Formulating the history matching problem with consistent error statistics

Geir Evensen[1,2]

Abstract

It is common to formulate the history-matching problem using Bayes' theorem. From Bayes', the conditional probability density function (pdf) of the uncertain model parameters is proportional to the prior pdf of the model parameters, multiplied by the likelihood of the measurements. The static model parameters are random variables characterizing the reservoir model while the observations include, e.g., historical rates of oil, gas, and water produced from the wells. The reservoir prediction model is assumed perfect, and there are no errors besides those in the static parameters. However, this formulation is flawed. The historical rate data only approximately represent the real production of the reservoir and contain errors. History-matching methods usually take these errors into account in the conditioning but neglect them when forcing the simulation model by the observed rates during the historical integration. Thus, the model prediction depends on some of the same data used in the conditioning. The paper presents a formulation of Bayes' theorem that considers the data dependency of the simulation model. In the new formulation, one must update both the poorly known model parameters and the rate-data errors. The result is an improved posterior ensemble of prediction models that better cover the observations with more substantial and realistic uncertainty. The implementation accounts correctly for correlated measurement errors and demonstrates the critical role of these correlations in reducing the update's magnitude. The paper also shows the consistency of the subspace inversion scheme by Evensen (Ocean Dyn. **54**, 539–560 2004) in the case with correlated measurement errors and demonstrates its accuracy when using a "larger" ensemble of perturbations to represent the measurement error covariance matrix.

Keywords Iterative ensemble smoother · Subspace EnRML · Correlated measurement errors · Reservoir history matching

1 Introduction

Reservoir simulators are typically forced by measured oil, gas, and water rates during the historical simulation to generate the predicted production [1]. This forcing can be applied directly using individual measured rates (e.g., the oil rate) or maybe more commonly by the produced reservoir volume as computed from the fluid rate data (i.e., the total volume of fluids produced by a well). Thus, one attempts to prescribe the observed production and injection from each well to the model, and if it is physically possible, the model will produce the prescribed rates. Sometimes the model is unable to satisfy the imposed production constraints. In such cases, the purpose of the conditioning is to update the reservoir description so that the model can produce the observed rates.

The approach described above is standard in most history-matching studies based on ensemble methods, see e.g., [4, 6, 21, 25]. Additionally, the inclusion of seismic information complements the information contained in the historical rates [5, 19]. Lately, the conditioning on seismic-information has become more efficient with adaptive localization schemes, such as the one proposed by Luo and Bhakta [20]. Petroleum companies now use ensemble conditioning methods operationally to manage their reservoir-model workflows [13, 23]. This paper aims to enhance the understanding of the formulation and solution of ensemble history-matching methods.

We will now formulate the history matching problem while allowing the historical rates to contain errors. Start by storing all the observed production and injection rates

✉ Geir Evensen
geir.evensen@norceresearch.no

1 NORCE (Norwegian Research Center), Bergen, Norway

2 Nansen Environmental and Remote Sensing Center, Bergen, Norway

of oil, gas, and water, for the whole history of the model, in the control vector, $\mathbf{u} \in \Re^{m_u}$, and define the $\mathbf{x} \in \Re^n$, to include all the uncertain parameters that characterize the reservoir. We can then write a model for the simulated rates of produced oil, water, and gas, i.e., the predicted measurements, $\mathbf{y} \in \Re^m$, as

$$\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{u}). \tag{1}$$

Thus, the model depends on the uncertain parameters in $\mathbf{x} \in \Re^n$ and the uncertain rates in the vector of controls, $\mathbf{u}$. Here m_u is the dimension of the control variable, $\mathbf{u}$, m is the dimension of the possible subset of rate data that we condition on, and n is the dimension of the vector of uncertain model parameters.

Now, characterize the random variables $\mathbf{x}$ and $\mathbf{u}$ by a prior joint probability density function (pdf), $f(\mathbf{x}, \mathbf{u})$. The model in Eq. 1 is then a stochastic equation starting from an uncertain set of parameters, $\mathbf{x}$, and forced by the uncertain controls, $\mathbf{u}$. A joint pdf for the model prediction and the poorly known inputs is

$$
\begin{aligned}
f(\mathbf{y}, \mathbf{x}, \mathbf{u}) &= f(\mathbf{x}, \mathbf{u}) \, f(\mathbf{y} - \mathbf{g}(\mathbf{x}, \mathbf{u})) \\
&= f(\mathbf{x}, \mathbf{u}) \, \delta(\mathbf{y} - \mathbf{g}(\mathbf{x}, \mathbf{u})),
\end{aligned} \tag{2}
$$

with δ being the Dirac-delta function. One traditionally neglects the stochastic noise in the measured rates, which results in an underestimated prediction uncertainty for $\mathbf{y}$. The impact of this noise can be significant, particularly in the case with time-correlated errors in the forcing $\mathbf{u}$, which yields significantly more substantial error-variance growth. In the current study, we neglect other model errors, such as those that may arise from numerical discretization. However, it is possible to include these additional errors using the algorithm from Evensen [10].

Define the relation between the measured rate data used in the conditioning, $\mathbf{d} \in \Re^m$, and the actual production, $\mathbf{y}$, by

$$\mathbf{d} = \mathbf{y} + \mathbf{e}, \tag{3}$$

where $\mathbf{e} \in \Re^m$ denote the errors in the measured rates, $\mathbf{d}$. We differentiate between the data conditioned on, $\mathbf{d}$, and the controls used to force the model, $\mathbf{u}$, (although $\mathbf{d}$ may typically include a subset of $\mathbf{u}$).

Now, write the following joint pdf:

$$
\begin{aligned}
f(\mathbf{y}, \mathbf{x}, \mathbf{u}, \mathbf{d}) &= f(\mathbf{y}, \mathbf{x}, \mathbf{u} \,|\, \mathbf{d}) f(\mathbf{d}) \\
&= f(\mathbf{d} \,|\, \mathbf{x}, \mathbf{y}, \mathbf{u}) f(\mathbf{y}, \mathbf{x}, \mathbf{u}).
\end{aligned} \tag{4}
$$

From this equation, one obtains Bayes' theorem

$$f(\mathbf{y}, \mathbf{x}, \mathbf{u} \,|\, \mathbf{d}) = \frac{f(\mathbf{d} \,|\, \mathbf{y}, \mathbf{x}, \mathbf{u}) f(\mathbf{y}, \mathbf{x}, \mathbf{u})}{f(\mathbf{d})}. \tag{5}$$

When deriving the ensemble methods, it is not required to include the normalizing denominator, and one can instead write (5) as

$$f(\mathbf{y}, \mathbf{x}, \mathbf{u} \,|\, \mathbf{d}) \propto f(\mathbf{d} \,|\, \mathbf{y}, \mathbf{x}, \mathbf{u}) f(\mathbf{y}, \mathbf{x}, \mathbf{u}). \tag{6}$$

The prior-term defined in Eq. 2 is straightforward to evaluate. The measurements, $\mathbf{d}$, only depends on $\mathbf{y}$, and the likelihood becomes $f(\mathbf{d} \,|\, \mathbf{x}, \mathbf{y}, \mathbf{u}) = f(\mathbf{d} \,|\, \mathbf{y})$, as usual. Still, it is necessary to separate the control variables used to force the model and the data on which we condition the posterior.

Then, using (2), it is possible to write (6) as

$$f(\mathbf{x}, \mathbf{y}, \mathbf{u} \,|\, \mathbf{d}) \propto f(\mathbf{d} \,|\, \mathbf{y}) \delta(\mathbf{y} - \mathbf{g}(\mathbf{x}, \mathbf{u})) f(\mathbf{x}, \mathbf{u}). \tag{7}$$

The marginal is obtained by integration over $\mathbf{y}$ and becomes

$$f(\mathbf{x}, \mathbf{u} \,|\, \mathbf{d}) \propto f(\mathbf{d} \,|\, \mathbf{g}(\mathbf{x}, \mathbf{u})) f(\mathbf{x}, \mathbf{u}), \tag{8}$$

and, together with Eq. 1 for the model prediction, constitutes the reformulated history-matching problem.

The formulation in Eq. 8 relates to the one discussed by Evensen [10], who formulated the history-matching problem when including model errors. Here the "model errors" correspond to the errors in the rate data. As in Evensen [10], we end up having to estimate both the parameters in $\mathbf{x}$ and the rate data in $\mathbf{u}$.

Note that when using ensemble methods, the parameter-space dimension is not of crucial importance. The ensemble methods compute the sensitivities between measurements and parameters, and parameters are updated only if the sensitivity is unequal to zero. Unimportant parameters will retain their prior variance. Thus, the addition of uncertain controls to the estimation does not increase the complexity of the problem.

2 Solution method with normal priors

Define the new random variable

$$\mathbf{z}^{\mathrm{T}} = (\mathbf{x}^{\mathrm{T}}, \mathbf{u}^{\mathrm{T}}) \in \Re^{n+m_u}, \tag{9}$$

and rewrite model as

$$\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{u}) = \mathbf{g}(\mathbf{z}). \tag{10}$$

It is then possible to write the problem stated by Eq. 8 as

$$f(\mathbf{z} \,|\, \mathbf{d}) \propto f(\mathbf{z}) f(\mathbf{d} \,|\, \mathbf{g}(\mathbf{z})). \tag{11}$$

Now, assume normal priors

$$f(\mathbf{z}) = \mathcal{N}(\mathbf{z}^{\mathrm{f}}, \mathbf{C}_{zz}), \tag{12}$$

$$f(\mathbf{d} \,|\, \mathbf{g}(\mathbf{z})) = f(\mathbf{e}) = \mathcal{N}(\mathbf{0}, \mathbf{C}_{dd}), \tag{13}$$

where

$$\mathbf{C}_{zz} = \begin{pmatrix} \mathbf{C}_{xx} & \mathbf{C}_{ux} \\ \mathbf{C}_{xu} & \mathbf{C}_{uu} \end{pmatrix}, \tag{14}$$

and Eq. 11 becomes

$$
\begin{aligned}
f(\mathbf{z} \,|\, \mathbf{d}) \propto \, &\exp\left\{ -\tfrac{1}{2}(\mathbf{z} - \mathbf{z}^{\mathrm{f}})^{\mathrm{T}} \mathbf{C}_{zz}^{-1}(\mathbf{z} - \mathbf{z}^{\mathrm{f}}) \right\} \\
&\times \exp\left\{ -\tfrac{1}{2}(\mathbf{g}(\mathbf{z}) - \mathbf{d})^{\mathrm{T}} \mathbf{C}_{dd}^{-1}(\mathbf{g}(\mathbf{z}) - \mathbf{d}) \right\}.
\end{aligned} \tag{15}
$$

Thus, Eq. 15 has the same form as in the traditional history-matching problem. The difference is that we solve for both the model parameters and the rate errors in $\mathbf{z}$ rather than only the model parameters, $\mathbf{x}$.

2.1 Cost funtion and ensemble formulation

Maximizing $f(\mathbf{z} \mid \mathbf{d})$ is equivalent to minimizing the cost function

$$\mathcal{J}(\mathbf{z}) = \left(\mathbf{z} - \mathbf{z}^{\text{f}}\right)^{\text{T}} \mathbf{C}_{zz}^{-1}\left(\mathbf{z} - \mathbf{z}^{\text{f}}\right) \\ + \left(\mathbf{g}(\mathbf{z}) - \mathbf{d}\right)^{\text{T}} \mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{z}) - \mathbf{d}\right). \tag{16}$$

We can approximately sample the posterior density by minimizing an ensemble of cost functions [17, 22],

$$\mathcal{J}(\mathbf{z}_j) = \left(\mathbf{z}_j - \mathbf{z}_j^{\text{f}}\right)^{\text{T}} \mathbf{C}_{zz}^{-1}\left(\mathbf{z}_j - \mathbf{z}_j^{\text{f}}\right) \\ + \left(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j\right)^{\text{T}} \mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j\right). \tag{17}$$

Thus, we have an ensemble of models forced by uncertain controls,

$$\mathbf{y}_j = \mathbf{g}(\mathbf{z}_j), \tag{18}$$

representing the prior information, and we have perturbed measurements,

$$\mathbf{d}_j = \mathbf{d} + \mathbf{e}_j, \tag{19}$$

representing the information from the likelihood.

2.2 Gradient and Hessian

The gradient of the cost function in Eq. 17 becomes

$$\nabla\mathcal{J}(\mathbf{z}_j) = \mathbf{C}_{zz}^{-1}\left(\mathbf{z}_j - \mathbf{z}_j^{\text{f}}\right) + \nabla\mathbf{g}(\mathbf{z}_j)\mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j\right), \tag{20}$$

and the approximate Hessian is

$$\nabla\nabla\mathcal{J}(\mathbf{z}_j) \approx \mathbf{C}_{zz}^{-1} + \nabla\mathbf{g}(\mathbf{z}_j)\mathbf{C}_{dd}^{-1}\left(\nabla\mathbf{g}(\mathbf{z}_j)\right)^{\text{T}}. \tag{21}$$

The above equations are similar to the ones obtained when deriving the Iterative Ensemble Smoother with model errors included [10, compare (20) and (21) with Eqs. 71 and 72 in].

2.3 Introduce ensemble averaged model sensitivity

Next, define a joint or averaged model sensitivity by the linear regression

$$\mathbf{G} = \mathbf{C}_{yz}(\mathbf{C}_{zz})^{-1}, \tag{22}$$

which represents the best least-squares-fit to the ensemble of model sensitivities

$$\mathbf{G}_j = \left(\nabla\mathbf{g}(\mathbf{z}_j)\right)^{\text{T}} \in \Re^{m \times n}. \tag{23}$$

One can then approximate the gradient and the Hessian from Eqs. 20 and 21 with

$$\nabla\mathcal{J}(\mathbf{z}_j) \approx \mathbf{C}_{zz}^{-1}\left(\mathbf{z}_j - \mathbf{z}_j^{\text{f}}\right) + \mathbf{G}^{\text{T}}\mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j\right) \tag{24}$$

Algorithm 1 Subspace EnRML algorithm. For the case with estimation and use of correlated rate errors one need to include the lines three and sixteen, and also the $\mathbf{E}_i$ in the model in lines 9 and 12, all marked in red.

1: Input: $\mathbf{X}_0 \in \Re^{n \times N}$ (prior model ensemble)	
2: Input: $\mathbf{D} \in \Re^{m \times N}$ (perturbed measurements)	
3: Input: $\mathbf{E}_0 \in \Re^{m_u \times N}$ (initial perturbations)	
4: $\mathbf{W}_0 = 0$	$\mathbf{W} \in \Re^{N \times N}$
5: $\boldsymbol{\Pi} = \left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\text{T}}\right)/\sqrt{N-1}$	$\boldsymbol{\Pi} \in \Re^{N \times N}$
6: $\mathbf{E} = \mathbf{D}\boldsymbol{\Pi}$	$\mathbf{E} \in \Re^{m \times N}$
7: i=0	
8: **repeat**	
9: $\quad\mathbf{Y}_i = \mathbf{g}(\mathbf{X}_i, \mathbf{E}_i)\boldsymbol{\Pi}$	$\mathbf{Y} \in \Re^{m \times N}$
10: $\quad\boldsymbol{\Omega}_i = \mathbf{I} + \mathbf{W}_i\boldsymbol{\Pi}$	$\boldsymbol{\Omega} \in \Re^{N \times N}$
11: $\quad\mathbf{S}_i = \mathbf{Y}_i\boldsymbol{\Omega}_i^{-1}$	$\mathbf{S} \in \Re^{m \times N}$
12: $\quad\mathbf{H}_i = \mathbf{S}_i\mathbf{W}_i + \mathbf{D} - \mathbf{g}(\mathbf{X}_i, \mathbf{E}_i)$	$\mathbf{H} \in \Re^{m \times N}$
13: $\quad\mathbf{W}_{i+1} = \mathbf{W}_i - \gamma\left(\mathbf{W}_i - \mathbf{S}_i^{\text{T}}\left(\mathbf{S}_i\mathbf{S}_i^{\text{T}} + \mathbf{E}\mathbf{E}^{\text{T}}\right)^{-1}\mathbf{H}_i\right)$	
14: $\quad\mathbf{T}_i = \left(\mathbf{I} + \mathbf{W}_{i+1}/\sqrt{N-1}\right)$	$\mathbf{T} \in \Re^{N \times N}$
15: $\quad\mathbf{X}_{i+1} = \mathbf{X}\mathbf{T}_i$	
16: $\quad\mathbf{E}_{i+1} = \mathbf{E}_0\mathbf{T}_i$	
17: i=i+1	
18: **until** convergence	

and

$$\nabla\nabla\mathcal{J}(\mathbf{z}_j) \approx \mathbf{C}_{zz}^{-1} + \mathbf{G}^{\text{T}}\mathbf{C}_{dd}^{-1}\mathbf{G}. \tag{25}$$

By replacing the model sensitivities, $\mathbf{G}_j$, with the average least-squares fit, $\mathbf{G}$, we introduce an approximation in the gradient in Eq. 20. We, therefore, also change the location of the minimum of the cost function. With the model adjoint's availability, it would be possible to compute the correct model sensitivity and thereby eliminate this approximation. The approximation introduced by using the averaged model sensitivity comes on top of the approximate use of the ensemble of cost functions in Eq. 17 to sample the posterior.

2.4 Subspace EnRML solution

For computing the Ensemble Smoother (ES) or Iterative Ensemble Smoother (IES) solutions, see [3, 9, 10, 12, 24], where the two last papers present new subspace-EnRML algorithm used in this work (see Algorithm 1). The method computes for each ensemble member $\mathbf{z}_j$, the following iteration for iterate number i,

$$\mathbf{z}_j^{i+1} = \mathbf{z}_j^i - \gamma\left(\nabla\nabla\mathcal{J}(\mathbf{z}_j^i)\right)^{-1}\nabla\mathcal{J}(\mathbf{z}_j^i), \tag{26}$$

which uses the definitions of the gradient and Hessian from Eqs. 24 and 25. With the step length, $\gamma = 1.0$, and only one iteration, the solution obtained is the ES solution. Otherwise, one needs to specify a value of $\gamma \in (0, 1]$, typically $\gamma = 0.5$ is a reasonable choice, and iterate until convergence.

The subspace Algorithm 1 takes as input the ensemble of model parameters, $\mathbf{X} = (\mathbf{x}_0, \ldots, \mathbf{x}_N) \in \Re^{n \times N}$, and the ensemble of perturbed measurements, $\mathbf{D} \in \Re^{m \times N}$. If we include rate errors in the estimation, we must also input the ensemble of rate perturbations $\mathbf{E}_0 \in \Re^{m_u \times N}$ (red line number 3).

From the input ensemble one can compute the ensemble of model predictions, $\mathbf{g}(\mathbf{X}) \in \Re^{m \times N}$. Furthermore, the algorithm uses the scaled matrices of model predictions anomalies, $\mathbf{Y} = \mathbf{g}(\mathbf{X})\mathbf{\Pi}$, and scaled measurement perturbations, $\mathbf{E} = \mathbf{D}\mathbf{\Pi}$. The projection $\mathbf{\Pi}$, as defined in line number 5 of Algorithm 1, subtracts the ensemble mean and then scales the resulting ensemble anomalies by $\sqrt{N-1}$.

In line number 11 of Algorithm 1, $\mathbf{S}_i$ is most easily computed by writing $\mathbf{\Omega}^{\mathrm{T}}\mathbf{S}_i^{\mathrm{T}} = \mathbf{Y}^{\mathrm{T}}$ and then computing the LU factorization of $\mathbf{\Omega}_i^{\mathrm{T}}$ followed by m back substitutions to a cost of $\mathcal{O}(mN^2)$ floating point operations. In line number 12, one needs to compute the matrix multiplication $\mathbf{S}_i \mathbf{W}_i$ to a cost of mN^2 operations.

It is possible to approximately compute the inversion in line number 13 to a cost of $\mathcal{O}(mN^2)$ operations by using the ensemble subspace algorithm discussed in [7, 8, 12] (and see also the discussion in the following section).

Finally, in line number 15, we obtain the updated model ensemble in nN^2 floating-point operations. In the case of uncertain rate errors, we must also update $\mathbf{E}_i$ in the red line number 16 to a cost of $m_u N^2$, and use $\mathbf{E}_i$ to force the model.

Thus, the most expensive computation in Algorithm 1 is likely the evaluation of the model-ensemble prediction $\mathbf{Y}_i = \mathbf{g}(\mathbf{X}_i)$ in each iteration. The cost of integrating the model is proportional to the number of time steps, n_t. With each time step requiring $n_c n$ computations, the total cost of the ensemble integration is $N n_t n_c n$. For nonlinear models, the number of timesteps is likely much larger than N, and the number of computations per time step can be several times n. All other computations in this algorithm are linear in the number of measurements, m, and the number of state variables, n. For the detailed derivation of the subspace EnRML algorithm, see [12, 24].

Thus, in the case where we include rate errors with uncertainty, we include the lines in red in Algorithm 1. The only required modification is that we need an unscaled version of the matrix of measurement perturbations stored in $\mathbf{E}_0$. Then the algorithm will use $\mathbf{E}_0$ to compute updated estimates of the rate errors, $\mathbf{E}_i$, in each iteration, and use these updated rates to force the model.

2.5 Ensemble subspace inversion

In Algorithm 1, we have represented the measurement error-covariance matrix by an ensemble of measurement perturbations, $\mathbf{E}$. The inversion is in this case computed using ensemble subspace scheme as was proposed by Evensen

[7] and further discussed in Evensen [8] and recently in Evensen [12]. This scheme projects the measurement error perturbations onto the ensemble subspace and computes the pseudo inverse of the following factorization

$$\left(\mathbf{S}\mathbf{S}^{\mathrm{T}} + \mathbf{E}\mathbf{E}^{\mathrm{T}}\right) \tag{27}$$

$$\approx \mathbf{S}\mathbf{S}^{\mathrm{T}} + (\mathbf{S}\mathbf{S}^+)\mathbf{E}\mathbf{E}^{\mathrm{T}}(\mathbf{S}\mathbf{S}^+)^{\mathrm{T}} \tag{28}$$

$$= \mathbf{U}\mathbf{\Sigma}\left(\mathbf{I}_N + \mathbf{\Sigma}^+ \mathbf{U}^{\mathrm{T}}\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{U}(\mathbf{\Sigma}^+)^{\mathrm{T}}\right)\mathbf{\Sigma}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}} \tag{29}$$

$$= \mathbf{U}\mathbf{\Sigma}\left(\mathbf{I}_N + \mathbf{Z}\mathbf{\Lambda}\mathbf{Z}^{\mathrm{T}}\right)\mathbf{\Sigma}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}} \tag{30}$$

$$= \mathbf{U}\mathbf{\Sigma}\mathbf{Z}\left(\mathbf{I}_N + \mathbf{\Lambda}\right)\mathbf{Z}^{\mathrm{T}}\mathbf{\Sigma}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}, \tag{31}$$

where we have defined the singular-value decomposition

$$\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}, \tag{32}$$

and the $\mathbf{I}_N \in \Re^{N \times N}$ is just the identity matrix. The eigenvalue decomposition in Eq. 30 is of the matrix product in Eq. 29. Note that this eigenvalue decomposition is most efficiently computed by a singular value decomposition of the product $\mathbf{\Sigma}^+ \mathbf{U}^{\mathrm{T}}\mathbf{E}$. The left singular vectors will then equal the eigenvectors in $\mathbf{Z}$ and the squares of the singular values will equal the eigenvalues in $\mathbf{\Lambda}$. Thus, the inversion becomes

$$\left(\mathbf{S}\mathbf{S}^{\mathrm{T}} + \mathbf{E}\mathbf{E}^{\mathrm{T}}\right)^{-1} \approx \left(\mathbf{U}(\mathbf{\Sigma}^+)^{\mathrm{T}}\mathbf{Z}\right)\left(\mathbf{I}_N + \mathbf{\Lambda}\right)^{-1}\left(\mathbf{U}(\mathbf{\Sigma}^+)^{\mathrm{T}}\mathbf{Z}\right)^{\mathrm{T}}. \tag{33}$$

The subspace inversion algorithm's properties will be discussed in more detail in Section 3.

3 EnKF update example

Before testing the proposed algorithm on a reservoir model, it is instructive to examine a single update step using a toy model. The purpose of this example is to illustrate the properties of the update scheme, in particular, when one uses the measurement perturbations to represent the measurement error covariance matrix. This example verifies the robustness of the projection of the measurement error covariance matrix onto the ensemble of predicted measurements. The update is identical to the EnRML algorithm's solution in the linear case, and as such, the learnings are representative for the iterative smoother update as well.

3.1 Example description

The test example uses a one-dimensional periodic domain with 1024 grid points and $\Delta x = 1$. In this domain, we simulate a smooth pseudo-random function with mean $\mu = 4$, variance $\sigma^2 = 1$, and decorrelation length $r_d = 40$, representing the unknown truth,

$$\mathbf{x}_{\text{true}} \leftarrow \mathcal{N}\left(\mu = 4, \sigma^2 = 1, r_d = 40\right). \tag{34}$$

The constant $\mu = 4$, is just added for plotting purposes.

The first guess solution is generated by simulating another realization $\mathbf{x} \leftarrow \mathcal{N}(0, 1, 40)$ and adding it to the truth, i.e.,

$$\mathbf{x}_{\text{fg}} = \frac{\mathbf{x} + \mathbf{x}_{\text{true}} - 4}{\sqrt{2}} + 4. \tag{35}$$

The factor $\sqrt{2}$ ensures that the variance of $\mathbf{x}_{\text{fg}}$ is equal to one.

The initial ensemble is created by adding random realizations $\mathbf{x}_i \leftarrow \mathcal{N}(0, 1, 40)$ to the first guess $\mathbf{x}_{\text{fg}}$,

$$\mathbf{x}_i^{\text{f}} = \mathbf{x}_{\text{fg}} + \mathbf{x}_i. \tag{36}$$

The measurements are distributed uniformly over the domain and sampled from a perturbed true solution according to

$$\mathbf{d}_i = \mathbf{M}(\mathbf{x}_{\text{true}} + \mathbf{e}_i), \tag{37}$$

where either uncorrelated, $\mathbf{e}_i \leftarrow \mathcal{N}(0, 0.25, 0)$, or Gaussian, $\mathbf{e}_i \leftarrow \mathcal{N}(0, 0.25, r_d)$, errors are used. Here, $\mathbf{M}$ is the linear measurement operator that extracts the measurements from the functions $\mathbf{x}_{\text{true}} + \mathbf{e}_i$.

The following experiments use the same reference-truth, measurements, initial ensemble, and random seed. When it is essential to eliminate sampling errors, we use extended ensemble sizes. Thus, although there are seed dependencies in the obtained solutions, the different methods should produce the same answer. We can, in most of the experiments, attribute differences in the solutions to the methods used. Thus, our approach is different from running multiple data assimilation experiments with varying seeds while attempting to estimate the correct answer.

3.2 Solution methods

In the following, we study two prominent cases, one with a diagonal measurement error-covariance matrix and another with correlated errors, $\mathbf{e}_i$, simulated from Eq. 37 and with $r_d = 40$. For the two cases of uncorrelated and correlated measurement errors, the EnKF computes the analysis using both an exactly specified measurement error-covariance matrix $\mathbf{C}_{dd}$ and while representing the measurement error covariance by the perturbations in $\mathbf{E}$.

The case with a full-rank measurement error-covariance matrix solves

$$\mathbf{X}^{\text{a}} = \mathbf{X}^{\text{f}} + \mathbf{A}\mathbf{S}^{\text{T}}(\mathbf{S}\mathbf{S}^{\text{T}} + \mathbf{C}_{dd})^{-1}(\mathbf{D} - \mathbf{M}\mathbf{X}), \tag{38}$$

where the ensemble perturbation matrix is $\mathbf{A} = \mathbf{X}\mathbf{\Pi}$. The matrix $\mathbf{C} = \mathbf{S}\mathbf{S}^{\text{T}} + \mathbf{C}_{dd}$ is formed and then inverted by computing an eigenvalue decomposition $\mathbf{C} = \mathbf{Z}\mathbf{\Lambda}\mathbf{Z}^{\text{T}}$. The inverse is just $\mathbf{C} = \mathbf{Z}\mathbf{\Lambda}^{+}\mathbf{Z}^{\text{T}}$ where the use of a pseudo inverse is needed in case the matrix $\mathbf{C}$ is poorly conditioned.

When using an ensemble representation for the measurement error-covariance matrix $\mathbf{C}_{dd} = \mathbf{E}\mathbf{E}^{\text{T}}$, one can solve for the update from

$$\mathbf{X}^{\text{a}} = \mathbf{X}^{\text{f}} + \mathbf{A}\mathbf{S}^{\text{T}}(\mathbf{S}\mathbf{S}^{\text{T}} + \mathbf{E}\mathbf{E}^{\text{T}})^{-1}(\mathbf{D} - \mathbf{M}\mathbf{X}). \tag{39}$$

In the examples below, the line labels used in the figures indicate the inversion scheme used to compute $(\mathbf{S}\mathbf{S}^{\text{T}} + \mathbf{C}_{dd})^{-1}$. The line label Cdd denotes using the standard EnKF analysis equation with a full rank measurement error-covariance matrix $\mathbf{C}_{dd}$, as explained above. The curves with line label EE correspond to the EnKF update when the samples in $\mathbf{E}$ replace the "exact" analytic measurement error covariance matrix $\mathbf{C}_{dd}$, and we use the ensemble subspace scheme.

Using $\mathbf{E}$ to represent the measurement error-covariance matrix introduces additional sampling errors. However, we will see below how it is possible to reduce these sampling errors to a negligible level with a simple algorithm modification. I.e., one uses a larger number of realizations in $\mathbf{E}$ to represent $\mathbf{C}_{dd}$ better. The code used is the test case from https://github.com/geirev/EnKF_Analysis.git.

### 3.3 Example 1 (large ensemble size)

The first example uses 50 measurements and a large ensemble size of 2000 to reduce sampling errors. Figure 1 shows the results for the two cases with either diagonal or correlated measurement errors.

The upper-left plot shows the EnKF estimates for the case with uncorrelated measurement errors. The two schemes, represented by the lines labeled Cdd and EE, give similar results in this case. The upper-right plot shows the prior and posterior error variances for the two updates, and again they are nearly identical. The lower-left plot presents the EnKF estimates for the case with correlated measurement errors. In this case, we also see that the results using the exact and approximate schemes (Cdd and EE) are nearly identical.

An apparent difference between the two cases is that, with uncorrelated errors, the measurements are scattered randomly about the correct solution. In contrast, with correlated measurement errors, successive measurements will have similar error values, and they follow a smooth curve.

The nonzero measurement correlations' role is to reduce the strength of the update, and the result is an update with a more substantial variance. By taking the measurement error correlations into account, we inform EnKF that neighboring measurements make the same error, and we reduce their accumulated impact.

Table 1 shows that for this case with a large ensemble, the approximation we introduce by using $\mathbf{E}\mathbf{E}^{\text{T}}$ in the subspace inversion scheme, instead of inverting the full $\mathbf{C}_{dd}$, is negligible. Thus, the learning from these two cases is that
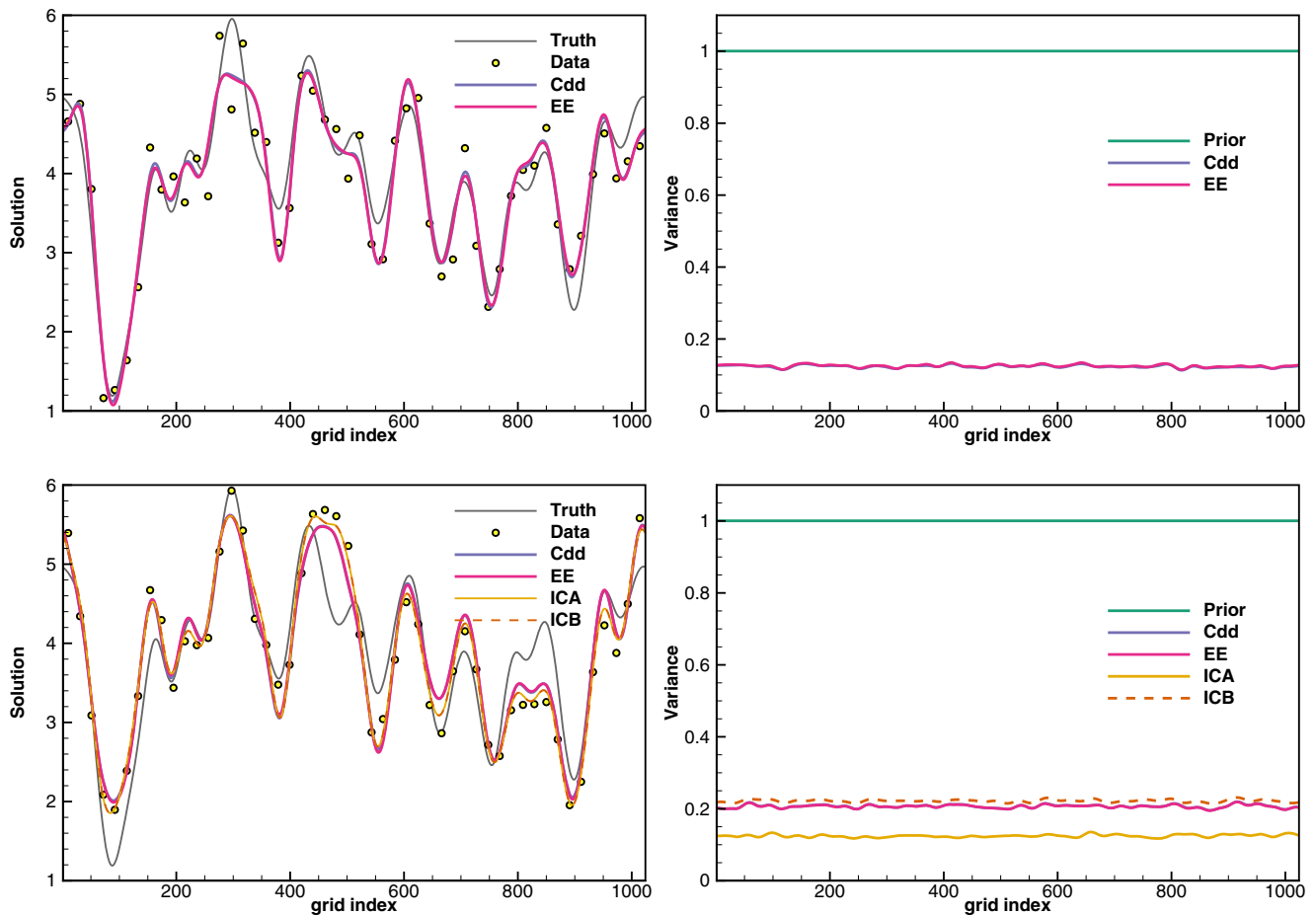
Fig. 1 Simple update example: The upper plots present the results for a case with uncorrelated measurement errors, while the lower plots give the results when using measurements with correlated errors and decorrelation length $r_d = 40$. The left plots show the results for the posterior ensemble means, while the right plots provide the associated error estimates. The line labels Cdd, EE, ICA, and ICB, denote different numerical implementations of the inversion scheme used, as is explained in the text. The ensemble size is 2000

correlated measurement errors reduce the measurements' impact on the update. Furthermore, the two schemes used to compute the inversion are mutually consistent and give approximately the same result with only minor sampling errors. At least, this result is accurate when the scales of the measurement perturbations are the same as in the predicted ensemble anomalies.

The lower plots in Fig. 1 include two inconsistent updates conditioned on the same measurements with correlated errors. In the first inconsistent update (named ICA), $\mathbf{C}_{dd} = \mathbf{I}$, thus neglecting the measurement-error correlations in $\mathbf{C}_{dd}$, and uncorrelated measurement error perturbations are sampled and used to form $\mathbf{D}$. This update is inconsistent, but it is the procedure that most practical EnKF applications use. I.e., one neglects the measurement-error correlations, both because they are challenging to prescribe but also because it allows for efficient computation of the update using (41) discussed below. We notice that the ICA estimate slightly overfits the measurements. The reason is that by neglecting the measurement error correlations in $\mathbf{C}_{dd}$, EnKF

gives "full weight" to all the measurements. Additionally, the ensemble variance becomes way too low and similar to the case with uncorrelated measurement errors. Thus, neglecting measurement error correlations leads to a too close fit to the observations and a too low variance estimate.

In the second inconsistent update (named ICB), we still set $\mathbf{C}_{dd} = \mathbf{I}$, but we retain the original correlated perturbations when forming $\mathbf{D}$. The surprising result is that we obtain a too large variance in this case. One can explain this larger variance: A set of measurements of a smooth "true" field, with correlated errors, will have similar values. By creating the measurement perturbation with correlated errors, one generates perturbed measurements of similar value for each realization. Then, using a diagonal $\mathbf{C}_{dd}$, each realization is pulled too strongly towards the common value of the perturbed measurements. Thus, the updated ensemble members will more closely resemble the variance of the measurement perturbations.

In Table 1, we see that cases ICA and ICB lead to significant errors, both for the estimated mean and the variance.

Table 1 Overview of the EnKF experiments

Experiment	N	m	n_e	r_d	RMSE (mean)	RMSE (variance)	Figure
Exp. 1	2000	50	1	0	0.007688	0.000635	Fig. 1, upper plots
Exp. 1	2000	50	1	40	0.004753	0.000189	Fig. 1, lower plots
Exp. 1 ICA	2000	50	1	40	0.029251	0.025474	Fig. 1, lower plots
Exp. 1 ICB	2000	50	1	40	0.029247	0.004105	Fig. 1, lower plots
Exp. 2	100	50	1	0	0.012850	0.002236	Fig. 2, upper plots
Exp. 2	100	50	1	40	0.016102	0.003404	Fig. 2, lower plots
Exp. 3	100	50	10	0	0.006946	0.001003	Fig. 3, upper plots
Exp. 3	100	50	10	40	0.010386	0.001135	Fig. 3, lower plots
Exp. 4	100	50	10	20	0.014163	0.001516	Fig. 4, upper plots
Exp. 4	100	50	10	80	0.004194	0.001642	Fig. 4, lower plots
Exp. 5	100	200	10	40	0.010219	0.001117	Fig. 5

Here N is the ensemble size, m is the number of measurements, n_e dimension inflation of the **E**-matrix, and r_d is the decorrelation length used for the measurement perturbations. The RMSE is the root mean square difference between the solutions (ensemble means and variances) computed using an exact inversion of $\mathbf{C}_{dd}$ and an approximation $\mathbf{EE}^{\mathrm{T}}$

These two inconsistent updates stress the importance of specifying the correct measurement error statistics in the ensemble update scheme.

3.4 Example 2 (ensemble size of 100)

We now repeat the previous experiment from Example 1 using a more common ensemble size of 100. The purpose is to illustrate the impact of sampling errors when using the measurement error perturbations in **E** to represent $\mathbf{C}_{dd}$. In Fig. 2, we observe that with 100 realizations, the additional sampling errors introduced by scheme EE lead to a slight deviation between the two estimates. More problematic is the underestimation of the ensemble variance. In a sequential data-assimilation context, this underestimation would have to be compensated for, e.g., by using inflation, to avoid possible filter divergence. In the next example, we will learn how to reduce these sampling errors to a negligible level.

3.5 Example 3 (augmenting realizations to E)

The benefit of using (39) over (38) is the reduced computational cost, but also the fact that it is easier to sample perturbations with accurate statistics than constructing a full-rank measurement error covariance matrix. An approach for reducing the sampling errors in scheme EE is to augment columns of new realizations of measurement perturbations to **E**. This modification only slightly increases the computational cost of the algorithm when computing $\mathbf{\Sigma}^+\mathbf{U}^{\mathrm{T}}\mathbf{E}$ in Eq. 29 and is a simple modification of the code. In Fig. 3, we see the results using 100 realizations and correlated measurement errors, and when using 1000 samples in **E**. From Table 1, we see that the augmentation of additional columns to **E** in Exp. 3 significantly reduces the errors in the

estimated means and variances for the two cases with correlated and uncorrelated measurement errors, when compared with the results from Exp. 2. It is clear that the two schemes Cdd and EE, solving (38) and (39) respectively, give almost identical results. In this case, the measurement error perturbations' projection onto the ensemble subspace does not significantly impact the results. Thus, the sampling errors introduced by using **E** to represent $\mathbf{C}_{dd}$ can be made negligible by increasing the sample size in **E** to only a minor additional cost.

3.6 Example 4 (different scales in E)

Example 4 repeats the experiment from Example 3 using measurements with two different error decorrelation lengths $r_d = 20$ and $r_d = 80$. These length scales are in contrast to the length scale of $r_d = 40$ used in the previous experiments. From the results shown in Fig. 4, we confirm that the subspace inversion algorithm works well also when the measurement error correlations have a different length scale than the model ensemble (i.e., $r_d = 40$).

The result from the case with smooth measurement-error correlations, $r_d = 80$, is slightly more accurate than the $r_d = 20$ result. If the measurement perturbations in **E** are smoother than the measured ensemble anomalies in **S**, then **E** should be well represented when projected on the space spanned by **S**.

However, in the case with $r_d = 20$, the measurement perturbations include small scales that are not represented by the predicted measurements' ensemble. Thus, when conditioning on measurements with small-scale noise in the measurements that cannot be well represented by the ensemble anomalies in **S**, the subspace projection introduces an approximation. The truncation of small scales in the
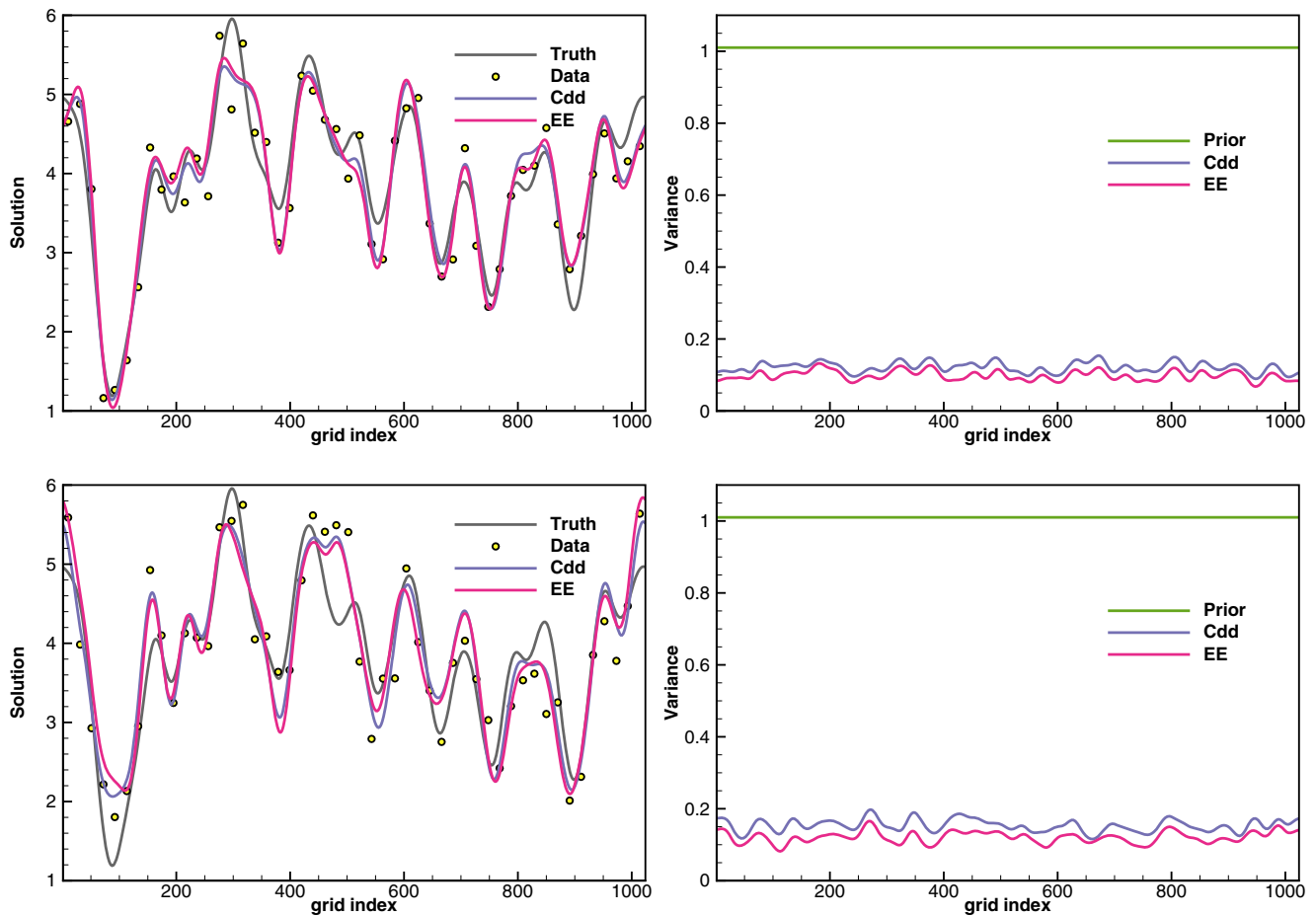
Fig. 2 Simple update example: Same as Fig. 1 but using an ensemble size of 100 realizations and excluding the cases ICA and ICB

measurement errors leads to a slight underestimation of the measurement error variance, as is observed by the slightly too large update and too low variance for the case EE in the bottom plots of Fig. 4. The residuals in Table 1 support this conclusion, where it is clear that the residual of the estimated mean increases for the case $r_d = 20$ and decreases for the case $r_d = 80$ compared to the $r_d = 40$ case.

3.7 Example 5 (large number of measurements)

The final example, shown in Fig. 5, increases the number of measurements from 50 to 200, i.e., twice the ensemble size. In this case, we apply a truncation at 99% of the variance when computing the inversion, which retained 29 singular values when computing the singular value decomposition of **S**. Again the results obtained are very similar using the two algorithms. It is also interesting to see how the measurements' impact reduces at the grid indices 400–500 in this plot. Note also that there is no indication of the so-called "ensemble collapse," and the analysis ensemble retains a significant variance. The posterior variance using 200 measurements is similar to the one obtained using

only 50 observations. This result indicates that including additional measurements does not introduce much new information in this example.

3.8 A note on the EnKF analysis equation

Most operational ensemble-based schemes apply an assumption of uncorrelated measurement errors and uses a diagonal $\mathbf{C}_{dd} = \mathbf{I}$, see, e.g., the reviews on data assimilation in the geosciences [2], weather prediction [15], and petroleum applications [1]. This assumption is emploied for simplicity for two reasons. First, the measurement error covariances are often not well known, and additionally, the update scheme (38) simplifies considerably. With $\mathbf{C}_{dd} = \mathbf{I}$, Eq. 38 becomes

$$\mathbf{X}^{\mathrm{a}} = \mathbf{X}^{\mathrm{f}} + \mathbf{A}\mathbf{S}^{\mathrm{T}}\left(\mathbf{S}\mathbf{S}^{\mathrm{T}} + \mathbf{I}\right)^{-1}\left(\mathbf{D} - \mathbf{M}\mathbf{X}\right), \qquad (40)$$

which makes it possible to use an efficient algorithm from Hunt et al. [16] where, by using a Woodbury identity, the EnKF update becomes

$$\mathbf{X}^{\mathrm{a}} = \mathbf{X}^{\mathrm{f}} + \mathbf{A}\left(\mathbf{S}^{\mathrm{T}}\mathbf{S} + \mathbf{I}\right)^{-1}\mathbf{S}^{\mathrm{T}}\left(\mathbf{D} - \mathbf{M}\mathbf{X}\right). \qquad (41)$$
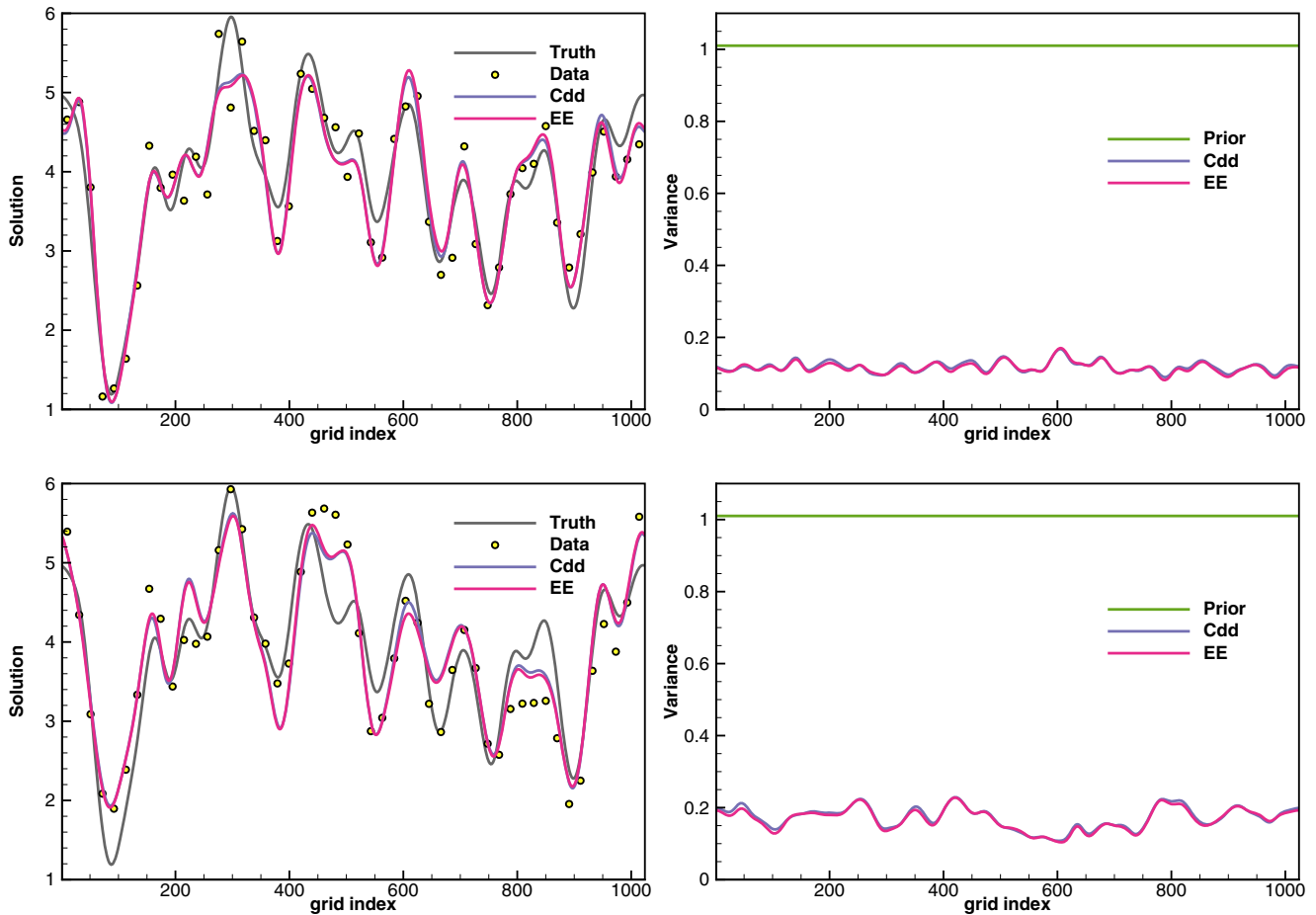
Fig. 3 Simple update example: Same as Fig. 2 but using an ensemble size of 1000 realizations to represent **E**

Thus, this modification reduces the size of the matrix to be inverted from $m \times m$ in Eq. 40 to $N \times N$ in Eq. 41. See also the discussion related to this particular implementation in Evensen et al. [12, Section 3.2].

Alternatively, it is possible to obtain an update equation like (40) if one has access to a factorization $\mathbf{C}_{dd} = \mathbf{C}_{dd}^{\frac{1}{2}}\mathbf{C}_{dd}^{\frac{1}{2}}$ with $\mathbf{C}_{dd}^{\frac{1}{2}}$ being a symmetrical square root of a full rank $\mathbf{C}_{dd}$. E.g., write the eigenvalue decomposition

$$\mathbf{C}_{dd} = \mathbf{Z}\boldsymbol{\Lambda}\mathbf{Z}^{\mathrm{T}} = \mathbf{Z}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{Z}^{\mathrm{T}}\,\mathbf{Z}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{Z}^{\mathrm{T}} = \mathbf{C}_{dd}^{\frac{1}{2}}\mathbf{C}_{dd}^{\frac{1}{2}} \tag{42}$$

and define the symmetrical square root

$$\mathbf{C}_{dd}^{\frac{1}{2}} = \mathbf{Z}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{Z}^{\mathrm{T}}, \tag{43}$$

and its' inverse

$$\mathbf{C}_{dd}^{-\frac{1}{2}} = \mathbf{Z}\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{Z}^{\mathrm{T}}. \tag{44}$$

Now, by scaling the predicted measurement anomalies and the innovations according to:

$$\hat{\mathbf{S}} \quad = \mathbf{C}_{dd}^{-\frac{1}{2}}\mathbf{S}, \tag{45}$$

$$\hat{\mathbf{D}} = \mathbf{C}_{dd}^{-\frac{1}{2}}\left(\mathbf{D} - \mathbf{MX}\right), \tag{46}$$

and with some algebra, Eq. 38 becomes

$$\mathbf{X}^{\mathrm{a}} = \mathbf{X}^{\mathrm{f}} + \mathbf{A}\hat{\mathbf{S}}^{\mathrm{T}}\big(\hat{\mathbf{S}}\hat{\mathbf{S}}^{\mathrm{T}} + \mathbf{I}\big)^{-1}\hat{\mathbf{D}}, \tag{47}$$

and using the Woodbury identity,

$$\mathbf{X}^{\mathrm{a}} = \mathbf{X}^{\mathrm{f}} + \mathbf{A}\big(\hat{\mathbf{S}}^{\mathrm{T}}\hat{\mathbf{S}} + \mathbf{I}\big)^{-1}\hat{\mathbf{S}}^{\mathrm{T}}\hat{\mathbf{D}}. \tag{48}$$

There are, however, significant numerical costs associated with the establishment of $\mathbf{C}_{dd}^{\frac{1}{2}}$ and the associated rescalings in Eqs. 45 and 46 which are both $\mathcal{O}(m^2N)$ operations. Additionally, $\mathbf{C}_{dd}^{\frac{1}{2}}$ needs to be of full rank or a formulation based on pseudo inverses must be employed. Thus, the discussion in this section justifies the use of the ensemble subspace projection scheme in Eq. 33 for computing consistent updates at the cost of $\mathcal{O}(mN^2)$, while taking measurement error-correlations into account.

4 Reservoir experiments

We will now discuss the application of the subspace EnRML method in Algorithm 1 for history matching a reservoir
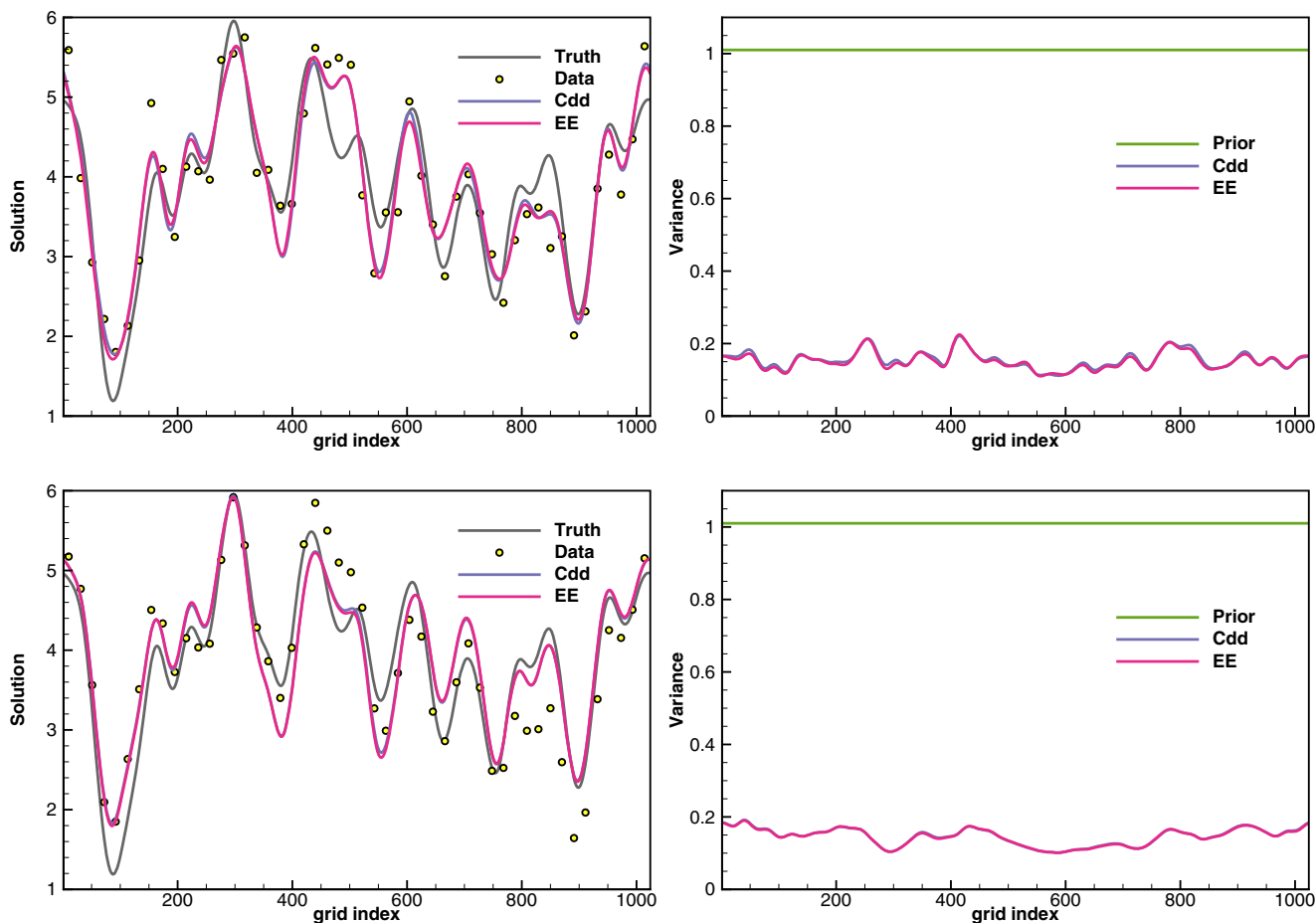
Fig. 4 Simple update example: Same as the lower plots in Fig. 3 but for two cases with different measurement correlations, upper plots $r_d = 20$ and lower plots $r_d = 80$

model. The objective is to assess the impact of including uncertainties in the historical rates used to force the model simulations.

The reservoir model is a simplification of a real reservoir model and provides a reasonably realistic reservoir case for testing new methods. The model has 27 755 active grid
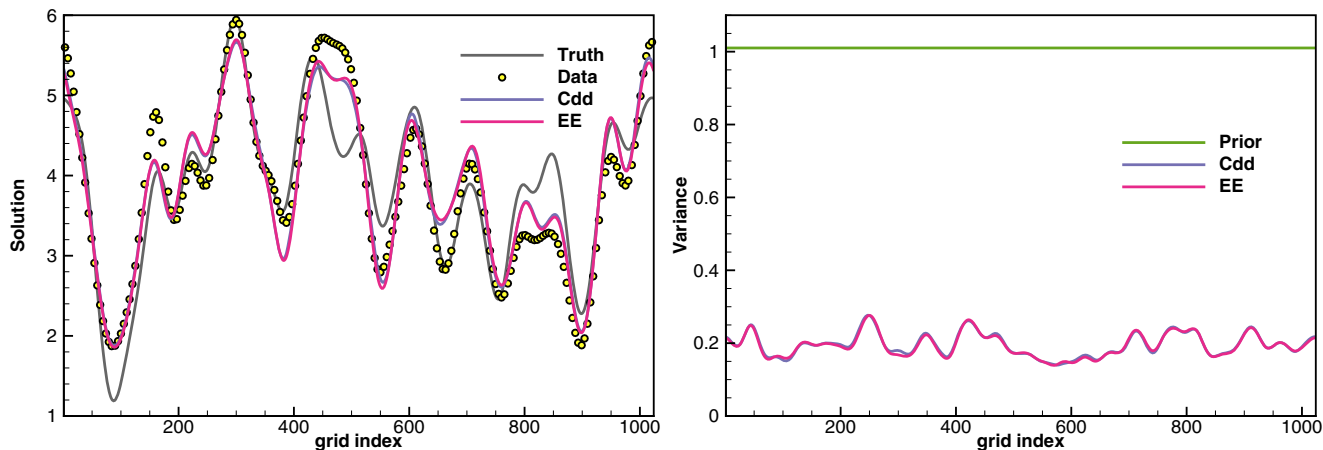
Fig. 5 Simple update example: Same as Fig. 3 but for a case with 200 measurements, which is twice the ensemble size, and with measurement-error correlations $r_d = 40$

cells on a 40-times-64 grid with 14 layers. The uncertain parameters include the three-dimensional porosity field and six fault multipliers F2–F7. There are five producing wells, OP1–OP5 and three injectors I1–I3. The model is the same as was used in Evensen et al. [12], and we will not describe the model at length here but rather focus on the impact of introducing correlated measurement errors for a reservoir case. The model was previously used also by Leeuwenburgh et al. [18],Hanea et al. [14], and Wang and Oliver [26], and we refer to these papers for a detailed presentation.

4.1 Overview of experiments

In the cases below, we condition on the monthly production rates. Thus, we have up to 36 triplets of the oil production rate (OPR), the gas production rate (GPR), and the water production rate (WPR) for each well (depending on the date the well is started). The error standard deviations for the rate measurements are five percent of the measured value of the OPR and 10 percent for the GPR and WPR. We didn't condition the model on the injection rates, and we didn't have access to pressure data. The measurements are real historical production rates from the reservoir.

We will present results from the experiments listed in Table 2. The tree first "PRED" experiments are ensemble predictions without conditioning the model on observations. We randomly picked one model realization from the ensemble and used this model's porosity and fault multipliers for all the ensemble members. The difference between the three ensemble-prediction experiments is the stochastic controls used to force the ensemble integration (white, red, or

Table 2 Overview of the three ensemble prediction runs and the four history matching experiments

Case	$\mathbf{C}_{dd}$	Noise Model	Update $\mathbf{E}$
PREDW		White	
PREDR		Red	
PREDB		Bias	
IES0	$\mathbf{I}$	White	no
IESnd	$\mathbf{EE}^{\mathrm{T}}$	Red	no
IESR	$\mathbf{EE}^{\mathrm{T}}$	Red	yes
IESB	$\mathbf{EE}^{\mathrm{T}}$	Bias	yes

The three PRED experiments are pure ensemble predictions with the same reservoir model used for all realizations. The IES experiments are model conditioning experiments. The noise model of the control variables (historical rates) is either uncorrelated (white), correlated (red), or a bias, in time. The first column gives the case name, while the second column denotes which form of $\mathbf{C}_{dd}$ we use in the inversion in the IES experiments. The last column says if we update the measurement perturbations for use in the schedule file or not

"bias"). In the experiment PREDW, the model forcing is the historical rate data contaminated by white noise, i.e., the measurement errors are uncorrelated in time. In PREDR, we set the time-decorrelation length of the measurement errors to 15 months, resulting in measurement errors that are smooth (red) in time. Finally, in PREDB, the measurement error perturbations are constant in time. Thus, the measurement errors are infinitely correlated, and for each realization, the errors act as a bias. These experiments' purpose is to assess the random forcing's impact on the prediction uncertainty while eliminating any contributions from model-parameter uncertainties.

Case IES0 is the default or standard case using a diagonal measurement error covariance matrix $\mathbf{C}_{dd} = \mathbf{I}$. Thus, there are no measurement error correlations and no stochastic noise added to the historical rate data used to force the reservoir model. Accordingly, there is no updating of the measurement perturbations. This case represents the traditional approach for solving the history-matching problem using ensemble methods. We computed the solution using Algorithm 1, with the red lines excluded.

The second case, IESnd, includes the effect of a nondiagonal $\mathbf{C}_{dd}$ represented by the product of correlated measurement perturbations $\mathbf{EE}^{\mathrm{T}}$. Like in IES0, there is no stochastic forcing of the model. This case allows us to examine the impact of accounting for the measurement error correlations in the analysis update.

The third case, IESR, is the full new Algorithm 1, including the red lines, using correlated measurement perturbations in $\mathbf{E}$ to perturb the measurements conditioned on, as well as taking measurement error correlations into account in the inversion. The algorithm also updates the ensemble of perturbations in $\mathbf{E}_i$ and uses these to force the model realizations. The initial forcing ensemble is the red one used in PREDR.

The last case, IESB, is similar to IESR, except that it uses the "bias" initial forcing ensemble used in PREDB.

The ensemble size is 100 for all experiments. Like in Section 3, the experiments also use the same initial ensemble and random seed to make it easier to compare cases.

4.2 Discussion of results

Figure 6 presents the resulting oil production for the OP2 well when forcing the model ensemble with a white, red, and a "biased" noise model. The uncertainty in the produced rates will depend on the spread in the historical rates used to force the model realizations. Thus, the prediction uncertainty is more apparent in the first half of the simulation than in the second half since the forcing-uncertainty
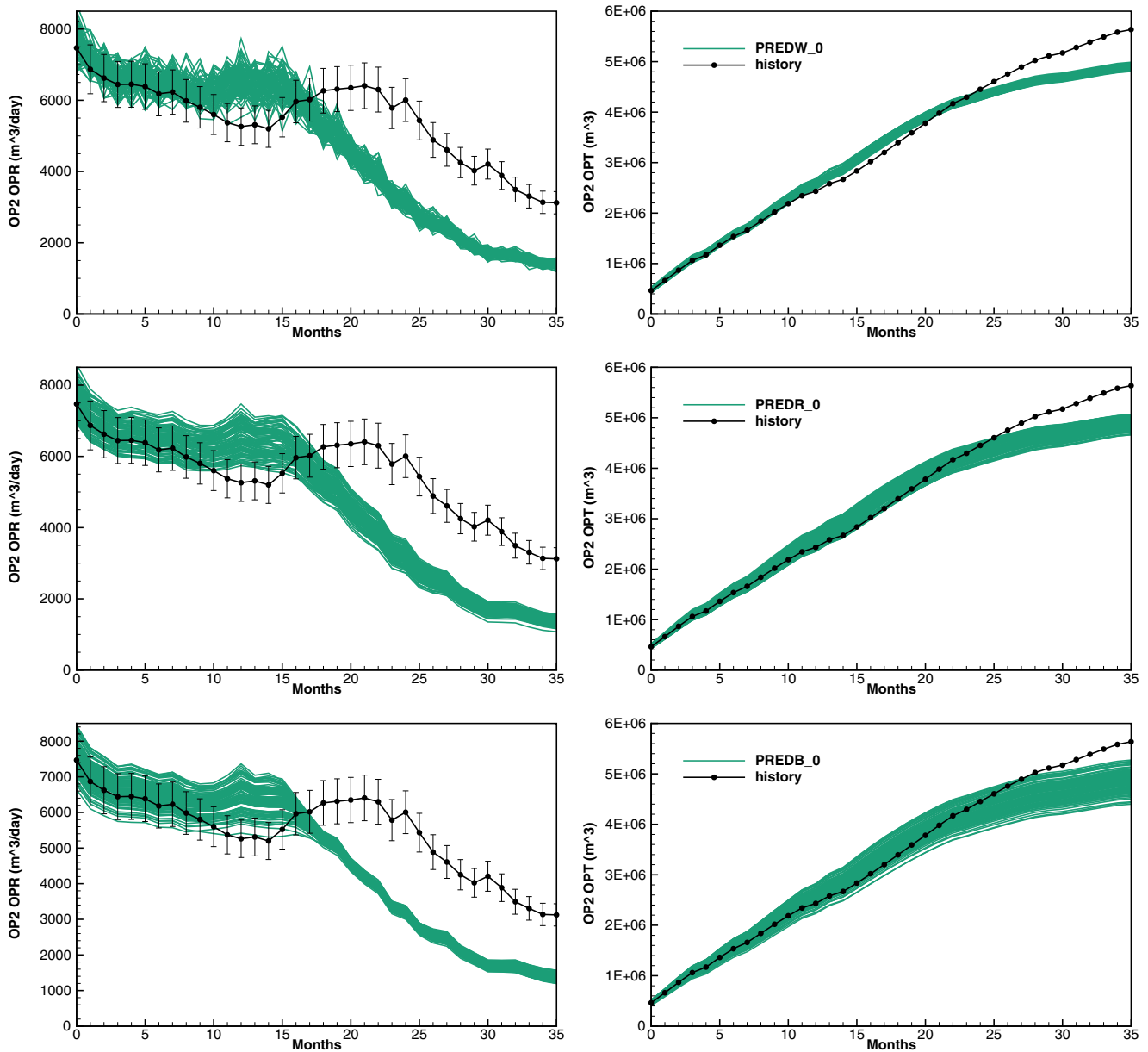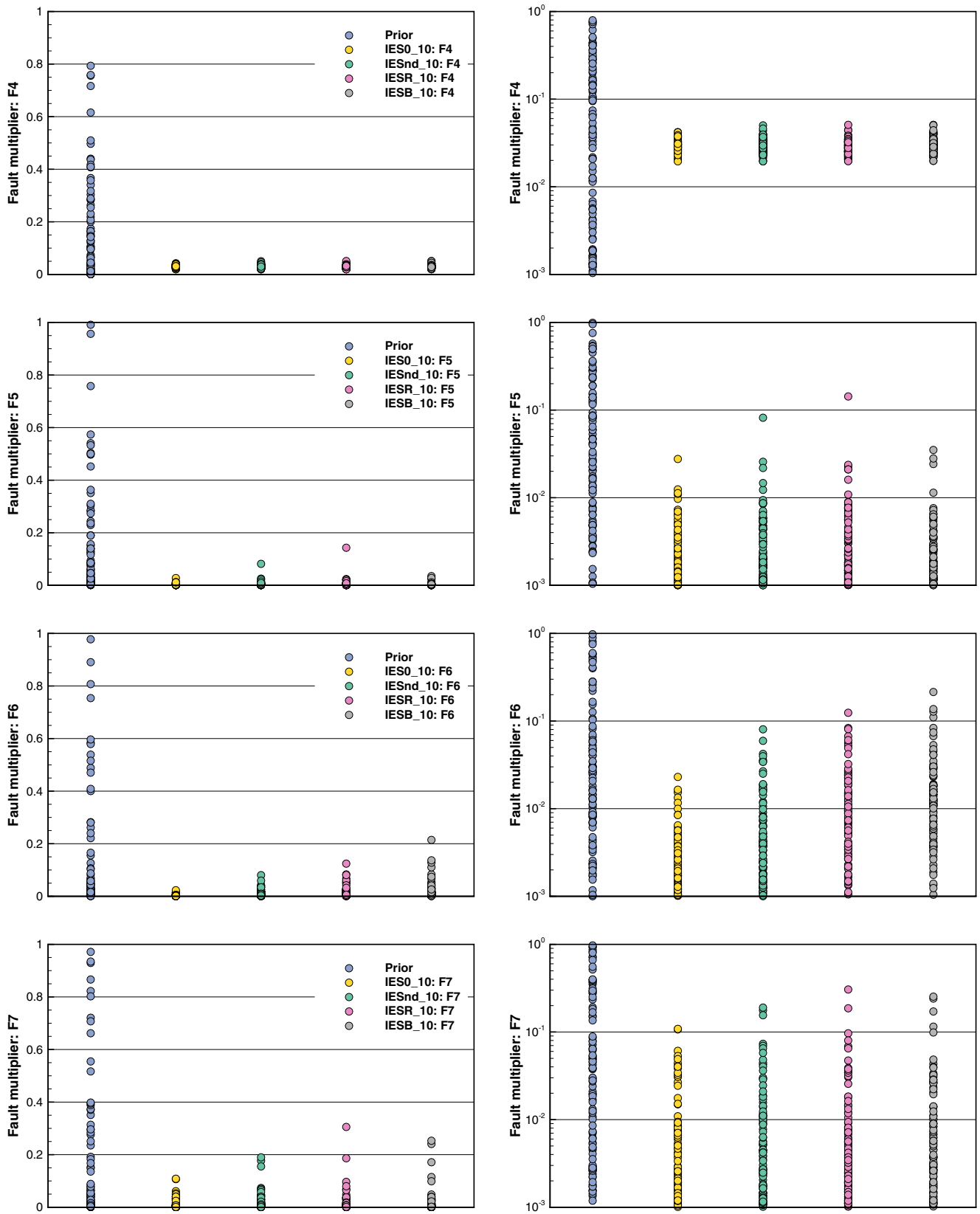
Fig. 6 Prediction experiments: The plots show oil production rate OPR (left), and accumulated oil production OPT (right) from producer OP2. The upper plots are for white noise (uncorrelated in time) forcing, the center plots show res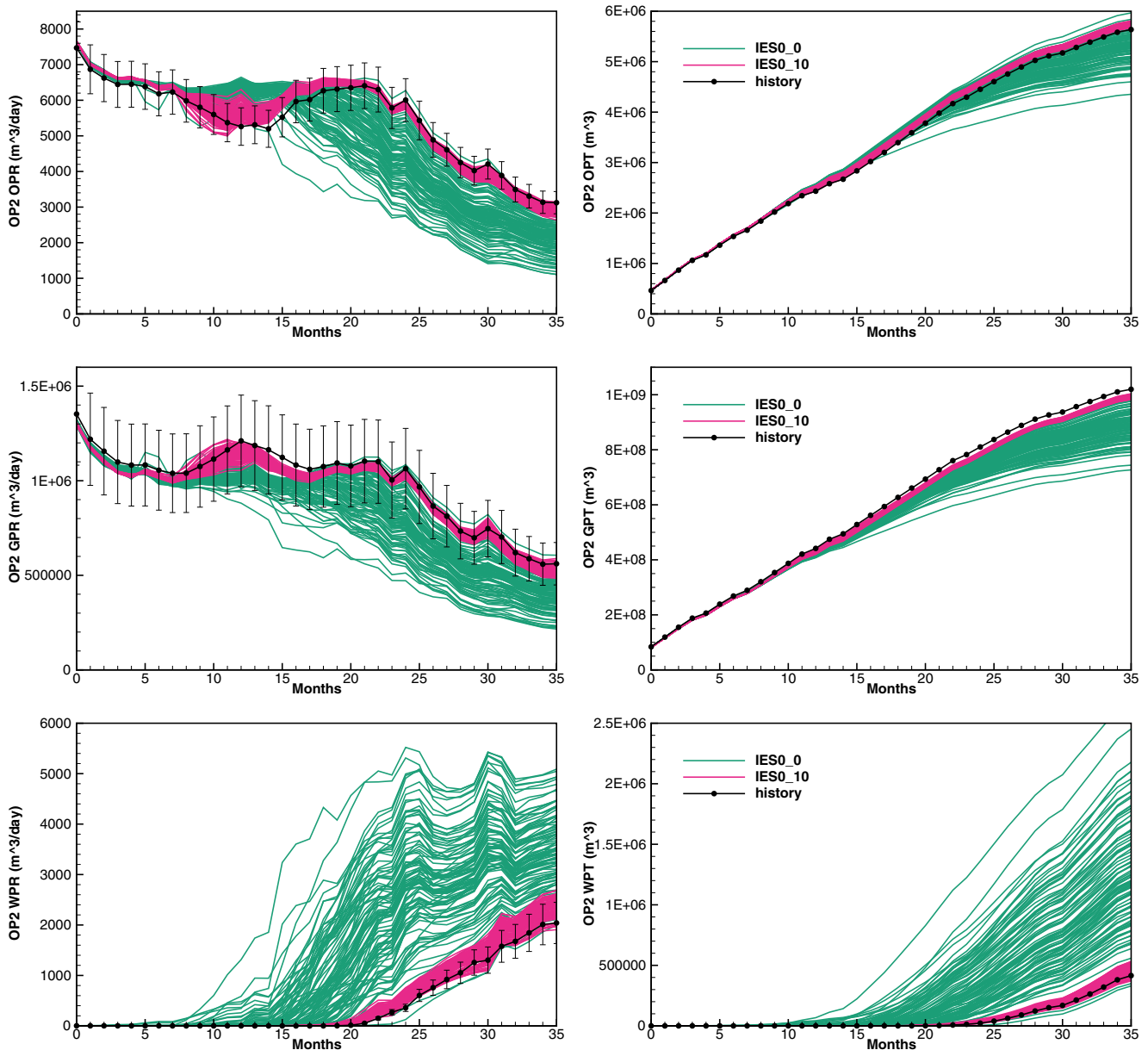ults for red noise (correlated in time), and the bottom plots corresponds to biased noise (constant in time). The error bars on the rate data indicate the plus-minus two standard deviations of the rate errors assumed in the assimilation experiments

scales with the data value. For PREDW with white-noise forcing, we notice the instant response in the model ensemble's predicted rates. Also, we see that the uncertainty in the accumulated production remains low throughout the integration. This "random-walk" process, where the rate uncertainties nearly cancel out over time, yields a near-linear variance increase in time (there is a nonlinearity contribution introduced by the model dynamics). In PREDR

and PREDB, we notice an increasing spread in the predicted ensembles' accumulated production. This uncertainty increase is most significant in PREDB. We can explain it by the constant-in-time errors added to the historical rates, which will continuously drive the different realizations towards different productions. In fact, for the PREDB case, the uncertainty in the total accumulated production for OP2 is almost of the same magnitude as the uncertainty of

Fig. 7 Prior and posterior ensembles of fault multipliers for the IES0, IESnd, IESR, and IESB cases, for four of the faults in the model. In the right column a log-axis is used to better visualize the ensemble spread of the multipliers
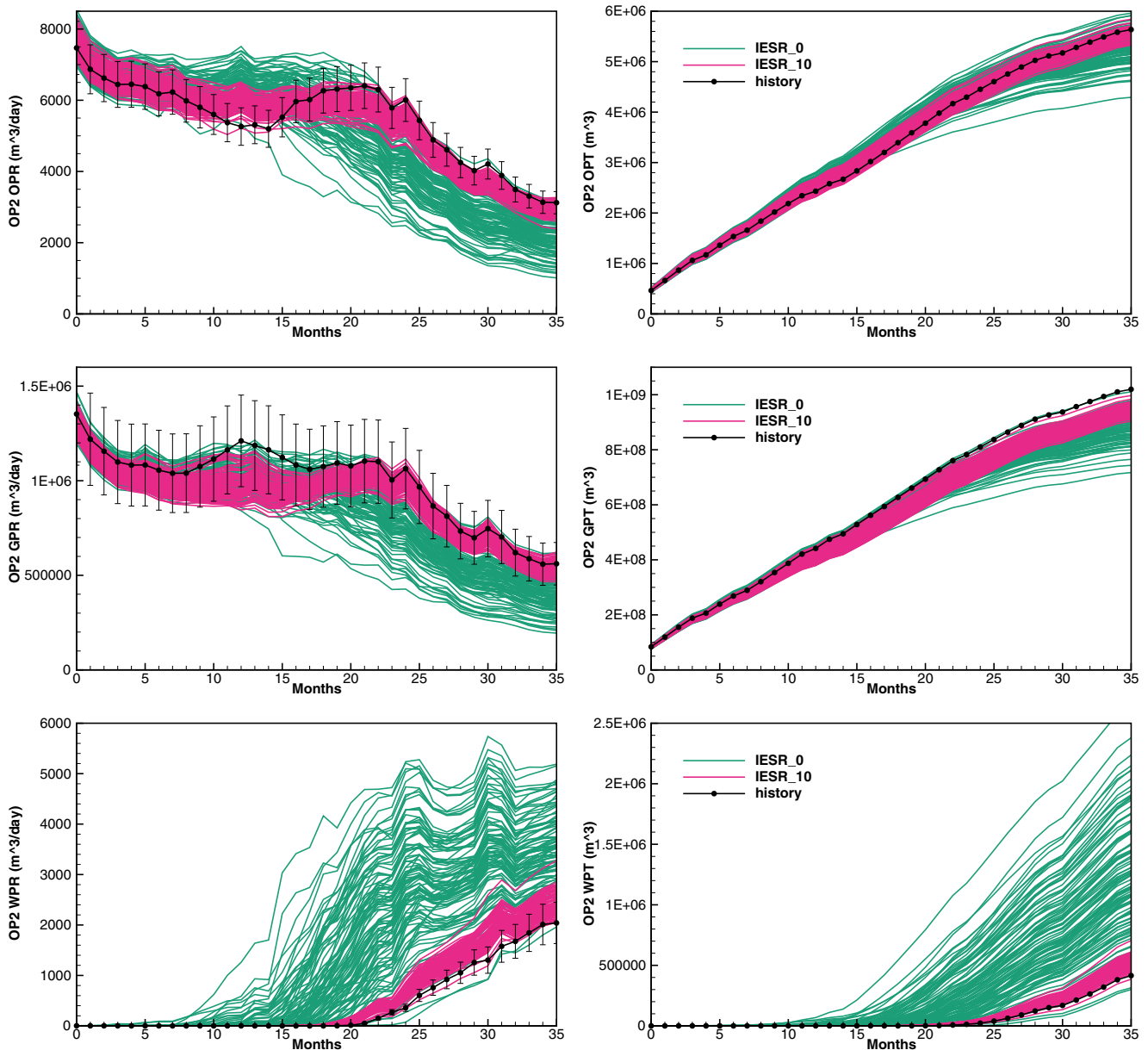
Fig. 8 Case IES0: Results from producer OP2. The upper plots show oil production rate OPR (left), and accumulated oil production OPT (right). Similarly, the middle plots are for gas production GPR and GPT, and the bottom plots show water production WPR and WPT. The error bars on the rate data indicate the plus-minus two standard deviations of the rate errors assumed in the assimilation experiments

the prior ensemble in the IES0 case (which has no stochastic forcing and only reflects the reservoir uncertainty). The inclusion of stochastic historical production rates is essential since it increases the predicted rates' variance, thereby enhancing the impact of observations on the update.

Figure 7 presents the prior and posterior ensembles of scalar fault multipliers for four of the six faults in the model, for the cases IES0, IESnd, IESR, and IESB. The typical behavior is that IES0 results in the most substantial

update, as we would expect from the discussion of the analysis scheme in Section 3. The introduction of correlated measurement errors in IESnd leads to a slight increase in the spread of the fault multipliers associated with a slightly weaker update. The additional stochastic forcing used in IESR and IESB leads to further weakening of the parameter updates. The reason is that the additional stochastic forcing reduces the correlations between the model parameters and the predicted production. Contrary, the stochastic forcing
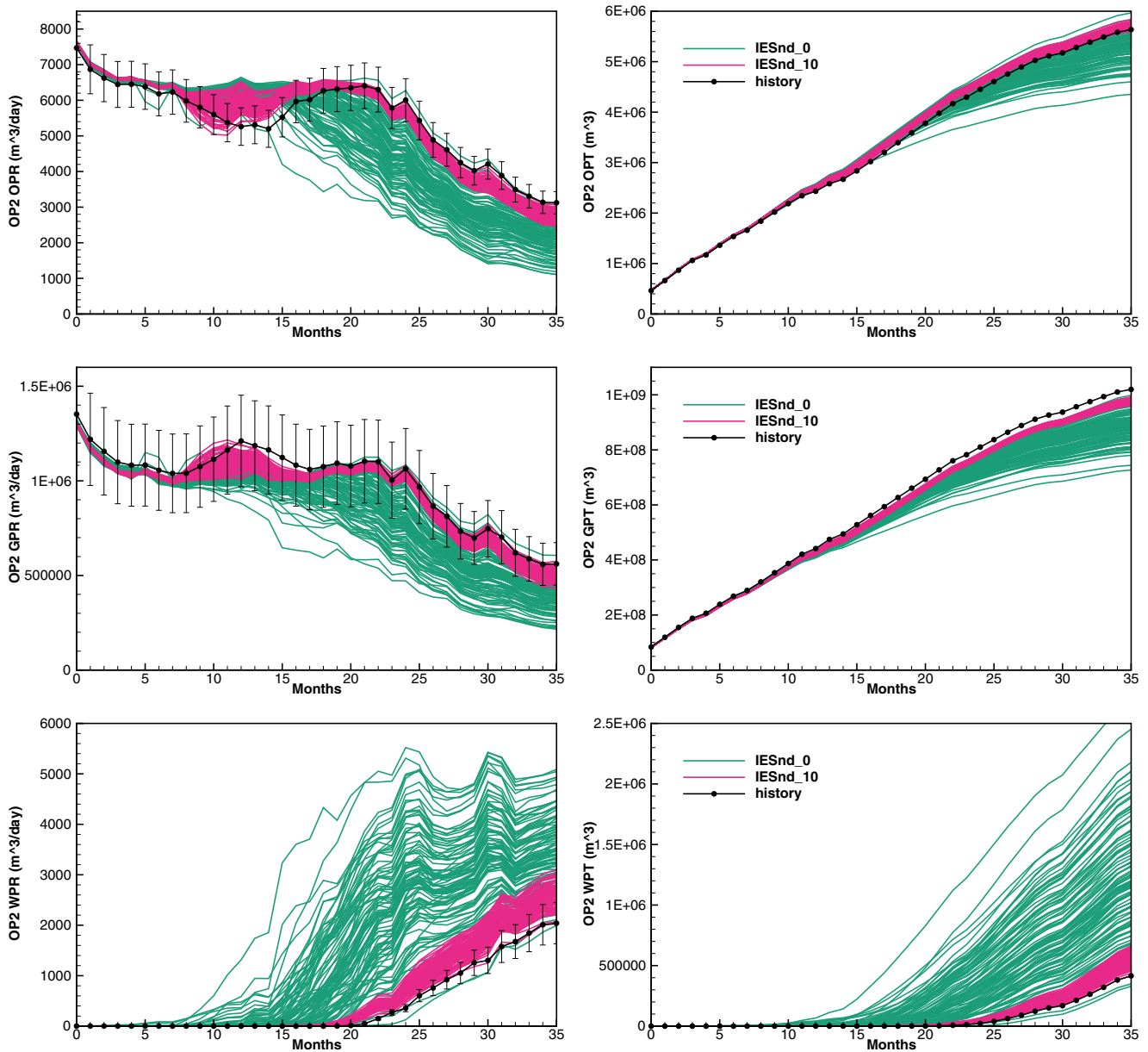
Fig. 9 Case IESnd: Results from producer OP2. The upper plots show oil production rate OPR (left), and accumulated oil production OPT (right). Similarly, the middle plots are for gas production GPR and GPT, and the bottom plots show water production WPR and WPT. The error bars on the rate data indicate the plus-minus two standard deviations of the rate errors assumed in the assimilation experiments

leads to an increased variance in the model-predicted rates, leading to an increased impact from the measurements. Thus, the total effect of including correlated rate errors is quite complicated to assess. The estimated porosity fields (not shown here) have similar behavior between the four examples, as seen for the fault multipliers.

For the experiments IES0, IESnd, IESR, and IESB, the Figs. 8, 9, 10 and 11 present the production rates and the accumulated production for oil, gas, and water for OP2,

and correspondingly the Figs. 12, 13, 14 and 15 give the results from well OP4. Additionally, Fig. 16 shows the prior and posterior estimates of the uncertain forcing rates for IESR. The producers, OP1, OP2, and OP3, had somewhat similar behavior while the match obtained for OP5 was nearly perfect in all the cases but with more substantial uncertainty in IESR and IESB. Thus, the figures include results from the wells OP2 and OP4, which appropriately illustrate the results.
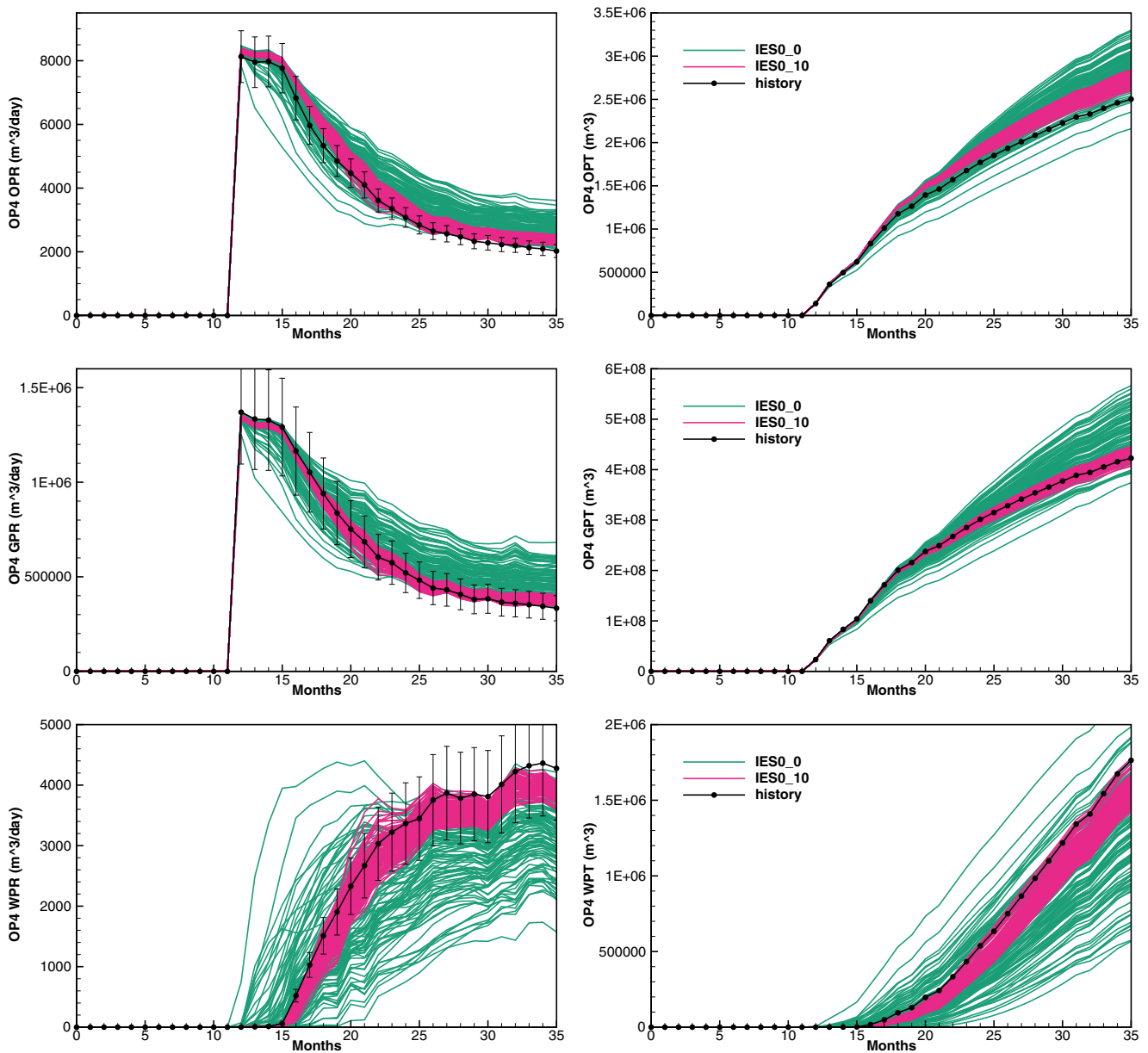
Fig. 10 Case IESR: Results from producer OP2. The upper plots show oil production rate OPR (left), and accumulated oil production OPT (right). Similarly, the middle plots are for gas production GPR and GPT, and the bottom plots show water production WPR and WPT. The error bars on the rate data indicate the plus-minus two standard deviations of the rate errors assumed in the assimilation experiments

For OP2, the posterior ensemble has very low uncertainty in IES0. The ensemble spread increases slightly in IESnd, but it is still too small compared to the measurement uncertainty. The small ensemble spread is particularly the case early in the simulation. However, in IESR, the ensemble spread is significantly increased, and the posterior ensemble represents better the historical oil production.

For OP4, the prior ensemble covers the measurements well, but the posterior from IES0 overestimates the oil production, and the posterior ensemble does not cover the historical rates. Thus, the posterior uncertainty is too low. The solution from IESnd slightly increases the ensemble spread. On the other hand, in IESR, the posterior ensemble has a more considerable variance, and the ensemble mean shifts towards lower production, in better agreement with the measurements.

For the case, IESR, the posterior ensemble of uncertain historical rates in Fig. 16, has a lower variance than the prior
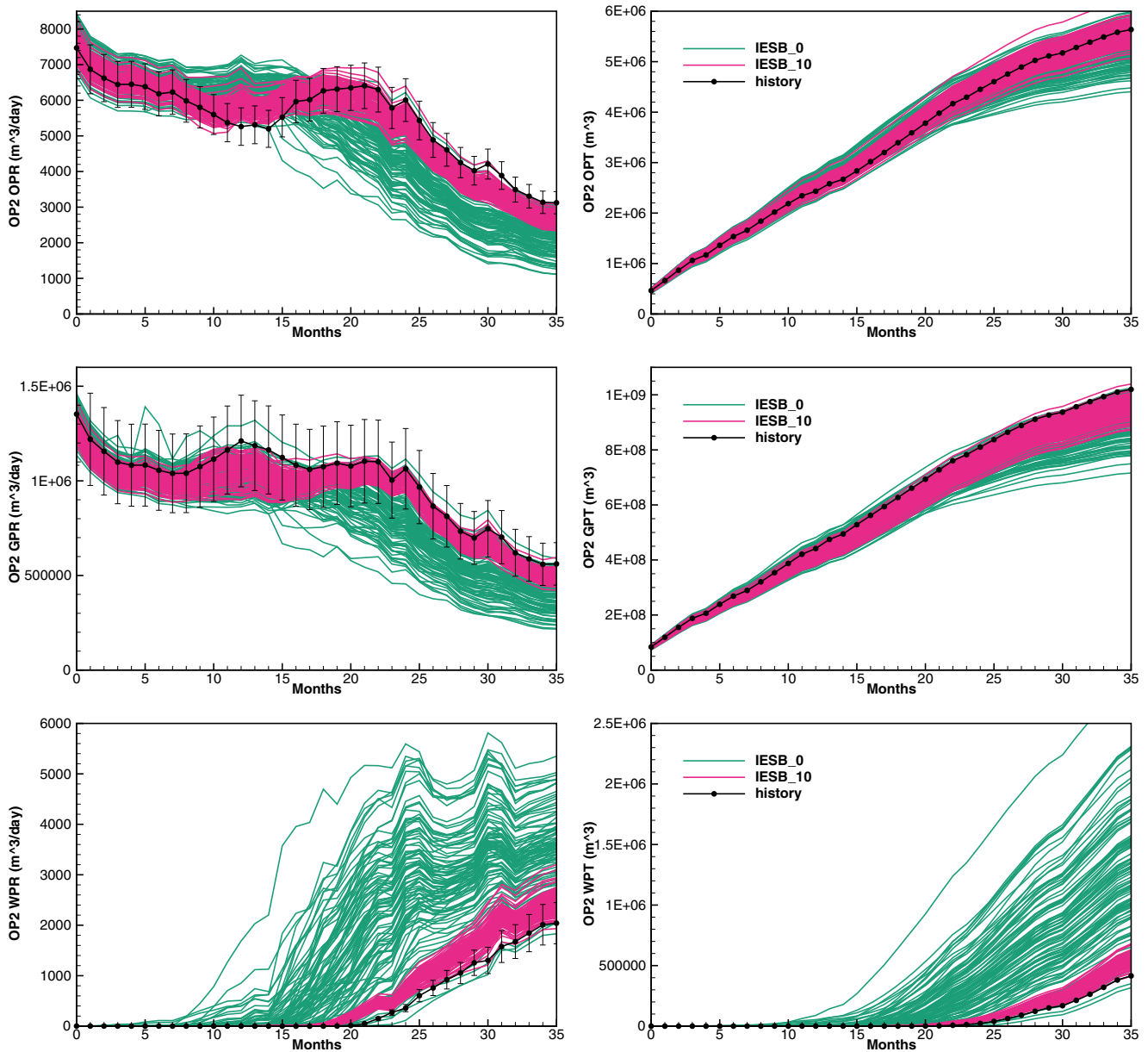
Fig. 11 Case IESB: Results from producer OP2. The upper plots show oil production rate OPR (left), and accumulated oil production OPT (right). Similarly, the middle plots are for gas production GPR and
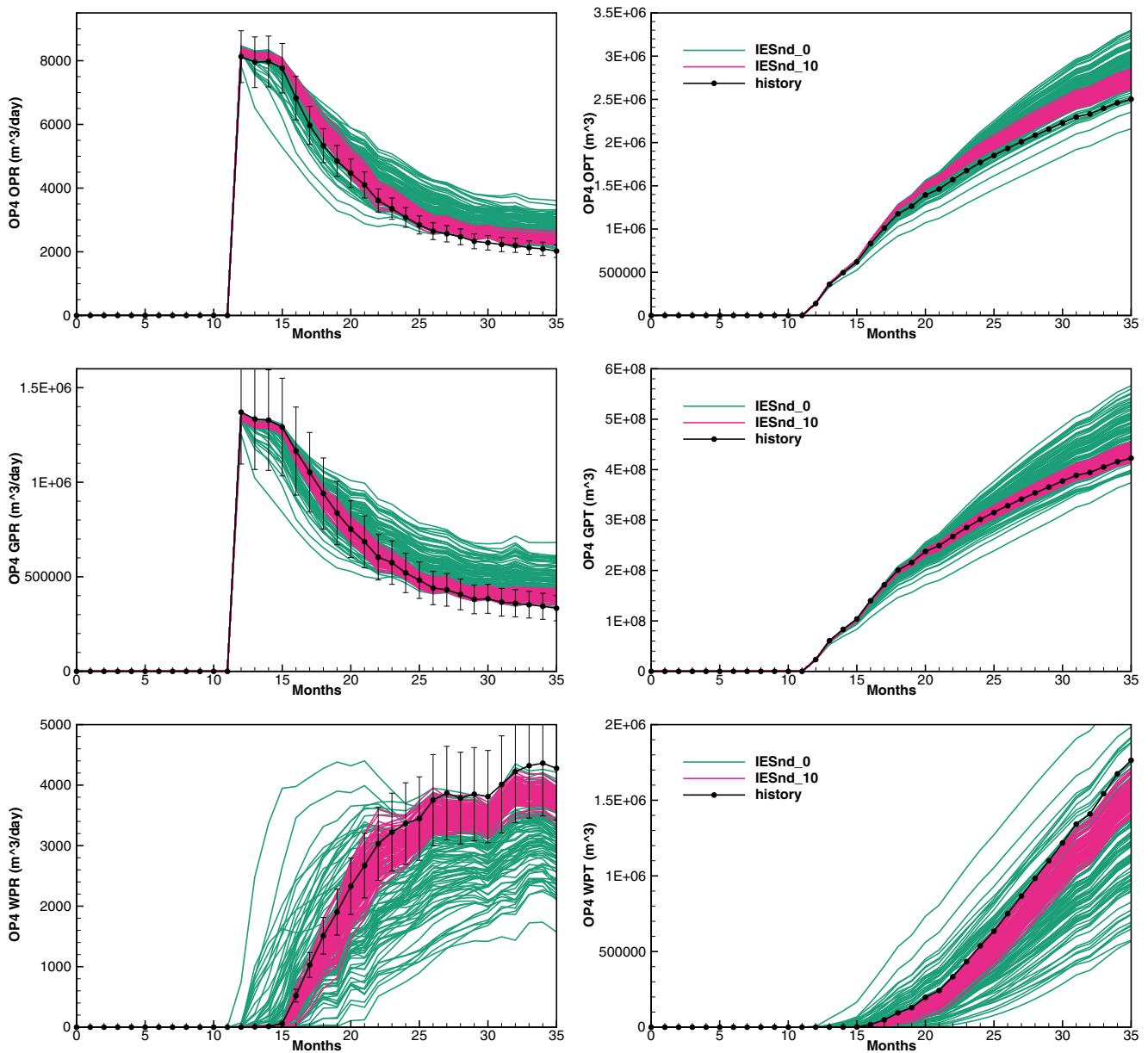
GPT, and the bottom plots show water production WPR and WPT. The error bars on the rate data indicate the plus-minus two standard deviations of the rate errors assumed in the assimilation experiments

one for all wells and variables. Also, we observe a shift in the historical rates, e.g., towards lower gas production for both wells OP2 and OP4, and lower water production for OP2. The magnitude of the update applied to the rates is modest. The update of the model parameters balances the rates' update, based on their relative impact on the solution and their prior variances.

A general observation is that case IES0 results in a significant improvement when compared to the prior ensemble.

Thus, the traditional history-matching approach, using the subspace EnRML, can generate a posterior ensemble solution with predictive skills. However, we notice that the posterior IES0 ensemble has too low uncertainty for most of the wells, with an ensemble spread that is much less than the variance in the historical measurements we condition on. This underestimation of the posterior variance results from using a diagonal measurement error-covariance matrix and ignoring the measurement errors' stochastic contribution in
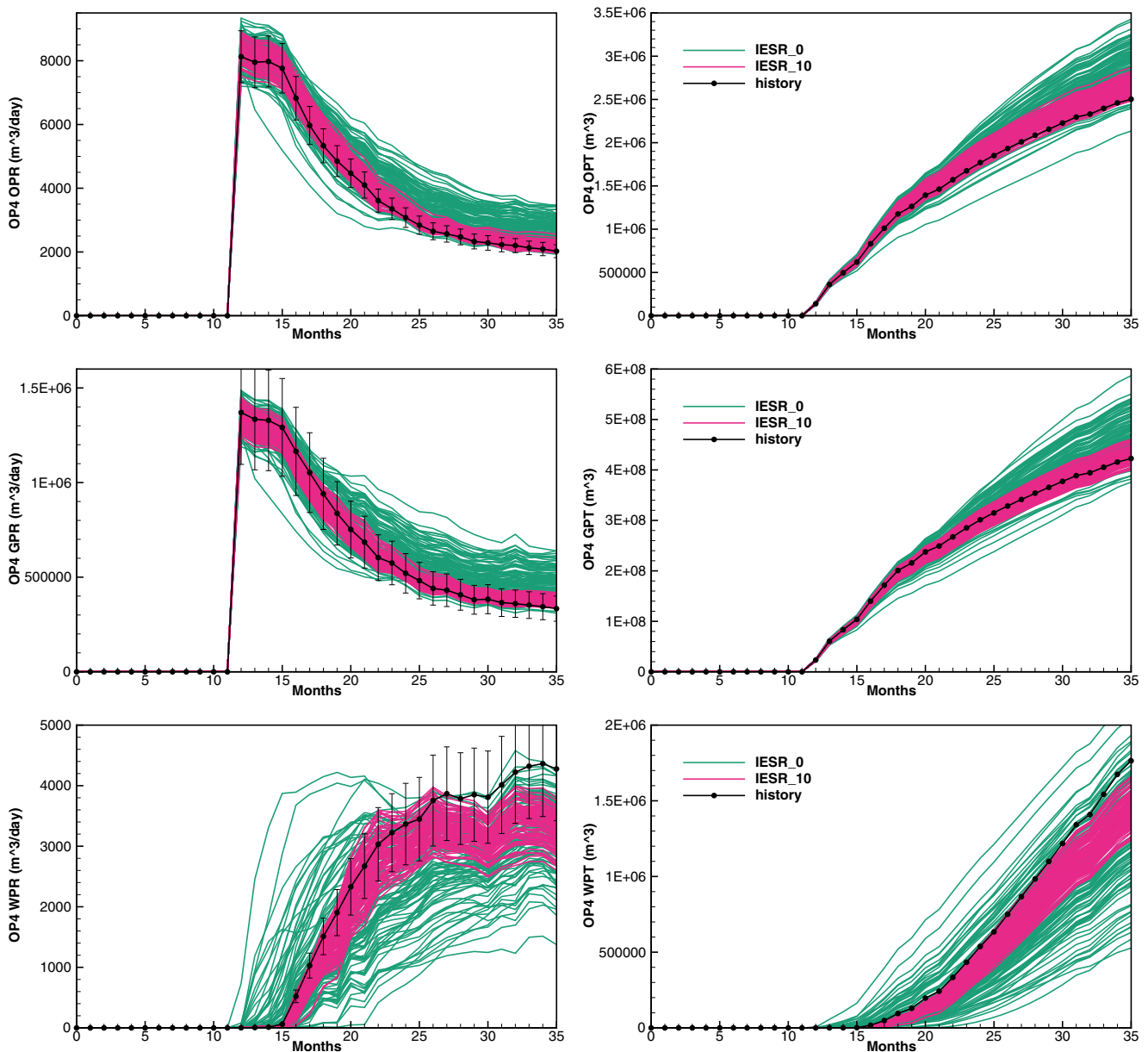
Fig. 12 Case IES0: Results from producer OP4. The upper plots show oil production rate OPR (left), and accumulated oil production OPT (right). Similarly, the middle plots are for gas production GPR and GPT, and the bottom plots show water production WPR and WPT. The error bars on the rate data indicate the plus-minus two standard deviations of the rate errors assumed in the assimilation experiments

the forward simulations. In several instances, the posterior ensemble is also unable to cover the observed history.

On the other hand, case IESR provides a slightly weaker update for the model parameters but compensates by updating the ensemble of perturbed historical rates. The result is a posterior prediction ensemble with significantly higher uncertainty, which in most instances, covers the historical data. The collective impact of updated rates and model parameters gives a better agreement with the observed

production. The result is a posterior ensemble consistent with the historical rates and more substantial and realistic uncertainty than in case IES0.

The final experiment, IESB, assumes "biased" rate errors for each realization. The numerical implementation of the analysis scheme, using the formulation from Section 2.5, handles this case with a rank-one **E** matrix. IESB has qualitatively similar behavior to IESR, but the more considerable spread of the prior IESB ensemble leads to higher

Fig. 13 Case IESnd: Results from producer OP4. The upper plots show oil production rate OPR (left), and accumulated oil production OPT (right). Similarly, the middle plots are for gas production GPR and GPT, and the bottom plots show water production WPR and WPT. The error bars on the rate data indicate the plus-minus two standard deviations of the rate errors assumed in the assimilation experiments

variance in the posterior IESB ensemble. We find that this ultimate increase of time correlations leads to a further weakening of the solution update. On the other hand, it seems that the overall match to the historical rate data may be better in IESB than IESR. Thus, maybe the assumption of infinite decorrelation time in the historical rate errors is accurate, e.g., like, if one extracts all a well's rate data using the same allocation table.

The uncertainty in the model parameters leads to biases in the individual realizations, and the posterior ensemble variance will grow with time, as we see from experiment IES0. However, the uncertainty in the controls in experiment IESR leads to more considerable uncertainty in the ensemble prediction early in the production period. However, after some time, the accumulated effect of the time-correlated rate-control errors starts canceling out. Thus, in the total

Fig. 14 Case IESR: Results from producer OP4. The upper plots show oil production rate OPR (left), and accumulated oil production OPT (right). Similarly, the middle plots are for gas production GPR and GPT, and the bottom plots show water production WPR and WPT. The error bars on the rate data indicate the plus-minus two standard deviations of the rate errors assumed in the assimilation experiments

production, we don't see a massive increase in the production uncertainty. This situation changes in experiment IESB where the rate-control errors act as a constant bias in time. An effect of the increased production uncertainty early in the simulation is that the first measurements will have a more substantial impact on the update.

In our experiments, we do not impose constraints on the total production of oil, gas, and water, as these quantities are uncertain. Thus, we allow the ensemble realizations

to produce different reservoir fluid quantities within their uncertainty range to span their uncertainty.

5 Summary and conclusions

The paper revisits the Bayesian formulation of the history-matching problem. It shows that we need to include stochastic errors in the rates used to force the ensemble
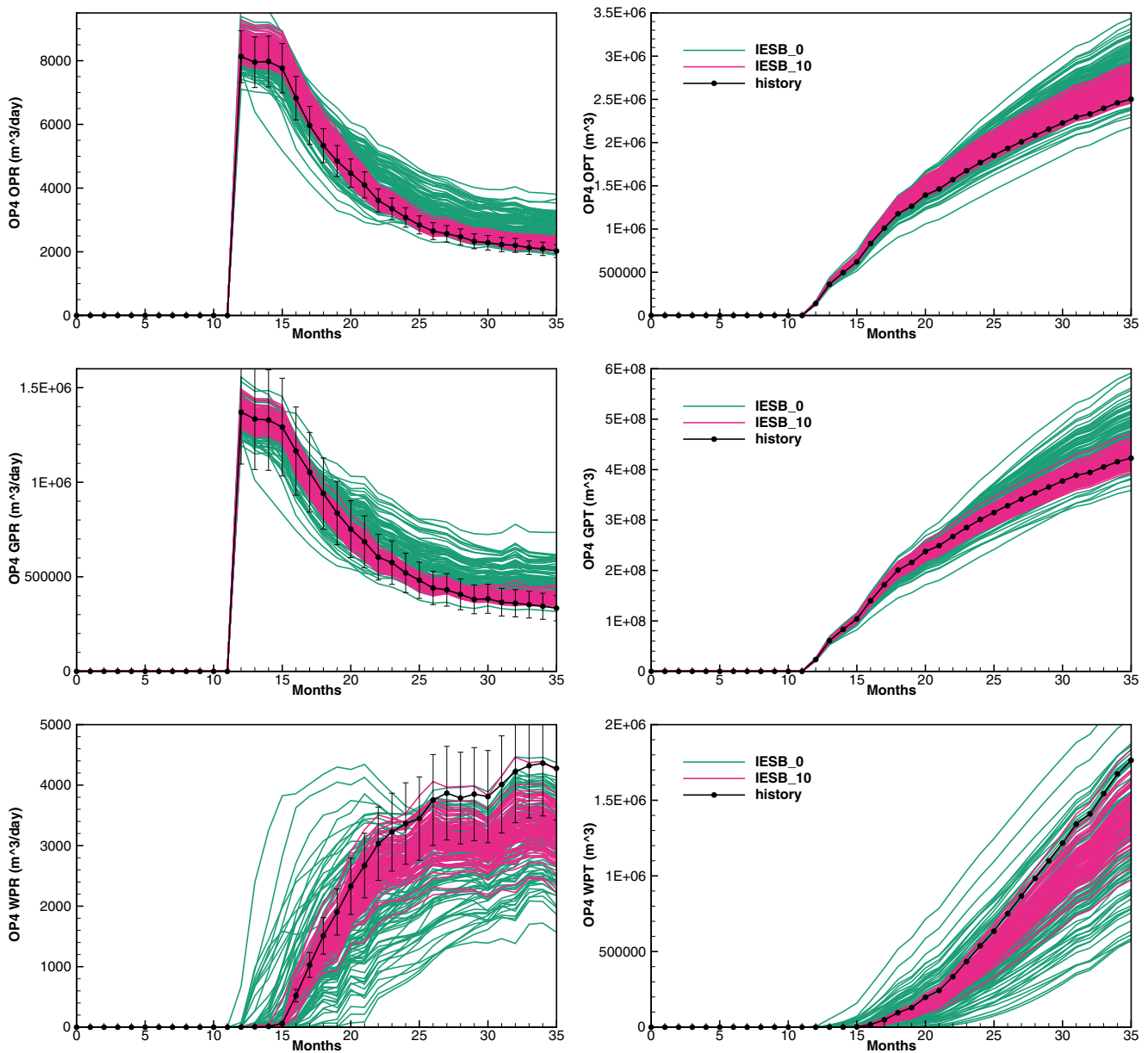
Fig. 15 Case IESB: Results from producer OP4. The upper plots show oil production rate OPR (left), and accumulated oil production OPT (right). Similarly, the middle plots are for gas production GPR and GPT, and the bottom plots show water production WPR and WPT. The error bars on the rate data indicate the plus-minus two standard deviations of the rate errors assumed in the assimilation experiments

simulations for a statistically consistent formulation. We must further augment these rates to the model state vector where they, together with the uncertain model parameters, are being updated.

The current manuscript complements the discussion in Evensen and Eikrem [11] on the impact of correlated rate errors and redundancy in rate measurements. They claimed that the negligence of correlated measurement errors would lead to an underestimate of the posterior variance. In extreme cases, with vast data sets, this underestimation may appear as a so-called "ensemble collapse" where all realizations become nearly identical, and the variance approaches zero. In Evensen and Eikrem [11] this problem was eluded by conditioning only on the total accumulated production for each fluid in each well. The current paper has used an alternative and statistically sound approach retaining all the data. However, it prescribes the correct measurement error statistics and implements a consistent
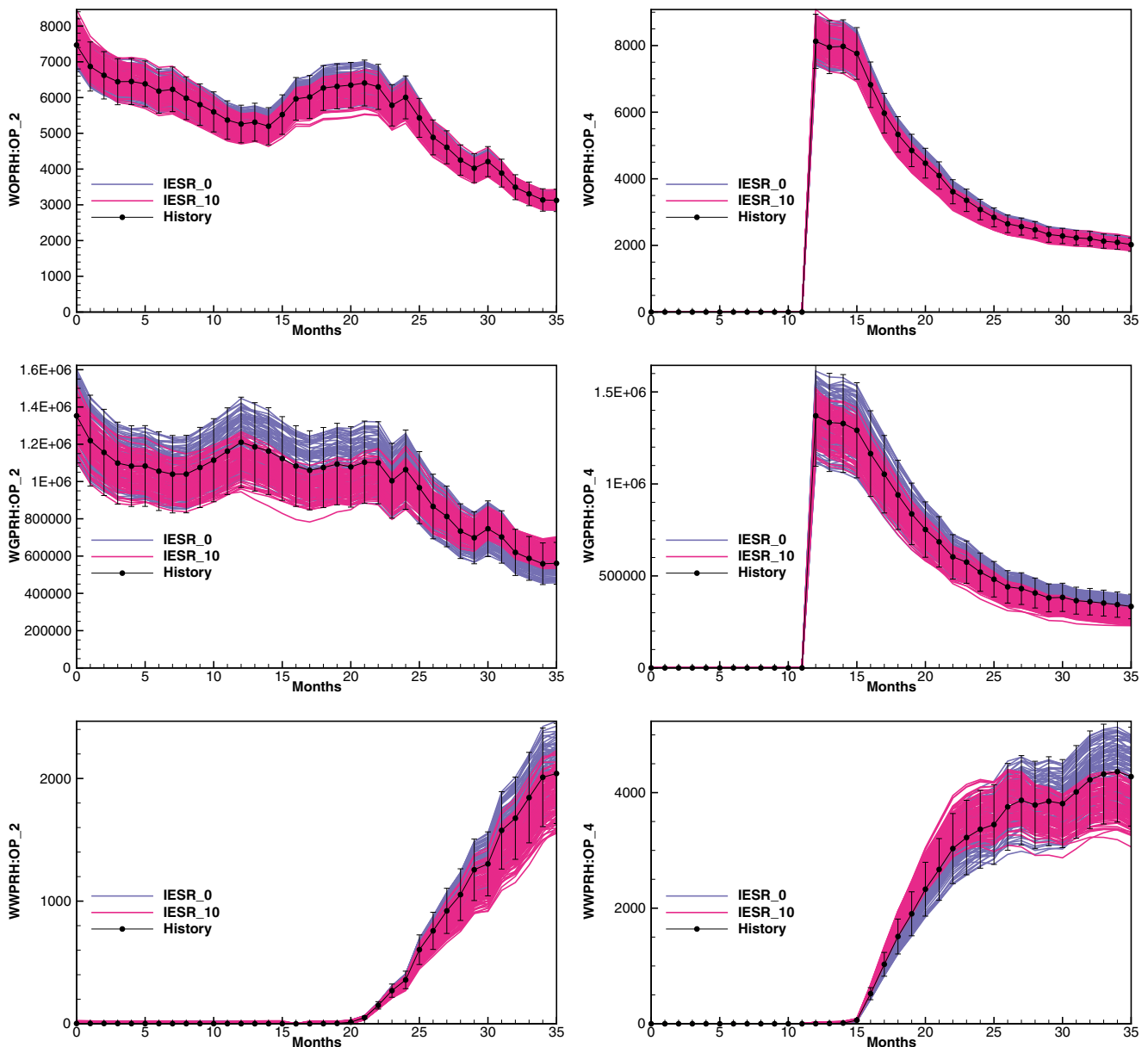
Fig. 16 IESR perturbed historic prior and posterior rates. The error bars on the rate data indicate the plus-minus two standard deviations of the rate errors assumed in the assimilation experiments

error treatment in the history-matching algorithm. The result is a posterior ensemble of models representing the real uncertainty and better agreeing with the historical rates.

The paper presents the new formulation's properties on a simple reservoir case that illustrates the benefits of using a complete measurement error-statistics treatment. The algorithm used to solve the conditioning problem is the subspace EnRML method described by Evensen et al. [12], combined with the procedure for correcting stochastic model errors in iterative smoothers presented by Evensen [10].

A section in the paper is dedicated to further examining and illustrating the consistency of the ensemble-subspace inversion algorithm [7, 8, 12] for cases with correlated measurement errors. This paper uses a larger ensemble of measurement perturbations to represent the measurement error-covariance matrix with reduced sampling errors. The article also examines the impact of the measurement errors' different correlation length scales compared to the ensemble members' smoothness. The examples show that in the case with shorter length scales in the measurement perturbations than in the ensemble realizations, the shorter

scales are truncated away when projecting the measurement perturbations onto the predicted measurements' ensemble. Thus, this is the only case where the new scheme introduces a significant approximation.

A practical implication of this work is that it resolves the reappearing question concerning the sampling frequency of rate data used in the conditioning. Using the current method, the sampling frequency of production rates does not matter as long as we specify realistic measurement-error correlations. The possibility of taking measurement error correlations into account will also be of vital importance when conditioning on sizeable seismic data sets. Seismic data often consists of gridded maps, and through the gridding procedure, one introduces additional error correlations. The proper formulation of measurement error statistics will also resolve the issues noted by Lorentzen et al. [19], related to the balanced impact of seismic data versus production data.

Algorithm 1, with the subspace inversion, is a computationally efficient approach for consistently solving the history matching problem. The algorithm's computational cost is linear in both the state dimension and the number of measurements, and the primary computation remains the integration of the model ensemble. Thus, Algorithm 1 is suitable for the conditioning of big models on big data.

The new algorithm's practical use will require reservoir engineers to model the rate data's error statistics properly. The common practice has been to set the error variance of the rate data to a fixed but sometimes arbitrary or subjective value and neglecting the presence of any measurement error correlations. In some situations, one has also inflated the measurement error variance to reduce the measurements' disproportionate impact when ignoring the measurement error correlations. Alternatively, it is possible to subsample the rate data in time to lessen their impact on the ensemble update. The recommendation is to specify as realistic as possible error variances for all the rate data, and there is no need for subsampling the data. The time correlations should reflect the procedure used to obtain the time series of rates. Unless one has access to a statistically sound method for estimating the time correlations of the measurement errors, the recommendation is to assume a time correlation with several months' decorrelation time. If we use allocation tables to obtain the production rates, then more than a year's decorrelation time is likely. Thus, for the reservoir engineers, the new algorithm doesn't pose a significant complication of use. In fact, by specifying one number, i.e., the time decorrelation of the errors in the production rates, one avoids previous issues with subsampling of data or inflation of measurement error variances to remedy formulation inconsistencies.

Finally, while the HM problem fits the model to past observations, the optimization problem maximizes future production or net-present value. The ensemble methods adapt well into a robust ensemble-based optimization framework [14, 18] where one can optimize future production and net-present value, and one can assess new optimal well-placements while taking the geological uncertainty into account.

Appendix: Implementation in ERT

The reservoir example adopts the Ensemble Reservoir Tool (ERT) available from Github https://github.com/equinor/ert.git for the management and conditioning of the model ensemble. The forward integration of the reservoir models is computed using the ECLIPSE reservoir simulator provided by Schlumberger. To be able to test the new formulation, some modifications were implemented in ERT and in the simulation job that runs ECLIPSE. The current version of ERT lacks functionality for simulating or specifying correlated measurement errors, and as most other assimilation implementations it works with a diagonal $\mathbf{C}_{dd}$. On the other hand, the actual implementation of the update scheme uses the same methods as discussed in the previous section, and has already full functionality for handling a nondiagonal measurement error-covariance matrix, $\mathbf{C}_{dd}$, or directly using the measurement perturbations, $\mathbf{E}$, see Evensen et al. [12]. The following sections briefly discuss how ERT was modified to accommodate for correlated measurement errors and stochastic forcing of the reservoir simulations. For now, the implementation uses a case-specific implementation with communication through files, but the plan is to upgrade ERT to generally support correlated measurement errors.

A.1 Convergence of the subspace EnRML

The step-length scheme used in Algorithm 2 (i.e., the value of γ), as implemented in ERT, is the following: One defines the maximum step length $t_1 = 0.5$, the minimum step length, $t_2 = 0.2$, and a step-length decline from one iteration to the next $t_3 = 2.5$, and we compute the value of γ_i, for iteration i, from

$$\gamma_i = t_2 + (t_1 - t_2)2^{\left(-(i-1)/(t_3-1)\right)}. \tag{49}$$

Here, γ_i follows a geometrical series starting from the value t_1 in the first iteration and then reducing geometrically with the number of iterations towards t_2. The formula in Eq. 49 allows us to define and test different step-length schemes easily. In real applications, it is desirable to start with a conservative step-length that works well "in most applications." If it turns out that the selected scheme becomes unstable, then the experiment should be restarted

with a new and more conservative step-length. There is no exact theory on how much we would need to reduce the step size, but a good starting point can be to reduce $t_1 = 0.5$ to $t_1 = 0.4$.

Algorithm 2 Subspace EnRML algorithm as modified in ERT to include estimation and use of correlated rate errors.

1: Input: $\mathbf{X}_0 \in \Re^{n \times N}$ (prior model ensemble)
2: Input: $\mathbf{D} \in \Re^{m \times N}$ (perturbed measurements)
3: $\mathbf{W}_0 = 0$ $\mathbf{W} \in \Re^{N \times N}$
4: $\mathbf{\Pi} = \left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}} \right) / \sqrt{N-1}$ $\mathbf{\Pi} \in \Re^{N \times N}$
5: Read from file: $\mathbf{E}_0 \in \Re^{m_u \times N}$ (simulated perturbations)
6: $\mathbf{D} = \mathbf{D}\frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}} + \mathbf{I}_e\mathbf{E}_0$ (mean of $\mathbf{D}$ plus new $\mathbf{E}_0$)
7: $\mathbf{E} = \mathbf{I}_e\mathbf{E}_0\,\mathbf{\Pi}$ (scaled $\mathbf{E}_0$ used in inversion)
8: i=0
9: **repeat**
10:     $\mathbf{Y}_i = \mathbf{g}(\mathbf{X}_i, \mathbf{E}_i)\,\mathbf{\Pi}$ $\mathbf{Y} \in \Re^{m \times N}$
11:     $\mathbf{\Omega}_i = \mathbf{I} + \mathbf{W}_i\mathbf{\Pi}$ $\mathbf{\Omega} \in \Re^{N \times N}$
12:     $\mathbf{S}_i = \mathbf{Y}_i\mathbf{\Omega}_i^{-1}$ $\mathbf{S} \in \Re^{m \times N}$
13:     $\mathbf{H}_i = \mathbf{S}_i\mathbf{W}_i + \mathbf{D} - \mathbf{g}(\mathbf{X}_i, \mathbf{E}_i)$ $\mathbf{H} \in \Re^{m \times N}$
14:     $\mathbf{W}_{i+1} = \mathbf{W}_i - \gamma \left(\mathbf{W}_i - \mathbf{S}_i^{\mathrm{T}}\left(\mathbf{S}_i\mathbf{S}_i^{\mathrm{T}} + \mathbf{E}\mathbf{E}^{\mathrm{T}}\right)^{-1}\mathbf{H}_i \right)$
15:     $\mathbf{T}_i = \left(\mathbf{I} + \mathbf{W}_{i+1}/\sqrt{N-1} \right)$ $\mathbf{T} \in \Re^{N \times N}$
16:     $\mathbf{X}_{i+1} = \mathbf{X}\mathbf{T}_i$
17:     $\mathbf{E}_{i+1} = \mathbf{E}_0\mathbf{T}_i$
18:     i=i+1
19: **until** convergence

The termination of the iterations can be based on the magnitude of the gradient, or one can use the relative change of the cost functions from one iteration to the next. For real reservoir applications, the time required per iteration can vary from hours to days, and the affordable number of iteration steps will always be limited. Thus, a practical procedure is to manually stop the iterations when one sees that the cost functions for the realizations are "almost" identical from one iteration to the next.

The rapid convergence of the assimilation algorithm for case IESR is illustrated in Fig. 17. Ten iterations were run, but after about five to six iterations there is only a marginal change in the parameters, and for practical applications five or six iterations should suffice for models of similar nonlinearity to the one used here.

A.2 Simulation of measurement perturbations

First, there is need of an application for simulating an ensemble of measurement perturbations for the time series of oil-production-rate (OPR), gas-production-rate (GPR), and water-production-rate (WPR), for each of the production wells. In the examples shown below, possible errors in the injection rates are ignored. The measurement perturbations can be white in time, red in time, or just a constant bias. The simulated ensemble of rate perturbations is

then stored in a file. The current implementation computes perturbations for all the rates given in the ECLIPSE schedule file, no matter if they are used or not in the conditioning. The code used for simulating the measurement perturbations is available from the Github repository https://github.com/geirev/EnKF_sampling/tree/ERTOBS/ERTsamp.

A.3 Importing E in the EnRML algorithm

In the current implementation, line 5 of Algorithm 2 illustrates how the subspace EnRML subroutine in ERT reads all the simulated rate perturbations from a file and store them in $\mathbf{E}_0$. The new measurement perturbations replace the original measurement perturbations supplied by ERT in $\mathbf{D}$ (line 6) and are input to the inversion in the iteration of $\mathbf{W}_i$ (line 14). The product $\mathbf{I}_e\mathbf{E}_0$ in lines 6 and 7, extracts the new simulated perturbations corresponding to the measurements contained in $\mathbf{D}$, which can typically be a subset of all the rates represented in $\mathbf{E}_0$. Thus, with this simple modification, it is possible to compute the EnRML update while allowing for correlated measurement errors.

A.4 Updating measurement perturbations Eᵢ

As soon as one has computed the transition matrix, $\mathbf{T}_i$, for iteration i, in line 15 of Algorithm 2, it is possible to augment the measurement perturbations to the model state vector and update them according to

$$\begin{pmatrix} \mathbf{X}_i \\ \mathbf{E}_i \end{pmatrix} = \begin{pmatrix} \mathbf{X}_0 \\ \mathbf{E}_0 \end{pmatrix} \mathbf{T}_i. \tag{50}$$
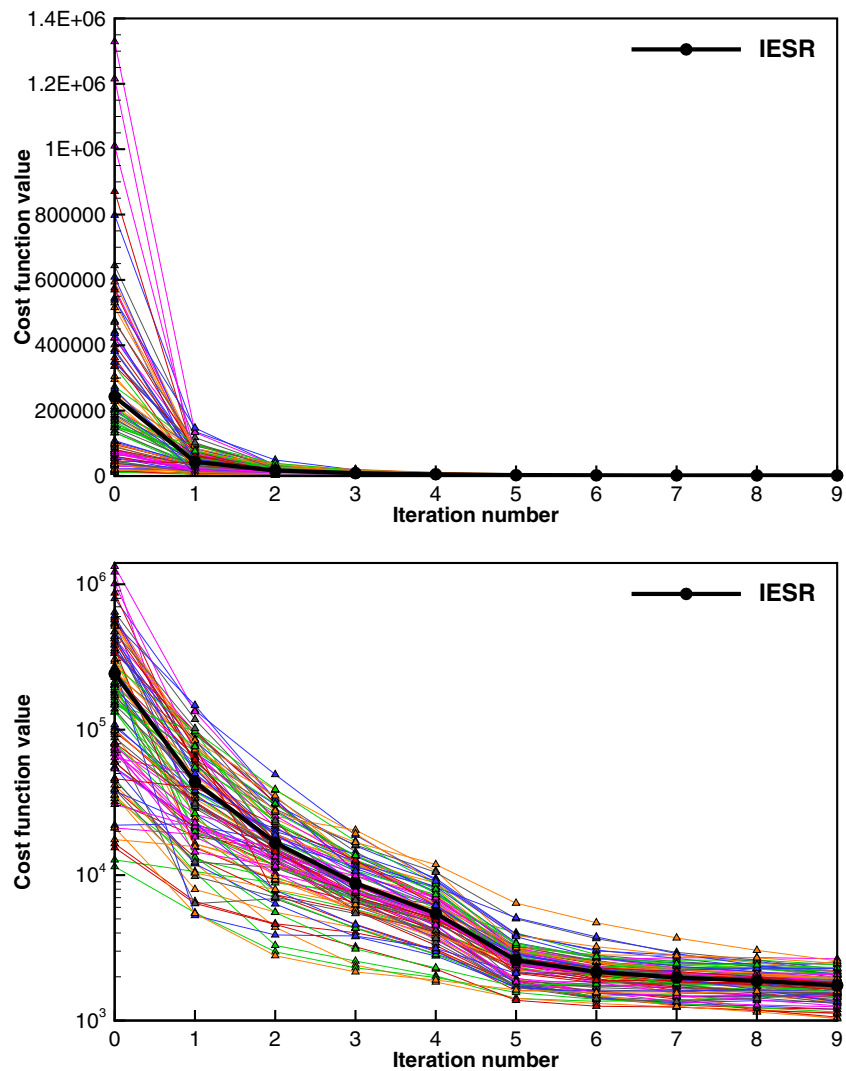
The beautiful property of the subspace EnRML algorithm is its independence of the model state vector, which only enters as input to the forward model. Thus, augmenting the measurement errors to the state vector does not impact the computational procedure used to obtain the transition matrix. However, when the measurement errors are input to the forward model, they will change the model predicted measurements, and thus lead to a different transition matrix.

A.5 Updating historical rates in the schedule file

In iteration i, the forward model is forced by the updated perturbations, $\mathbf{g}(\mathbf{X}_i, \mathbf{E}_i)$. Practically, one needs to update the Eclipse schedule file with new historical OPR, GPR, and WPR rates, by adding the updated perturbations to the corresponding "WCONHIST" rates. ERT handles this schedule modification by running a simple program, which is available from https://github.com/geirev/Schedule_parser.git. This program reads the updated perturbations in $\mathbf{E}_i$ from file and adds them to the historical rates defined in the original schedule file. ERT runs this program immediately before starting the Eclipse simulation for each realization.

Fig. 17 Convergence of each
realization using the subspace
EnRML Algorithm 1 for the
IESR case. The thick black line
is the mean of the ensemble of
cost functions. The lower plot is
the same as the upper one, but
using a log scale for the cost
function value

References

1. Aanonsen, S.I., Naevdal, G., Oliver, D.S., Reynolds, A., Valles, B.: Ensemble Kalman filter in reservoir engineering – a review. SPE J. SPE-117274-PA **14**(3), 393–412 (2009). https://doi.org/10.21188/117274-PA

2. Carrassi, A., Bocquet, M., Bertino, L., Evensen, G.: Data Assimilation in the Geosciences: An overview on methods, issues and perspectives. Wires Clim. Change **9**(5), 50 (2018). https://doi.org/10.1002/wcc.535

3. Chen, Y., Oliver, D.S.: Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. Computat. Geosci. **17**, 689–703 (2013)

4. Chen, Y., Oliver, D.S.: History matching of the Norne full-field model using an iterative ensemble smoother. SPE Reserv. Eval. Eng. **17**(2), 244–256 (2014)

5. Emerick, A.A.: Analysis of performance of ensemble-based assimilation of production and seismic data. J. Petroleum Sci. Eng. **139**, 219–239 (2016)

6. Emerick, A.A., Reynolds, A.C.: Ensemble smoother with multiple data assimilation. Comput. Geosci. **55**, 3–15 (2013)

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

8. Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, 2nd edn. Springer (2009)

9. Evensen, G.: Analysis of iterative ensemble smoothers for solving inverse problems. Computat. Geosci. **22**(3), 885–908 (2018). https://doi.org/10.1007/s10596-018-9731-y

10. Evensen, G.: Accounting for model errors in iterative ensemble smoothers. Computat. Geosci **23**(4), 761–775 (2019). https://doi.org/10.1007/s10596-019-9819-z

11. Evensen, G., Eikrem, K.S.: Strategies for conditioning reservoir models on rate data using ensemble smoothers. Computat. Geosci. **22**(5), 1251–1270 (2018). https://doi.org/10.1007/s10596-018-9750-8

12. Evensen, G., Raanes, P., Stordal, A., Hove, J.: Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching. Front. Appl. Math. Stat. **5**, 47 (2019). https://doi.org/10.3389/fams.2019.00047, https://www.frontiersin.org/article/10.3389/fams.2019.00047

13. Hanea, R., Evensen, G., Hustoft, L., Ek, T., Chitu, A., Wilschut, F.: Reservoir management under geological uncertainty using fast model update. In: SPE Reservoir Simulation Symposium, Society of Petroleum Engineers, p. 12, https://doi.org/10.2118/173305-MS (2015)

14. Hanea, R.G., Casanova, P., Hustoft, L., Bratvold, R.B., Nair, R., Hewson, C., Leeuwenburgh, O., Fonseca, R.M.: Drill and learn: A decision making workflow to quantify value of learning. In: SPE Reservoir Simulation Conference. Society of Petroleum Engineers (2017)

15. Houtekamer, P.L., Zhang, F.: Review of the ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. **144**, 4489–4533 (2016)

16. Hunt, B.R., Kostelich, E.J., Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Physica D **230**, 112–126 (2007)

17. Kitanidis, P.K.: Quasi-linear geostatistical theory for inversing. Water Resour. Res. **31**(10), 2411–2419 (1995)

18. Leeuwenburgh, O., Chitu, A.G., Nair, R., Egberts, P.JP., Ghaz-aryan, L., Feng, T., Hustoft, L.: Ensemble-based methods for well drilling sequence and time optimization under uncertainty. In: ECMOR XV-15th European Conference on the Mathematics of Oil Recovery (2016)

19. Lorentzen, R., Luo, X., Bhakta, T., Valestrand, R.: History matching the full norne field model using seismic and production data. SPE J. SPE-194205-MS **24**, 1452–1467 (2019). https://doi.org/10.2118/194205-PA

20. Luo, X., Bhakta, T.: Automatic and adaptive localization for ensemble-based history matching. Petroleum Science and Engineering, p. 184. https://doi.org/10.1016/j.petrol.2019.106559 (2020)

21. Oliver, D., Reynolds, A., Liu, N.: Inverse Theory for Petroleum Reservoir Characterization and History Matching. Cambridge University Press, Cambridge (2008)

22. Oliver, D.S., He, N., Reynolds, A.C.: Conditioning permeability fields to pressure data. In: ECMOR V-5th European Conference on the Mathematics of Oil Recovery, p. 11, https://doi.org/10.3997/ 2214-4609.201406884 (1996)

23. Perrone, A., Pennadoro, F., Tiani, A., Della Rossa, E., Saetrom, J.: Enhancing the geological models consistency in ensemble based history matching an integrated approach. Society of Petroleum Engineers. https://doi.org/10.2118/186049-MS (2017)

24. Raanes, P.N., Stordal, A.S., Evensen, G.: Revising the stochastic iterative ensemble smoother. Nonlin Processes Geophys **26**, 325–338 (2019). https://doi.org/10.5194/npg-2019-10

25. Skjervheim, J.A., Evensen, G., Hove, J., Vabø, J.: An ensemble smoother for assisted history matching. SPE, 141929 (2011)

26. Wang, L., Oliver, D.S.: Efficient optimization of well drilling sequence with learned heuristics. SPE J **24**(5), 2111–2134 (2019). https://doi.org/10.2118/195640-PA