



Marginalized iterative ensemble smoothers for data assimilation

Andreas S. Stordal^{1,2} · Rolf J. Lorentzen¹ · Kristian Fossum¹

Received: 2 November 2022 / Accepted: 18 July 2023
© The Author(s) 2023

Abstract

Data assimilation is an important tool in many geophysical applications. One of many key elements of data assimilation algorithms is the measurement error that determines the weighting of the data in the cost function to be minimized. Although the algorithms used for data assimilation treat the measurement uncertainty as known, it is in many cases estimated or set based on some expert opinion. Here we treat the measurement uncertainty as a hyperparameter in a fully Bayesian hierarchical model and derive a new class of iterative ensemble methods for data assimilation where the measurement uncertainty is integrated out. The proposed algorithms are compared with the standard iterative ensemble smoother on a 2D synthetic reservoir model.

Keywords Measurement uncertainty · Ensemble methods · Bayesian inversion · Data assimilation · History matching · Hierarchical models

1 Introduction

Data assimilation is the process of combining model predictions with real measurements for estimating unknown parameters and states of a model. In statistics, estimating the hidden (random) states of a model is commonly known as filtering or smoothing, whereas the parameter estimation problem is known as inversion or inverse problem in mathematical literature. The parameter estimation problem can be formulated either as a (regularized) maximum likelihood problem, or as a Bayesian inversion. In petroleum engineering, the inverse problem is commonly referred to as history matching [16]. There are multiple ways to address data assimilation, and here we focus on a particular approximate Bayesian approach known as the iterative ensemble methods [2, 18] since exact Bayesian methods such as e.g. MCMC methods are too expensive for large scale models. The iterative ensemble methods are well suited for large scale data assimilation since they treat the model as a black box and do not require the model to be differentiated.

In addition to a numerical model and the data itself, Bayesian data assimilation algorithms require prior informa-

tion in terms of the prior distribution for both the parameters and states of the model as well as the likelihood function for the data, i.e. the measurement bias (if any) and uncertainty. For most standard Bayesian methods the prior distribution, as well as the likelihood, is assumed to be known. In many situations these parameters, often denoted hyperparameters, are unknown and can be modelled using hierarchical Bayes [8]. In principle there are several ways to handle hyperparameters with the largest class being the empirical Bayesian methods [3] where the hyperparameters are estimated from the data either by some point estimates or by some iterative procedure that alternates between computing the posterior of the parameters and states given the hyperparameters and computing the posterior of the hyperparameters given the states and parameters. The first approach typically ignores some of the uncertainty in the posterior while the latter requires multiple applications of the data assimilation method. The fully Bayesian hierarchical analysis, however, is to integrate out all the hyperparameters to obtain the full posterior. This approach often involves approximation of integrals via nested Laplace transformation (see e.g. [22]), but in some cases the integral can be solved analytically. The focus here is on the measurement uncertainty and since the likelihood is frequently assumed to be Gaussian in data assimilation methods, we can choose prior densities such that the likelihood can be analytically integrated over the measurement error variances.

The error of a measurement typically stems from model error or representation error and not necessarily from the measurement instrument alone. There are many sources of

✉ Andreas S. Stordal
asto@norceresearch.no

¹ NORCE, Norwegian Research Center, Nygårdsporten 112, 5008 Bergen, Norway

² Department of Mathematics, University of Bergen, Postboks 7803, N-5020 Bergen, Norway

measurement error, but regardless of how they are derived, (almost) all methods used for data assimilation have one thing in common: they all assume that the true uncertainty is specified. Naturally this is a strong assumption in many real world applications. Several techniques to estimate unknown errors online are available, such as adaptive filtering [14, 15]. However, the assumptions required are usually not satisfied in real applications. Robust approaches such as H_∞ filters [23] or Levy filters [24] serve as alternative methods, but the application in data assimilation setting is very limited to the best of our knowledge.

In geophysical models, it is quite often difficult to estimate the measurement error due to simplified physics, simplified parameterization or lack of knowledge of the measurement process [21]. Most approaches within atmospheric science are based on the work of [5, 6, 10], and [7], using the difference between observation and error forecasts (see e.g. [4]). Other approaches such as e.g. [1] use multiple data sets in combination with theory from [9]. A Bayesian approach is presented in [26] where the measurement uncertainty is estimated by maximizing an approximate log-posterior at each time step. [12] estimate the covariance of a water balance model in a weak constrained ENKF, where a variational Bayesian approach is used in a second update step of the algorithm, and an online approach for estimating colored observation noise has recently been introduced in [19].

For petroleum application there is not a lot of discussion of the measurement error of production data. It can in principle be obtained from the measuring instrument, but this would require a perfect model. For time-lapse seismic observations the noise can be estimated from the difference in data between different surveys in domains where the seismic signal is expected to be unchanged. [13] used techniques from image de-noising such as local and non-local moving average and sparse wavelet representation of the observations. When combining production data and seismic data, specifying the uncertainty becomes even more troublesome as there are two or more data terms in the likelihood function. Furthermore, due to imperfect models and pre-processing of certain data, the observation error is more than just noise related to the survey equipment (see [17] and references therein).

Regardless of how the measurement error is estimated or specified, and whether or not bias correction is included, the uncertainty in the estimate should be taken into account when applying Bayes' rule. This is commonly neglected in the data assimilation literature. From a Bayesian point of view, it is possible to incorporate the measurement uncertainty as a hyperparameter, and as far as the data assimilation algorithm is concerned, the measurement error variances can be considered like any of the physical model parameter. However, despite the interdependence (in the sense that the distribution of one is conditioned on the other), there is no obvious

hierarchical relationship to the physical parameters, unlike, e.g., hyperparameters of the prior. Instead, they could, or perhaps should, be considered 'nuisance parameters', i.e. parameters that are of no inherent interest but must be taken into account with their uncertainties in order to determine the physical model parameters of interest. Importantly, it is not necessary to reconstruct the probability distribution of the nuisance parameters, as long as their impact on the physical model parameters is accounted for. Here, an iterative ensemble smoother (iES) is proposed, where the likelihood function is marginalized over the measurement error. That is compliant with the fully Bayesian approach discussed above where the hyperparameters are integrated out of the hierarchical model. The resulting algorithm, denoted marginalized iterative ensemble smoother (miES), requires only a few modifications of iES algorithm. The theory is presented for a single measurement variance, but the more general case with several data types is briefly discussed and also presented in an example.

In the next section we define the data assimilation and inverse problem and give a short summary of the subspace version of the iES. The marginalized version is introduced in Section 3 and in Section 4 we provide examples and comparison with iES. The paper is concluded with a summary and discussion of further work.

2 Bayesian framework

Our quantity of interest, X , is considered as an unknown random variable with a prior probability density, $f_X(x)$. The observation, Y , is linked to X via the likelihood function $f_{Y|X}(y|x)$. The conceptual solution of the inverse problem is given by the posterior density

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{\int f_{Y|X}(y|x)f_X(x) dx}.$$

We have omitted any potential time dependence for simplicity.

2.1 Data assimilation

In most data assimilation algorithms, and ensemble based methods in particular, it is common to assume that both the prior (after a possible transformation of variables) and the likelihood function are Gaussian and that the observations are given as a nonlinear mapping of X with additive noise

$$Y = \mathcal{H}(X) + \epsilon.$$

The posterior is then known up to the normalizing constant

$$-\log(f_{X|Y}(x|y)) = \frac{1}{2} \|x - \mu\|_{\mathbf{P}}^2 + \frac{1}{2} \|y - \mathcal{H}(x)\|_{\mathbf{\Sigma}}^2 + C, \tag{1}$$

where μ is the prior mean, \mathbf{P} is the prior covariance matrix, $\mathbf{\Sigma}$ is the observation error covariance matrix and C is the logarithm of the normalizing constant. The iterative ensemble methods discussed in later sections seek to minimize the objective function Eq. 1 or a stochastic version of it. When $\mathbf{\Sigma}$ is uncertain, this will be reflected in the posterior and hence the objective function to be minimized.

2.2 Marginalizing the likelihood

In the following, we assume that the measurement error covariance matrix is known up to a constant, although the more general situation with block matrices with unknown variances is also presented. The measurement error is assumed to be Gaussian with a covariance matrix $\mathbf{\Sigma} = \sigma^2 \mathbf{R}$, where \mathbf{R} is a known $M \times M$ matrix. The likelihood function, given σ , is

$$f_{Y|X}(y|x, \sigma^2) = (2\pi\sigma^2)^{-M/2} |\mathbf{R}|^{-1/2} e^{-\frac{1}{2\sigma^2} \|y - \mathcal{H}(x)\|_{\mathbf{R}}^2}. \tag{2}$$

In a Bayesian framework, it is natural to view σ^2 as a hyper parameter with a prior density $f_{\sigma^2}(\sigma^2)$ and then marginalize the likelihood w.r.t. σ^2 [25]. If no information is available, a non-informative prior on σ^2 defined by $f_{\sigma^2}(\sigma^2) = \sigma^{-2}$ can be used. This is known as Jeffrey’s prior and represents the state of no prior information on the scale parameter σ [11]. With this choice we can analytically compute $\int f_{Y|X}(y|x, \sigma^2) p(\sigma^2) d\sigma^2$ to get the marginalized likelihood (see e.g. [25])

$$f_{Y|X}(y|x) \propto \left(\|y - \mathcal{H}(x)\|_{\mathbf{R}}^2 \right)^{-M/2}, \tag{3}$$

with corresponding log-likelihood given by

$$\ell(y|x) = -\frac{M}{2} \log \left(\|y - \mathcal{H}(x)\|_{\mathbf{R}}^2 \right). \tag{4}$$

More often than not, the measurement error is specified either from an expert opinion or from some type of estimation. In either case it is uncertain, and this uncertainty should be accounted for.

Both the estimate of measurement error and its uncertainty can easily be incorporated into a prior distribution. To this end, a frequently used prior for the variance in Bayesian statistics is the scale inverse χ^2 distribution defined by

$$f_{\sigma^2}(\sigma^2) \propto (\sigma^2)^{-v/2-1} e^{-vs^2/(2\sigma^2)}, \tag{5}$$

where s^2 is the estimate of the measurement error variance σ^2 and v is the degrees of freedom. In this case the marginal likelihood is proportional to

$$f_{Y|X}(y|x) \propto (\|y - \mathcal{H}(x)\|_{\mathbf{R}}^2 / (vs^2) + 1)^{-(M+v)/2}, \tag{6}$$

which is a multivariate t distribution with $t = \|y - \mathcal{H}(x)\|_{\mathbf{R}}^2 / s^2$ and v degrees of freedom. The log-likelihood is then

$$\ell(y|x) = -\frac{M+v}{2} \log(\|y - \mathcal{H}(x)\|_{\mathbf{R}}^2 / (vs^2) + 1). \tag{7}$$

The two marginalized likelihoods Eqs. 3 and 6 can easily be extended to K data types of dimension M_k with different variance, the log likelihood is then

$$\ell(y_1, y_2, \dots, y_K|x) = -\frac{1}{2} \sum_{k=1}^K M_k \log \left(\|y - \mathcal{H}(x)\|_{\mathbf{R}_k}^2 \right), \tag{8}$$

for the non-informative prior and

$$\ell(y_1, y_2, \dots, y_K|x) = -\frac{1}{2} \sum_{k=1}^K (M_k + v_k) \log \left(\|y - \mathcal{H}(x)\|_{\mathbf{R}_k}^2 / (v_k s_k^2) + 1 \right), \tag{9}$$

for the scale inverse χ^2 prior. The marginalized likelihoods can be used in Bayes’ formulation for filtering, data assimilation and Bayesian inversion. Here we focus on large scale inversion, but the algorithms can also be directly implemented for sequential data assimilation as they are presented. In the following, we develop a subspace square root iES algorithm [2]. For the ease of notation we focus on the single data type case, but some examples presented later contains multiple data types.

2.3 The new objective functions

Since solving the fully Bayesian inversion problem is too computationally costly for large scale models, we seek an approximate solution via maximization of the posterior (MAP) with the associate uncertainty provided by the inverse of the approximate Hessian matrix [2]. The cost function to be minimized in the miES is defined as the negative log posterior density. Hence, for Jeffrey’s prior we have

$$J(x) = \frac{1}{2} \|x - \mu\|_{\mathbf{P}}^2 + \frac{1}{2} M \log \left(\|y - \mathcal{H}(x)\|_{\mathbf{R}}^2 \right). \tag{10}$$

The gradient and approximate Hessian (using a first order model expansion) are given by

$$\nabla J(x) = \mathbf{P}^{-1}(x - \mu) - \frac{M}{\chi_y} \mathbf{H}_x^\top \mathbf{R}^{-1}(y - \mathcal{H}(x)), \quad (11)$$

$$\mathbf{C}_x = \mathbf{P}^{-1} + \frac{M}{\chi_y} \mathbf{H}_x^\top \mathbf{R}^{-1} \mathbf{H}_x, \quad (12)$$

where $\chi_y = \|y - \mathcal{H}(x)\|_{\mathbf{R}}^2$ and \mathbf{H}_x is the tangent linear operator of \mathcal{H} evaluated at x .

For the case with scale inverse χ^2 prior, the cost function is

$$J(x) = \frac{1}{2} \|x - \mu\|_{\mathbf{P}}^2 + \frac{1}{2}(M + \nu) \log \left(\|y - \mathcal{H}(x)\|_{\mathbf{R}}^2 / (\nu s^2) + 1 \right), \quad (13)$$

with corresponding gradient and approximate Hessian

$$\nabla J(x) = \mathbf{P}^{-1}(x - \mu) - \frac{M + \nu}{\chi_y + \nu s^2} \mathbf{H}_x^\top \mathbf{R}^{-1}(y - \mathcal{H}(x)), \quad (14)$$

$$\mathbf{C}_x = \mathbf{P}^{-1} + \frac{M + \nu}{\chi_y + \nu s^2} \mathbf{H}_x^\top \mathbf{R}^{-1} \mathbf{H}_x, \quad (15)$$

where $\chi_y = \|y - \mathcal{H}(x)\|_{\mathbf{R}}^2$ and \mathbf{H}_x is the tangent linear operator of \mathcal{H} evaluated at x . A quasi Gauss-Newton algorithm for minimizing the cost functions is given by

$$x_{k+1} = x_k - \mathbf{C}_x^{-1} \nabla J(x_k), \quad (16)$$

where k is the iteration index. If there are several data types, the likelihood term of the gradients in Eqs. 11 and 14 are replaced by a sum of gradient terms derived from Eqs. 8 and 9. In the next section an ensemble subspace approximation of the quasi Gauss-Newton scheme is presented.

2.4 Ensemble subspace formulation

In most applications of ensemble methods, the dimension of the state, M , is greater than the number of ensembles N . In this case we can use the simplified subspace ensemble version of iES with the new cost functions which avoids computing the pseudo inverse of matrices of dimension $M \times N$ [2, 18]. Let \mathbf{X} be the matrix of prior ensemble anomalies (samples with the mean subtracted from each column). The variable x can then be expressed in the ensemble subspace as

$$x(\omega) = \bar{x} + \mathbf{X}\omega, \quad (17)$$

where \mathbf{X} is the initial ensemble anomalies, \bar{x} is the ensemble mean and ω is the $N \times 1$ weight vector initially set to 0.

Inserting the change of variables into the Jeffery's prior gives the following objective function

$$J_n(\omega) = \frac{1}{2} \|\omega\|_{\frac{1}{N-1} \mathbf{I}_N}^2 + \frac{1}{2} M \log \|y - \mathcal{H}(\bar{x} + \mathbf{X}\omega)\|_{\mathbf{R}}^2, \quad (18)$$

where N is the ensemble size and \mathbf{I}_N is the identity matrix of size N . The gradient and first order Hessian approximation is

$$\nabla J(\omega) = (N - 1)\omega - \frac{M}{\chi_y} \mathbf{X}^\top \mathbf{H}_x^\top \mathbf{R}^{-1}(y - \mathcal{H}(\bar{x} + \mathbf{X}\omega)), \quad (19)$$

$$\mathbf{C}_\omega = (N - 1)\mathbf{I}_N + \frac{M}{\chi_y} \mathbf{X}^\top \mathbf{H}_x^\top \mathbf{R}^{-1} \mathbf{H}_x \mathbf{X}. \quad (20)$$

Let \mathbf{E} be the matrix with ensemble member i as its i 'th column and let \bar{x} denote the column mean. The ensemble anomaly matrix \mathbf{X} is defined by $\mathbf{X} = \mathbf{E} - \bar{x}\mathbf{1}^\top$, where $\mathbf{1}$ is the N -dimensional column vector of ones. The ensemble can be written $\mathbf{E} = \bar{x}\mathbf{1}^\top + \mathbf{X}\mathbf{W}$, where \mathbf{W} is initially set to the identity matrix of dimension $N \times N$. For each ensemble member, i.e. each column of \mathbf{E} , the forward model is run and the output is collected in the matrix denoted $\mathcal{H}(\mathbf{E})$. The linearization we use to linearize \mathcal{H} using the ensemble is given by $\mathbf{Y} = \mathcal{H}(\mathbf{E})\mathbf{W}^{-1}\mathbf{\Pi}_1^\top$, where $\mathbf{\Pi}_1^\perp = \mathbf{I}_N - \mathbf{1}\mathbf{1}^\top/N$, \mathbf{I}_N is the N -dimensional identity matrix. This implementation avoids using the chain rule by direct regression of the composite function $\mathcal{H}(x(\omega))$. For more details on the linearization we refer to [18]. We approximate Eq. 19 using the ensemble with Jeffrey's prior as

$$\mathbf{Y} = \mathcal{H}(\mathbf{E})\mathbf{W}^{-1}\mathbf{\Pi}_1^\perp \quad (21)$$

$$\chi_y = \left\| y - \overline{\mathcal{H}(\mathbf{E})} \right\|_{\mathbf{R}}^2 \quad (22)$$

$$\nabla_\omega J^{\text{lkhd}} = -\mathbf{Y}^\top \frac{M}{\chi_y} \mathbf{R}^{-1} [y - \overline{\mathcal{H}(\mathbf{E})}] \quad (23)$$

$$\nabla_\omega J^{\text{prior}} = (N - 1)\omega \quad (24)$$

$$\mathbf{C}_\omega = \frac{M}{\chi_y} \mathbf{Y}^\top \mathbf{R}^{-1} \mathbf{Y} + (N - 1)\mathbf{I}_N \quad (25)$$

Finally, the ensemble subspace formulation of Eq. 16, at iteration k , is then given by

$$\omega_{k+1} = \omega_k - \gamma_k \mathbf{C}_\omega^{-1} [\nabla_{\omega_k} J^{\text{prior}} + \nabla_{\omega_k} J^{\text{lkhd}}] \quad (26)$$

$$x_{k+1} = \bar{x} + \mathbf{X}\omega_{k+1} \quad (27)$$

$$\mathbf{W}_{k+1} = (\mathbf{C}_\omega / (N - 1))^{-1/2} \quad (28)$$

$$\mathbf{E}_{k+1} = x_{k+1}\mathbf{1}^\top + \mathbf{X}\mathbf{W}_{k+1} \quad (29)$$

where γ_k is the step size. Often step size is not implemented in Gauss Newton, however, since we are using so many

approximations in our scheme, we have seen that there can be advantages of not taking a full step ($\gamma < 1$) in some cases.

For the inverse χ^2 prior the objective function (with the change of variables) is given by

$$J(\omega) = \frac{1}{2} \|\omega\|_{\frac{1}{N-1}\mathbf{I}_N}^2 + \frac{1}{2}(M + \nu) \log(\|y - \mathcal{H}(\bar{x} + \mathbf{X}\omega)\|_{\mathbf{R}}^2 / (\nu s^2) + 1). \tag{30}$$

$$\nabla J(\omega) = (N - 1)\omega - \frac{M + \nu}{\chi_y + \nu s^2} \mathbf{X}^\top \mathbf{H}_x^\top \mathbf{R}^{-1} (y - \mathcal{H}(x)), \tag{31}$$

$$\mathbf{C}_\omega = (N - 1)\mathbf{I}_N + \frac{M + \nu}{\chi_y + \nu s^2} \mathbf{X}^\top \mathbf{H}_x^\top \mathbf{R}^{-1} \mathbf{H}_x \mathbf{X}. \tag{32}$$

The following defines our subspace formulation marginalized iterative Ensemble smoother with an inverse χ^2 prior.

$$\mathbf{Y} = \mathcal{H}(\mathbf{E})\mathbf{W}^{-1}\mathbf{\Pi}_1^\perp \tag{33}$$

$$\chi_y = \left\| y - \overline{\mathcal{H}(\mathbf{E})} \right\|_{\mathbf{R}}^2 \tag{34}$$

$$\nabla_\omega J^{\text{lkhd}} = -\mathbf{Y}^\top \frac{M + \nu}{\chi_y + \nu s^2} \mathbf{R}^{-1} [y - \overline{\mathcal{H}(\mathbf{E})}] \tag{35}$$

$$\nabla_\omega J^{\text{prior}} = (N - 1)\omega \tag{36}$$

$$\mathbf{C}_\omega = \frac{M + \nu}{S + \nu s^2} \mathbf{Y}^\top \mathbf{R}^{-1} \mathbf{Y} + (N - 1)\mathbf{I}_N \tag{37}$$

$$\omega_{k+1} = \omega_k - \gamma_k \mathbf{C}_{\omega_k}^{-1} [\nabla_{\omega_k} J^{\text{prior}} + \nabla_{\omega_k} J^{\text{lkhd}}] \tag{38}$$

$$x_{k+1} = \bar{x} + \mathbf{X}\omega_{k+1} \tag{39}$$

$$\mathbf{W}_{k+1} = (\mathbf{C}_\omega / (N - 1))^{-1/2} \tag{40}$$

$$\mathbf{E}_{k+1} = x_{k+1} \mathbf{1}^\top + \mathbf{X}\mathbf{W}_{k+1}. \tag{41}$$

Remark 1 The gradient and Hessian approximations involves terms of the form $(y - \mathcal{H}(\bar{x} + \mathbf{X}\omega))$. Since we run the ensemble, and not the ensemble mean, this quantity is not available

to us. There are two ways to deal with this issue. First, it is possible to run $N - 1$ ensemble members and the mean to obtain the desired quantity without increasing the computational cost, alternatively we can approximate $\mathcal{H}(\bar{x} + \mathbf{X}\omega)$ with $\overline{\mathcal{H}(\mathbf{E})}$, as is done in [2]. The latter approach is implemented here.

In the next section the two algorithms are tested on a synthetic reservoir model and compared with a square root implementation of the iES.

3 Reservoir example

The test case is a 2D reservoir of 60×60 grid cells with a two phase flow consisting of oil and water operated with six producing wells and one injection well. There are three different data types: oil production rates, water production rates and water injection rates, so the log likelihood for the miES is a sum of three terms ($K = 3$ in Eqs. 8 and 9). We assume that the porosity is known (fixed at 0.2) and that the permeability field is unknown and modeled using a Gaussian prior distribution with a constant mean of 5, a constant variance of 1, and with a spherical variogram of range 30, anisotropy of 0.33, and with an angle of -45 degrees. The true permeability field is a realization from the prior shown in Fig. 1 along with the prior mean.

The historical production data are obtained by running the The Open Porous Media Flow simulator [20] on bottom hole pressure control. The measurements are generated by adding a zero mean Gaussian measurement error to the oil production rates and water production rates in the six producers as well as the water injection rates in a single injector every 150 days for a total of 12000 days (80 observation time points) except for producer 4 which opens after 6000 days and producers 5 and 6 which open after 9000 days. The standard deviation of the measurement error is set to 10, 20 and

Fig. 1 True permeability field and prior mean

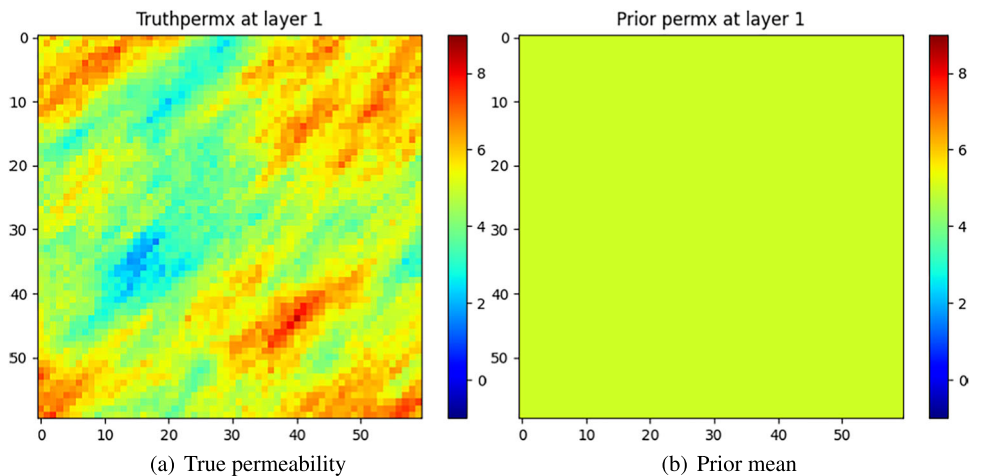
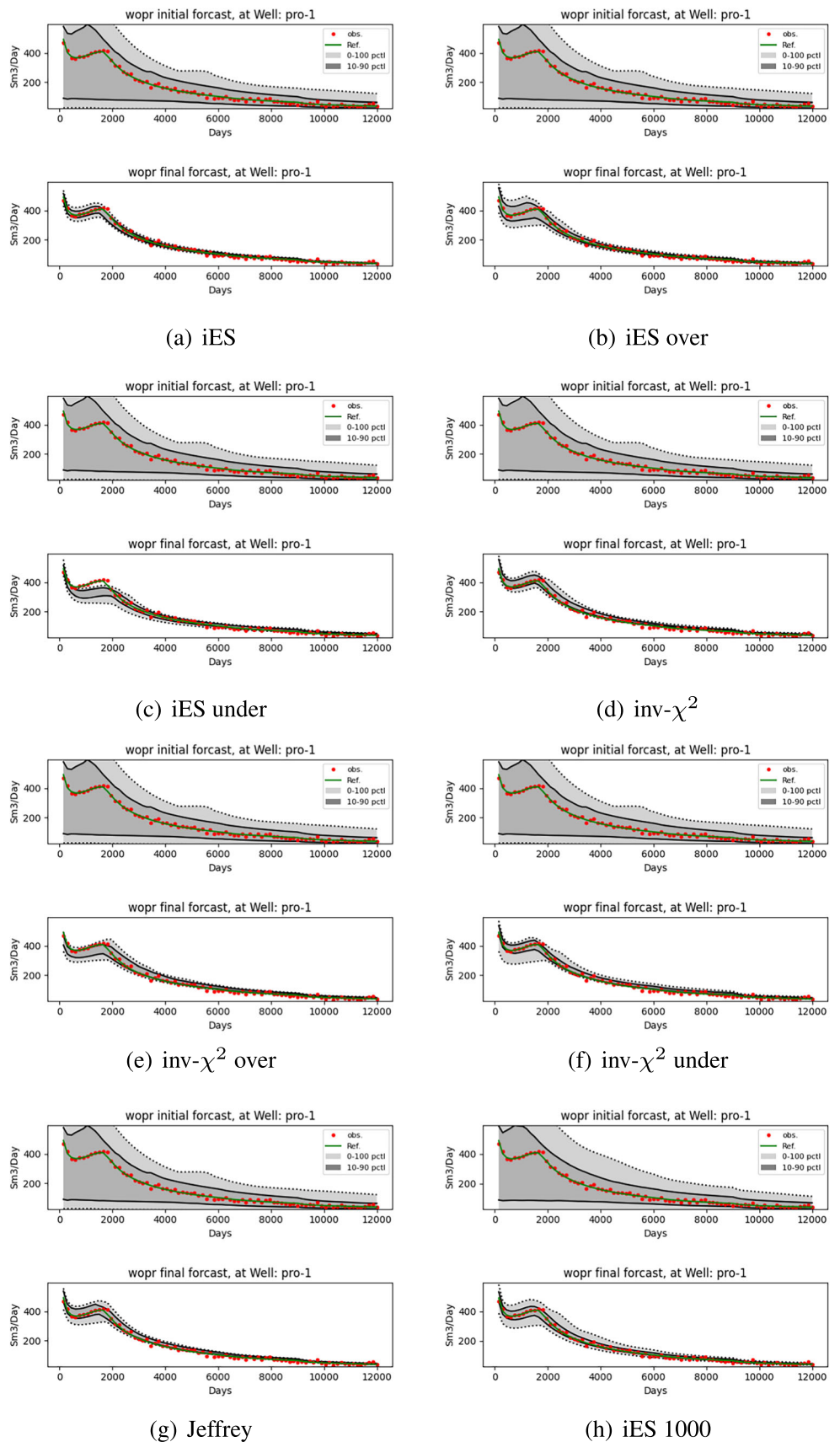


Fig. 2 Oil production rates, prior and posterior, for producer 1



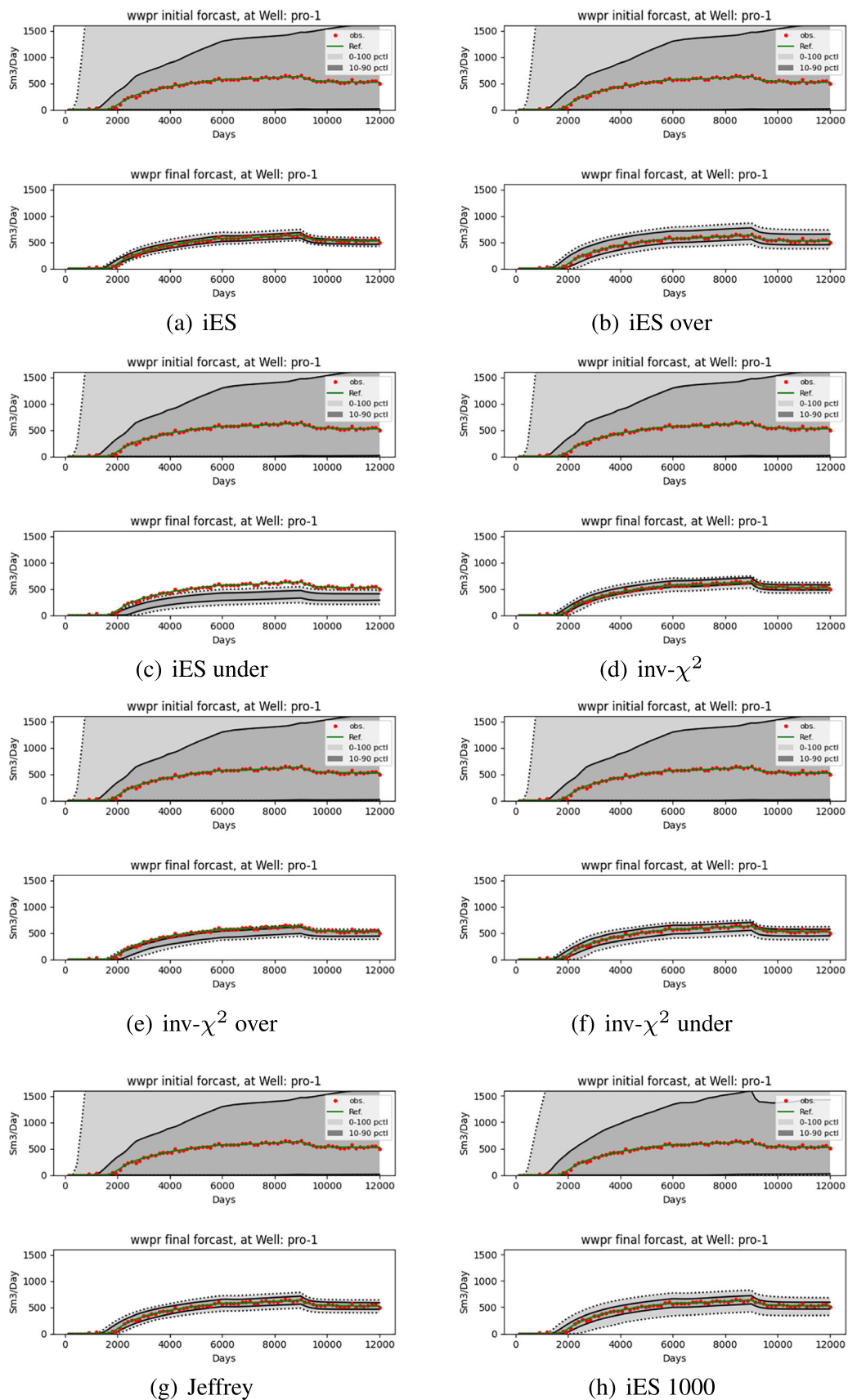


Fig. 3 Water production rates, prior and posterior, for producer 1

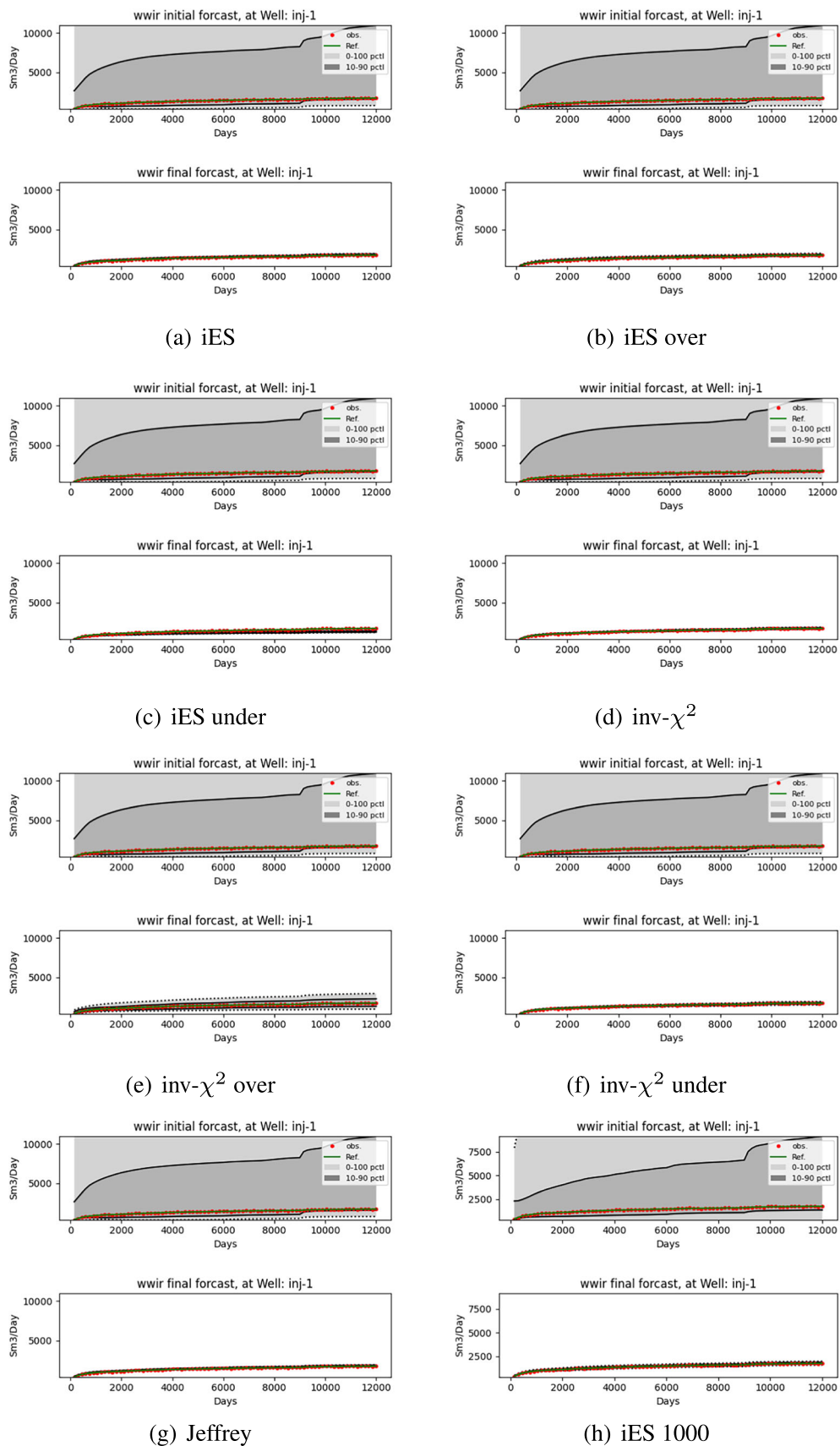


Fig. 4 Water injection rates, prior and posterior, for injector 1

Fig. 5 Posterior permeability fields

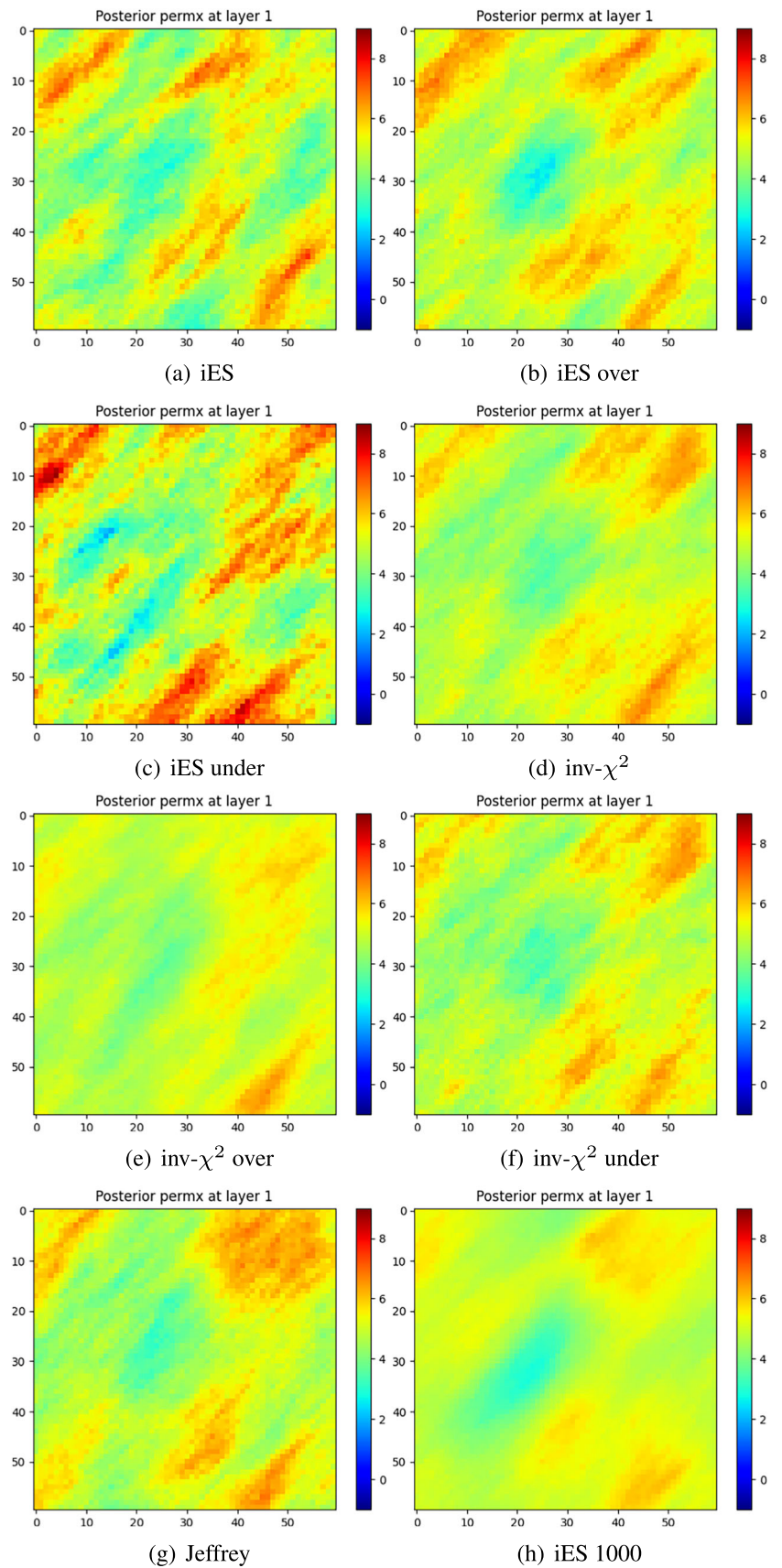
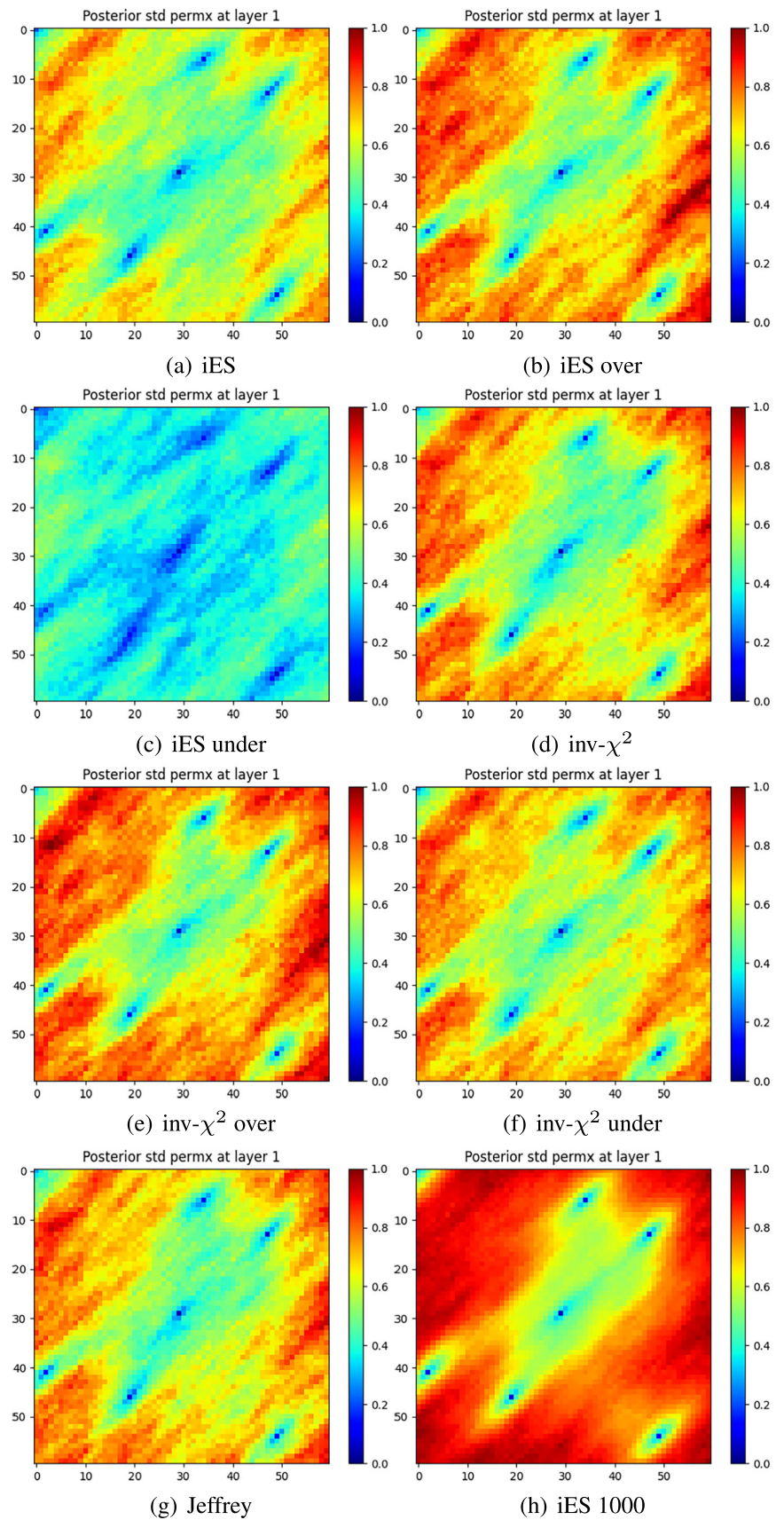


Fig. 6 Posterior standard deviations



30, for oil production, water production and water injection respectively.

We run seven different cases with three different ensemble methods. The iES and miES with inverse χ^2 prior are run with three different inputs (the standard deviation for iES and s for the miES). Once with the correct values for the standard deviations (10,20,30), once with a lower value ($\sqrt{40}$) for all three data types and once with a higher value ($\sqrt{1000}$) for all three data types. In all runs the degrees of freedom in miES, ν , is set equal to the number of measurements (80). Since the miES with Jeffrey's prior does not require any input it is only run once. The ensemble size is 100. In addition, we run the iES with the correct standard deviations for the measurement error and an ensemble size of 1000 that will serve as a reference solution. That is, all results should be compared with the result from this run as it is the best solution we can get for an iES algorithm.

Normally some type of localization should be applied when using ensemble methods, but for the sake of comparison of the impact of the different methods and inputs, we did not apply any localization in this example. In Figs. 2,3,4 the prior and posterior oil and water production is shown for the first producer as well as the water injection for the injector. All the methods provide a fairly good match with the data, with the exception of the iES when the measurement error standard deviations are set to a lower value than the true standard deviations.

Figure 5 shows the mean of the posterior permeability fields, where we particularly notice the robustness of the miES in the sense that the end results are very similar irrespective of the input for the measurement variance, unlike the iES where the results are much more sensitive to the input. Finally in Fig. 6 the posterior standard deviations in each grid cell are shown. Again we notice the similarity for the miES results, as well as the high sensitivity of iES to the input values. The results clearly demonstrate the advantage of the miES: it is by construction far less sensitive to the input than the iES, yielding a more robust algorithm for real field cases. From the results it seems as the miES has a 'localization' effect on the ensemble when compared with the iES with 1000 members.

4 Summary and future work

This paper has presented a new class of iterative ensemble methods called the marginalized iterative ensemble smoother (miES). The new algorithms are derived from maximization of the posterior density where the measurement error variance parameters in the likelihood are treated as hyperparameters and integrated out. Two different priors for the hyperparameters were studied. The first was the non-informative or Jeffrey's prior, which requires no user input.

Hence the miES algorithm requires no input for the measurement error variances. The second was the scaled inverse chi-square prior where an offline estimate or an expert opinion can be used as input. The resulting miES algorithm can then be used with the same input as the standard iES algorithm. A test on a synthetic reservoir model clearly demonstrated potential advantages of the new method when compared to the iES. The solution is less sensitive to the input and produced solutions that were more in line with the iES with an ensemble size of 1000 and correctly specified standard deviations than the iES with 100 ensemble members. The miES can handle different variances for different datatypes and it is also possible to use different priors for different data. The algorithm is of the same complexity as the standard iES.

In situations where the measurements are correlated, but with unknown correlation structure, further work is required to use the miES. A direct generalization using non-informative or an inverse Wishart prior density is possible assuming that the entire measurement error covariance is unknown, but this may lead to a very uncertain likelihood function and it is perhaps more interesting to study a known correlation structure with unknown parameters. Both these topics will be investigated in future research.

Acknowledgements The authors acknowledge financial support from the NORCE research project "Assimilating 4D Seismic Data: Big Data Into Big Models" which is funded by industry partners, Equinor Energy AS, Repsol Norge AS, Shell Global Solutions International B.V., Total-EnergiesEP Norge AS, WintershallDeaNorve AS and AkerBP ASA, as well as the Research Council of Norway (PETROMAKS2).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Anthes, R., Rieckh, T.: Estimating observation and model error variances using multiple data sets. *Atmos. Meas. Tech.* **11**(7),

- 4239–4260 (2018). <https://doi.org/10.5194/amt-11-4239-2018>. (<https://amt.copernicus.org/articles/11/4239/2018/>)
2. Bocquet, M., Sakov, P.: An iterative ensemble Kalman smoother. *Q. J. R. Meteorol. Soc.* **140**(682), 1521–1535 (2012)
 3. Casella, G.: An introduction to empirical Bayes data analysis. *Am. Stat.* **39**(2), 83–87 (1985). ISSN 00031305. <http://www.jstor.org/stable/2682801>
 4. Chen, S.-Y., Huang, C.-Y., Kuo, Y.-H., Sokolovskiy, S.: Observational error estimation of formosat-3/cosmic GPS radio occultation data. *Mon. Weather. Rev.* **139**(3), 853–865 (2011). <https://doi.org/10.1175/2010MWR3260.1>. (<https://journals.ametsoc.org/view/journals/mwre/139/3/2010mwr3260.1.xml>)
 5. Daley, R.: Atmospheric Data Analysis. Cambridge University Press, Cambridge Atmospheric and Space Science Series (1993). ISSN 00031305. <http://www.jstor.org/stable/2682801>
 6. Desroziers, G., Ivanov, S.: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Q. J. R. Meteorol. Soc.* **127**(574), 1433–1452 (2001). <https://doi.org/10.1002/qj.49712757417>. (<https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712757417>)
 7. Desroziers, G., Berre, L., Chapnik, B., Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **131**(613), 3385–3396 (2005). <https://doi.org/10.1256/qj.05.108>. (<https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.108>)
 8. Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin, D.: Bayesian Data Analysis (3rd ed.). Chapman and Hall/CRC (2013)
 9. Gray, J., Allan, D.: A method for estimating the frequency stability of an individual oscillator. In: 28th Annual Symposium on Frequency Control, pp. 243–246. (1974). <https://doi.org/10.1109/FREQ.1974.200027>
 10. Hollingsworth, A., Lönnberg, P.: The statistical structure of short-range forecast errors as determined from radiosonde data. part I: The wind field. *Tellus A* **38A**(2), 111–136 (1986). <https://doi.org/10.1111/j.1600-0870.1986.tb00460.x>. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.1986.tb00460.x>
 11. Jaynes, E.: Probability Theory: The Logic of Science. Cambridge University Press Cambridge (2003)
 12. Khaki, M., Ait-El-Fquih, B., Hoteit, I., Forootan, E., Awange, J., Kuhn, M.: Unsupervised ensemble Kalman filtering with an uncertain constraint for land hydrological data assimilation. *J. Hydrol.* **564**, 175–190 (2018). ISSN 0022–1694. <https://doi.org/10.1016/j.jhydrol.2018.06.080>. <https://www.sciencedirect.com/science/article/pii/S002216941830502X>
 13. Luo, X., Bhakta, T.: Estimating observation error covariance matrix of seismic data from a perspective of image denoising. *Comput. Geosci.* **21**(2), 205–222 (2017)
 14. Mehra, R.: On the identification of variances and adaptive Kalman filtering. *IEEE Trans. Autom. Control.* **15**(2), 175–184 (1970). <https://doi.org/10.1109/TAC.1970.1099422>
 15. Mehra, R.: On the identification of variances and adaptive Kalman filtering. *IEEE Trans. Autom. Control.* **15**(2), 175–184 (1970). <https://doi.org/10.1109/TAC.1970.1099422>
 16. Oliver, D.S., Reynolds, A.C., Liu, N.: Inverse theory for petroleum reservoir characterization and history matching. Cambridge, (2008)
 17. Oliver, D.S., Fossum, K., Bhakta T., Sandø, I., Nævdal, G., Lorentzen, R.J.: 4d seismic history matching. *J. Petrol. Sci. Eng.* **207**, 109119 (2021). ISSN 0920-4105. <https://doi.org/10.1016/j.petrol.2021.109119>. <https://www.sciencedirect.com/science/article/pii/S0920410521007750>
 18. Raanes, P.N., Stordal, A.S., Evensen, G.: Revising the stochastic iterative ensemble smoother. *Nonlinear Process. Geophys.* **26**(3), 325–338 (2019). <https://doi.org/10.5194/npg-26-325-2019>. (<https://npg.copernicus.org/articles/26/325/2019/>)
 19. Raboudi, N.F., Ait-El-Fquih, B., Hoteit, I.: Online estimation of colored observation-noise parameters within an ensemble Kalman filtering framework. *Q. J. R. Meteorol. Soc.* n/a, (n/a) (2023). <https://doi.org/10.1002/qj.4484>. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4484>
 20. Rasmussen, A.F., Sandve, T. H., Bao, K., Lauser, A., Hove, J., Skaffestad, B., Klöfkorn, R., Blatt, M., Rustad, A.B., Sævareid, O., Lie, K.-A., Thune, A.: The open porous media flow reservoir simulator. *Comput Math Appl* **81**, 159–185 (2021). ISSN 0898-1221. <https://doi.org/10.1016/j.camwa.2020.05.014>. <https://www.sciencedirect.com/science/article/pii/S0898122120302182>. Development and Application of Open-source Software for Problems with Numerical PDEs
 21. Rawlinson, N., Fichtner, A., Sambridge, M., Young, M.K.: Seismic tomography and the assessment of uncertainty. *Adv. Geophys.* **55**, 1–76 (2014). ISSN 0065–2687. <https://doi.org/10.1016/bs.agph.2014.08.001>. <https://www.sciencedirect.com/science/article/pii/S0065268714000028>
 22. Rue, H., Martino, S., Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71**(2), 319–392 (2009). <https://doi.org/10.1111/j.1467-9868.2008.00700.x>. <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00700.x>
 23. Simon, D.: Optimal state estimation: Kalman, H infinity, and non-linear approaches. John Wiley & Sons (2006)
 24. Sornette, D., Ide, K.: The Kalman–Lévy filter. *Physica D Nonlinear Phenom* **151**(2), 142–174 (2001). ISSN 0167-2789. [https://doi.org/10.1016/S0167-2789\(01\)00228-7](https://doi.org/10.1016/S0167-2789(01)00228-7). <https://www.sciencedirect.com/science/article/pii/S0167278901002287>
 25. Tilmann, F.J., Sadeghisorkhani, H., Mauerberger, A.: Another look at the treatment of data uncertainty in Markov chain Monte Carlo inversion and other probabilistic methods. *Geophys. J. Int.* **222**, 388–405, 05 (2020). ISSN 0956-540X. <https://doi.org/10.1093/gji/ggaa168>
 26. Ueno, G., Nakamura, N.: Bayesian estimation of the observation-error covariance matrix in ensemble-based filters. *Q. J. R. Meteorol. Soc.* **142**(698), 2055–2080 (2016). <https://doi.org/10.1002/qj.2803>. (<https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2803>)

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