-space inversion procedure. *Computational Geosciences*, Online. <https://doi.org/10.1007/s10596-017-9633-4> (2017a)
53. Sun, W., Vink, J.C., Gao, G.: A practical method to mitigate spurious uncertainty reduction in history matching workflows with imperfect reservoir models. In: Proceedings of the SPE Reservoir Simulation Conference, Montgomery, Texas, number SPE-182599-MS. <https://doi.org/10.2118/182599-MS> (2017b)
54. Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, Philadelphia (2005)
55. van Leeuwen, P.J., Evensen, G.: Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Weather Rev.* **124**:2898–2913. [https://doi.org/10.1175/1520-0493\(1996\)124<2898:DAAIMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2) (1996)
56. Zhang, Y., Oliver, D.S.: Improving the ensemble estimate of the Kalman gain by bootstrap sampling. *Math. Geosci.* **42**, 327–345 (2010). <https://doi.org/10.1007/s11004-010-9267-8>

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