ORIGINAL PAPER

Accounting for model errors in iterative ensemble smoothers

Geir Evensen[1,2]

Abstract

In the strong-constraint formulation of the history-matching problem, we assume that all the model errors relate to a selection of uncertain model input parameters. One does not account for additional model errors that could result from, e.g., excluded uncertain parameters, neglected physics in the model formulation, the use of an approximate model forcing, or discretization errors resulting from numerical approximations. If parameters with significant uncertainties are unaccounted for, there is a risk for an unphysical update, of some uncertain parameters, that compensates for errors in the omitted parameters. This paper gives the theoretical foundation for introducing model errors in ensemble methods for history matching. In particular, we explain procedures for practically including model errors in iterative ensemble smoothers like ESMDA and IES, and we demonstrate the impact of adding (or neglecting) model errors in the parameter-estimation problem. Also, we present a new result regarding the consistency of using the sample covariance of the predicted nonlinear measurements in the update schemes.

Keywords Model errors · Iterative ensemble smoothers · History matching · Data assimilation · IES · ESMDA

1 Introduction

It is standard to assume the model to be perfect when we use ensemble smoothers for solving inverse problems. This paper addresses the problem of consistently including the additional effect of stochastic model errors in different ensemble smoothers. In particular, we consider methods such as Ensemble Smoother (ES) [12, 16], Iterative ES (IES) [5, 6], and Ensemble Smoother with Multiple Data Assimilations (ESMDA) [9].

There is a vast literature on solving the data assimilation problem in the presence of model errors (see, e.g., the reviews by Carrassi and Vannitsem [4] and Harlim [17]). We traditionally characterize the data-assimilation problem as being either a weak-constraint or a strong-constraint problem, dependent on whether we include the dynamical model as a strong or a weak constraint in the cost function. The classical book by Bennett [1] is entirely devoted to solving the weak-constraint inverse problem, and it

illustrates that in the weak-constraint case, we need to invert simultaneously for the state vector as a function of both space and time. In Eknes and Evensen [7], the variational formulation was solved for a weak-constraint parameter and state estimation problem using the representer method by Bennett [1], and in Evensen [11], different methods including ensemble methods were used to solve a weak-constraint state- and parameter-estimation problem.

A conclusion from these works is that, if model errors are present, we need to increase the dimension of the problem by either updating the model solution as a function of space and time or by estimating the actual model errors and thereafter solve for the model solution. Also, when using Ensemble Kalman Filter (EnKF) by Evensen [10] or ES, it is relatively easy to include stochastic model errors as long as we update both the parameters and the state variables simultaneously [11, 13].

At this point, we should mention that we do not distinguish between model errors and model bias. In fact, with time-correlated stochastic model errors, and if the correlation becomes perfect (equal to one), then the model errors become equivalent to a constant bias. Fortunately, the procedure outlined in this paper can be used to estimate both model errors and model bias.

In a recent paper by Sakov et al. [23], the Iterative EnKF was reformulated to allow for additive model errors. However, for the history-matching problem, we need to

✉ Geir Evensen
 geir.evensen@norceresearch.no

1 Norwegian Research Centre (NORCE), Bergen, Norway

2 Nansen Environmental and Remote Sensing Center (NERSC), Bergen, Norway

 Springer

account for more general representations of the error term since the solution of the reservoir simulator depends nonlinearly on the errors in the rate data used to force the model.

We will start by defining the standard strong-constraint history-matching problem, and after that, we move on to the general weak-constraint formulation. We then formally derive the ES, ESMDA, and IES, in the presence of model errors. The different smoother methods are used in a simple example to demonstrate the consistency of the formulation and to illustrate the impact of model errors on the parameter-estimation problem.

2 Standard history-matching problem

The strong-constraint formulation given by Evensen [14] is attractive because it simplifies the inverse problem and efficient ensemble smoothers can be defined. A first fundamental assumption is that we have a perfect deterministic forward model where the prediction $\mathbf{y}$ only depends on the input model parametrization $\mathbf{x}$,

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

In a reservoir history-matching problem, the model operator is the reservoir simulation model, which predicts the observed production of oil, water, and gas, from the reservoir. Thus, given the true parametrization of $\mathbf{x}$, the true prediction of $\mathbf{y}$ is precisely determined by the model in Eq. 1. Also, we have measurements $\mathbf{d}$ of the true prediction $\mathbf{y}$ given as

$$\mathbf{d} = \mathbf{y} + \mathbf{e}. \tag{2}$$

From evaluating the model operator $\mathbf{g}(\mathbf{x})$, given a realization of the uncertain model parameters $\mathbf{x} \in \Re^n$, we uniquely determine a realization of predicted measurements $\mathbf{y} \in \Re^m$ (corresponding to the real measurements $\mathbf{d} \in \Re^m$). Here, n is the number of parameters and m the number of measurements. We want to use the measurements to estimate the parameters $\mathbf{x}$, and the measurements are assumed to contain random errors $\mathbf{e} \in \Re^m$.

In history matching, it is common to define a prior distribution for the uncertain parameters since we usually will have more degrees of freedom in the parameters, than we have independent information in the measurements. Bayes' theorem gives the joint posterior pdf for $\mathbf{x}$ and $\mathbf{y}$ as

$$\begin{aligned} f(\mathbf{x}, \mathbf{y} \mid \mathbf{d}) &\propto f(\mathbf{x}, \mathbf{y}) f(\mathbf{d} \mid \mathbf{y}) \\ &= f(\mathbf{x}) f(\mathbf{y} \mid \mathbf{x}) f(\mathbf{d} \mid \mathbf{y}). \end{aligned} \tag{3}$$

In the case of no model errors, the transition density $f(\mathbf{y} \mid \mathbf{x})$ becomes the Dirac delta function, and we can write

$$f(\mathbf{x}, \mathbf{y} \mid \mathbf{d}) \propto f(\mathbf{x})\delta(\mathbf{y} - \mathbf{g}(\mathbf{x})) f(\mathbf{d} \mid \mathbf{y}). \tag{4}$$

We are interested in the marginal pdf for $\mathbf{x}$, which we obtain by integrating (4) over $\mathbf{y}$, giving

$$\begin{aligned} f(\mathbf{x} \mid \mathbf{d}) &\propto \int f(\mathbf{x})\delta(\mathbf{y} - \mathbf{g}(\mathbf{x})) f(\mathbf{d} \mid \mathbf{y}) \mathrm{d}\mathbf{y} \\ &= f(\mathbf{x}) f(\mathbf{d} \mid \mathbf{g}(\mathbf{x})). \end{aligned} \tag{5}$$

When introducing the normal priors

$$f(\mathbf{x}) = \mathcal{N}(\mathbf{x}^{\mathrm{f}}, \mathbf{C}_{xx}), \tag{6}$$

$$f(\mathbf{d} \mid \mathbf{g}(\mathbf{x})) = f(\mathbf{e}) = \mathcal{N}(\mathbf{0}, \mathbf{C}_{dd}), \tag{7}$$

we can write Eq. 5 as

$$\begin{aligned} f(\mathbf{x} \mid \mathbf{d}) \propto &\exp\left\{ -\tfrac{1}{2}\left(\mathbf{x} - \mathbf{x}^{\mathrm{f}}\right)^{\mathrm{T}} \mathbf{C}_{xx}^{-1}\left(\mathbf{x} - \mathbf{x}^{\mathrm{f}}\right) \right\} \\ &\times \exp\left\{ -\tfrac{1}{2}\left(\mathbf{g}(\mathbf{x}) - \mathbf{d}\right)^{\mathrm{T}} \mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{x}) - \mathbf{d}\right) \right\}. \end{aligned} \tag{8}$$

Note that the posterior pdf in Eq. 8 is non-Gaussian due to the nonlinear model $\mathbf{g}(\mathbf{x})$. Maximizing $f(\mathbf{x} \mid \mathbf{d})$ is equivalent to minimizing the cost function

$$\begin{aligned} \mathscr{J}(\mathbf{x}) = &\left(\mathbf{x} - \mathbf{x}^{\mathrm{f}}\right)^{\mathrm{T}} \mathbf{C}_{xx}^{-1}\left(\mathbf{x} - \mathbf{x}^{\mathrm{f}}\right) \\ &+ \left(\mathbf{g}(\mathbf{x}) - \mathbf{d}\right)^{\mathrm{T}} \mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{x}) - \mathbf{d}\right). \end{aligned} \tag{9}$$

Most methods for history matching apply the assumptions of a perfect model and Gaussian priors, and they attempt to solve either one of Eqs. 8 or 9.

For this strong-constraint problem, Evensen [14] explained how (8) or (9) can be approximately solved using the ES, ESMDA, and IES. We can interpret these methods to approximately sample the posterior pdf in Eq. 8, and we can easily derive them as an ensemble of minimizing solutions of the cost function in Eq. 9 written for each realization as

$$\begin{aligned} \mathscr{J}(\mathbf{x}_j) = &\left(\mathbf{x}_j - \mathbf{x}_j^{\mathrm{f}}\right)^{\mathrm{T}} \mathbf{C}_{xx}^{-1}\left(\mathbf{x}_j - \mathbf{x}_j^{\mathrm{f}}\right) \\ &+ \left(\mathbf{g}(\mathbf{x}_j) - \mathbf{d}_j\right)^{\mathrm{T}} \mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{x}_j) - \mathbf{d}_j\right). \end{aligned} \tag{10}$$

This approach relates to the papers on Ensemble Randomized Likelihood (EnRML) [18, 20]. Note that the minimizing solutions will not precisely sample the posterior non-Gaussian distribution.

3 ES solution

We start by defining the covariance matrix $\mathbf{C}_{xy}$ between two vectors $\mathbf{x}$ and $\mathbf{y}$ as the expectation,

$$\mathbf{C}_{xy} = \mathbb{E}\left[\left(\mathbf{x} - \mathbb{E}[\mathbf{x}]\right)\left(\mathbf{y} - \mathbb{E}[\mathbf{y}]\right)^{\mathrm{T}} \right], \tag{11}$$

where $\mathbb{E}[\mathbf{x}]$ and $\mathbb{E}[\mathbf{y}]$ define the expectations or means of $\mathbf{x}$ and $\mathbf{y}$. In the ensemble methods, we introduce the sample mean

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{j=1}^{N} \mathbf{x}_j, \tag{12}$$

and the sample covariance between two arbitrary vectors $\mathbf{x}$ and $\mathbf{y}$ as

$$\overline{\mathbf{C}}_{xy} = \frac{1}{N-1} \sum_{j=1}^{N} \left(\mathbf{x}_j - \overline{\mathbf{x}}\right)\left(\mathbf{y}_j - \overline{\mathbf{y}}\right)^{\mathrm{T}}. \tag{13}$$

We obtain the ES solution by first sampling the prior parameters $\mathbf{x}_j^{\mathrm{f}}$ and the perturbed measurements $\mathbf{d}_j$ from

$$\mathbf{x}_j^{\mathrm{f}} \sim \mathcal{N}(\mathbf{x}^{\mathrm{f}}, \mathbf{C}_{xx}), \tag{14}$$
$$\mathbf{d}_j \sim \mathcal{N}(\mathbf{d}, \mathbf{C}_{dd}). \tag{15}$$

We then compute the model predictions $\mathbf{y}_j^{\mathrm{f}}$ from

$$\mathbf{y}_j^{\mathrm{f}} = \mathbf{g}(\mathbf{x}_j^{\mathrm{f}}). \tag{16}$$

Evensen [14] derived the ES update equations for the parameters $\mathbf{x}_j^{\mathrm{a}}$ and the predictions $\mathbf{y}_j^{\mathrm{a}}$ as

$$\mathbf{x}_j^{\mathrm{a}} = \mathbf{x}_j^{\mathrm{f}} + \overline{\mathbf{C}}_{xy} \left(\overline{\mathbf{C}}_{yy} + \mathbf{C}_{dd}\right)^{-1} \left(\mathbf{d}_j - \mathbf{y}_j^{\mathrm{f}}\right), \tag{17}$$
$$\mathbf{y}_j^{\mathrm{a}} = \mathbf{g}(\mathbf{x}_j^{\mathrm{a}}), \tag{18}$$

but see also the alternative derivation in Sections 4.3–4.6 below regarding the consistency of using the sample covariance $\overline{\mathbf{C}}_{yy}$ in Eq. 17.

Thus, to compute the ES solution, we start from an initial ensemble of parameters $\mathbf{x}_j^{\mathrm{f}}$ and perturbed measurements $\mathbf{d}_j$. We generate a prior ensemble prediction $\mathbf{y}_j^{\mathrm{f}}$ by evaluating the model in Eq. 16 for each realization $\mathbf{x}_j^{\mathrm{f}}$. Then we compute the prior ensemble covariance between the parameters and the predicted measurements $\overline{\mathbf{C}}_{xy} \in \mathfrak{R}^{n \times m}$ and the ensemble covariance of the predicted measurements $\overline{\mathbf{C}}_{yy} \in \mathfrak{R}^{m \times m}$ and use them in the ES update in Eq. 17. Finally, we can recompute the model prediction using the updated parameters $\mathbf{x}_j^{\mathrm{a}}$ in Eq. 18.

As an alternative to solving the model in Eq. 18 for $\mathbf{y}_j^{\mathrm{a}}$, it is possible to compute the updated prediction directly from an update equation

$$\mathbf{y}_j^{\mathrm{a}} = \mathbf{y}_j^{\mathrm{f}} + \overline{\mathbf{C}}_{yy} \left(\overline{\mathbf{C}}_{yy} + \mathbf{C}_{dd}\right)^{-1} \left(\mathbf{d}_j - \mathbf{y}_j^{\mathrm{f}}\right), \tag{19}$$

and in the case with a linear model, the result would be identical to solving (18). However, due to the nonlinearity of the deterministic model, the Eqs. 18 and 19 will give different results for $\mathbf{y}_j^{\mathrm{a}}$. Equation 19 is just the standard ES update $\mathbf{y}_j^{\mathrm{a}}$ given the prior forecast ensemble for $\mathbf{y}_j^{\mathrm{f}}$ and the perturbed measurements $\mathbf{d}_j$ of $\mathbf{y}$. We will later show that, for nonlinear models, there may be a benefit of computing $\mathbf{y}_j^{\mathrm{a}}$ indirectly through integration of the model in Eq. 18, initialized with $\mathbf{x}_j^{\mathrm{a}}$, rather than using the direct update in Eq. 19. Also, the indirect update in Eq. 18 allows for the use of iterations.

4 General weak-constraint problem

Lets now look at a formulation where we assume that the model depends nonlinearly on the model errors $\mathbf{q}$ as well as the parameters $\mathbf{x}$, i.e.,

$$\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{q}). \tag{20}$$

The nonlinear model operator is quite general and can represent any numerical or mathematical model, including a recursion or time steps. The model error is a vector of noise components that can represent any unresolved physical process, errors from the numerical computation, or errors in model forcing data. An example of the latter is the time-correlated noise in rate data used to force a reservoir simulation model. Thus, there is a significant difference between $\mathbf{x}$ and $\mathbf{q}$. Firstly, $\mathbf{x}$ is a static parameter that does not change with time. Thus, once the parameters $\mathbf{x}$ are estimated, we can use them in a future prediction simulation. The model errors $\mathbf{q}$, on the other hand, will vary in time and are only estimated for the period of the inverse calculation. Only in the case of time-correlated errors or when $\mathbf{q}$ represents a bias, can the estimated $\mathbf{q}$ be used as a forcing in the future prediction.

We assume that we have prior pdfs for the parameters and the model errors, $f(\mathbf{x})$ and $f(\mathbf{q})$. It is common, but not necessary, to assume that the model errors and parameters are independent, and we can then write the joint prior pdf as $f(\mathbf{x}, \mathbf{q}) = f(\mathbf{x})f(\mathbf{q})$. The transition density for the model evolution is

$$f(\mathbf{y}|\mathbf{x}, \mathbf{q}) = \delta\left(\mathbf{y} - \mathbf{g}(\mathbf{x}, \mathbf{q})\right), \tag{21}$$

and we obtain the joint pdf by multiplying the prior with the transition density,

$$f(\mathbf{y}, \mathbf{x}, \mathbf{q}) = \delta\left(\mathbf{y} - \mathbf{g}(\mathbf{x}, \mathbf{q})\right)f(\mathbf{x})f(\mathbf{q}). \tag{22}$$

The likelihood function for the measurements given the prediction $\mathbf{y}$ is $f(\mathbf{d}|\mathbf{y})$; thus, the posterior conditional pdf becomes

$$f(\mathbf{y}, \mathbf{x}, \mathbf{q}|\mathbf{d}) \propto f(\mathbf{d}|\mathbf{y})\delta\left(\mathbf{y} - \mathbf{g}(\mathbf{x}, \mathbf{q})\right)f(\mathbf{x})f(\mathbf{q}). \tag{23}$$

Since $\mathbf{y}$ is given by the model in Eq. 20 as soon as we know $\mathbf{x}$ and $\mathbf{q}$, we can compute the marginal density

$$\begin{aligned} f(\mathbf{x}, \mathbf{q}|\mathbf{d}) &\propto \int f(\mathbf{d}|\mathbf{y})\delta\left(\mathbf{y} - \mathbf{g}(\mathbf{x}, \mathbf{q})\right)f(\mathbf{x})f(\mathbf{q})\mathrm{d}\mathbf{y} \\ &= f\left(\mathbf{d}|\mathbf{g}(\mathbf{x}, \mathbf{q})\right)f(\mathbf{x})f(\mathbf{q}). \end{aligned} \tag{24}$$

4.1 Gaussian priors

We now assume Gaussian priors

$$f(\mathbf{x}) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{x} - \mathbf{x}^{\mathrm{f}}\right)^{\mathrm{T}}\mathbf{C}_{xx}^{-1}\left(\mathbf{x} - \mathbf{x}^{\mathrm{f}}\right)\right\}, \tag{25}$$

and

$$f(\mathbf{q}) \propto \exp\left\{ -\frac{1}{2}(\mathbf{q} - \mathbf{q}^{\mathrm{f}})^{\mathrm{T}} \mathbf{C}_{qq}^{-1}(\mathbf{q} - \mathbf{q}^{\mathrm{f}}) \right\}, \tag{26}$$

where $\mathbf{q}^{\mathrm{f}}$ would normally be zero. The likelihood for the measurements becomes

$$\begin{aligned} f(\mathbf{d}|\mathbf{g}(\mathbf{x}, \mathbf{q})) = \\ \exp\left\{ -\frac{1}{2}(\mathbf{g}(\mathbf{x}, \mathbf{q}) - \mathbf{d})^{\mathrm{T}} \mathbf{C}_{dd}^{-1}(\mathbf{g}(\mathbf{x}, \mathbf{q}) - \mathbf{d}) \right\}, \end{aligned} \tag{27}$$

and we write the marginal posterior as

$$\begin{aligned} f(\mathbf{x}, \mathbf{q}|\mathbf{d}) &\propto \\ \exp\Big\{ &-\frac{1}{2}(\mathbf{x} - \mathbf{x}^{\mathrm{f}})^{\mathrm{T}} \mathbf{C}_{xx}^{-1}(\mathbf{x} - \mathbf{x}^{\mathrm{f}}) \\ &-\frac{1}{2}(\mathbf{q} - \mathbf{q}^{\mathrm{f}})^{\mathrm{T}} \mathbf{C}_{qq}^{-1}(\mathbf{q} - \mathbf{q}^{\mathrm{f}}) \\ &-\frac{1}{2}(\mathbf{g}(\mathbf{x}, \mathbf{q}) - \mathbf{d})^{\mathrm{T}} \mathbf{C}_{dd}^{-1}(\mathbf{g}(\mathbf{x}, \mathbf{q}) - \mathbf{d}) \Big\}. \end{aligned} \tag{28}$$

Maximizing the posterior pdf in Eq. 28 is equivalent to minimizing the cost function

$$\begin{aligned} \mathscr{J}(\mathbf{x}, \mathbf{q}) = &(\mathbf{x} - \mathbf{x}^{\mathrm{f}})^{\mathrm{T}} \mathbf{C}_{xx}^{-1}(\mathbf{x} - \mathbf{x}^{\mathrm{f}}) \\ &+ (\mathbf{q} - \mathbf{q}^{\mathrm{f}})^{\mathrm{T}} \mathbf{C}_{qq}^{-1}(\mathbf{q} - \mathbf{q}^{\mathrm{f}}) \\ &+ (\mathbf{g}(\mathbf{x}, \mathbf{q}) - \mathbf{d})^{\mathrm{T}} \mathbf{C}_{dd}^{-1}(\mathbf{g}(\mathbf{x}, \mathbf{q}) - \mathbf{d}). \end{aligned} \tag{29}$$

As in the strong-constraint case, we sample the priors and define a cost function for each sample realization, and we obtain the weak-constraint analog to the strong-constraint cost function in Eq. 10, i.e.,

$$\begin{aligned} \mathscr{J}(\mathbf{x}_j, \mathbf{q}_j) = &(\mathbf{x}_j - \mathbf{x}_j^{\mathrm{f}})^{\mathrm{T}} \mathbf{C}_{xx}^{-1}(\mathbf{x}_j - \mathbf{x}_j^{\mathrm{f}}) \\ &+ (\mathbf{q}_j - \mathbf{q}_j^{\mathrm{f}})^{\mathrm{T}} \mathbf{C}_{qq}^{-1}(\mathbf{q}_j - \mathbf{q}_j^{\mathrm{f}}) \\ &+ (\mathbf{g}(\mathbf{x}_j, \mathbf{q}_j) - \mathbf{d}_j)^{\mathrm{T}} \mathbf{C}_{dd}^{-1}(\mathbf{g}(\mathbf{x}_j, \mathbf{q}_j) - \mathbf{d}_j). \end{aligned} \tag{30}$$

4.2 Stationarity condition

To develop a consistent set of equations for the different methods, it is simpler to rewrite the cost function for an augmented variable $\mathbf{z}^{\mathrm{T}} = (\mathbf{x}^{\mathrm{T}}, \mathbf{q}^{\mathrm{T}})$. We have now redefined n to be the dimension of $\mathbf{z}$. We can then define the covariance

$$\mathbf{C}_{zz} = \begin{pmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xq} \\ \mathbf{C}_{qx} & \mathbf{C}_{qq} \end{pmatrix}, \tag{31}$$

where we allow for correlations between $\mathbf{x}$ and $\mathbf{q}$ since this correlation becomes important in the iterative methods, and we can rewrite the cost function in Eq. 30 as

$$\begin{aligned} \mathscr{J}(\mathbf{z}_j) = &(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}})^{\mathrm{T}} \mathbf{C}_{zz}^{-1}(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}}) \\ &+ (\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j)^{\mathrm{T}} \mathbf{C}_{dd}^{-1}(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j). \end{aligned} \tag{32}$$

This cost function is slightly more general than the one in Eq. 30 since we do not require independence between $\mathbf{q}$ and

$\mathbf{x}$. By taking the gradient with respect to $\mathbf{z}_j$, we obtain the following closed system of nonlinear equations for $\mathbf{z}_j$

$$\mathbf{C}_{zz}^{-1}(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}}) + \nabla_{\mathbf{z}}\mathbf{g}(\mathbf{z}_j)\mathbf{C}_{dd}^{-1}(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j) = 0. \tag{33}$$

In the case of a linear model, the solution is the standard Kalman filter update equation for the mean. However, the presence of the nonlinear function $\mathbf{g}(\mathbf{z}_j)$ makes the solution more elaborate. All of ES, ESMDA, and IES are developed to solve this system of equations.

4.3 Derivation of ES update equations

To derive the ES solution, we will use a Taylor expansion and approximation that allows us to obtain a closed form solution for each realization of $\mathbf{z}_j$, i.e.,

$$\mathbf{g}(\mathbf{z}_j) = \mathbf{g}(\mathbf{z}_j^{\mathrm{f}}) + \mathbf{G}_j(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}}), \tag{34}$$

where we have defined

$$\mathbf{G}_j^{\mathrm{T}} = \nabla_{\mathbf{z}}\mathbf{g}(\mathbf{z})\big|_{\mathbf{z}=\mathbf{z}_j^{\mathrm{f}}}. \tag{35}$$

Thus, we approximate the nonlinear function $\mathbf{g}(\mathbf{z}_j)$ with its linearization in Eq. 34 around $\mathbf{z} = \mathbf{z}_j^{\mathrm{f}}$ and in addition evaluate the gradient in Eq. 33 at $\mathbf{z}_j^{\mathrm{f}}$. We now have the gradient of the model, $\mathbf{G}_j$ defined in Eq. 35, which differ for each realization. We wish to replace the individual model gradients with an "averaged" gradient that is common for all realizations, and for now, we denote it $\mathbf{G}$, which allows us to write Eq. 33 as

$$\left(\mathbf{C}_{zz}^{-1} + \mathbf{G}^{\mathrm{T}}\mathbf{C}_{dd}^{-1}\mathbf{G}\right)(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}}) = \mathbf{G}^{\mathrm{T}}\mathbf{C}_{dd}^{-1}(\mathbf{d}_j - \mathbf{g}(\mathbf{z}_j^{\mathrm{f}})). \tag{36}$$

By solving for $\mathbf{z}_j$ and using the matrix identity

$$\left(\mathbf{G}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{G} + \mathbf{C}^{-1}\right)^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{D}^{-1} = \mathbf{C}\mathbf{G}^{\mathrm{T}}\left(\mathbf{G}\mathbf{C}\mathbf{G}^{\mathrm{T}} + \mathbf{D}\right)^{-1}, \tag{37}$$

(which can be derived from the Woodbury identity), where we substitute $\mathbf{C}_{zz}$ for $\mathbf{C}$, and $\mathbf{C}_{dd}$ for $\mathbf{D}$, we obtain the solution for $\mathbf{z}_j$ as

$$\mathbf{z}_j = \mathbf{z}_j^{\mathrm{f}} + \mathbf{C}_{zz}\mathbf{G}^{\mathrm{T}}\left(\mathbf{G}\mathbf{C}_{zz}\mathbf{G}^{\mathrm{T}} + \mathbf{C}_{dd}\right)^{-1}\left(\mathbf{d}_j - \mathbf{g}(\mathbf{z}_j^{\mathrm{f}})\right). \tag{38}$$

4.4 Linear regression for G

Evensen [14] used a Taylor expansion of $\mathbf{g}(\mathbf{z}_j)$ around the ensemble mean $\bar{\mathbf{z}} = \mathbf{z}^{\mathrm{f}}$ and could replace the individual gradients in Eq. 35 evaluated at $\mathbf{z}_j$ with the gradient $\mathbf{G}_{\bar{\mathbf{z}}}$ evaluated at the ensemble mean $\bar{\mathbf{z}}$. He then showed that

$$\overline{\mathbf{C}}_{yz} \approx \mathbf{G}_{\bar{\mathbf{z}}}\overline{\mathbf{C}}_{zz}, \tag{39}$$

and

$$\overline{\mathbf{C}}_{yy} \approx \mathbf{G}_{\bar{\mathbf{z}}}\overline{\mathbf{C}}_{zz}\mathbf{G}_{\bar{\mathbf{z}}}^{\mathrm{T}}, \tag{40}$$

Evensen [14, Eqs. 29 and 30], but see also Reynolds et al. [21]. He could then replace the analytical gradients with sample covariances. This approach is mathematically consistent, but the combined use of these two approximations introduces an inconsistency in the update equation which is discussed below.

Rather than defining $\mathbf{G}$ as the tangent-linear model evaluated at the ensemble mean, we will here use an interpretation based on linear regression (see also Reynolds et al. [21] and Chen and Oliver [6]) where we start by defining

$$\mathbf{G} \triangleq \mathbf{C}_{yz}\mathbf{C}_{zz}^{-1}. \tag{41}$$

Thus, we view $\mathbf{G}$ as the sensitivity matrix in a linear regression between $\mathbf{y}$ and $\mathbf{z}$, as

$$\mathbf{G}\mathbf{C}_{zz} = \mathbf{C}_{yz}. \tag{42}$$

Also, using Eq. 41, we can write

$$\mathbf{G}\mathbf{C}_{zz}\mathbf{G}^{\mathrm{T}} = \mathbf{C}_{yz}\mathbf{C}_{zz}^{-1}\mathbf{C}_{zy}. \tag{43}$$

These two equations are mutually consistent for any definition of $\mathbf{G}$ and they are not derived from Taylor expansions as was used to obtain (39) and (40). The right-hand side of Eq. 43 is not generally equal to $\mathbf{C}_{yy}$ as is often assumed.

We can replace the gradient $\mathbf{G}$ in the update (38), using Eqs. 42 and 43 to obtain

$$\mathbf{z}_j^{\mathrm{a}} = \mathbf{z}_j^{\mathrm{f}} + \mathbf{C}_{zy}\left(\mathbf{C}_{yz}\mathbf{C}_{zz}^{-1}\mathbf{C}_{zy} + \mathbf{C}_{dd}\right)^{-1}\left(\mathbf{d}_j - \mathbf{g}(\mathbf{z}_j^{\mathrm{f}})\right). \tag{44}$$

The solution of this equation is identical to the solution of Eq. 36 with $\mathbf{G}$ defined by Eq. 41. However, if we replace $\mathbf{C}_{yz}\mathbf{C}_{zz}^{-1}\mathbf{C}_{zy}$ with $\mathbf{C}_{yy}$ in Eq. 44, the solutions of Eqs. 36 and 44 will differ in the nonlinear case.

4.5 Ensemble representation

We now introduce the ensemble-anomaly matrices

$$\mathbf{Y} = \frac{1}{\sqrt{N-1}}\left(\mathbf{y}_1 - \overline{\mathbf{y}}, \ldots, \mathbf{y}_N - \overline{\mathbf{y}}\right) \in \Re^{m \times N}, \tag{45}$$

$$\mathbf{Z} = \frac{1}{\sqrt{N-1}}\left(\mathbf{z}_1 - \overline{\mathbf{z}}, \ldots, \mathbf{z}_N - \overline{\mathbf{z}}\right) \in \Re^{n \times N}, \tag{46}$$

and write ensemble representations of Eqs. 42 and 43 as

$$\mathbf{G}\overline{\mathbf{C}}_{zz} \approx \overline{\mathbf{C}}_{yz} \quad \text{or} \quad \mathbf{G}\mathbf{Z} \approx \mathbf{Y}, \tag{47}$$

and

$$\mathbf{G}\overline{\mathbf{C}}_{zz}\mathbf{G}^{\mathrm{T}} \approx \overline{\mathbf{C}}_{yz}\overline{\mathbf{C}}_{zz}^{+}\overline{\mathbf{C}}_{zy} = \mathbf{Y}(\mathbf{Z}^{+}\mathbf{Z})(\mathbf{Z}^{+}\mathbf{Z})\mathbf{Y}^{\mathrm{T}}, \tag{48}$$

where the superscript $^{+}$ denotes pseudo inverse. The symmetric matrix $\mathbf{Z}^{+}\mathbf{Z}$ defines the orthogonal projection onto the range of $\mathbf{Z}^{\mathrm{T}}$.

We will now consider three cases:

4.5.1 Linear model

For a linear model (and linear measurement operator), we have $\mathbf{Y} \equiv \mathbf{G}\mathbf{Z}$, and Eq. 48 becomes

$$\begin{aligned}\overline{\mathbf{C}}_{yz}\overline{\mathbf{C}}_{zz}^{+}\overline{\mathbf{C}}_{zy} &= \mathbf{G}\mathbf{Z}(\mathbf{Z}^{+}\mathbf{Z})(\mathbf{Z}^{+}\mathbf{Z})\mathbf{Z}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}} \\ &= \mathbf{G}\mathbf{Z}(\mathbf{G}\mathbf{Z})^{\mathrm{T}} = \mathbf{Y}\mathbf{Y}^{\mathrm{T}} = \overline{\mathbf{C}}_{yy}.\end{aligned} \tag{49}$$

Hence, Eq. 49 is consistent with the definition of the covariance matrix $\overline{\mathbf{C}}_{yy} = \mathbf{Y}\mathbf{Y}^{\mathrm{T}}$ for linear models and all combinations of n and N.

4.5.2 Nonlinear model and $n \geq N - 1$

In the case with a nonlinear model and $n \geq N - 1$ the rank of $\mathbf{Z}$ is $N - 1$, and the projection $\mathbf{Z}^{+}\mathbf{Z} = \left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)$ with $\mathbf{1} \in \Re^{N}$ being a vector with all elements equal to one (see the Appendix in Sakov et al. [22]). This result is seen from the fact that $\mathbf{Z}$ has only one singular value equal to zero corresponding to the right singular vector $\mathbf{1}/\sqrt{N}$. The projection $\mathbf{Z}^{+}\mathbf{Z}$ is then just the subtraction of the ensemble mean. Since we have already removed the ensemble mean from $\mathbf{Y}$, we can write Eq. 48 as

$$\begin{aligned}\overline{\mathbf{C}}_{yz}\overline{\mathbf{C}}_{zz}^{+}\overline{\mathbf{C}}_{zy} &= \mathbf{Y}\left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\mathbf{Y}^{\mathrm{T}} \\ &= \mathbf{Y}\mathbf{Y}^{\mathrm{T}} = \overline{\mathbf{C}}_{yy},\end{aligned} \tag{50}$$

and as in the linear case, Eq. 48 exactly corresponds to the definition of the ensemble covariance $\overline{\mathbf{C}}_{yy}$. This nonlinear case with $n \geq N - 1$ is the most considered case for history matching, data assimilation with nonlinear measurement operators, and iterative smoothers used for sequential data assimilation in nonlinear models. Thus, for most applications of ensemble methods, we can replace the product $\overline{\mathbf{C}}_{yz}\overline{\mathbf{C}}_{zz}^{+}\overline{\mathbf{C}}_{zy}$ with the sample covariance $\overline{\mathbf{C}}_{yy}$.

4.5.3 Nonlinear model and $n < N - 1$

In the case of a nonlinear model and $n < N - 1$, which applies for the example considered in Section 6, the expression in Eq. 48 is not equal to $\mathbf{Y}\mathbf{Y}^{\mathrm{T}}$ and we must include the projection and redefine the sample covariance as

$$\widetilde{\mathbf{C}}_{yy} \triangleq \mathbf{Y}(\mathbf{Z}^{+}\mathbf{Z})\left(\mathbf{Y}(\mathbf{Z}^{+}\mathbf{Z})\right)^{\mathrm{T}}, \tag{51}$$

i.e., we compute the covariance of the predicted measurement anomalies projected onto the range of $\mathbf{Z}^{\mathrm{T}}$. This case also applies for *nonlinear models in the limit of infinite ensemble size*. Thus, we must use the definition (51) to evaluate $\widetilde{\mathbf{C}}_{yy}$ to ensure consistency in the derivation of the update equation. The importance of including the projection

is further discussed in the next section, and illustrated in the examples in Section 6

4.6 ES algorithm

By representing the covariances using a finite sample size and sample covariance matrices, we get the update equation for the finite ensemble of N realizations as

$$\mathbf{z}_j^{\text{a}} = \mathbf{z}_j^{\text{f}} + \overline{\mathbf{C}}_{zy}\left(\widetilde{\mathbf{C}}_{yy} + \mathbf{C}_{dd}\right)^{-1}\left(\mathbf{d}_j - \mathbf{g}(\mathbf{z}_j^{\text{f}})\right), \tag{52}$$

where $\widetilde{\mathbf{C}}_{yy}$ is defined in Eq. 51. If we omit the projection in Eq. 51, then the solution computed by Eq. 44 (using sample covariances) and Eq. 52 will differ in the case of a nonlinear model and $n < N - 1$, also when N becomes infinitely large. Thus, our update will be biased.

To compute the ES update, we start by sampling the Gaussian prior variables for the parameters $\mathbf{x}_j^{\text{f}}$, the model errors $\mathbf{q}_j^{\text{f}}$, and the measurement perturbations $\mathbf{e}_j$,

$$\mathbf{x}_j^{\text{f}} \sim \mathcal{N}(\mathbf{x}^{\text{f}}, \mathbf{C}_{xx}^{\text{f}}), \tag{53}$$

$$\mathbf{q}_j^{\text{f}} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{qq}^{\text{f}}), \tag{54}$$

$$\mathbf{e}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{dd}). \tag{55}$$

The vector $\mathbf{q}$ contains all stochastic model errors over the time interval of the model integration, and the errors can also have correlations in time. We obtain the model prediction from the model written on the form

$$\mathbf{y}_j^{\text{f}} = \mathbf{g}(\mathbf{x}_j^{\text{f}}, \mathbf{q}_j^{\text{f}}), \tag{56}$$

where the model operator depends nonlinearly on the model-error term. Next, we can compute the sample covariances $\overline{\mathbf{C}}_{xy}^{\text{f}}$, and $\overline{\mathbf{C}}_{qy}^{\text{f}}$, from the ensembles of $\mathbf{y}_j^{\text{f}}$, $\mathbf{x}_j^{\text{f}}$, and $\mathbf{q}_j^{\text{f}}$, and $\widetilde{\mathbf{C}}_{yy}^{\text{f}}$ from the definition in Eq. 51.

Equation 52 defines the final update equations for $\mathbf{x}_j^{\text{a}}$ and $\mathbf{q}_j^{\text{a}}$ which becomes

$$\mathbf{x}_j^{\text{a}} = \mathbf{x}_j^{\text{f}} + \overline{\mathbf{C}}_{xy}^{\text{f}}\left(\widetilde{\mathbf{C}}_{yy}^{\text{f}} + \mathbf{C}_{dd}\right)^{-1}\left(\mathbf{d}_j - \mathbf{y}_j^{\text{f}}\right), \tag{57}$$

$$\mathbf{q}_j^{\text{a}} = \mathbf{q}_j^{\text{f}} + \overline{\mathbf{C}}_{qy}^{\text{f}}\left(\widetilde{\mathbf{C}}_{yy}^{\text{f}} + \mathbf{C}_{dd}\right)^{-1}\left(\mathbf{d}_j - \mathbf{y}_j^{\text{f}}\right). \tag{58}$$

We then rerun the model using the updates $\mathbf{x}_j^{\text{a}}$ and $\mathbf{q}_j^{\text{a}}$ to get

$$\mathbf{y}_j^{\text{a}} = \mathbf{g}(\mathbf{x}_j^{\text{a}}, \mathbf{q}_j^{\text{a}}). \tag{59}$$

Alternatively, we can also compute the update of the predicted measurements from

$$\mathbf{y}_j^{\text{a}} = \mathbf{y}_j^{\text{f}} + \overline{\mathbf{C}}_{yy}^{\text{f}}\left(\overline{\mathbf{C}}_{yy}^{\text{f}} + \mathbf{C}_{dd}\right)^{-1}\left(\mathbf{d}_j - \mathbf{y}_j^{\text{f}}\right), \tag{60}$$

and in the case of a linear model the result would be identical to that obtained by integrating the model in Eq. 59.

5 Iterative smoothers in the presence of model errors

The critical approximations used in the derivation of ES are, firstly, the linearization in Eq. 34 of the model about $\mathbf{x}_j^{\text{f}}$ meaning that large updates will have large errors, and secondly, that an averaged statistical ensemble gradient replaces the exact analytic gradients. Only a single linear update step is computed, and with strong nonlinearities, these approximations may lead to poor results as was discussed in Evensen [14].

The minimization problem in Eq. 10 can be solved using iterative methods like IES, ESMDA, and IEnKF. The iterative ensemble smoother (IES) by Chen and Oliver [5, 6] minimizes the ensemble of cost functions by direct minimization using an approximate ensemble gradient. Alternatively, the Ensemble Smoother with Multiple Data Assimilations (ESMDA) by Emerick and Reynolds [8] rewrites the ES update equation as a sequence of recursive updates by inflating the measurement errors, and thereby reduces the approximation introduced by the linearization in the ES update.

Sakov et al. [22] and Bocquet and Sakov [3] derived the iterative EnKF (IEnKF) and iterative Ensemble Kalman Smoother (IEnKS) to better handle nonlinearities in the dynamical model and the observation operator. The focus was on state estimation where the model state at the time t_i is updated using measurements of the state at time t_{i+1}. IEnKF solves the same kind of problem as given by the marginal conditional pdf in Eq. 5 or the cost function in Eq. 9. In a recent paper by Sakov et al. [23], the IEnKF was extended to account for additive model errors, and the method should also be applicable for the history-matching problem in the presence of additive model errors.

In the following, we will present variants of ESMDA and IES that take more general model errors into account as is required when solving the weak-constraint history-matching problem.

5.1 ESMDA

As explained in Evensen [14], ESMDA solves the standard ES update equations using a predefined number of recursive steps. In each step, the measurement error covariance and associated measurement perturbations are inflated to reduce the impact of the measurements. With correctly chosen inflation factors and linear model and observation operators, the ESMDA update precisely replicates the ES update. When the model or observation operators are nonlinear, it turns out that the use of multiple short update steps reduces the errors and improves the solution as compared to using one long update step in ES.

From the previous discussion, it is clear that, in the presence of model errors, we need to recursively update both

the parameters and the model errors. It is easiest to derive ESMDA by using a tempering of the likelihood function [19] which leads to a recursive minimization of a sequence of N_{mda} cost functions, [14, 24],

$$
\begin{aligned}
\mathscr{J}\left(\mathbf{z}_{j,i+1}\right) = {} & \left(\mathbf{z}_{j,i+1} - \mathbf{z}_{j,i}\right)^{\mathrm{T}}\left(\mathbf{C}_{zz}^{i}\right)^{-1}\left(\mathbf{z}_{j,i+1} - \mathbf{z}_{j,i}\right) \\
& + \left(\mathbf{g}(\mathbf{z}_{j,i+1}) - \mathbf{d} - \sqrt{\alpha_{i+1}}\mathbf{e}_{j,i}\right)^{\mathrm{T}} \\
& \times \left(\alpha_{i+1}\mathbf{C}_{dd}\right)^{-1}\left(\mathbf{g}(\mathbf{z}_{j,i+1}) - \mathbf{d} - \sqrt{\alpha_{i+1}}\mathbf{e}_{j,i}\right),
\end{aligned}
\tag{61}
$$

where we evaluate $\mathbf{C}_{zz}^{i}$ at the ith iterate $\mathbf{z}_{i}$, and we must have

$$
\sum_{i=1}^{N_{\text{mda}}} \frac{1}{\alpha_i} = 1. \tag{62}
$$

Similarly to the derivation of the ES in the previous section, we obtain the recursive update equations for ESMDA given by Eqs. 68 and 69 in the algorithm below. As in ES, the update direction is computed based on a linearization around the prior realizations of each update step. Thus, we can interpret the ES update as taking one long Euler step of length $\Delta\tau = 1$ in pseudo time τ, while in ESMDA, we take a predefined number of shorter Euler steps of step length $\Delta\tau_i = 1/\alpha_i$ that satisfy (62) (see, e.g., the discussion in Evensen [14]).

To compute the ESMDA solution, we start by sampling the initial ensembles from Eqs. 53 and 54 to initialize the recursion in ESMDA

$$
\mathbf{x}_{j,0} \sim \mathcal{N}(\mathbf{x}^{\mathrm{f}}, \mathbf{C}_{xx}), \tag{63}
$$

$$
\mathbf{q}_{j,0} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{qq}). \tag{64}
$$

Then the model is integrated according to Eq. 56 to obtain the prior ensemble prediction for the first ESMDA step,

$$
\mathbf{y}_{j,0} = \mathbf{g}(\mathbf{x}_{j,0}, \mathbf{q}_{j,0}), \tag{65}
$$

and we compute recursively the following for each iteration $i = 0, \ldots, N_{\text{mda}} - 1$:

We construct the sample covariances $\overline{\mathbf{C}}_{xy}^{i}$, and $\overline{\mathbf{C}}_{qy}^{i}$, from the ensembles of $\mathbf{y}_{j,i}$, $\mathbf{x}_{j,i}$, and $\mathbf{q}_{j,i}$, and $\widetilde{\mathbf{C}}_{yy}^{i}$ from the definition in Eq. 51, and we sample the measurement perturbations

$$
\mathbf{e}_{j,i} \sim \mathcal{N}(\mathbf{0}, \alpha_{i+1}\mathbf{C}_{dd}), \tag{66}
$$

used to generate the perturbed measurements

$$
\mathbf{d}_{j,i} = \mathbf{d} + \mathbf{e}_{j,i}. \tag{67}
$$

We then compute the updates

$$
\mathbf{x}_{j,i+1} = \mathbf{x}_{j,i} + \overline{\mathbf{C}}_{xy}^{i}\left(\widetilde{\mathbf{C}}_{yy}^{i} + \alpha_{i+1}\mathbf{C}_{dd}\right)^{-1}\left(\mathbf{d}_{j,i} - \mathbf{y}_{j,i}\right), \tag{68}
$$

$$
\mathbf{q}_{j,i+1} = \mathbf{q}_{j,i} + \overline{\mathbf{C}}_{qy}^{i}\left(\widetilde{\mathbf{C}}_{yy}^{i} + \alpha_{i+1}\mathbf{C}_{dd}\right)^{-1}\left(\mathbf{d}_{j,i} - \mathbf{y}_{j,i}\right), \tag{69}
$$

and rerun the model to obtain the updated prediction

$$
\mathbf{y}_{j,i+1} = \mathbf{g}(\mathbf{x}_{j,i+1}, \mathbf{q}_{j,i+1}), \tag{70}
$$

for step $i + 1$. We repeat this procedure until $i = N_{\text{mda}} - 1$, which results in the ESMDA solution for $\mathbf{x}_j$, $\mathbf{q}_j$ and $\mathbf{y}_j$.

5.2 IES

In IES, we use a gradient-based minimization method, and we need to evaluate the first and second order derivatives of the cost function in Eq. 32 with respect to $\mathbf{z}$. The gradient of the cost function in Eq. 32 is already derived above as Eq. 33

$$
\nabla_{\mathbf{z}}\mathscr{J}(\mathbf{z}_j) = \mathbf{C}_{zz}^{-1}(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}}) + \nabla_{\mathbf{z}}\mathbf{g}(\mathbf{z}_j)\mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j\right). \tag{71}
$$

An approximation to the Hessian of the cost function is obtained by operating again by $\nabla_{\mathbf{z}}$ on the gradient in Eq. 71 to obtain

$$
\nabla_{\mathbf{z}}\nabla_{\mathbf{z}}\mathscr{J}(\mathbf{z}_j) \approx \mathbf{C}_{zz}^{-1} + \nabla_{\mathbf{z}}\mathbf{g}(\mathbf{z}_j)\mathbf{C}_{dd}^{-1}\left(\nabla_{\mathbf{z}}\mathbf{g}(\mathbf{z}_j)\right)^{\mathrm{T}}, \tag{72}
$$

where we have neglected the second derivatives or Hessian of the vector function $\mathbf{g}(\mathbf{z})$, i.e., $\nabla_{\mathbf{z}}\nabla_{\mathbf{z}}\mathbf{g}(\mathbf{z})$. We can then write a Gauss-Newton iteration

$$
\mathbf{z}_{j,i+1} = \mathbf{z}_{j,i} - \gamma \Delta\mathbf{z}_{j,i}, \tag{73}
$$

where we define $\Delta\mathbf{z}$ as the gradient normalized by the approximate Hessian as follows

$$
\begin{aligned}
\Delta\mathbf{z}_{j,i} = {} & \left(\nabla_{\mathbf{z}}\nabla_{\mathbf{z}}\mathscr{J}(\mathbf{z}_{j,i})\right)^{-1}\nabla_{\mathbf{z}}\mathscr{J}(\mathbf{z}_{j,i}) \\
= {} & \left(\mathbf{C}_{zz}^{-1} + \mathbf{G}_{j,i}^{\mathrm{T}}\mathbf{C}_{dd}^{-1}\mathbf{G}_{j,i}\right)^{-1}\mathbf{C}_{zz}^{-1}(\mathbf{z}_{j,i} - \mathbf{z}_j^{\mathrm{f}}) \\
& + \left(\mathbf{C}_{zz}^{-1} + \mathbf{G}_{j,i}^{\mathrm{T}}\mathbf{C}_{dd}^{-1}\mathbf{G}_{j,i}\right)^{-1}\mathbf{G}_{j,i}^{\mathrm{T}}\mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{z}_{j,i}) - \mathbf{d}_j\right),
\end{aligned}
\tag{74}
$$

and we define

$$
\mathbf{G}_{i,j}^{\mathrm{T}} = \nabla_{\mathbf{z}}\mathbf{g}(\mathbf{z}_{i,j}) \tag{75}
$$

as the gradient of the model, evaluated at iteration i and for ensemble member j. Equations 73 and 74 define the Ensemble Randomized Likelihood method [18, 20].

Since we are not computing the analytical gradient of the model we will need to approximate the ensemble of analytical gradients with an averaged gradient like $\mathbf{G}$ from Eq. 41, or we can evaluate the gradient at the ensemble average for the local iterate $\mathbf{G}_{\overline{\mathbf{z}}}$ as was explained by Evensen [14]. The same model gradient is now used for all realizations, and this leads to a different solution than the solution of the originally posed problem.

Equation 74 is exactly Eq. (2) in Chen and Oliver [6]. Also, Chen and Oliver [6] suggested using the state covariance in the Hessian evaluated at the local iterate to simplify further computations, since changing the Hessian does not change the gradient and thus the final converged solution (although it changes the step lengths in each iteration).

Thus, we can rewrite (74) with the averaged model gradient and introduce the state covariance for the local iterate in the Hessian, to obtain

$$\Delta \mathbf{z}_{j,i} = \left((\mathbf{C}_{zz}^i)^{-1} + \mathbf{G}_i^{\mathrm{T}} \mathbf{C}_{dd}^{-1} \mathbf{G}_i \right)^{-1} \mathbf{C}_{zz}^{-1} (\mathbf{z}_{j,i} - \mathbf{z}_j^{\mathrm{f}})$$
$$+ \left((\mathbf{C}_{zz}^i)^{-1} + \mathbf{G}_i^{\mathrm{T}} \mathbf{C}_{dd}^{-1} \mathbf{G}_i \right)^{-1} \mathbf{G}_i^{\mathrm{T}} \mathbf{C}_{dd}^{-1} (\mathbf{g}(\mathbf{z}_{j,i}) - \mathbf{d}_j). \tag{76}$$

Then using the corollaries

$$\left(\mathbf{C}^{-1} + \mathbf{G}^{\mathrm{T}} \mathbf{D}^{-1} \mathbf{G} \right)^{-1} = \mathbf{C} - \mathbf{C} \mathbf{G}^{\mathrm{T}} (\mathbf{G} \mathbf{C} \mathbf{G}^{\mathrm{T}} + \mathbf{D})^{-1} \mathbf{G} \mathbf{C}, \tag{77}$$

$$\left(\mathbf{G}^{\mathrm{T}} \mathbf{D}^{-1} \mathbf{G} + \mathbf{C}^{-1} \right)^{-1} \mathbf{G}^{\mathrm{T}} \mathbf{D}^{-1} = \mathbf{C} \mathbf{G}^{\mathrm{T}} (\mathbf{G} \mathbf{C} \mathbf{G}^{\mathrm{T}} + \mathbf{D})^{-1}, \tag{78}$$

which are derived from the Woodbury identity, we can write (76) as

$$\Delta \mathbf{z}_{j,i} = \mathbf{C}_{zz}^i \mathbf{C}_{zz}^{-1} (\mathbf{z}_{j,i} - \mathbf{z}_j^{\mathrm{f}})$$
$$- \mathbf{C}_{zz}^i \mathbf{G}_i^{\mathrm{T}} \left(\mathbf{G}_i \mathbf{C}_{zz}^i \mathbf{G}_i^{\mathrm{T}} + \mathbf{C}_{dd} \right)^{-1} \tag{79}$$
$$\times \left(\mathbf{G}_i \mathbf{C}_{zz}^i \mathbf{C}_{zz}^{-1} (\mathbf{z}_{j,i} - \mathbf{z}_j^{\mathrm{f}}) - (\mathbf{g}(\mathbf{z}_{j,i}) - \mathbf{d}_j) \right).$$

Now, from Eqs. 42 and 43, we can write (79) as

$$\Delta \mathbf{z}_{j,i} = \mathbf{C}_{zz}^i \mathbf{C}_{zz}^{-1} (\mathbf{z}_{j,i} - \mathbf{z}_j^{\mathrm{f}})$$
$$- \mathbf{C}_{zy}^i \left(\mathbf{C}_{yz}^i (\mathbf{C}_{zz}^i)^{-1} \mathbf{C}_{zy}^i + \mathbf{C}_{dd} \right)^{-1} \tag{80}$$
$$\times \left(\mathbf{C}_{yz}^i \mathbf{C}_{zz}^{-1} (\mathbf{z}_{j,i} - \mathbf{z}_j^{\mathrm{f}}) - (\mathbf{g}(\mathbf{z}_{j,i}) - \mathbf{d}_j) \right).$$

In the original algorithm, the expression $\mathbf{C}_{yz}^i (\mathbf{C}_{zz}^i)^{-1} \mathbf{C}_{zy}^i$ was replaced with the covariance $\mathbf{C}_{yy}^i$. But, as we have seen, this will break the consistency between (76) and (80) in the nonlinear case with $n < N - 1$, and we need to use the definition in Eq. 51 to replace this expression.

The numerical solution method for this equation is discussed in more detail by Chen and Oliver [6]. It is clear that it is the introduction of low-rank ensemble representations of the covariances that makes it possible to compute the update steps $\Delta \mathbf{z}_{j,i}$, and the computation requires the use of singular-value decompositions and pseudo inversions.

6 Examples

To verify the new algorithms, we will use the scalar example from Evensen [14]. The example resembles the use of conditioning methods in history matching, i.e., there is a parameter x that serves as an input to a forward model to predict $y = g(x, q)$. We assume an initial state x and a prediction y, given by the model

$$y = g(x, q)$$
$$= x(1 + \beta x^2) + q. \tag{81}$$

Here, β is a parameter that determines the nonlinearity of the model. In the current example, we have used $\beta = 0.0$

for the linear cases and $\beta = 0.2$ for the nonlinear cases. Clearly, in this case, the model error is additive to make the linear case completely linear. If we have a product of x and q, the problem becomes nonlinear, and we can not test if the different methods converge to the same solution in the entirely linear case where the convergence partly constitutes our proof of consistency.

The model error q is a random variable sampled from $\mathcal{N}(0, C_{qq})$ with $C_{qq} = 0.0$ in the case with no model errors and $C_{qq} = 0.25$ in the case including model errors.

In all the four cases, we sample the prior ensemble for x from a Gaussian distribution with mean $x^{\mathrm{f}} = 1$ and variance $C_{xx} = 1$ and we sample the perturbed observations of y from a Gaussian error distribution with mean $d = -1$ and variance $C_{dd} = 1$. Thus, in the current example, x represents the initial state or the model parameter, while y is the predicted observation. The goal is to estimate x given a measurement of y and then to recompute the correct prediction of y subject to model errors consistently with Bayes theorem.

In this example, we use a sufficiently large number of samples, i.e., 10^7, to generate accurate estimates of the probability density functions and this allows us to work directly with the pdfs and to examine the converged solutions of the methods.

6.1 Results from the linear case

In Figs. 1 and 2, we show the results from the linear cases without and with model errors. In Fig. 1, we plot the joint pdfs for the prior and the updated solutions, and in Fig. 2, we plot the corresponding marginal pdfs.

The joint pdfs illustrate the effect of including stochastic model errors. Without model errors, there is a unique mapping from x to y, and the pdf is zero except along the curve (or line in the linear case) defined by the model function $y = g(x)$. The prior joint pdf has a maximum value located at $(x, y) = (1, 1)$ while the posterior joint pdf has shifted the maximum value to $(x, y) = (0, 0)$ for all the methods. When we introduce the model errors, we notice that we obtain a stronger update in y and weaker update in x, than in the case without model errors. Still, we observe that all the smoother methods give a result that is identical to the Bayesian update. We can better visualize these results when we examine the marginal pdfs plotted in Fig. 2.

In the case without model errors, we see that the prediction pdf for y and the measurement pdf have the same variance and only differ in the value of the means. The measurement is at $y = -1$ while the mean prediction is located at $y = 1$. The update from ES, ESMDA, and IES, exactly matches the Bayesian update in this case and is centered between the measurement and prediction pdf as we would expect.
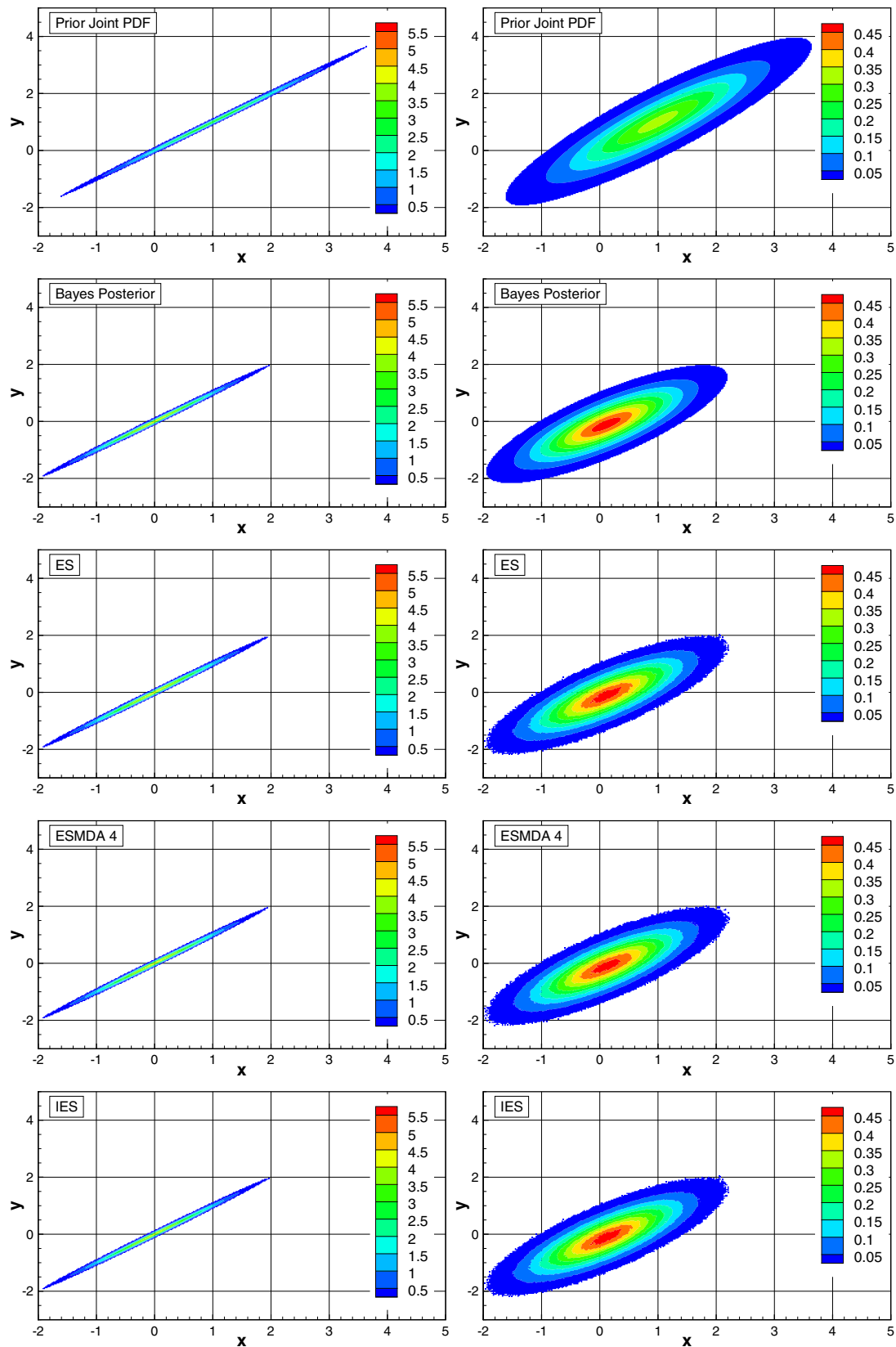
Fig. 1 The plots show joint pdfs for the linear case. In the left column, the model error is set to zero (a small model error is retained in the final prediction for plotting purposes), while in the right column, we include a model error with standard deviation equal to 0.5. The two upper rows are the analytical prior and posterior, while the three lower rows show the results from ES, ESMDA with four steps, and IES
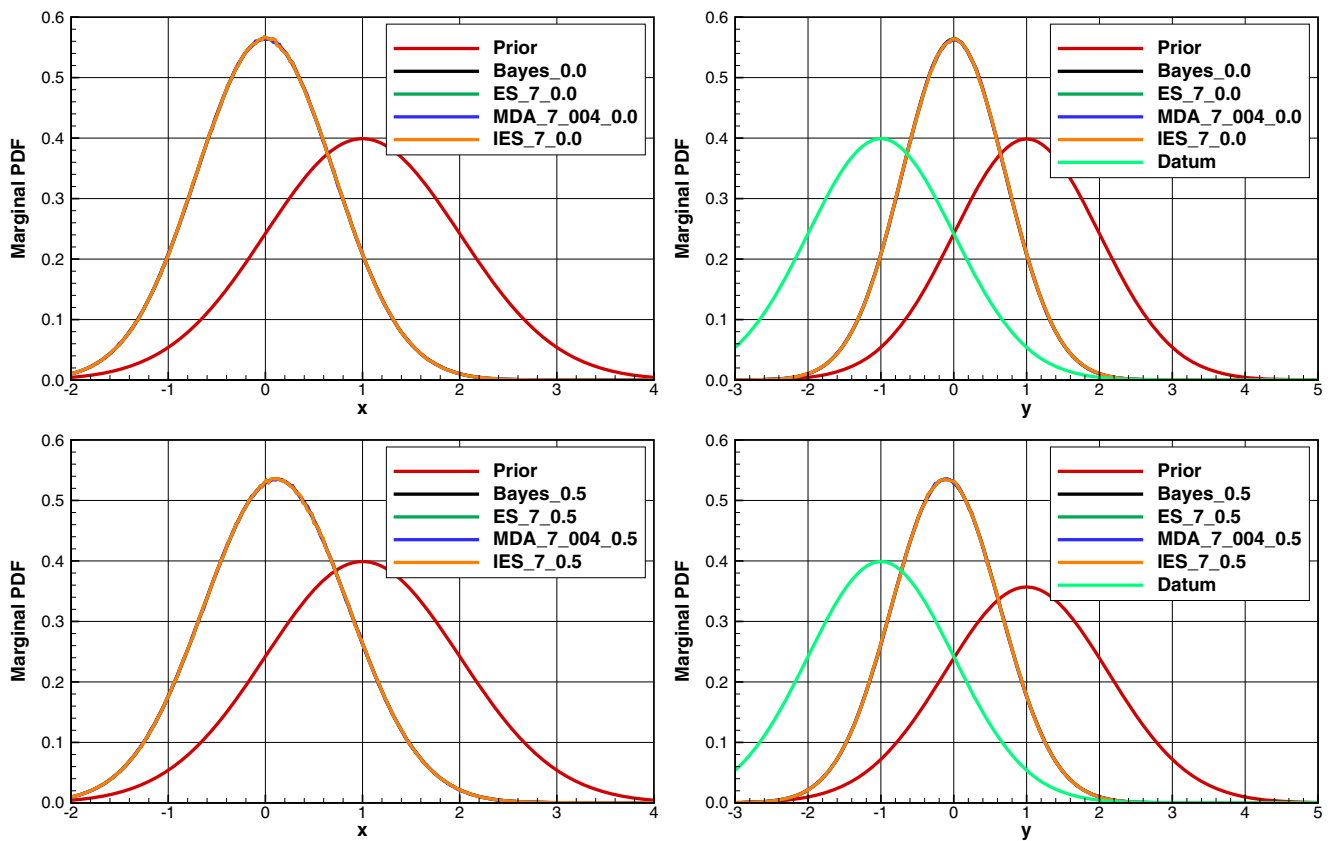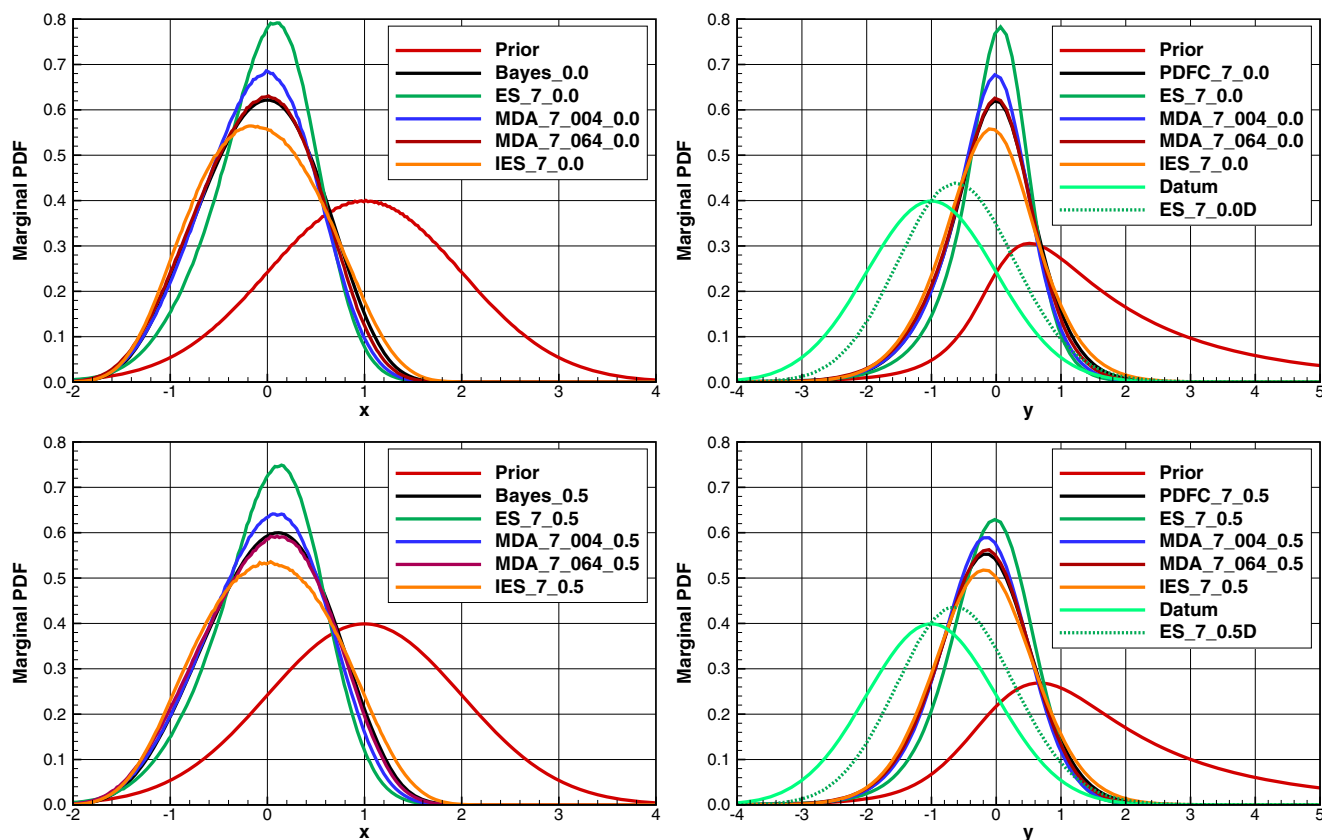
Fig. 2 The plots show the marginal pdfs for x in the left column and the marginal pdfs for y in the right column for the linear case corresponding to Fig. 1. The upper row is the case with zero model error, while the lower row shows the result when we include a model error with standard deviation equal to 0.5. In all the plots, the legends, e.g.,

MDA_7_004_0.5 denote ESMDA with 10^7 ensemble members, 4 MDA steps, and a model error with variance equal to 0.5. Note that, in the linear case, all the methods give results that are identical to the Bayesian posterior, and the posterior pdfs are indistinguishable

When we include model errors, the effect is that the prediction gets a higher variance, although the mean is the same (in this particular case). Due to the higher variance, we give more weight to the measurement in the update,
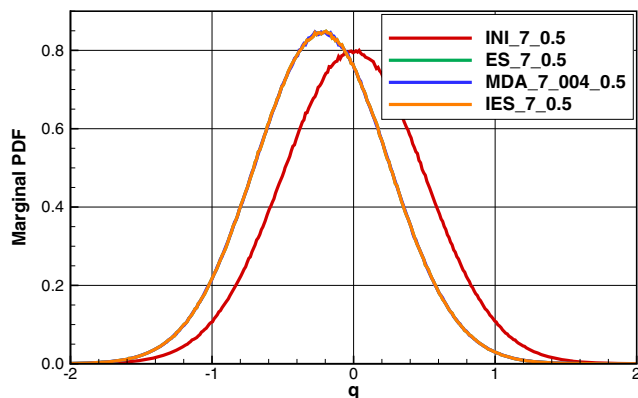
Fig. 3 Model error distributions for the linear case. It is not possible to distinguish between the pdfs for the estimated model errors from the different methods

and the update for *y* is stronger than in the case without model errors. On the other hand, the update for *x* is weaker in this case, since the addition of model errors reduces the correlation between the predicted measurement *y* and the prior *x*.

So, how can the update for *y* be shifted towards the observation in this case? After all, we compute *y* as a prediction from *x*. Here, the inclusion of the model errors in the inversion plays a vital role. We simultaneously update the ensemble for *x* and the estimate of the model errors *q*. In Fig. 3, we see how we shift the model errors towards negative values. Thus, when we integrate the model forward from the updated *x*, the forcing from *q* compensates for the weaker update of *x* and also the additional shift of *y* towards the measured value.

This example illustrates how model errors impact the updates of *x* and *y* as well as how we also need to include the model errors as a parameter in the estimation and then use it in the prediction to obtain the correct estimate of *y*. Finally, we also demonstrate that in the linear case with and without model errors, ES, IES, and ESMDA, all converge to the correct Bayesian solution.
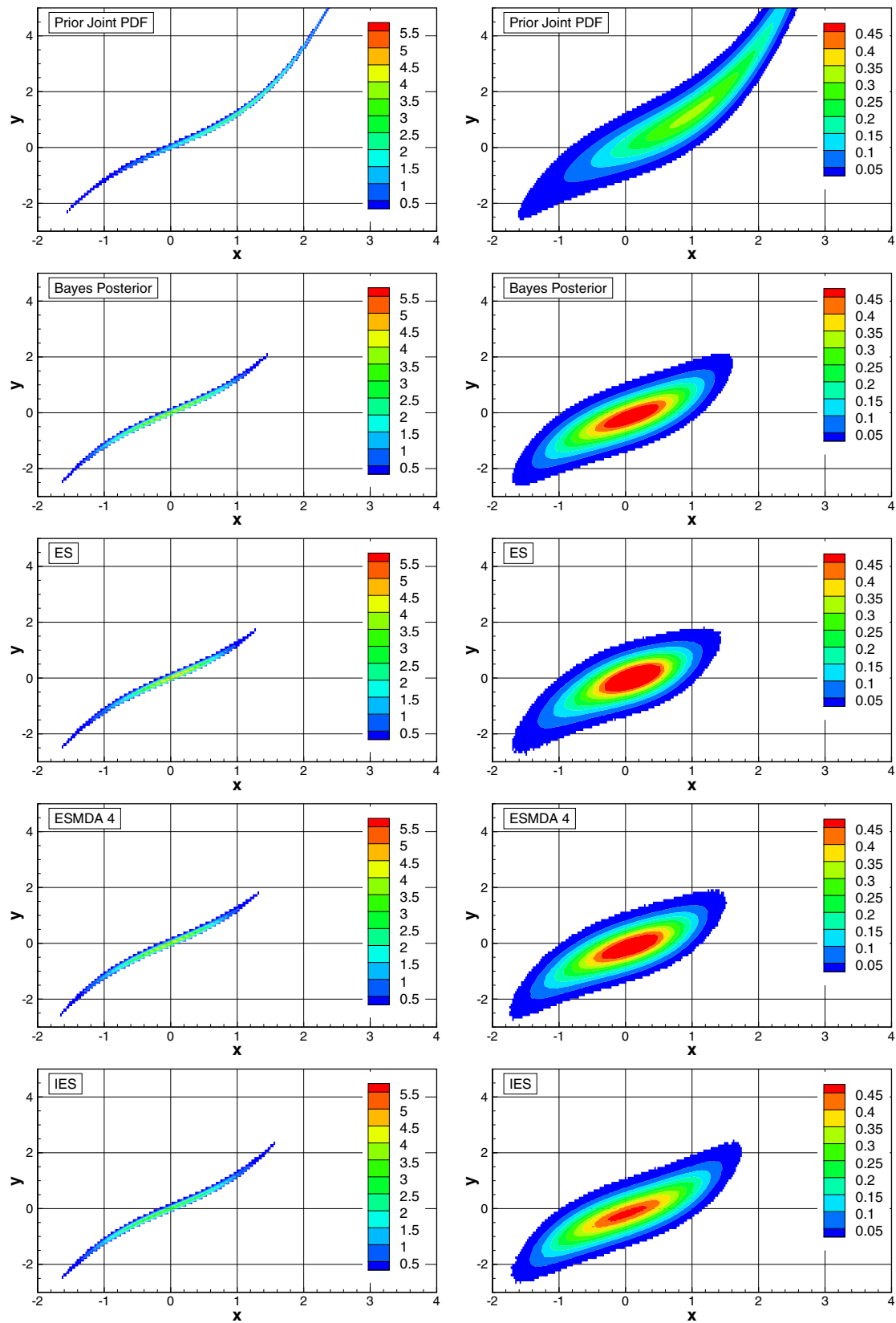
Fig. 4 The plots show joint pdfs for the nonlinear case. In the left column, the model error is set to zero, while in the right column, we include a model error with standard deviation equal to 0.5. The two upper rows are the analytical prior and posterior, while the three lower rows show results from ES, ESMDA with four steps, and IES

Fig. 5 The plots show the marginal pdfs for *x* in the left column and the marginal pdfs for *y* in the right column for the nonlinear case corresponding to Fig. 4. The upper row is the case with zero model error, while the lower row shows the result when we include a model error with standard deviation equal to 0.5. In all the plots, the legends,

e.g., MDA_7_004_0.5 denote ESMDA with 10^7 ensemble members, 4 MDA steps, and a model error with variance equal to 0.5. The legends ES_7_0.0D and ES_7_0.5D refers to an ES case where the prediction y_j^a is updated directly by Eq. 60

6.2 Results from the nonlinear case

In Figs. 4 and 5, we show the results from the nonlinear cases with and without model errors, where Fig. 4 plots the joint pdfs for the prior and the updated solutions, and in Fig. 5, we plot the corresponding marginal pdfs.

From the joint pdfs, we notice that the various smoother methods give different results both with and without the inclusion of model errors, although the general shape and locations of the pdfs are reasonably consistent with the theoretical solution as given by Bayes theorem.

We get a clearer picture from the marginal pdfs in Fig. 5. As for the linear case, we get a weaker update of *x* and a stronger update of *y*. We also notice that the introduction of model errors is handled well by the iterative methods, and the results are somewhat better and more consistent with the theoretical solution than in the case without model errors. ES is still the poorest estimator, and the iterative smoothers provide a significant improvement in the estimate, also in the case including model errors. We show the corresponding updates of the model error *q* in Fig. 6 and the different smoother methods all give slightly different results.

The dashed green line in the plots for *y* in Fig. 5 is the direct ES update of *y* using the predicted ensemble for *y* and the measurement. It is clear that the update of *x* followed by an integration of the model to obtain *y* gives a better result than a direct update of *y*. Furthermore, the additional use of iterations improves the estimate of *y* even further. This result is the motivation for introducing IEnKF in sequential data assimilation [22, 23] and also the iterative smoother
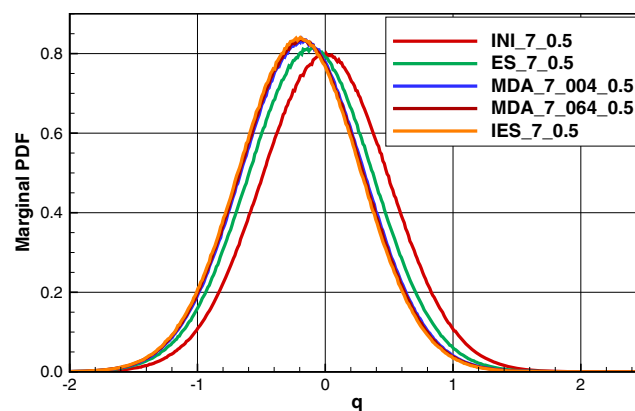

Fig. 6 Model error distributions for the nonlinear case

by Bocquet and Sakov [2, 3]. In history matching, we are primarily interested in estimating **x**, and the prediction **y** is just the result given the model parameters in **x**. But also in history matching, the ultimate value comes from accurate predictions of **y**.

The impact of using (51) for evaluating $\widetilde{\mathbf{C}}_{yy}$ in the update schemes is illustrated in Fig. 7, where we show results including and excluding the projection. The impact is most pronounced when using ES and ESMDA with few update steps where the use of $\overline{\mathbf{C}}_{yy}$ instead of $\widetilde{\mathbf{C}}_{yy}$ severely impacts the computation of the long linear update steps. In IES, we must include the projection to ensure that the gradients defined in Eqs. 76 and 80 are identical but in the current case the relative difference in the estimated mean when including or excluding the projection is only around one percent.

7 Including model errors in history matching

The need for including model errors in iterative ensemble smoothers became apparent while working with the paper [15], which considered the conditioning of reservoir models on production-rate data. Typically, in history matching, we assign errors to the rate data used in the conditioning step, while we neglect these errors when the same data are used to force the reservoir simulation model during the historical simulation.

We consider the errors in rate data to be a significant component of the total model error when we exclude errors in the estimated model parameters. Also, Evensen and Eikrem [15] pointed out that there are strong time correlations in the errors in rate data due to the rate allocation procedure used. When we include a stochastic model forcing using time-correlated errors, we will experience a significantly stronger impact than when the errors are white in time (see Chap. 12 in Evensen [12]).

The functional form $\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{q})$ can represent the prediction of the produced rates (that we observe) from
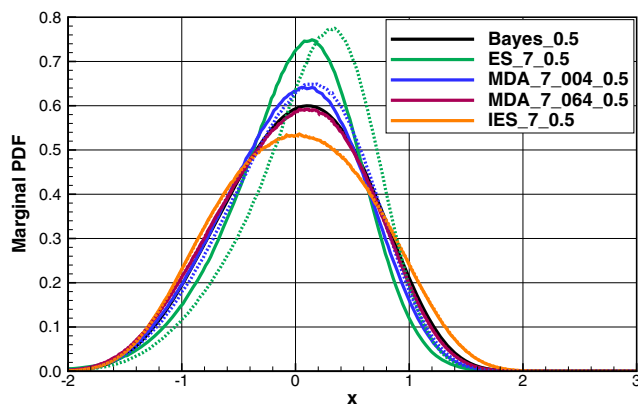

Fig. 7 The plot compares solutions including (solid lines) and excluding (dashed lines) the projection in Eq. 51

a reservoir simulation model. Note that using ensemble methods, we do not need to explicitly construct the functional form $\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{q})$ since we represent the gradients using ensemble covariances. We only need access to a numerical reservoir simulation model. The inputs **x** then contain all the uncertain parameters of the model, such as, e.g., porosity and permeability fields, various transmissibilities, and even parameters defining the reservoir structure. The model errors can be the errors in the rates used to force the model in the historical period. These errors are not additive since the model prediction depends nonlinearly on the specified rates.

If we associate the dominant model errors with the rates used to force the simulation model, then the size of the vector of model errors **q** is equal to the number of rate data used to force the model. A typical frequency of historic rate data for a well can be 12 data points per year, i.e., if we force the model using monthly reservoir-volume rates.

The prior error statistics for the rate data used to force the model should be the same as is used for the rate data in the update step. Thus, we can simulate a prior ensemble of time series with mean zero, a specified variogram in time, and a specified variance, to represent the model errors. These time series are then defining the vectors $\mathbf{q}_j$ of model errors for each realization. An ensemble model-error covariance $\overline{\mathbf{C}}_{qq}$ is then defined by the ensemble $\mathbf{q}_j$.

The conventional procedure of deriving the production rates from rate allocation tables, which we only update in connection with separator tests, often several years apart, means that the errors in the rate data will be nearly perfectly correlated in between each separator test. Note also that the inclusion of time correlations significantly reduces the degrees of freedom in the model errors and simplifies the estimation of the model errors in the conditioning step.

The expected impact of including model errors is first a larger and more realistic spread of the prior ensemble. Second, we will obtain a weaker and more correct update of the reservoir parameters in conditioning step. Furthermore, the posterior ensemble will give a more accurate and consistent prediction at the end of the history-matching period since the posterior realizations are forced by updated and improved estimated rates. Thus, we have an improved basis for making future predictions. We also expect that the information contained in the improved estimates of rates, in some cases may be used to correct for biases in the rates used in future predictions.

8 A note on the specification of model errors

The model written in the form (20) is quite general and can in principle represent any functional relation between a set of inputs **x** and **q** and a prediction **y**. The functional form

$\mathbf{g}(\mathbf{x}, \mathbf{q})$ can represent, e.g., a reservoir simulation model as is referred to in this paper, or it can represent the measured outputs of a weather prediction model run over a specific time interval.

The representation of model errors in $\mathbf{q}$ is also quite general and covers both additive errors and errors on which the model depend nonlinearly. The vector $\mathbf{q}$ can contain a time sequence of errors applied at different time steps, and the errors can represent perturbations to any poorly modeled process. Furthermore, $\mathbf{q}$ can be any vector of any length containing any uncertain error parameter. We can decompose $\mathbf{q}$ into additive parts and parts on which the model depends nonlinearly. We can model errors due to missing physics or errors in the discretization as an additive error term.

The specification of the prior model error statistics is often seen as the primary obstacle when accounting for model errors, but this is not an argument for entirely ignoring the model errors. The prior specification of the model error covariance is nothing different from specifying the error statistics of uncertain initial conditions, uncertain parameters, and uncertain measurements. We restrict ourselves to Gaussian priors, so we need to prescribe the mean and the covariance. If the covariance is isotropic for each error type or component and different components are uncorrelated, we only need to specify the variogram length for each component. The variogram length would then depend on the physical model and the particular type of error term. Note also that we never need to compute the full error covariance $\mathbf{C}_{qq}$ since only realizations of the model errors are required. A stepwise approach for introducing model errors with gradually increasing statistical complexity could be the following:

1. Assume unbiased and uncorrelated errors, thus prescribe only the variance when sampling the realizations of the prior error terms.
2. Specify correlations for different types of errors (or error components) independently.
3. Introduce possible correlations between different error components.

The posterior results from the inversion should be verified to check the consistency with the prescribed priors.

9 Summary

In this paper, we have given consistent formulations of iterative ensemble smoothers when we include model errors. We demonstrate the consistency by showing that the ensemble smoothers all converge to the Bayesian posterior in the linear case. The main result is that the model errors need to be treated as another set of unknown parameters and

estimated in the same way as the input parameters to the model. The proposed approach allows for the inclusion of general nonlinear errors that can be both red and white in time. Thus, we can apply the iterative methods for history matching of reservoir models forced with uncertain rate data having time-correlated errors, as well as for the sequential data assimilation problem with general model-error terms.

We demonstrate that the inclusion of model errors leads to a weaker update of the input parameters to the model, but a stronger update of the measured model prediction. Vice verse, the negligence of model errors that should be present, will lead to a too substantial update of the model input parameters with an associated underestimated uncertainty and also a too weak update of the prediction.

Thus, the results open for a more consistent solution of the history-matching problem, given that significant model errors are neglected in all previous history-matching applications with iterative ensemble smoothers.

We also briefly discuss an inconsistency of the linearizations in the analysis scheme that appear for nonlinear operators and when the state dimension is less than the ensemble size minus one, and we show that we must include an additional projection of the predicted measurements for consistency in the derivation of the final update equations.

References

1. Bennett, A.F.: Inverse Methods in Physical Oceanography. Cambridge University Press (1992)
2. Bocquet, M., Sakov, P.: Joint state and parameter estimation with an iterative ensemble Kalman smoother. Nonlinear. Processes. in. Geophysics., European Geosciences Union (EGU) **20**(5), 803–818 (2013)

3. Bocquet, M., Sakov, P.: An iterative ensemble Kalman smoother. Q.. J.. R.. Meteorol.. Soc.. **140**, 1521–1535 (2014)

4. Carrassi, A., Vannitsem, S.: Deterministic treatment of model error in geophysical data assimilation. In: Ancona, F., Cannarsa, P., Jones, C., Portaluri, A. (eds.) Mathematical Paradigms of Climate Science, Springer INdAM Series, vol. 15, pp. 175–213. Springer, Cham (2016)

5. Chen, Y., Oliver, D.S.: Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math. Geosci. **44**, 1–26 (2012)

6. 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)

7. Eknes, M., Evensen, G.: Parameter estimation solving a weak constraint variational formulation for an Ekman model. J. Geophys. Res. **102**(C6), 12.479–12.491 (1997)

8. Emerick, A.A., Reynolds, A.C.: History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. Computat. Geosci. **16**(3), 639–659 (2012)

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

10. Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. **99**(C5), 10.143–10.162 (1994)

11. Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean. Dyn. **53**, 343–367 (2003)

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

13. Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation. IEEE Control. Syst. Mag. **29**(3), 83–104 (2009b)

14. 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

15. 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

16. Evensen, G., van Leeuwen, P.J.: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasi-geostrophic model. Mon. Weather. Rev. **124**, 85–96 (1996)

17. Harlim, J.: Model error in data assimilation. In: Franzke, C. (ed.) Nonlinear and Stochastic Climate Dynamics. Cambridge University Press, Cambridge (2017). also available as arXiv:1311.3579

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

19. Neal, R.M.: Sampling from multimodal distributions using tempered transitions. Stat. Comput. **6**(4), 353–366 (1996)

20. Oliver, D.S., He, N., Reynolds, A.C.: Conditioning permeability fields to pressure data. In: ECMOR – 5th European Conference on the Mathematics of Oil Recovery (1996)

21. Reynolds, A.C., Zafari, M., Li, G.: Iterative forms of the Ensemble Klman Filter. In: ECMOR – 10th European Conference on the Mathematics of Oil Recovery (2006)

22. Sakov, P., Oliver, D.S., Bertino, L.: An iterative EnKF for strongly nonlinear systems. MonWeather. Rev. **140**, 1988–2004 (2012)

23. Sakov, P., Haussaire, J.M., Bocquet, M.: An iterative ensemble Kalman filter in the presence of additive model error. Q. J. R. Meteorol. Soc. https://doi.org/10.1002/qj.3213 (2018)

24. Stordal, A., Elsheikh, A.H.: Iterative ensemble smoothers in the annealed importance sampling framework. Adv. Water. Resour. **86**, 231–239 (2015)