Hybrid Iterative Ensemble Smoother for History Matching of Hierarchical Models

Dean S. Oliver[1]

Abstract The choice of a prior model can have a large impact on the ability to assimilate data. In standard applications of ensemble-based data assimilation, all realizations in the initial ensemble are generated from the same covariance matrix with the implicit assumption that this covariance is appropriate for the problem. In a hierarchical approach, the parameters of the covariance function, for example, the variance, the orientation of the anisotropy and the ranges in two principal directions, may all be uncertain. Thus, the hierarchical approach is much more robust against model misspecification. In this paper, three approaches to sampling from the posterior for hierarchical parameterizations are discussed: an optimization-based sampling approach (randomized maximum likelihood, RML), an iterative ensemble smoother (IES), and a novel hybrid of the previous two approaches (hybrid IES). The three approximate sampling methods are applied to a linear-Gaussian inverse problem for which it is possible to compare results with an exact "marginal-then-conditional" approach. Additionally, the IES and the hybrid IES methods are tested on a two-dimensional flow problem with uncertain anisotropy in the prior covariance. The standard IES method is shown to perform poorly in the flow examples because of the poor representation of the local sensitivity matrix by the ensemble-based method. The hybrid method, however, samples well even with a relatively small ensemble size.

Keywords Hierarchical · Data assimilation · Uncertainty quantification · Bayesian

✉ Dean S. Oliver
dean.oliver@norceresearch.no

1   Technology Department, NORCE Norwegian Research Centre, Nygårdsgaten 112, 5008 Bergen, Norway

1 Introduction

In Bayesian methods of data assimilation, it is necessary to specify the prior joint probability density for the model parameters (the "prior" for short). The prior for parameters of a subsurface reservoir model (properties such as permeability and porosity) may be based partly on data that have already been assimilated, such as log or core measurements and seismic surveys. The choice of a prior joint probability is also often influenced by the joint distributions of properties observed in modern geological analogues of ancient depositional environments.

The choice of the prior is a challenge in Bayesian inference (Scales and Tenorio 2001). In almost all applications of ensemble Kalman-based data assimilation methods, the prior is chosen to be multivariate normal with fully specified prior mean and prior covariance. In these cases, the prior mean largely determines the average properties in regions where the data are not sensitive to parameters of the model, while the choice of the covariance determines the smoothness of the spatial distributions, the variability of magnitudes, and the orientation and the range of correlation of the spatially distributed parameter fields. In most subsurface applications, it is difficult to select these parameters (Malinverno and Briggs 2004).

In many applications of the Bayesian data assimilation methods to field cases, the prior is found to have been too narrowly specified —it does not allow for the possibility of events that could (and perhaps will) occur in the future, and is often inconsistent with historical measurements of flow and transport. If a prior model that is inconsistent with data is used for data assimilation, the result will be biased estimates and forecasts, implausible parameter values, and unjustified reduction in uncertainty (Moore and Doherty 2005; Oliver and Alfonzo 2018). At the end of an expensive model-building and calibration exercise, it will be apparent that the model is inadequate and will need to be rebuilt.

Although allowing for uncertainty in the prior mean is not common in ensemble-based history matching, its usefulness has been demonstrated in both synthetic and field cases. Li et al. (2010) allowed for a uniform adjustment to the mean, while Zhang and Oliver (2011) allowed the prior mean to be characterized by a trend surface whose coefficients were uncertain. In some cases (such as the Brugge benchmark case), the use of a hierarchical model with uncertainty in the mean permeability was not necessary for matching production data, but the hierarchical model provided a more reasonable explanation of the bias in the initial forecasts (Chen and Oliver 2010).

Uncertainty in the prior covariance is widely recognized as being a key aspect of overall uncertainty, but careful treatment of covariance uncertainty in history matching is rare. A practical approach to parameterizing the uncertainty in the covariance is to assume a family of covariance functions with hyperparameters that control the smoothness, the covariance range, the variance, and the anisotropy. If these hyperparameters are fixed at inappropriate values, which are then used in data assimilation, the ability to assimilate flow data will often be limited. One approach to treating uncertainty in the hyperparameters is to use the concept of scenarios to represent the possibilities of a discrete number of alternative values for the hyperparameters. Park et al. (2013) demonstrated the usefulness of this approach for a case in which the orientation of the anisotropy was assumed to be one of two possible angles. The

approach used by Emerick (2016) was somewhat similar, except that weights for the scenarios were computed using an iterative ensemble smoother. Although Malinverno and Briggs (2004) did not consider uncertainty in anisotropy, they allowed uncertainty in five hyperparameters of the covariance for a problem of inferring compressional wave slowness in a one-dimensional earth model.

If the covariance is allowed to be uncertain in an ensemble Kalman-based approach, the parameterization of the model and the choice of the covariance model are both important to successful data assimilation. Chada et al. (2018) investigated the effect of both centered and non-centered parameterizations (described in Sect. 2.1) of a hierarchical spatial model for use with ensemble Kalman iteration (EKI). In their experiments, which included nonstationary hyperparameters, they concluded that using a hierarchical approach with a non-centered parameterization was significantly better than a hierarchical approach with a centered parameterization and that both hierarchical approaches were better than a non-hierarchical approach. Subsequently, Dunlop et al. (2020) showed that for maximum a posteriori (MAP) estimation of the hyperparameters in a linear inverse problem, the centered parameterization is to be preferred when the goal is MAP estimation as opposed to uncertainty quantification.

2 The Hierarchical Model

Gaussian priors are often used in data assimilation because they are relatively tolerant to errors in misspecification, but the choice of the parameters of the covariance still have an influence on the quality of the final data match. In Sect. 4.2, the feasibility of the hybrid IES to assimilate data in a model with an uncertain prior covariance is investigated. In that example, the "true" model that generated the data has an anisotropic covariance with a longer range in one direction. If one were fortunate enough to know the covariance model from which the truth was sampled, it would not be necessary to use a hierarchical model as a good average total squared data mismatch (e.g. $S_d^o = 365$ for the correct covariance in the two-dimensional flow problem) could be achieved without consideration of uncertainty in hyperparameters. On the other hand, if the prior covariance is mistakenly fixed at an incorrect orientation (off by $\pi/2$), the misfit to data would be substantially worse ($S_d^o = 2142$), and the ability to accurately forecast future behavior would also suffer. Uncertainty in the covariance can be easily accounted for by introducing a hierarchical model for which the anisotropy and ranges are uncertain. By applying the hierarchical model in this problem and using the hybrid iterative ensemble smoother for data assimilation, it is possible to obtain a mean squared data mismatch ($S_d^o = 1151$) that is intermediate between the value obtained using the correct covariance and a covariance with incorrect orientation. When hierarchical modeling is an appropriate approach to describing uncertainty, it provides more robust priors for history matching and reduces the need for rebuilding the model and re-history matching. Unfortunately, hierarchical models are more nonlinear than non-hierarchical models, and data assimilation or history matching is more difficult. A straightforward application of an iterative ensemble smoother with localization to the hierarchical problem with uncertain anisotropy results in an extremely poor match to data ($S_d^o = 13,000$). It appears that in order to use a hierarchical model with an

ensemble-based data assimilation, modification of the data assimilation methodology will generally be required, as presented in Sect. 3.3.

The probability distribution of Gaussian random variables (denoted m) is completely determined by the mean and the covariance of the variable. In this study, we will assume that the mean m_{pr} is known but that the covariance C_m is uncertain. There are a number of parameters that could describe the uncertainty in C_m; for two-dimensional fields, we will focus on correlation range ρ, orientation of the anisotropy ϕ, and the ratio of range in two principal directions. For a one-dimensional field, we will focus on problems in which the variance and the correlation range are uncertain.

2.1 Hierarchical Model Parameterization: Centered and Non-centered

There are two common parameterizations for Gaussian hierarchical problems: the centered (or natural) parameterization and the non-centered parameterization (Papaspiliopoulos et al. 2007). In the centered parameterization, the natural variables (denoted m for model parameters) are augmented with a set of hyperparameters θ that characterize the covariance and perhaps the mean. With the augmented parameters, the set of parameters that must be sampled can be written as $x = (m, \theta)$. The centered parameterization utilizes the conditional independence of the data d and the hyperparameters of the distribution θ given the observable parameters m, so it is relatively easy to implement a Gibbs sampler for a centered parameterization. The disadvantage of the centered parameterization for the ensemble form of data assimilation, however, is that the objective function contains the term $(m - m_{\text{pr}})^{\text{T}} C_m^{-1} (m - m_{\text{pr}})$. The matrix C_m depends on the parameters θ, so that an ensemble approximation of C_m is not appropriate.

In the non-centered parameterization, the relationship between the natural Gaussian variable m and the non-centered parameters can be written

$$m = m_{\text{pr}} + L(\theta)z, \tag{1}$$

where L is a "square root" of the model covariance matrix, i.e., $LL^{\text{T}} = C_m$, and z is a vector of independent standard normal deviates with the same dimension as m. In the non-centered parameterization, θ and z are a priori independent. Thus, for the non-centered parameterization, we can write the entire set of parameters $x = (z, \theta)$. Note that the prior probability density function (pdf) for x is Gaussian in the centered parameterization, making it suitable for an ensemble form of data assimilation.

Although Eq. 1 is used in this paper to generate realizations of m, Eq. 1 will only be feasible for large grids if the square root of the covariance, L, is compact (as it is for the spherical covariance), or if the correlation range is sufficiently short that most elements of L are effectively zero. Other methods could be more efficient for large grids. In particular, it may often be advantageous to use either the inverse of the covariance matrix (the precision matrix) or a factorization of the precision matrix (Stojkovic et al. 2017) or a Matérn–Whittle covariance model (Gneiting et al. 2010) for which stochastic partial differential equations can be used efficiently to generate a random field (Roininen et al. 2019; Zhou et al. 2018). Another relatively common approach for

describing the uncertainty in Gaussian hierarchical models is to assume that the prior distribution for the covariance model matrix is inverse Wishart (Myrseth and Omre 2010; Tsyrulnikov and Rakitko 2017). While this approach has some computational advantages, it seems less likely to describe the prior uncertainty in the covariance.

3 Data Assimilation for Gaussian Hierarchical Models

For large geoscience inverse problems, iterative ensemble smoothers are often an effective approach. These are all based on the original development of the ensemble Kalman filter (Evensen 1994), which has several advantages over Kalman filters or extended Kalman filters: a low-rank approximation of the covariance matrix is used instead of the full covariance, and the linearization of the relationship between predicted data and model parameters is approximated without requiring adjoints. For geoscience inverse problems, it has been found that it is more efficient to use a "smoother" to update the parameters of the inverse problem using all the data simultaneously. On the other hand, the problem of parameter estimation becomes more nonlinear when all data are assimilated simultaneously, so iteration is almost always required when a smoother is used for history matching. Although there are many variants of iterative ensemble smoothers, they can generally be classified into one of two approaches. In multiple data assimilation (MDA), the same data are assimilated multiple times with an inflated observation error (Reich 2011; Emerick and Reynolds 2013). This approaches reduces the nonlinearity in the update, but also requires updating the approximation of the covariance matrix at each iteration. That seems to be more difficult for hierarchical models than for models in which the prior covariance is assumed to be known.

The second class of iterative ensemble smoothers is based on the randomized maximum likelihood (RML) approach to approximate sampling from the posterior (Kitanidis 1995; Oliver et al. 1996, 2008). In the RML approach, samples from the prior are updated to become approximate samples from the posterior by minimizing a stochastic objective function. Like MDA, this method is exact for linear Gaussian data assimilation problems, but this method does not require updating of the covariance at each iteration. In the iterative ensemble smoother form of RML (Chen and Oliver 2012), an average sensitivity, computed from the ensemble of samples, is used to approximate the downhill direction. Because an ensemble average sensitivity will not provide an accurate sensitivity when the problem is highly nonlinear, a hybrid data assimilation method that is a blend of RML and IES is introduced in which some derivatives are computed analytically, while other derivatives are estimated from the ensemble. Finally, for the linear observation case, results will be compared with the marginal-then-conditional (MTC) approach introduced by Fox and Norton (2016).

The unnormalized posterior pdf for the model parameters in the non-centered parameterization can be written as

$$
p(z, \theta \mid d) \propto \exp\left(-\frac{1}{2}(d^o - g(m(\theta, z)))^{\mathrm{T}} C_d^{-1}(d^o - g(m(\theta, z)))\right)
$$
$$
\times \exp\left(-\frac{1}{2}(\theta - \bar{\theta})^{\mathrm{T}} C_\theta^{-1}(\theta - \bar{\theta}) - \frac{1}{2} z^{\mathrm{T}} z\right) \tag{2}
$$

or more simply

$$p(x \mid d) \propto \exp\left(-\frac{1}{2}(d^o - g(m(x)))^{\mathrm{T}} C_d^{-1}(d^o - g(m(x)))\right)$$
$$\exp\left(-\frac{1}{2}(x - \bar{x})^{\mathrm{T}} C_x^{-1}(x - \bar{x})\right). \tag{3}$$

The non-centered parameterization appears to be well suited to data assimilation using an iterative ensemble smoother when the prior pdf for both z and θ have been assumed Gaussian (perhaps after transformation), as the pdf for x (Eq. 3) is identical in form to the pdf for m in traditional Bayesian history matching. Nonlinearity is a result either of the relationship $d = g(m)$ being nonlinear, as is the case if the data are water rates and the model parameters are porosity and log-permeability, or a result of nonlinearity in the relationship between m and θ. Although Eq. 3 is written for the case in which the prior for the hyperparameters is Gaussian, when the uncertainty in the orientation of the anisotropy is moderately large, a Gaussian approximation is not appropriate. We discuss that case in Appendix A.

3.1 Data assimilation: RML

The randomized maximum likelihood approach begins, like the perturbed observation form of the ensemble Kalman filter (Burgers et al. 1998; Houtekamer and Mitchell 1998), by drawing samples (x_i', ϵ_i'), $i = 1, \ldots, N_s$, from the Gaussian distribution

$$q_{X'\epsilon'}(x', \epsilon') \propto \exp\left(-\frac{1}{2}(x - \bar{x})^{\mathrm{T}} C_x^{-1}(x - \bar{x})\right) \exp\left(-\frac{1}{2}(\epsilon')^{\mathrm{T}} C_\epsilon^{-1}\epsilon'\right) \tag{4}$$

for a given $\bar{x}$. The ith approximate sample from the posterior is obtained by computing the minimizer of the cost functional

$$J_i(x) = \frac{1}{2}\left(x - x_i'\right)^{\mathrm{T}} C_x^{-1}\left(x - x_i'\right)$$
$$+ \frac{1}{2}\left(g(m(x)) + \epsilon_i' - d^o\right)^{\mathrm{T}} C_d^{-1}\left(g(m(x)) + \epsilon_i' - d^o\right). \tag{5}$$

This is usually done by solving

$$\nabla_x J_i(x) = 0 \tag{6}$$

for x. If Levenberg–Marquardt with a Gauss–Newton approximation of the Hessian is used for the minimization, the ℓth update is of the form

$$\delta x_\ell = \frac{x_i' - x_\ell}{1 + \lambda_\ell} - C_x G_\ell^{\mathrm{T}}\left[(1 + \lambda_\ell)C_d + G_\ell C_x G_\ell^{\mathrm{T}}\right]^{-1}\left[(g(x_\ell) + \epsilon_i' - d^o) - \frac{G_\ell(x_\ell - x_i')}{1 + \lambda_\ell}\right], \tag{7}$$

where $G^T = \nabla_x(g^T)$, and λ_ℓ is the Levenberg–Marquardt regularization parameter. This method of sampling from the posterior distribution is only exact if the relationship between the data and the model parameters is linear. It will sample accurately from multimodal distributions in some situations, but exact sampling using RML requires computation of additional critical points and weighting of solutions (Ba et al. 2022). For geoscience applications, the standard unweighted RML is almost always used.

3.2 Data assimilation: IES

One disadvantage of RML is the need for computation of the gradient of the objective function. For many forward models of the data such as reservoir production simulation, the derivatives are not readily available. In those cases, the iterative ensemble smoothers offer an alternative that avoids the need to compute G directly. The basic idea is that terms that appear in Eq. 7 can often be computed efficiently using ensemble approximations. In an ensemble-based approach to sampling from the posterior, Eq. 7 is replaced by the following update step (Chen and Oliver 2013)

$$
\begin{aligned}
\delta x_{\ell+1} = &-\frac{1}{(1+\lambda_\ell)}\Delta x_\ell \, \Delta x_\ell^T C_x^{-1}(x_\ell - x') \\
&- \Delta x_\ell \, \Delta d_\ell^T \left((1+\lambda_\ell)C_d + \Delta d_\ell \Delta d_\ell^T\right)^{-1} \\
&\times \left(g(m_\ell) + \epsilon_i' - d^o - \frac{1}{(1+\lambda)}\Delta d_\ell \Delta x_\ell^T C_x^{-1}(x_\ell - x_i')\right),
\end{aligned}
\tag{8}
$$

where $\Delta x_\ell = \frac{(X_\ell - \bar{X}_\ell)}{\sqrt{(N-1)}}$ and similar for Δd_ℓ. If this approach were applied directly, it can be seen that any update is restricted to the space spanned by the initial ensemble, and the number of degrees of freedom available for calibration is $N_e - 1$ where N_e is the number of realizations in the initial ensemble. To avoid this limitation, localization is nearly always applied in large problems, but for clarity, localization has been omitted in Eq. 8. The main disadvantage of an ensemble-based methodology is the failure to handle strong nonlinearity well—primarily because the same average sensitivity is used for all samples.

3.3 Data assimilation: hybrid IES

The hybrid IES is a method that takes advantage of the ability of the RML to use different gain matrices for each sample and the ability of the IES to avoid the need for adjoint systems. As in the RML method, the (transpose of) sensitivity of data to model parameters in a non-centered parameterization is

$$
G^T = G(x)^T = \begin{bmatrix} \nabla_x g_1 & \nabla_x g_2 & \dots & \nabla_x g_{N_d} \end{bmatrix} = \nabla_x(g^T)
\tag{9}
$$

and similarly define the sensitivity of data to the observable model parameters m,

$$G_m^{\mathrm{T}} = G(m)^{\mathrm{T}} = \begin{bmatrix} \nabla_m g_1 & \nabla_m g_2 & \cdots & \nabla_m g_{N_d} \end{bmatrix} = \nabla_m (g^{\mathrm{T}}). \tag{10}$$

These two sensitivity matrices are related

$$G^{\mathrm{T}} = \nabla_x (g^{\mathrm{T}}) = \nabla_x (m^{\mathrm{T}}) \cdot \nabla_m (g^{\mathrm{T}}) = \nabla_x (m^{\mathrm{T}}) \cdot G_m^{\mathrm{T}} \tag{11}$$

or

$$G = G_m \cdot \left(\nabla_x (m^{\mathrm{T}}) \right)^{\mathrm{T}}. \tag{12}$$

For notational simplicity, the Jacobian of the observable model parameters with respect to the non-centered hierarchical parameters is written as

$$M_x = \left(\nabla_x (m^{\mathrm{T}}) \right)^{\mathrm{T}}, \tag{13}$$

where the dimension of M_x is $N_m \times N_x$. Substituting $G = G_m M_x$ into the RML update expression (Eq. 7) with ensemble representation of G_m results in a hybrid IES data assimilation approach

$$\begin{aligned} \delta x &= -\frac{1}{(1+\lambda)}(x - x') - C_x M_x^{\mathrm{T}} G_m^{\mathrm{T}} \left((1+\lambda)C_d + G_m M_x C_x M_x^{\mathrm{T}} G_m^{\mathrm{T}} \right)^{-1} \\ &\quad \times \left(g(m) + \epsilon' - d^o - \frac{1}{(1+\lambda)} G_m M_x (x - x') \right) \\ &= -\frac{1}{(1+\lambda)}(x - x') \\ &\quad - C_x M_x^{\mathrm{T}} (\Delta m)^{-\mathrm{T}} \Delta d^{\mathrm{T}} \left((1+\lambda)C_d + \Delta d (\Delta m)^{-1} M_x C_x M_x^{\mathrm{T}} (\Delta m)^{-\mathrm{T}} \Delta d^{\mathrm{T}} \right)^{-1} \\ &\quad \times \left(g(m) + \epsilon' - d^o - \frac{1}{(1+\lambda)} \Delta d (\Delta m)^{-1} M_x (x - x') \right). \end{aligned} \tag{14}$$

Note that in the hybrid IES method, each ensemble member has its own Kalman gain matrix since the sensitivity M_x of the physical model parameters m to the non-centered parameters x is specific to an ensemble member; assuming that m is determined from the known mean m_{pr}, the uncertain covariance C_m and from the stochastic variable z, i.e., $m = m_{pr} + L(\sigma, a)z$ where $LL^T = C_m$.

3.3.1 Factorization of Covariance Functions

The hybrid approach will only be useful if computation of the terms in Eq. 14 is practical. First, note that C_x is diagonal, so that it can be eliminated by scaling of the variables. Computing M_x this way relies on factorability of the covariance matrix. For

functions in one dimension, the equivalent to defining $C_m = LL^T$ is to define

$$C(r) = f * f^T = \int_{-\infty}^{+\infty} f(s)f(r+s)\mathrm{d}s. \tag{15}$$

For illustration, consider the one-dimensional exponential covariance function $C(r) = \sigma^2 e^{-|r|/a}$ where r is the distance between two variables, a is a measure of the range of the correlation, and σ^2 is the variance. In one dimension, a symmetric factorization is $L_s(r) = \frac{\sigma\sqrt{2}}{\sqrt{a\pi}} K_0(|r|/a)$, where K_0 is the modified Bessel function of the second kind of order 0. The factorizations are not unique, however, and it may be beneficial to use a factorization for which the "square root" is not singular. A one-sided factorization that is always finite is $L_a(r) = \sigma\sqrt{2/a} H(r)\exp(-r/a)$ where $H(r)$ is the Heaviside function. Similarly, the symmetric factorization of the squared exponential or Gaussian covariance $C(r) = \sigma^2 e^{-r^2/a^2}$ can be shown to be $L_s(x) = \sigma \left(\frac{4}{a^2\pi}\right)^{1/4} \exp(-2r^2/a^2)$.

Similar factorizations in two and three dimensions can easily be derived for other covariance functions such as the spherical covariance in three dimensions, the circular covariance in two dimensions, and Whittle's covariance (Oliver 1995). For example, in a two-dimensional stationary field with geometric anisotropy and positions x, we define $r = \sqrt{(x-x')^T H (x-x')}$ and compute convolution square roots of the covariance as described above. In this paper, the square root of a two-dimensional Gaussian covariance model is used. Alternatively, one can start with a specific covariance "square root" and compute the corresponding covariance function (Gaspari and Cohn 1999). In this study, the factorization of the covariance function has been used to obtain an approximation of the factorization of the covariance matrix. A scale space implementation might be justified if additional accuracy was required (Lindeberg 1990).

4 Numerical Examples

Two numerical examples are presented to illustrate data assimilation for Gaussian models with uncertainty in the covariance. In the first example, the model is defined on a one-dimensional grid, and the observations are of m. In the second example, the uncertain permeability field in a two-dimensional porous medium is estimated by assimilation of a time series of water cut observations at six producing wells.

4.1 One-Dimensional Linear Gaussian

Consider a simple one-dimensional Gaussian random field, discretized on the interval $[0, 1]$, into $n_m = 150$ lattice points. The random variable m is observed at every fourth lattice point, with independent measurement noise characterized by $\sigma_d = 0.01$. The data-generating model has correlation range $a^{tr} = 0.100$ and standard deviation $\sigma_m^{tr} = 1.08$. The noisy observations are shown in Fig. 1a.
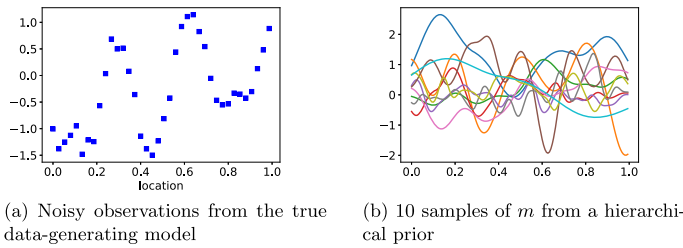
(a) Noisy observations from the true data-generating model

(b) 10 samples of m from a hierarchical prior

Fig. 1 The one-dimensional linear test problem with hierarchical prior

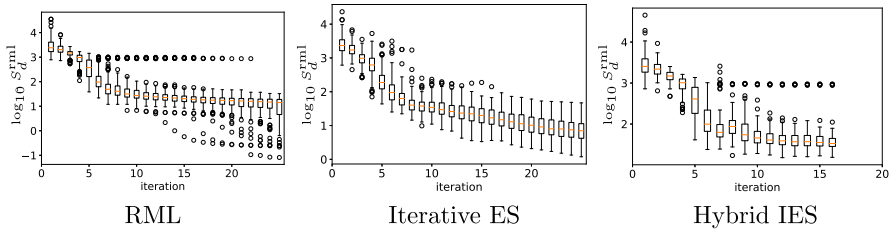RML   Iterative ES   Hybrid IES

Fig. 2 Reduction of squared data mismatch to perturbed data (Eq. 17) for three data assimilation algorithms

The form of the covariance of the data-generating model C_m (squared Gaussian) is assumed to be known—only the correlation range a and model variance σ^2 are uncertain. Because both parameters are required to be positive, log-normal distributions are assumed for each: $\theta_1 \equiv \log \sigma_m \sim N(-0.22, 0.5^2)$ and $\theta_2 \equiv \log a \sim N(-2.3, 0.6^2)$. Ten unconditional samples of m from the prior are shown in Fig. 1b.

The posterior pdf can therefore be written as

$$p(z, a, \sigma \mid d) \propto \exp\left(-\frac{1}{2}(d - g(m(u, z)))^T C_d^{-1}(d - g(m(u, z)))\right.$$
$$\left. -\frac{1}{2}(\theta - \theta_{pr})^T C_\theta^{-1}(\theta - \theta_{pr}) - \frac{1}{2}z^T z\right), \tag{16}$$

and the corresponding objective function for a minimization-based sampling approach is given by Eq. 5. Results from the three different data assimilation approaches described in Sect. 3 (RML, IES, and hybrid IES) are compared with the exact distribution of hyperparameters obtained using marginalization for the centered parameterization (Rue and Held 2005; Rue and Martino 2007; Fox and Norton 2016) and with the true values of the hyperparameters.

Note that although RML does not use information from the ensemble for data assimilation—each realization is generated independently—100 RML samples were generated for the comparison. It is necessary to set a few algorithmic parameters for the minimizations. For RML with Levenberg–Marquardt minimization, an initial value of $\lambda = 5000$ was used. It was decreased by a factor of 4 if the objective decreased. Otherwise, it was increased by a factor of 4. For each minimization, the initial model was chosen to be the sample from the prior. For IES and hybrid IES, the initial value of
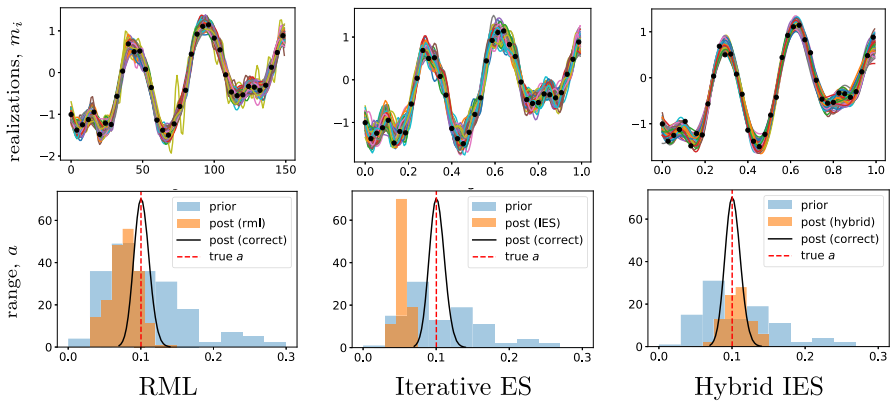
Fig. 3 Samples of m from the posterior for three data assimilation methods (top row) and comparison of prior and posterior distributions of hyperparameters (bottom row)

λ was based on the initial mean value of the data mismatch objective function (Chen and Oliver 2013). An ensemble size of 200 was used for the IES, and an ensemble size of 100 was used for the hybrid IES.

The reduction in the squared mismatch of simulated data to perturbed observations (Eq. 17) was initially rapid for all three methods, although a few realizations with long initial correlation ranges failed to converge for both the RML and the hybrid IES (Fig. 2). The smallest mismatch values with perturbed observations were achieved by the RML, with values of the realizations of the data objective function achieving values

$$S_d^{\text{rml}}(m) = \frac{1}{2} \sum_{i=1}^{N_d} \frac{(g(m) - (d^o + \epsilon))^2}{\sigma_d^2} \tag{17}$$

smaller than one in several cases. While this objective is useful for monitoring convergence, for model checking it is more useful to evaluate the squared data mismatch between predicted data from the calibrated models and the actual data; the expected value of that metric is

$$E[S_d^{obs}(m)] = \frac{1}{2} \sum_{i=1}^{N_d} \frac{(g(m) - d^o)^2}{\sigma_d^2} = \frac{1}{2} N_d = 19. \tag{18}$$

The average values from all three methods are close to the expected value.

The a posteriori realizations of m from all three methods also look generally plausible (top row Fig. 3), although it appears that realizations from the RML and IES have correlation lengths that are generally shorter that the correlation length in the data-generating model (Fig. 3). Qualitatively, it appears that the standard deviation of the posterior model realizations is somewhat larger for the IES method than for the other two methods. Figure 4 confirms this quantitatively, although the differences

Fig. 4 Posterior estimates of standard deviation of the model as functions of position, for three data assimilation methods applied to the one-dimensional linear problem
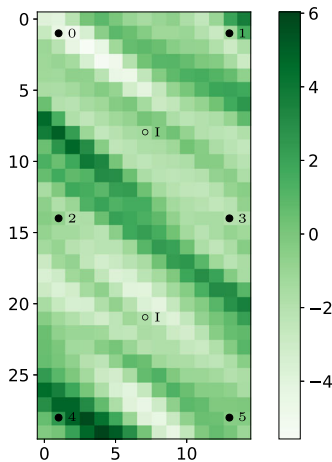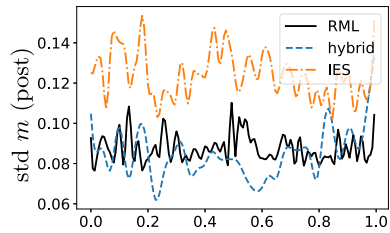
(a) true $\ln K$
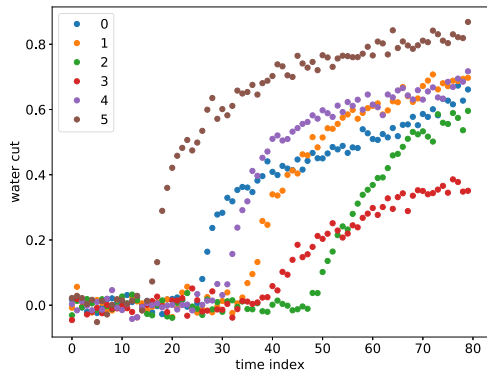
(b) Noisy observations of water cut in 6 producing wells.

Fig. 5 Two-dimensional reservoir flow problem with uncertain prior covariance

between methods are relatively small. Note that the standard error in the observations is 0.1, and the prior standard deviation for m was uncertain.

4.2 Fluid Flow Example

In the second example, the data-generating system is a two-dimensional (30×15) anisotropic porous medium with two-phase (oil and water), immiscible, incompressible flow driven by two injectors and six producing wells. The permeability field that generates the data is a draw from a prior model in which the angle of the principal axis of anisotropy is 0.93 radians, the range parameter for the longest correlation length is 1.0, and the ratio of correlation ranges in the two principal directions is 6.0. The covariance type for log-permeability is set to be "Gaussian," i.e., squared exponential (Eq. B6) with standard deviation 2.0. The porosity is uniform.

The data are measurements of water cut (fraction of produced fluid that is water) as a function of time in the producing wells. Although the permeability is highly variable, the wells are controlled by total flow rate which is identical for all producers. Figure 5a shows the permeability field that was used to generate the observations, and Fig. 5b shows the noisy observations of water cut at each of the wells. The
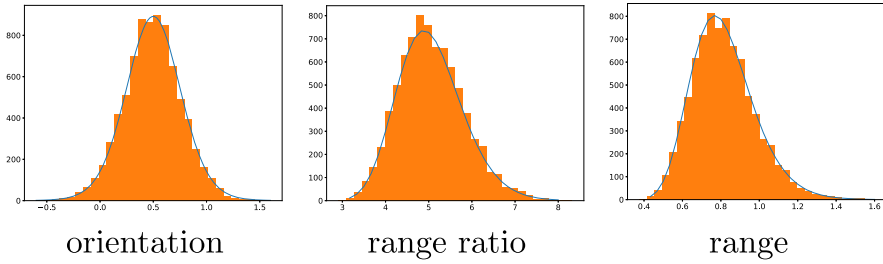
Fig. 6 Prior uncertainty in each of the hyperparameters of the hierarchical model for the two-dimensional flow problem

errors in the observations are independent Gaussian with standard error 0.02. For data assimilation, the prior covariance for log-permeability has uncertainty in the orientation of the principal axes of anisotropy and in the range of the correlation in each of the two principal directions. The prior uncertainties for each of the hyperparameters of the covariance are shown as histograms of sampled values in Fig. 6. The priors for correlation length and ratio of correlation ranges are both assumed to be log-normal. The prior for orientation of the principal axes of the covariance for permeability is Gauss–von Mises (Eq. A1), which is close to Gaussian when the variance is relatively small.

4.3 Data Assimilation Results

Data assimilation with the hierarchical parameterization was performed using two methods: a standard Levenberg–Marquardt iterative ensemble smoother and a hybrid IES, also with Levenberg–Marquardt regularization. In both approaches, the parameters being updated are z and the three hyperparameters of the covariance function. Recall that z has the same dimension as m (450) and is defined in Eq. 1. The updates of z were localized in the IES approach using the Gaspari–Cohn correlation function with a taper range that was the same as the true correlation range. Localization was not used in the hybrid IES approach. An ensemble size of 200 was used for the IES, while a smaller ensemble ($N_e = 100$) was used for the hybrid IES approach. In both approaches, the starting value of the Levenberg–Marquardt parameter λ was selected based on the magnitude of the initial data mismatch (Chen and Oliver 2013). If the average squared data mismatch decreased in an iteration, the value of λ was decreased by a factor of 4. If the average square data mismatch increased, then λ was multiplied by a factor of 4. Iterations were stopped if λ increased in two successive iterations, or if the number of iterations exceeded 25, or if the magnitude of the reduction in data mismatch was too small.

Prior and posterior predictive distributions of water cut are compared with the observed water cut in Fig. 7 for five of the six producers. The prior predictive distribution (top row) "covers" the observations, although the prior distribution for producer 3 appears to be inconsistent with trends in the data. The posterior predictive distribution computed using the IES (middle row) has reduced spread compared to the prior
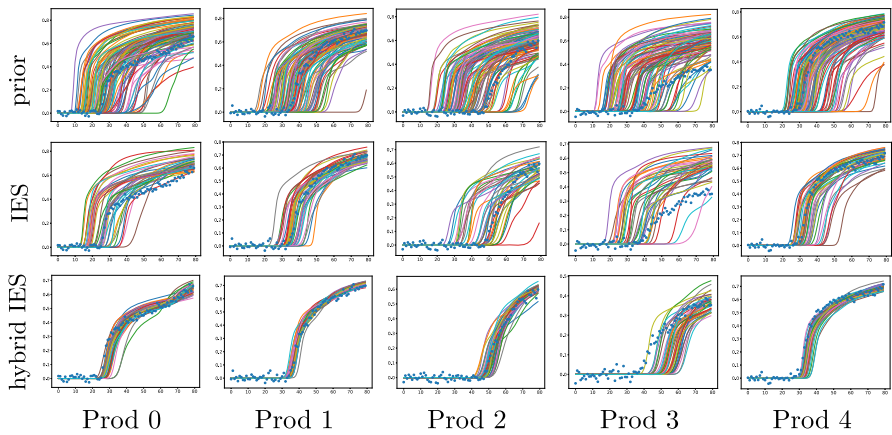
Fig. 7 Data assimilation with the IES and hybrid IES. Watercut versus time for 40 realizations from the prior and from the posterior for five producers. Blue dots are noisy observations of water cut
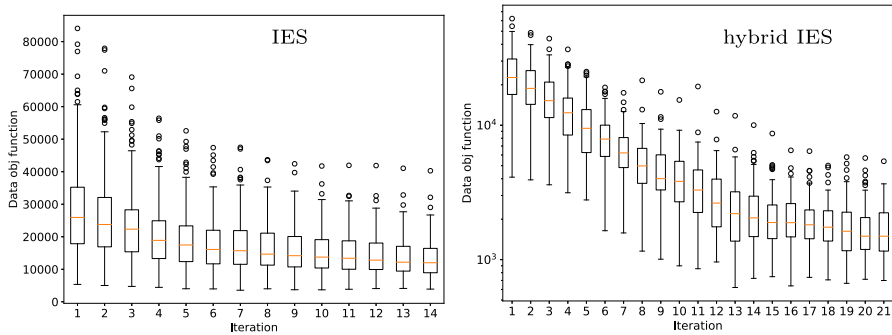

Fig. 8 Reduction in squared mismatch with observed (not perturbed) data. The expected value at convergence is 240. Note that different scales are used for the two plots

distribution, but the mismatch with actual observations (blue dots) is much larger than would be expected for conditional realizations. The hybrid IES (bottom row) generates a posteriori predictions that are generally consistent with both the observations and with the measurement error, with the exception that the match with data in producer 3 is poor.

The improved match to the data obtained using the hybrid IES approach is confirmed quantitatively in Fig. 8. In both subfigures, the spread of squared data mismatch values is shown as a box which represents the range containing the central 50% of the values. Note that the reduction in the squared data mismatch is slow for the IES and that the final mean value (13,000) is much greater than the expected value for a properly calibrated ensemble (240). The hybrid IES converges much faster and ends at a much smaller mean value of data mismatch (1,151). Although the final value is larger than expected, it is largely a result of the poor match to water cut observations in producer 3.
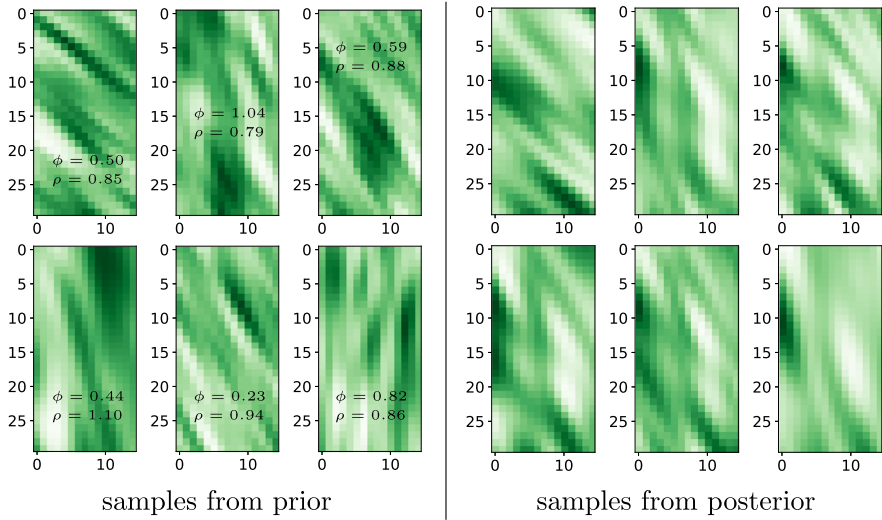
Fig. 9 The first six realizations of the log-permeability field from the prior (left) and the corresponding realizations from the posterior (right). Darker greens correspond to larger values of log-permeability
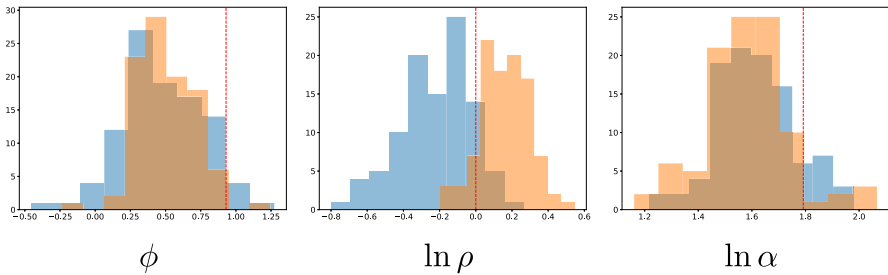

Fig. 10 Modification of distributions of hyperparameters after assimilation of flow data. Blue is prior distribution. Orange is posterior distribution. Red dashed lines show the values used in the data-generating model

Figure 9 shows the first six ensemble members from the prior and the corresponding ensemble members from the posterior after data assimilation in the hierarchical model using the hybrid IES method. Six realizations are not sufficient to illustrate the results of the data assimilation, but by comparing corresponding realizations, it appears that the spread in orientation of the anisotropy is reduced and that the principal range is largely maintained. The effect of data assimilation on the hyperparameters is illustrated more quantitatively in Fig. 10 which compares the prior distribution for the three hyperparameters (blue), with the posterior marginal distributions (orange), and the values used in the data-generating model (dashed red line). The posterior distribution of orientations (ϕ) is narrower than the prior distribution, but the mean is nearly unchanged from the prior (Fig. 10 (left)). The range parameter (ρ) shows the largest influence of data assimilation. Short correlation ranges have been eliminated from the posterior, while the mean of the posterior distribution has been shifted to a value that
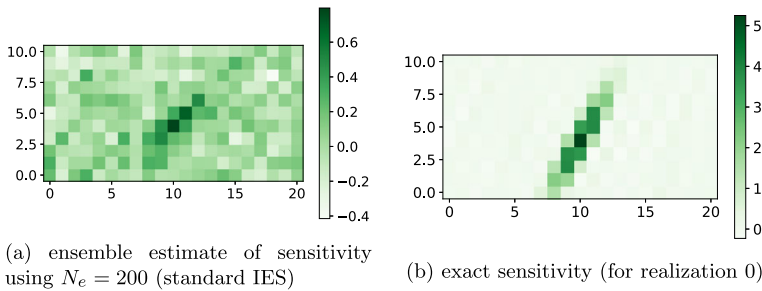
(a) ensemble estimate of sensitivity using $N_e = 200$ (standard IES)

(b) exact sensitivity (for realization 0)

Fig. 11 Sensitivity of observation of m at (10,4) to values of z at every cell. Note difference in color scales

is larger than both the prior mean and the value that was used in the data-generating model (Fig. 10 (center)). If the relationship of data to model hyperparameters was linear, one would expect the posterior mean to lie between the prior and the data-generating value, which is not what is seen here. The mean value of the ratio of correlation ranges in the principal directions (α) has shifted to a slightly smaller value after data assimilation (Fig. 10 (right)), meaning that the mean correlation range in the second principal direction has increased slightly, but less than the range in the first principal direction.

4.4 Discussion: Benefit of the Hybrid Approach

In Sect. 4.3, the IES method failed to assimilate data properly in a hierarchical model for which the anisotropy in the covariance was uncertain. The problem of estimating the permeability field from flow observations in this case is clearly nonlinear, but IES often works well for highly nonlinear problems, so that does not appear to be the primary explanation. In this type of hierarchical problem, it is relatively straightforward to see why a standard IES may not converge and why the hybrid IES works very well. To illustrate the relative performance of the two algorithms, consider a simpler two-dimensional hierarchical problem in which an ensemble of realizations of m are updated from partial noisy observations of m. The prior covariance in this example has uncertain orientation and ranges, just as in the two-dimensional flow problem but with simpler observation and sensitivity. It is instructive in this case to look at the ensemble estimate of the sensitivity, $\nabla_z m$, between observation of m at a location near the center of the domain and z for an ensemble with 200 realizations (Fig. 11a) and compare it to the exact sensitivity for one particular realization (Fig. 11b).

The ensemble estimate of sensitivity (Fig. 11a) suffers from several problems. The first is that although the ensemble size (200) is larger than typically used for data assimilation, the magnitudes of the spurious correlations are comparable to the estimated magnitude at the observation location. Also, the magnitude of the ensemble estimate of the sensitivity at the observation location is almost an order of magnitude smaller than the correct value. Finally, the region of significant sensitivity in the ensemble estimate is different from the region in the exact calculation. Increasing ensemble size would reduce the spurious correlations and improve the magnitude of sensitivity at the
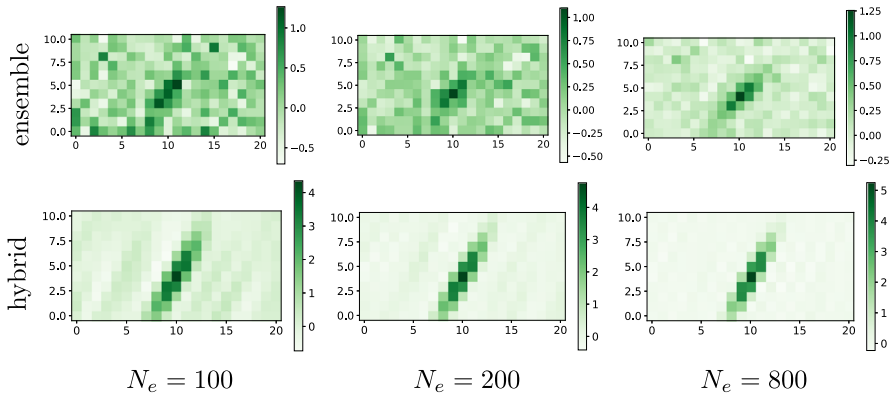
Fig. 12 Comparison of purely ensemble-based estimates of the cross-covariance of observation at (10,4) to the values of z (top row) to hybrid estimates of cross-covariance (bottom row)

peak, but the region of sensitivity would not improve because the average orientation is different from the orientation in realization 0.

The effect of ensemble size on the estimation of the cross-covariance between the observation of m at (10,4) and values of z at every cell is shown in Fig. 12. Because the covariance matrix for z is the identity matrix, the cross-covariance should be the same as the sensitivity matrix if both are computed accurately. It can be seen, in fact, that the hybrid IES estimate of the cross-covariance for large ensemble size (Fig. 12) is identical to the exact sensitivity shown in Fig. 11. Importantly, the estimates from the hybrid IES method with $N_e = 100$ is much better than the standard IES estimate with ensemble size $N_e = 800$. And again, the bias in the cross-covariance from the ensemble is large, indicating that the global estimate is not representative of the local value in this highly nonlinear problem.

The reasons for the remarkably good results for the hybrid IES in this example are twofold. First, $C_z = I$ for the non-centered hierarchical parameterization. The hybrid IES method uses the exact C_z, while the traditional IES uses a low-rank approximation of C_z, which suffers from spurious correlations. Second, the hybrid method uses the chain rule to compute part of the sensitivity analytically. This eliminates the bias that appears when the mean hyperparameters do not match the hyperparameters of the individual realization.

4.4.1 Summary

In an attempt to increase the robustness of data assimilation against model misspecification, a hierarchical Gaussian model was introduced for history matching of flow data. The non-centered parameterization used here is relatively simple, but allows uncertainty in several important parameters of the prior covariance that are typically fixed during history matching. Because we assumed that the prior covariance was stationary, the total number of uncertain parameters was only slightly larger than the number of uncertain parameters in standard data assimilation, but the effective number

of uncertain parameters is much larger, as the grid-based variables are independent and identically distributed (iid) in the non-centered hierarchical parameterization.

Unfortunately, updating model parameters using a standard ensemble Kalman-like method was not effect when the number of model cells was relatively large compared to the size of the ensemble, even though it is common in data assimilation for the number of model parameters to be much larger than the ensemble size. There appears to be two reasons for the increased difficulty with the use of data assimilation for the hierarchical model. The first is that the effect of spurious correlations is much larger with the non-centered parameterization, and localization based on the range of the prior covariance was not useful when the grid-based parameters were iid. Secondly, the addition of hyperparameters makes the problem more nonlinear so that the use of average sensitivities (or Jacobians) computed from the ensemble do not represent local sensitivities accurately. Although the problem of spurious correlations could conceptually be solved by increasing the ensemble size, that does not solve the problem of incorrect sensitivities.

To solve these problems while retaining the advantages of the iterative ensemble smoothers (no need to derive the adjoint system and no need to store very large matrices), a hybrid IES was developed in which the "difficult sensitivities" such as the sensitivities of water production rate to permeability are still computed approximately using an ensemble of model parameters and the ensemble of model predictions. The "easy sensitivities" such as the sensitivity of the permeability to the hyperparameters or to the latent iid variable z are computed analytically. In a two-dimensional problem with uncertainty in the orientation of the principal axes of the covariance, hybrid estimates of G and CG^T were much less noisy than estimates computed directly from the ensemble. Additionally, the hybrid estimates of sensitivities G and cross-covariances CG^T are specific to a realization and will ill allow convergence to the correct value as ensemble size increases for linear observation operators. Finally, since each realization has its own Kalman gain matrix, the methodology is less limited by Gaussian assumptions in the standard IES approaches. The hybrid method provides greater robustness against nonlinearity if some of the nonlinearity is in the relationship between z and m, as it is for the hierarchical model and also for the truncated plurigaussian model (Oliver and Chen 2018).

In this paper, the hybrid IES method was tested on a one-dimensional linear problem and on a two-dimensional-flow problem with hierarchical parameterizations. In the one-dimensional problem, the hierarchical parameterization provided uncertainty in the range and variance of the prior covariance. The number of parameters was small compared to the ensemble size in this case, so all methods performed similarly, although the standard IES tended to produce correlation ranges that were too short. In the two-dimensional flow problem, there was uncertainty in the orientation of the anisotropic covariance and in the correlation range in the two principle directions. The data assimilation using hybrid IES gave results that were much better than the IES. In this problem, the number of model parameters was much greater than the number of ensemble members. The hierarchical model with hybrid IES data assimilation provided updated models with data mismatch magnitudes that were close to the expected values in all wells except one.

The comparison of data assimilation with a hierarchical prior to data assimilation with a "good" and a "bad" prior showed results with the hierarchical parameterization were slightly worse than results with the good prior and much better than results with the bad prior. But the good results for the hierarchical method were only obtained in conjunction with the use of the hybrid IES method of data assimilation. When the standard IES was used with the hierarchical parameterization, the mismatch to data was nearly as large as the prior mismatch.

Declaration

Appendix A Gauss–Newton update with circular hyperparameters

In Sect. 3.3, the hybrid IES data assimilation algorithm is described in somewhat general terms, assuming that the prior densities for the hyperparameters of the covariance were Gaussian. Here, additional background is provided for the case of geometric anisotropy in which the prior for the orientation is Gauss–von Mises. The physical model parameters m for the two-dimensional flow problem in Sect. 4.2 are determined from the mean m_{pr} (assumed known), the uncertain covariance function C_m (which is a function of correlation range ρ, ratio of ranges α, and orientation of anisotropy, ϕ), and from the stochastic variable z

$$m = m_{\text{pr}} + L(\rho, \alpha, \phi)z.$$

Here, $LL^{\text{T}} = C_m$, $\rho \in (0, \infty)$ is the range parameter of the covariance, $\alpha \in (0, \infty)$ is the stretch parameter, and $\phi \in [-\pi/2, \pi/2]$ is the orientation of the principal axis of the covariance.

Because ρ and α are required to be positive, I have chosen log-normal prior distributions for $u_1 \equiv \ln \rho$,

$$p(u_1) \propto \exp\left(-\frac{1}{2}\frac{(u_1 - \mu_{u_1})^2}{\sigma_{u_1}^2}\right),$$

and for $u_2 \equiv \ln \alpha$,

$$p(u_2) \propto \exp\left(-\frac{1}{2}\frac{(u_2 - \mu_{u_2})^2}{\sigma_{u_2}^2}\right).$$

The prior distribution for the orientation ϕ cannot be Gaussian as it is defined on the interval $[-\pi/2, \pi/2]$. A Gauss–von Mises (GVM) distribution with parameters μ and κ is a reasonable choice for the probability density function of a random variable θ on the circle

$$g(\phi; \mu, \kappa) = \frac{1}{\pi I_0(\kappa)} \exp[\kappa \cos 2(\phi - \mu_\phi)]. \tag{A1}$$

where $I_0(\kappa)$ is a modified Bessel function of the first kind of order 0 (Jammalamadaka and Sengupta 2001). Note that the prior for ϕ is the only prior that is not Gaussian (although for large values of κ it is approximately Gaussian or a mixture of Gaussians). Finally, the prior for z is

$$p(z) \propto \exp\left(-\frac{1}{2}z^\mathrm{T} z\right),$$

and hence the posterior pdf for ρ, α, ϕ, z can be written as

$$p(z, u, \phi \mid d) \propto \exp\left(-\frac{1}{2}(d - g(m(u, \phi, z)))^\mathrm{T} C_d^{-1}(d - g(m(u, \phi, z)))\right)$$
$$\times \exp\left(-\frac{1}{2}(u - \mu_u)^\mathrm{T} C_u^{-1}(u - \mu_u) - \frac{1}{2}z^\mathrm{T} z\right) \exp[\kappa \cos 2(\phi - \mu_\phi)],$$

where I have defined $u = [u_1, u_2]$, and $C_u = \begin{bmatrix} \sigma_{u_1}^2 & 0 \\ 0 & \sigma_{u_1}^2 \end{bmatrix}$. Computation of the maximum a posteriori point can be found from the maximizer of the negative logarithm of the posterior pdf,

$$S(z, u, \phi) = \frac{1}{2}(d - g(m(u, \phi, z)))^\mathrm{T} C_d^{-1}(d - g(m(u, \phi, z)))$$
$$+ \frac{1}{2}u^\mathrm{T} C_u^{-1} u + \frac{1}{2}z^\mathrm{T} z - \kappa \cos 2(\phi - \mu_\phi).$$

For an RML-like approximation to sampling, minimizers of

$$
\begin{aligned}
S^*(z, u, \phi) &= \frac{1}{2}(d - (g(m) + \epsilon^*))^\mathrm{T} C_d^{-1}(d - (g(m) + \epsilon^*)) \\
&+ \frac{1}{2}(u - u^*)^\mathrm{T} C_u^{-1}(u - u^*) \\
&+ \frac{1}{2}(z - z^*)^\mathrm{T}(z - z^*) - \kappa \cos 2(\phi - \phi^*)
\end{aligned}
$$

are computed, where $\epsilon^* \sim N[0, C_d]$ is a sample of the observation error, $u^* \sim N[\mu_u, C_u]$, etc.

An efficient way to find a minimizer is to solve for $\nabla S = 0$, where

$$
\nabla S = -G^\mathrm{T} C_d^{-1}(d - (g(m) + \epsilon^*)) + \begin{bmatrix} (z - z^*) \\ C_u^{-1}(u - u^*) \\ 2\kappa \sin 2(\phi - \phi^*) \end{bmatrix}, \qquad \text{(A2)}
$$

where $G = D_{z,u,\phi} g$ instead of the sensitivity with respect to m, as is more typical. The Gauss–Newton approximation of the Hessian is

$$
\nabla(\nabla S)^\mathrm{T} \approx G^\mathrm{T} C_d^{-1} G + \begin{bmatrix} I_z & 0 & 0 \\ 0 & C_u^{-1} & 0 \\ 0 & 0 & 4\kappa \cos 2(\phi - \phi^*) \end{bmatrix},
$$

but

$$
\nabla(\nabla S)^\mathrm{T} \approx G^\mathrm{T} C_d^{-1} G + \begin{bmatrix} I_z & 0 & 0 \\ 0 & C_u^{-1} & 0 \\ 0 & 0 & 4\kappa \end{bmatrix}
$$

is used for data assimilation, as it is always positive definite. Define the prior covariance matrix for the model parameters as

$$
\mathbf{C}_x = \begin{bmatrix} I_z & 0 & 0 \\ 0 & C_u & 0 \\ 0 & 0 & 1/(4\kappa) \end{bmatrix}.
$$

With this definition of the prior covariance, Eq. A2 can be rewritten as

$$
\nabla S = G^\mathrm{T} C_d^{-1}(g(m) + \epsilon^* - d) + \mathbf{C}_x^{-1} \begin{bmatrix} (z - z^*) \\ (u - u^*) \\ \frac{1}{2} \sin 2(\phi - \phi^*) \end{bmatrix}.
$$

After substitution, the ℓth Levenberg–Marquardt update for x can then be written as

$$
\delta x_{\ell+1} = -\frac{1}{(1+\lambda_\ell)} \begin{bmatrix} (z_\ell - z^*) \\ (u_\ell - u^*) \\ \frac{1}{2}\sin 2(\phi_\ell - \phi^*) \end{bmatrix}
$$

$$
- C_x G_\ell^{\mathrm{T}} \left((1+\lambda_\ell) C_d + G_\ell C_x G_\ell^{\mathrm{T}} \right)^{-1} \tag{A3}
$$

$$
\times \left(g(m_\ell) + \epsilon^* - d - \frac{1}{(1+\lambda)} G_\ell \begin{bmatrix} (z_\ell - z^*) \\ (u_\ell - u^*) \\ \frac{1}{2}\sin 2(\phi_\ell - \phi^*) \end{bmatrix} \right)
$$

which corresponds to Eq. 7, the only difference being the explicit treatment of uncertainty in orientation as a Gauss–von Mises distribution.

Appendix B Sensitivity of m to parameters: geometric anisotropy

In the hybrid method, derivatives of the square root L of the prior covariance matrix with respect to the hyperparameters are required. In order to obtain numerical results, attention is restricted to the case of geometric anisotropy (Chilès and Delfiner 1999, sec. 2.5.2), in which the anisotropy is defined by rotation and stretching of the coordinate system; assuming that in the new coordinate system, the covariance can be written as $\mathrm{cov}(x, x') = \mathrm{cov}(r)$, where r is the distance between points x and x' after transformation.

B.0.1 The coordinate transformation for geometric anisotropy

The new coordinates x^* are defined such that $x^* = Ax$ where

$$
A = \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}
$$

$$
= \begin{bmatrix} \cos\phi & \sin\phi \\ -\alpha\sin\phi & \alpha\cos\phi \end{bmatrix}.
$$

The distance (for covariance computation) is thus of the form $r = \sqrt{(x^*)^{\mathrm{T}} x^*} = \sqrt{x^{\mathrm{T}} H x}$ where

$$
H = \begin{bmatrix} \cos\phi & -\alpha\sin\phi \\ \sin\phi & \alpha\cos\phi \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \\ -\alpha\sin\phi & \alpha\cos\phi \end{bmatrix}
$$

$$
= \begin{bmatrix} \cos^2\phi + \alpha^2\sin^2\phi & (1-\alpha^2)\sin\phi\cos\phi \\ (1-\alpha^2)\sin\phi\cos\phi & \sin^2\phi + \alpha^2\cos^2\phi \end{bmatrix}. \tag{B4}
$$

Recall that $m = m^{\mathrm{pr}} + Lz$, where $C_x = LL^{\mathrm{T}}$, and

$$
M_x = \begin{bmatrix} L & \left(\frac{\partial}{\partial u_1} L\right)z & \left(\frac{\partial}{\partial u_2} L\right)z & \left(\frac{\partial}{\partial \phi} L\right)z \end{bmatrix}. \tag{B5}
$$

For the flow problem in Sect. 4.2, it was assumed that the covariance took the form

$$C(r) = \sigma^2 e^{-3r^2/\rho^2}. \tag{B6}$$

The convolution "square root" in two dimensions is

$$f(r) = \frac{2\sigma\sqrt{3}}{\rho\sqrt{\pi}} \exp\left(-\frac{6r^2}{\rho^2}\right).$$

The matrix square root, L, will simply be the discrete version of this functional form, and effects of boundaries will be ignored.

B.1 Derivative of f wrt ϕ

$$\begin{aligned}
\frac{\partial f}{\partial \phi} &= \frac{2\sigma\sqrt{3}}{\rho\sqrt{\pi}}\left(-\frac{6}{\rho^2}\right)\exp\left(-\frac{6r^2}{\rho^2}\right)\frac{\partial r^2}{\partial \phi} \\
&= -\frac{4\sigma\sqrt{27}}{\rho^3\sqrt{\pi}}\exp\left(-\frac{6r^2}{\rho^2}\right)\left(x^{\mathrm{T}}\frac{\partial H}{\partial \phi}x\right) \\
&= -\frac{6}{\rho^2}\left(x^{\mathrm{T}}\frac{\partial H}{\partial \phi}x\right)f(r).
\end{aligned}$$

From Eq. B4, it is straightforward to show that

$$\frac{\partial H}{\partial \phi} = (1-\alpha^2)\begin{bmatrix} -\sin 2\phi & \cos 2\phi \\ \cos 2\phi & \sin 2\phi \end{bmatrix}.$$

(Note that if $\alpha = 1$, the covariance is isotropic and there is no sensitivity to orientation.)

B.2 Derivative of f wrt α

$$\begin{aligned}
\frac{\partial f}{\partial \alpha} &= \frac{2\sigma\sqrt{3}}{\rho\sqrt{\pi}}\left(-\frac{6}{\rho^2}\right)\exp\left(-\frac{6r^2}{\rho^2}\right)\frac{\partial r^2}{\partial \alpha} \\
&= -\frac{4\sigma\sqrt{27}}{\rho^3\sqrt{\pi}}\exp\left(-\frac{6r^2}{\rho^2}\right)\left(x^{\mathrm{T}}\frac{\partial H}{\partial \alpha}x\right),
\end{aligned} \tag{B7}$$

where

$$\frac{\partial H}{\partial \alpha} = \alpha\begin{bmatrix} 1-\cos 2\phi & -\sin 2\phi \\ -\sin 2\phi & 1+\cos 2\phi \end{bmatrix}.$$

B.3 Derivative of f wrt ρ

$$\frac{\partial f}{\partial \rho} = \left(\frac{12r^2}{\rho^2} - 1 \right) \left(\frac{2\sigma\sqrt{3}}{\rho^2\sqrt{\pi}} \right) \exp\left(-\frac{6r^2}{\rho^2} \right).$$

References

Ba Y, de Wiljes J, Oliver DS, Reich S (2022) Randomized maximum likelihood based posterior sampling. Comput Geosci 26(1):217–239. https://doi.org/10.1007/s10596-021-10100-y

Burgers G, van Leeuwen PJ, Evensen G (1998) Analysis scheme in the ensemble Kalman filter. Mon Weather Rev 126(6):1719–1724. https://doi.org/10.1175/1520-0493(1998)1261719:ASITEK2.0.CO;2

Chada NK, Iglesias MA, Roininen L, Stuart AM (2018) Parameterizations for ensemble Kalman inversion. Inverse Probab 34(5):055009. https://doi.org/10.1088/1361-6420/aab6d9

Chen Y, Oliver DS (2010) Parameterization techniques to improve mass conservation and data assimilation for ensemble Kalman filter (SPE 133560). In: SPE Western Regional Meeting, 27–29 May 2010, Anaheim, California, USA. https://doi.org/10.2118/133560-MS

Chen Y, Oliver DS (2012) Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math Geosci 44(1):1–26. https://doi.org/10.1007/s11004-011-9376-z

Chen Y, Oliver DS (2013) Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. Comput Geosci 17(4):689–703. https://doi.org/10.1007/s10596-013-9351-5

Chilès J-P, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York

Dunlop MM, Helin T, Stuart AM (2020) Hyperparameter estimation in Bayesian MAP estimation: parameterizations and consistency. SMAI J. Comput. Math. 6:69–100. https://doi.org/10.5802/smai-jcm.62

Emerick AA (2016) Analysis of the performance of ensemble-based assimilation of production and seismic data. J Pet Sci Eng 139:219–239. https://doi.org/10.1016/j.petrol.2016.01.029

Emerick AA, Reynolds AC (2013) Ensemble smoother with multiple data assimilation. Comput Geosci 55:3–15. https://doi.org/10.1016/j.cageo.2012.03.011

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 99(C5):10143–10162. https://doi.org/10.1029/94JC00572

Fox C, Norton RA (2016) Fast sampling in a linear-Gaussian inverse problem. SIAM/ASA J Uncertain Quantif 4(1):1191–1218. https://doi.org/10.1137/15M1029527

Gaspari G, Cohn SE (1999) Construction of correlation functions in two and three dimensions. Q J R Meteorol Soc 125(554):723–757

Gneiting T, Kleiber W, Schlather M (2010) Matérn cross-covariance functions for multivariate random fields. J Am Stat Assoc 105(491):1167–1177. https://doi.org/10.1198/jasa.2010.tm09420

Houtekamer PL, Mitchell HL (1998) Data assimilation using an ensemble Kalman filter technique. Mon Weather Rev 126(3):796–811. https://doi.org/10.1175/1520-0493(1998)1260796:DAUAEK2.0.CO;2

Jammalamadaka SR, Sengupta A (2001) Topics in circular statistics. Multivariate analysis, 5th edn. World Scientific Publishing, Singapore

Kitanidis PK (1995) Quasi-linear geostatistical theory for inversing. Water Resour Res 31(10):2411–2419. https://doi.org/10.1029/95WR01945

Li G, Han M, Banerjee R, Reynolds AC (2010) Integration of well-test pressure data into heterogeneous geological reservoir models. SPE Reserv Eval Eng 13(03):496–508. https://doi.org/10.2118/124055-PA

Lindeberg T (1990) Scale-space for discrete signals. IEEE Trans Pattern Anal Mach Intell 12(3):234–254. https://doi.org/10.1109/34.49051

Malinverno A, Briggs VA (2004) Expanded uncertainty quantification in inverse problems: hierarchical Bayes and empirical Bayes. Geophysics 69(4):1005–1016. https://doi.org/10.1190/1.1778243

Moore C, Doherty J (2005) Role of the calibration process in reducing model predictive error. Water Resour Res 41(5):W05020. https://doi.org/10.1029/2004WR003501

Myrseth I, Omre H (2010) Hierarchical ensemble Kalman filter. SPE J 15(2):569–580. https://doi.org/10.2118/125851-PA

Oliver DS (1995) Moving averages for Gaussian simulation in two and three dimensions. Math Geol 27(8):939–960. https://doi.org/10.1007/BF02091660

Oliver DS, Alfonzo M (2018) Calibration of imperfect models to biased observations. Comput Geosci 22(1):145–161. https://doi.org/10.1007/s10596-017-9678-4

Oliver DS, Chen Y (2018) Data assimilation in truncated plurigaussian models: impact of the truncation map. Math Geosci 50(8):867–893. https://doi.org/10.1007/s11004-018-9753-y

Oliver DS, He N, Reynolds AC (1996) Conditioning permeability fields to pressure data. In: Proceedings of the European conference on the mathematics of oil recovery, V, pp 1–11 . https://doi.org/10.3997/2214-4609.201406884

Oliver DS, Reynolds AC, Liu N (2008) Inverse theory for petroleum reservoir characterization and history matching. Cambridge University Press, Cambridge

Papaspiliopoulos O, Roberts GO, Sköld M (2007) A general framework for the parametrization of hierarchical models. Stat Sci 22(1):59–73. https://doi.org/10.1214/088342307000000014

Park H, Scheidt C, Fenwick D, Boucher A, Caers J (2013) History matching and uncertainty quantification of facies models with multiple geological interpretations. Comput Geosci 17(4):609–621. https://doi.org/10.1007/s10596-013-9343-5

Reich S (2011) A dynamical systems framework for intermittent data assimilation. BIT Numer Math 51(1):235–249. https://doi.org/10.1007/s10543-010-0302-4

Roininen L, Girolami M, Lasanen S, Markkanen M (2019) Hyperpriors for Matérn fields with applications in Bayesian inversion. Inverse Probl Imaging 13(1):1–29. https://doi.org/10.3934/ipi.2019001

Rue H, Held L (2005) Gaussian Markov random fields: theory and applications. Chapman and Hall/CRC, New York. https://doi.org/10.1201/9780203492024

Rue H, Martino S (2007) Approximate Bayesian inference for hierarchical Gaussian Markov random field models. J Stat Plan Inference 137(10):3177–3192. https://doi.org/10.1016/j.jspi.2006.07.016

Scales JA, Tenorio L (2001) Prior information and uncertainty in inverse problems. Geophysics 66(2):389–397. https://doi.org/10.1190/1.1444930

Stojkovic I, Jelisavcic V, Milutinovic V, Obradovic Z (2017) Fast sparse Gaussian Markov random fields learning based on Cholesky factorization. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence (IJCAI-17), pp 2758–2764 . https://doi.org/10.24963/ijcai.2017/384

Tsyrulnikov M, Rakitko A (2017) A hierarchical Bayes ensemble Kalman filter. Phys D 338:1–16. https://doi.org/10.1016/j.physd.2016.07.009

Zhang Y, Oliver DS (2011) History matching using a multiscale stochastic model with the ensemble Kalman filter: a field case study. SPE J 16(2):307–317. https://doi.org/10.2118/118879-PA

Zhou Q, Liu W, Li J, Marzouk YM (2018) An approximate empirical Bayesian method for large-scale linear-Gaussian inverse problems. Inverse Probl 34(9):095001. https://doi.org/10.1088/1361-6420/aac287