

# Ensemble Kalman Filtering in a Bayesian Regression Framework

Jon Sætrum and Henning Omre

**Abstract** Bayesian methods have in recent years become popular when considering problems in geosciences, such as sequential data assimilation in high dimensions. Ensemble based Monte Carlo methods, such as the Ensemble Kalman Filter (EnKF), are attractive to use because they are easy to implement and computationally fast. However, high computational demands will often restrict the ensemble size. Problems resulting from estimation uncertainty and dependencies between the ensemble members can therefore occur. As a result, the traditional EnKF updating schemes can lead to unreliable predictions with a severe underestimation of the prediction interval. In this paper, we introduce alternative EnKF updating schemes based on Bayesian regression techniques. The main idea is to replace the traditional plug-in estimate of the Kalman gain matrix with individual realisations from a matrix-variate distribution for each updated ensemble member. We evaluate the performance of the suggested schemes through simulation on synthetic case studies. The results reveal that we can dramatically improve the accuracy of the forecast and predictions intervals, especially for small ensemble sizes.

## 1 Introduction

The Ensemble Kalman Filter (EnKF) is a Monte Carlo method for solving non-linear spatiotemporal inverse problems in high dimensions [Evensen, 2009]. Applications include numerical weather prediction [Houtekamer and Mitchell, 2001], oceanography [Keppenne and Rienecker, 2003], hydrology [Moradkhani et al., 2005] and petroleum reservoir characterisation [Aanonsen et al., 2009]. The EnKF is based on the traditional Kalman Filter (KF) [Kalman, 1960], which provides

---

Jon Sætrum

Statoil ASA, e-mail: jonsat@statoil.com,

Henning Omre

The Norwegian University of Science and Technology, e-mail: omre@math.ntnu.no

Ninth International Geostatistics Congress, Oslo, Norway, June 11. – 15., 2012

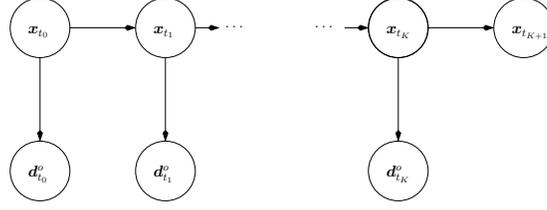
an analytical solution for the posterior probability density function (pdf) of interest, assuming linear system dynamics and linear, Gaussian assumptions, termed the Gauss-linear model.

Analytical tractability of the posterior pdf will be lost for non-linear system dynamics and non-Gaussian distributions. Techniques such as Markov chain Monte Carlo or importance sampling can therefore be used to generate realisations correctly from the posterior pdf [Doucet et al., 2000, Liu, 2001, West and Harrison, 1999]. High dimensional problems will, however, restrict the tractability of these methods because of high computational demands. Moreover, approximate solutions such as the extended KF [Gelb, 1974, Jazwinski, 1970], where the forward model is linearised, can lead to an unstable solution, in addition to potentially high computational demands [Evensen, 1992].

The EnKF solution to the spatiotemporal forecast problem is based on a Monte Carlo approach with sequential forecasting and assimilation of available observations. The updating step is based on a linearisation where the required unknown covariances are assessed from the Monte Carlo ensemble. From these estimates we can estimate the weights in the linearisation, referred to as the Kalman gain matrix. This entails that for the Gauss-linear model, the EnKF will be consistent with the KF as the ensemble size tends to infinity [Mardia et al., 1979]. When the ensemble size is finite, however, problems resulting from estimation uncertainty and ensemble collinearities are known to occur [Furrer and Bengtsson, 2007, Houtekamer and Mitchell, 1998, Sacher and Bartello, 2008]. As a consequence, the updated ensemble members will fail to correctly represent the statistical properties of the posterior distribution.

From classical multivariate statistics [Anderson, 2003b], it is well known that estimation of the unknown Kalman gain matrix can be equally formulated as a multivariate linear regression problem [Anderson, 2003a, Sætrom and Omre, 2011]. Hence, the EnKF updating scheme can be reformulated based on known regression techniques aiming at improving both the accuracy and variance estimates of the forecasts. Examples are shrinkage regression techniques for collinear data, where the unbiased least squares estimator is replaced by biased alternatives having improved predictive capabilities [Farrar and Glauber, 1967, Hastie et al., 2009]. Indeed, such an approach can lead to considerable improvements compared to the standard EnKF algorithm for small ensemble sizes [Sætrom and Omre, 2011]. However, the ensemble members will, similar to the standard EnKF, be coupled over time because the same Kalman gain estimate is used to update every ensemble member.

In Sætrom and Omre [2012], theoretical results for the bias and covariances in the forecast based on the classical updating scheme were presented, taking into account the uncertainty of the unknown Kalman gain matrix. These results reveal that using one common plug-in estimate of the Kalman gain matrix will lead to a positive correlation between the conditioned ensemble members, which necessarily will lead to an underestimation of the prediction variance. In this paper we therefore apply Bayesian regression techniques to reduce the ensemble coupling. In this approach, each ensemble member is updated based on a Kalman gain matrix independently generated from a matrix variate distribution, rather than using one common plug-



**Fig. 1** Stochastic Directed Acyclic Graph (DAG) of the model considered.

in estimate. We consider both conjugate and non-informative prior distributions, and an approximate dimension reducing scheme for high dimensional models is suggested. A synthetic example inspired by petroleum reservoir evaluation problems is used to empirically evaluate the performance of the suggested procedures.

## 2 Notation and problem formulation

Throughout this paper the notation  $\mathbf{x} \in \mathbb{R}^{n_x \times 1}$  will be used to denote that  $\mathbf{x}$  is an  $n_x$ -dimensional column vector in the real space and  $\mathbf{x}^T$  will denote its transpose. Similarly, a matrix  $\mathbf{A}$  in the real space containing  $a$  rows and  $b$  columns will be denoted by  $\mathbf{A} \in \mathbb{R}^{a \times b}$ . For simplicity, the same notation will be used for random vectors and matrices.

Consider the stochastic Directed Acyclic Graph (DAG) outlined in Figure 1. Here  $\mathbf{x}_{t_k} \in \mathbb{R}^{n_x \times 1}$  denotes the state of the unknown random vector of interest at time step  $k$  and time  $t_k$ , and similarly  $\mathbf{d}_{t_k}^o \in \mathbb{R}^{n_d \times 1}$  denotes the vector of observed data. For notational convenience, we will from now on drop the subscript  $t_k$ , and write  $\mathbf{x}_k$ ,  $\mathbf{d}_k^o$ . Moreover,  $\mathbf{x}$  and  $\mathbf{d}$  will be referred to as the state and observation vector respectively.

Let  $f(\mathbf{x}_0)$  denote the prior pdf of the state vector at the initial time step. Through the Markov properties of a stochastic DAG, we have conditional independence between  $\mathbf{x}_{k+1}$  and  $\mathbf{x}_l$ ,  $l = 0, \dots, k-1$  given  $\mathbf{x}_k$ , which implies

$$f(\mathbf{x}_{k+1} | \mathbf{x}_k, \dots, \mathbf{x}_0) = f(\mathbf{x}_{k+1} | \mathbf{x}_k).$$

In general assume that

$$\mathbf{x}_{k+1} = \omega(\mathbf{x}_k, \boldsymbol{\varepsilon}_{\mathbf{x}_k}), \quad k = 0, \dots, K, \quad (1)$$

where  $\omega : (\mathbb{R}^{n_x \times 1} \times \mathbb{R}^{n_x \times 1}) \rightarrow \mathbb{R}^{n_x \times 1}$  is a known, possibly non-linear forward function. Here  $\boldsymbol{\varepsilon}_{\mathbf{x}_k}$  represents random model errors and/or numerical errors in the forward model, assumed to follow a known probability distribution. This implies that we implicitly get a fully specified prior model of the unknown state vector through  $f(\mathbf{x}_0)$  and  $\omega(\cdot, \cdot)$ . The function:

$$\mathbf{d}_k^o = \zeta(\mathbf{x}_k, \boldsymbol{\varepsilon}_{\mathbf{d}_k}), \quad k = 0, \dots, K, \quad (2)$$

connecting the observed data  $\mathbf{d}_k^o$  to  $\mathbf{x}_k$ , where  $\zeta : (\mathbb{R}^{n_x \times 1} \times \mathbb{R}^{n_d \times 1}) \rightarrow \mathbb{R}^{n_d \times 1}$ , is a known, possibly non-linear, function and  $\boldsymbol{\varepsilon}_{\mathbf{d}_k}$  represents the likelihood model and observation errors, again assumed to follow a known pdf.

For notational convenience, let  $\mathbf{x}_k^c \sim f(\mathbf{x}_k | \mathbf{d}_{0:k}^o)$  and  $\mathbf{x}_{k+1}^u \sim f(\mathbf{x}_{k+1} | \mathbf{d}_{0:k}^o)$ , for  $k = 1, \dots, K$ , where we use  $\mathbf{d}_{0:l}^o$  to denote the sequence  $\mathbf{d}_0^o, \dots, \mathbf{d}_l^o$ . The objective in this model setting is to solve the spatiotemporal forecast problem of predicting  $\mathbf{x}_k$  given observations  $\mathbf{d}_{0:(k-1)}^o$ , for  $k = 1, \dots, K+1$ . Bayesian inversion provides a sequential solution to this problem. Through Bayes rule and the Markov properties of the prior model, we have:

$$\begin{aligned} f(\mathbf{x}_k | \mathbf{d}_{0:k}^o) &\propto f(\mathbf{x}_k | \mathbf{d}_{0:(k-1)}^o) f(\mathbf{d}_k^o | \mathbf{x}_k) \\ f(\mathbf{x}_{k+1} | \mathbf{d}_{0:k}^o) &= \int f(\mathbf{x}_{k+1} | \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{d}_{0:k}^o) d\mathbf{x}_k. \end{aligned} \quad (3)$$

Note that the conditional pdf  $f(\mathbf{x}_{k+1} | \mathbf{x}_k)$  and  $f(\mathbf{d}_k^o | \mathbf{x}_k)$  are implicitly defined through  $\omega(\mathbf{x}_k, \boldsymbol{\varepsilon}_{\mathbf{x}_k})$  and  $\zeta(\mathbf{x}_k, \boldsymbol{\varepsilon}_{\mathbf{d}_k})$  defined in Eq. (1) and (2) respectively.

In the general case, we only know the conditional distributions defined in Eq. (3) up to an unknown normalising constant. Computationally demanding techniques, such as Markov chain Monte Carlo (McMC) or rejection sampling, can then be used to generate realisations correctly from the posterior distribution [Doucet et al., 2000]. However, for spatiotemporal problems in high dimensions, such as petroleum reservoir evaluation and weather forecasting, these techniques are computationally prohibitive. This follows because even a single evaluation of  $\omega(\mathbf{x}_k, \boldsymbol{\varepsilon}_{\mathbf{x}_k})$ , which involves solving non-linear partial differential equations in dimensions of order  $10^6 - 10^9$ , can take several hours or even days. An approximate solution can be defined by assuming that  $\mathbf{x}_k^u$  and  $\mathbf{d}_k^o$  follow a distribution that ensures analytical tractability of  $f(\mathbf{x}_k | \mathbf{d}_{0:k}^o)$ , for example the multivariate Gaussian distribution. These model assumptions are equivalent to those made in the EnKF algorithm [Evensen, 2009], which we will consider next.

## 2.1 The Ensemble Kalman Filter

Let  $\mathbf{x}_0^{c(i)}$  be a realisation from the unspecified conditional distribution at the initial time step,  $f(\mathbf{x}_0 | \mathbf{d}_0^o)$ . Further let  $\mathbf{x}_k^{u(i)} = \omega(\mathbf{x}_{k-1}^{c(i)}, \boldsymbol{\varepsilon}_{\mathbf{x}_k}^{(i)})$  and  $\mathbf{d}_k^{(i)} = \zeta(\mathbf{x}_k^{u(i)}, \boldsymbol{\varepsilon}_{\mathbf{d}_k}^{(i)})$ , for  $k > 0$ , where we at each time step, as an approximation, assume that the joint state and data vector follows as the Gaussian distribution:

$$\begin{bmatrix} \mathbf{x}_k^{u(i)} \\ \mathbf{d}_k^{(i)} \end{bmatrix} \sim \text{Gauss}_{n_y} \left( \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}_k^u} \\ \boldsymbol{\mu}_{\mathbf{d}_k} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{x}_k^u} & \boldsymbol{\Sigma}_{\mathbf{x}_k^u, \mathbf{d}_k} \\ \boldsymbol{\Sigma}_{\mathbf{d}_k, \mathbf{x}_k^u} & \boldsymbol{\Sigma}_{\mathbf{d}_k} \end{bmatrix} \right), \quad (4)$$

with  $n_y = n_x + n_d$ . For notational convenience, we will from now on omit the subscript  $k$ , because the focus is on updating at a single time step.

Under the Gaussian assumption in Eq. (4), the posterior pdf  $f(\mathbf{x}_k | \mathbf{d}_{0:k}^o)$  is Gaussian with analytically obtainable mean:

$$\mu_{\mathbf{x}^u | \mathbf{d}} = \mu_{\mathbf{x}^u} + \Sigma_{\mathbf{x}^u, \mathbf{d}} \Sigma_{\mathbf{d}}^{-1} (\mathbf{d}^o - \mu_{\mathbf{d}}), \quad (5)$$

and covariance matrix:

$$\Sigma_{\mathbf{x}^u | \mathbf{d}} = \Sigma_{\mathbf{x}^u} - \Sigma_{\mathbf{x}^u, \mathbf{d}} \Sigma_{\mathbf{d}}^{-1} \Sigma_{\mathbf{d}, \mathbf{x}^u}, \quad (6)$$

where  $\mathbf{d}^o$  is the vector of observed data. Furthermore, it is straightforward to show that

$$\mathbf{x}^{c(i)} = \mathbf{x}^{u(i)} + \Sigma_{\mathbf{x}^u, \mathbf{d}} \Sigma_{\mathbf{d}}^{-1} (\mathbf{d}^o - \mathbf{d}^{(i)}), \quad (7)$$

is one realisation from the Gaussian posterior distribution with mean and covariance given in Eq. (5) and (6) [Mardia et al., 1979].

Eq. (7) involves two model parameters, namely  $\Sigma_{\mathbf{x}^u, \mathbf{d}}$  and  $\Sigma_{\mathbf{d}}$  forming the Kalman gain matrix,

$$\mathbf{K} = \Sigma_{\mathbf{x}^u, \mathbf{d}} \Sigma_{\mathbf{d}}^{-1} \in \mathbb{R}^{n_x \times n_d}. \quad (8)$$

For the Gauss-linear model,  $\mathbf{K}$  will be analytically tractable and given by the Kalman recursions [Furrer and Bengtsson, 2007, Kalman, 1960]. Analytical tractability is, however, lost in the general model setting we consider here. The EnKF solution to this problem is to use an ensemble of  $n_e$  realisations  $\{(\mathbf{x}^{u(1)}, \mathbf{d}^{(1)}), \dots, (\mathbf{x}^{u(n_e)}, \mathbf{d}^{(n_e)})\}$ , from which we can obtain empirical estimate of the unknown covariance matrices.

Let  $\mathbf{X} = [\mathbf{x}^{u(1)}, \dots, \mathbf{x}^{u(n_e)}] \in \mathbb{R}^{n_x \times n_e}$  and  $\mathbf{D} = [\mathbf{d}^{(1)}, \dots, \mathbf{d}^{(n_e)}] \in \mathbb{R}^{n_d \times n_e}$  denote the ensemble matrices for the state- and observation vector respectively. Thus, the expression

$$\hat{\mathbf{K}} = \hat{\Sigma}_{\mathbf{x}^u, \mathbf{d}} \hat{\Sigma}_{\mathbf{d}}^{-1} = \mathbf{X} \mathbf{H}_{n_e} \mathbf{D}^T (\mathbf{D} \mathbf{H}_{n_e} \mathbf{D}^T)^{-1}, \quad (9)$$

defines the classical estimate of the Kalman gain matrix in a general setting [Anderson, 2003b], where  $\mathbf{H}_{n_e} = \mathbf{I}_{n_e} - 1/n_e \mathbf{1}_{n_e,1} \mathbf{1}_{n_e,1}^T \in \mathbb{R}^{n_e \times n_e}$  is the idempotent centring matrix. Here  $\mathbf{I}_{n_e}$  is the  $n_e$ -dimensional identity matrix and  $\mathbf{1}_{n_e,1}$  is a  $n_e$ -dimensional vector where all the entries are equal to one.

Replacing the unknown Kalman gain matrix in Eq. (7), with the empirical estimate in Eq. (9) defines the EnKF updating scheme in a general form:

$$\mathbf{x}^{c(i)} = \mathbf{x}^{u(i)} + \hat{\mathbf{K}} (\mathbf{d}^o - \mathbf{d}^{(i)}), \text{ for } i = 1, \dots, n_e. \quad (10)$$

We will from now on refer to the replacement of the unknown Kalman gain matrix with the empirical estimate in Eq. (9) as the Full EnKF. Note that we here motivate the EnKF updating scheme by a Gaussian prior model assumption. However, because the conditioned state vector,  $\mathbf{x}^{c(i)}$ , at each updating step is given as a linear combination of  $\mathbf{x}^{u(i)}$ ,  $\mathbf{d}^o$  and  $\mathbf{d}^{(i)}$ , with the model parameters only involved in the Kalman gain matrix, we can still capture non-Gaussian properties present in the general posterior distribution at the initial time step,  $f(\mathbf{x}_0 | \mathbf{d}_0^o)$ , at the later time steps

[Evensen, 2009]. Note also that the Full EnKF scheme presented here is slightly different from the schemes presented in the EnKF literature, where the likelihood is either Gauss-linear, or includes an additive Gaussian noise term [Burgers et al., 1998].

Under the Gauss-linear model, we have convergence in distribution of the EnKF updating scheme based on  $\hat{\mathbf{K}}$  towards the classical KF scheme as  $n_e \rightarrow \infty$  [Furrer and Bengtsson, 2007]. However, for finite ensemble sizes the Full EnKF updating scheme will produce a conditional ensemble which underrepresents the variability of  $\mathbf{x}^c$  [Furrer and Bengtsson, 2007, van Leeuwen, 1999, Sacher and Bartello, 2008]. In Sætrom and Omre [2012], known results from classical linear regression theory were used to establish theoretical results for the statistical properties of the EnKF scheme. The results revealed a critical weakness with the EnKF: Using the same estimated Kalman gain matrix to update all ensemble members, leads to correlated ensemble members after one updating step. Consequently, by using the classical covariance matrix estimates based on the updated ensemble, we underestimate the posterior covariance matrix. More troubling, however, is that we amplify the ensemble coupling as we perform sequential data assimilation. Hence, it is not surprising that the updated ensemble can ultimately collapse into one single realisation when applying the EnKF updating scheme.

Myrseth and Omre [2010a], suggested a hierarchical Bayesian approach, termed the Hierarchical EnKF (HEnKF), as an alternative to using empirical estimates of the covariance matrix. The HEnKF replace the estimated covariance matrices with realisations from the analytically tractable posterior distribution, assuming conjugate prior distributions for the mean and covariance in Eq. (4). Hence, the coupling between the updated ensemble will be reduced, leading to an improved representation of the prediction uncertainty. In the current paper, we have extended the HEnKF to a more general setting, using results known from Bayesian regression [Box and Tiao, 1992, Press, 1982]. This entails sampling the Kalman gain matrix directly for each ensemble member through analytically tractable matrix variate pdfs.

### 3 Bayesian regression

We will now consider the posterior distribution of the Kalman gain matrix,  $\mathbf{K}$ , through a hierarchical Bayesian approach using both a natural conjugate prior, similar to Myrseth and Omre [2010a], and a non-informative Jeffreys' prior [Press, 1982].

#### 3.1 Informative prior distribution

Consider the prior model assumption in Eq. (4), and let in addition  $[\mu_{\mathbf{y}} | \Sigma_{\mathbf{y}}] \sim \text{Gauss}_{n_y}(\eta_{\mathbf{y}}, \xi^{-1} \Sigma_{\mathbf{y}})$  and  $\Sigma_{\mathbf{y}} \sim W_{n_y}^{-1}(\Psi_{\mathbf{y}}^{-1}, \nu)$ , where  $\mathbf{y} = [\mathbf{x}^u; \mathbf{d}] \in \mathbb{R}^{n_y \times 1}$  is the

stacked state and data vector. Here,  $W_{(\cdot)}^{-1}(\cdot)$  denotes the inverted Wishart distribution [Gupta and Nagar, 2000], and  $\eta_{\mathbf{y}}$ ,  $\Psi_{\mathbf{y}}$ ,  $\xi$  and  $\mathbf{v}$  are known prior hyperparameters. Let  $\mathbf{Y} = [\mathbf{X}; \mathbf{D}] \in \mathbb{R}^{n_y \times n_e}$  denote the ensemble matrix for the stacked state- and observation vector,  $\mathbf{y}$ , then [Anderson, 2003b]:

$$\begin{aligned} [\mu_{\mathbf{y}} | \Sigma_{\mathbf{y}}, \mathbf{Y}] &\sim \text{Gauss}_{n_y} \left( \eta_{\mathbf{y}}^c, \frac{1}{\xi^c} \Sigma_{\mathbf{y}} \right) \\ [\Sigma_{\mathbf{y}} | \mathbf{Y}] &\sim W_{n_y}^{-1}(\Psi_{\mathbf{y}}^{c-1}, \mathbf{v}^c), \end{aligned}$$

with

$$\begin{aligned} \eta_{\mathbf{y}}^c &= \begin{bmatrix} \eta_{\mathbf{x}^u}^c \\ \eta_{\mathbf{d}}^c \end{bmatrix} = \frac{n_e \hat{\mu}_{\mathbf{y}} + \xi \eta_{\mathbf{y}}}{\xi^c} \\ \Psi_{\mathbf{y}}^c &= \begin{bmatrix} \Psi_{\mathbf{x}^u}^c & \Psi_{\mathbf{x}^u, \mathbf{d}}^c \\ \Psi_{\mathbf{d}, \mathbf{x}^u}^c & \Psi_{\mathbf{d}}^c \end{bmatrix} = \Psi_{\mathbf{y}} + (n_e - 1) \hat{\Sigma}_{\mathbf{y}} + \frac{n_e \xi}{\xi^c} (\hat{\mu}_{\mathbf{y}} - \eta_{\mathbf{y}})(\hat{\mu}_{\mathbf{y}} - \eta_{\mathbf{y}})^T, \end{aligned} \quad (11)$$

$\mathbf{v}^c = \mathbf{v} + n_e$  and  $\xi^c = \xi + n_e$ . Here,  $\hat{\mu}_{\mathbf{y}}$  is the average and  $\hat{\Sigma}_{\mathbf{y}}$  is the traditional unbiased estimate of the covariance matrix  $\Sigma_{\mathbf{y}}$  of the ensemble  $\mathbf{Y}$ . Using the results in Gupta and Nagar [2000], it can be shown that the Kalman gain matrix given the joint ensemble is distributed as

$$[\mathbf{K} | \mathbf{Y}] \sim \text{MatrixT}_{n_x, n_d} \left( \Gamma_{\mathbf{K}}^c, (\Psi_{\mathbf{d}}^c)^{-1}, \Psi_{\mathbf{x}^u | \mathbf{d}}^c, \mathbf{v}^c - n_x + 1 \right), \quad (12)$$

where  $\Gamma_{\mathbf{K}}^c = \Psi_{\mathbf{x}^u, \mathbf{d}}^c \Psi_{\mathbf{d}}^{c-1}$  and  $\Psi_{\mathbf{x}^u | \mathbf{d}}^c = \Psi_{\mathbf{x}^u}^c - \Gamma_{\mathbf{K}}^c \Sigma_{\mathbf{d}}^c \Gamma_{\mathbf{K}}^{cT}$ . Replacing the estimated Kalman gain matrix,  $\hat{\mathbf{K}}$  in Eq. (10) with independent realisations from the matrix- $t$  distribution in Eq. (12) thus defines the EnKF updating scheme.

### 3.2 Non-Informative Prior Distribution

In the hierarchical Bayesian approach outlined above, we need to select prior hyperparameters for both the unknown mean and covariance. When considering high dimensional spatiotemporal inverse problems, this task can be far from trivial. A non-informative prior distribution can be used to avoid this problem.

Consider the multivariate linear regression problem:

$$\mathbf{X} = \mathbf{K} \mathbf{D} + \mathbf{R}, \quad (13)$$

where we for notational convenience assume that  $\mathbf{X} \in \mathbb{R}^{n_x \times n_e}$  and  $\mathbf{D} \in \mathbb{R}^{n_d \times n_e}$  are centred ensemble matrices,  $\mathbf{K} \in \mathbb{R}^{n_x \times n_d}$  is the matrix of regression coefficients, or the Kalman gain, and  $\mathbf{R} \in \mathbb{R}^{n_x \times n_e}$  is the matrix of regression model error independent of  $\mathbf{K}$  and  $\mathbf{D}$ . Under the EnKF model assumptions in Eq. (4), we have  $\mathbf{D} \sim \text{MatrixG}_{n_d, n_e}(\mathbf{0}_{n_d, n_e}, \mathbf{I}_{n_e}, \Sigma_{\mathbf{d}})$ . Furthermore, let the regression error term be

$\mathbf{R} \sim \text{MatrixG}_{n_x, n_e}(\mathbf{0}_{n_x, n_e}, \mathbf{I}_{n_e}, \Sigma_{\mathbf{r}})$ , where the matrix-variate Gaussian distribution is defined in Gupta and Nagar [2000]. As shown in Mardia et al. [1979], the conditional likelihood of  $\mathbf{X}$  given  $\mathbf{D}$  can be written as:

$$f(\mathbf{X}|\mathbf{D}, \mathbf{K}, \Sigma_{\mathbf{r}}) \propto |\Sigma_{\mathbf{r}}|^{-n_e/2} \exp\left\{-\frac{1}{2}\text{tr}(\Psi_{\mathbf{K}}\Sigma_{\mathbf{r}}^{-1})\right\}, \quad (14)$$

where  $\Psi_{\mathbf{K}} = n_e \left[ (\mathbf{K} - \hat{\mathbf{K}}) \hat{\Sigma}_{\mathbf{d}} (\mathbf{K} - \hat{\mathbf{K}})^T + \hat{\Sigma}_{\mathbf{r}} \right]$ . Here  $\hat{\mathbf{K}}$ , given in Eq. (9), and  $\hat{\Sigma}_{\mathbf{r}} = 1/n_e (\mathbf{X} - \hat{\mathbf{K}}\mathbf{D})(\mathbf{X} - \hat{\mathbf{K}}\mathbf{D})^T$  are the maximum likelihood (ML) estimates of the Kalman gain and residual covariance matrix respectively, while  $\hat{\Sigma}_{\mathbf{d}} = 1/n_e \mathbf{D}\mathbf{D}^T$ .

We will now consider the posterior distribution of  $\mathbf{K}$  for the non-informative Jeffreys' prior [Jeffreys, 1946, Press, 1982]:

$$f(\mathbf{K}, \Sigma_{\mathbf{r}}) \propto |\Sigma_{\mathbf{r}}|^{-(n_x+1)/2}. \quad (15)$$

By Bayes' rule using the pdfs in Eq. (14) and (15) we get [Press, 1982]:

$$[\mathbf{K}|\mathbf{Y}] \sim \text{MatrixT}_{n_x, n_d} \left( \hat{\mathbf{K}}, \hat{\Sigma}_{\mathbf{d}}^{-1}, \hat{\Sigma}_{\mathbf{r}}, n_e - (n_x + n_d - 1) \right) \quad (16)$$

An EnKF updating scheme is then defined by replacing the Kalman gain matrix in Eq. (10) with independent realisations from the matrix- $t$  distribution in Eq. (16).

### 3.3 The Kalman Gain Posterior Distributions

Similar to the approach in Section 3.2, we can derive the posterior distribution based on a fully specified conjugate prior model for  $\mathbf{K}$  and  $\Sigma_{\mathbf{r}}$ . That is [Press, 1982]:

$$f(\mathbf{K}, \Sigma_{\mathbf{r}}) = f(\mathbf{K}|\Sigma_{\mathbf{r}})f(\Sigma_{\mathbf{r}}),$$

with  $[\mathbf{K}|\Sigma_{\mathbf{r}}] \sim \text{MatrixG}_{n_x, n_d}(\Gamma_{\mathbf{K}}, \Theta, \Sigma_{\mathbf{r}})$ , and  $\Sigma_{\mathbf{r}} \sim W_{n_x}^{-1}(\Psi_{\mathbf{r}}, \nu)$ . Such an approach does, however, make it less clear how the unknown hyperparameters should be specified. In addition, the approach outlined in Section 3.1 makes the posterior distribution consistent with the posterior distribution using a non-informative prior when  $n_e \rightarrow \infty$ . Moreover, the presented approach increases the degrees of freedom in the prior covariance matrix of  $\mathbf{K}$  from  $\frac{1}{2}(n_x(n_x + 1) + n_d(n_d + 1))$  to  $\frac{1}{2}(n_x + n_d)(n_x + n_d + 1)$  when directly defining a prior distribution for the Kalman gain. This is due to the structure of the Kronecker product [Drèze and Richard, 1983, Press, 1982].

It should be noted that choosing conjugate or Jeffreys' prior distributions in Bayesian linear regression is not without issues [Broemeling, 1985, Brown, 1993, Dawid, 1988, Press, 1982, Robert, 2001]. In this study, however, we have adopted these priors mainly because of their analytical properties. Of course we might in-

stead consider approaches based on generalised natural conjugate priors [Press, 1982], or selecting non-conjugate prior distributions [Brown et al., 1999].

Because the same prior distributions are used in the HEnKF [Myrseth and Omre, 2010a] and the approach outlined in Section 3.1, we should expect that there is a close connection between the Bayesian regression approach based on the informative prior and the HEnKF. Indeed, there are many similarities between the two approaches because the covariance hyperparameter matrix,  $\Psi_{\mathbf{y}}^c$ , is used in both posterior distributions. The major difference, however, lies in the sampling of the respective Kalman gain matrices. For the HEnKF we start by sampling  $\Sigma_{\mathbf{y}}^{(i)} \sim W_{n_{\mathbf{y}}}^{-1}(\Psi_{\mathbf{y}}^c, \mathbf{v}^c)$ , thus generating a realisation of the Kalman gain as  $\mathbf{K}^{(i)} = \Sigma_{\mathbf{x}, \mathbf{d}}^{(i)} \Sigma_{\mathbf{d}}^{(i)-1}$ . For  $\mathbf{K} \sim \text{MatrixT}_{n_x, n_d} \left( \Gamma_{\mathbf{K}}^c, (\Psi_{\mathbf{d}}^c)^{-1}, \Psi_{\mathbf{x}, \mathbf{d}}^c, \mathbf{v}^c - n_x + 1 \right)$ , however,  $\mathbf{K}^{(i)} = \Gamma_{\mathbf{K}}^c + \Gamma_{\mathbf{K}}^{(i)}$ , where  $\Gamma_{\mathbf{K}}^c$  is the posterior mean  $\text{E}[\mathbf{K}^{(i)}]$ , and  $\Gamma_{\mathbf{K}}^{(i)} \in \mathbb{R}^{n_x \times n_d}$  is a random matrix, given as a product of the square root of a matrix following the inverted Wishart distribution and a matrix following the matrix Gaussian distribution as explained in Property 1, Appendix 1. Hence, the latter approach will have improved computational properties compared with the HEnKF in the general case, because we do not need to invert the matrix  $\Sigma_{\mathbf{d}}^{(i)}$  for each realisation. Note, however, that the updated realisations based on the HEnKF updating scheme are expected to be less coupled. This follows because we avoid using the constant matrix  $\Gamma_{\mathbf{K}}^c$  for all updated ensemble members, which for the Bayesian approaches considered here involves the empirically estimated covariance matrix  $\hat{\Sigma}_{\mathbf{y}}$ .

### 3.4 Approximate matrix- $t$ Distribution

When we consider a non-informative prior distribution, the posterior distribution of  $\mathbf{K}$ , defined in Eq. (16), will not exist for  $n_e < n_x + n_d$ . Moreover, if we consider high dimensional reservoir models, generating realisations from the matrix- $t$  distribution can be highly computationally demanding. We will therefore consider an approximate version of the singular matrix- $t$  distribution, inspired by Principal Component Analysis (PCA) [Mardia et al., 1979] and shrinkage methods in a multivariate regression setting [Brown, 1993, Hastie et al., 2009]. This is motivated by the property that the matrix- $t$  distribution is closed under linear transformations  $\mathbf{U} = \mathbf{A}\mathbf{T}\mathbf{B}$ , for constant matrices  $\mathbf{A}$  and  $\mathbf{B}$  of proper dimensions [Gupta and Nagar, 2000].

As shown in Díaz-García and Gutiérrez-Jáimez [2006], for the case with the Jeffreys' prior distribution with  $\text{rank}(\hat{\Sigma}_{\mathbf{d}}) = \text{rank}(\hat{\Sigma}_{\mathbf{r}}) = n_e - 1$ ,  $[\mathbf{K}|\mathbf{Y}] \sim \text{MatrixT}_{n_x, n_d}^{n_e-1, n_e-1} \left( \hat{\mathbf{K}}, \hat{\Sigma}_{\mathbf{d}}^-, \hat{\Sigma}_{\mathbf{r}}, 1 \right)$ , where the singular matrix- $t$  distribution is defined in Appendix 2 and  $(\hat{\Sigma}_{\mathbf{d}})^-$  is the Moore-Penrose inverse [Strang, 1988]. To ensure that the degrees of freedom is larger than one, we therefore propose to approximate the singular matrix- $t$  variate distribution by selecting the dimension of the respective reduced order subspaces  $p$  and  $q$ , smaller than the rank of  $\hat{\Sigma}_{\mathbf{d}}$  and  $\hat{\Sigma}_{\mathbf{r}}$ . This corresponds to PCA, where the assumption is that a small set of

variables, termed principal components, explain most of the variability in the data [Mardia et al., 1979]. To improve the predictive capabilities of the regression model [Sætrum and Omre, 2011], we can replace the ML estimate of the Kalman gain matrix in Eq. (9) with e.g. the Principal Component Regression (PCR) estimate:  $\hat{\mathbf{K}}_{\text{PCR}} = \mathbf{X}\mathbf{V}_{\mathbf{D}}^q\mathbf{S}_{\mathbf{D}}^{q-1}\mathbf{U}_{\mathbf{D}}^{qT}$ . Here  $\mathbf{D} \approx \mathbf{U}_{\mathbf{D}}^q\mathbf{S}_{\mathbf{D}}^q\mathbf{V}_{\mathbf{D}}^{qT}$  is the truncated Singular Value Decomposition (SVD) of  $\mathbf{D}$ , keeping only the first  $q$  components corresponding to the  $q$  largest singular values [Golub and van Loan, 1996]. This implies that for the approximate matrix- $t$  distribution based on PCR,

$$\mathbf{K} \sim \text{MatrixT}_{n_x, n_d}^{p, q} \left( \hat{\mathbf{K}}_{\text{PCR}}, \left( \hat{\Sigma}_{\mathbf{d}}^q \right)^{-1}, \hat{\Sigma}_{\mathbf{r}, n_e - (q - p - 1)}^p \right), \quad (17)$$

where  $\hat{\Sigma}_{\mathbf{d}}^q$  and  $\hat{\Sigma}_{\mathbf{r}}^p$  denotes the low rank approximation of the two matrices based on a truncated eigenvalue decomposition keeping  $q \leq \min\{n_d, n_e - 1\}$  and  $p \leq \min\{n_x, n_e - 1\}$  non-zero eigenvalues respectively. Alternative estimators for the Kalman gain matrix based on dimension reduction techniques such as Partial Least Squares Regression (PLSR) [Rännér et al., 1994], can be used as well. Finally note that we use Cross-Validation (CV) to select the subspace dimensions  $p$  and  $q$  [Seber and Lee, 2003, Sætrum and Omre, 2011].

### 3.5 Computational Properties

As shown in Appendix 1, we can generate a realisation from the matrix- $t$  distribution using  $n_x^2 n_d / 2 + n_d^2 n_x / 2 + (2/3 + 1/3)n_x^3 + \mathcal{O}(n_x \max\{n_x, n_d\})$  floating point operations (flops), when the Cholesky factorisation of the respective matrices are given. Here the first two terms are the number of flops required for multiplying an upper diagonal matrix with a full matrix of dimension  $n_x \times n_d$ , and  $(2/3 + 1/3)n_x^3$  are the leading terms for computing the matrix inverse based on back substitution and multiplication of two upper triangular matrices [Golub and van Loan, 1996]. For a singular matrix- $t$  distributed random matrix, the computational demands are  $\mathcal{O}(n_x n_d \max\{p, q\})$ , given the SVD of the respective matrices and the estimated Kalman gain matrix.

When considering the suggested procedure in an EnKF setting, however, only  $n_d^2 + 2n_x n_d + 2n_x^2 + \mathcal{O}(\max\{n_x, n_d\})$  flops are required to update one realisation in the fully specified prior setting. For the singular matrix- $t$  distributed Kalman gain, the computational demand is  $\mathcal{O}(n_e \max\{n_x, n_d\})$  flops, which is the same as for the Full EnKF scheme presented in Section 2.1 for the rank deficient case [Evensen, 2003]. The memory requirement for the informative prior case is  $\mathcal{O}(\max\{n_x^2, n_d^2\})$ , which implies that approximate techniques such as localisation [Evensen, 2009] must be used when  $n_x$  or  $n_d$  is large. For the singular matrix- $t$  distribution defined through the non-informative Jeffreys' prior, the memory requirement is  $\mathcal{O}(n_e \max\{n_x, n_d\})$ , which corresponds to the memory requirement of the Full EnKF.

Note that only one single Cholesky and SVD factorisation of the respective matrices is required for all ensemble members. This requires  $\mathcal{O}(n_x^3) + \mathcal{O}(n_d^3)$  flops for the conjugate prior and  $\mathcal{O}(\max\{n_x, n_d\}n_e^2)$  flops for the non-informative Jeffreys' prior in a reduced order subspace. The former can be quite costly for high dimensional problems, while the latter corresponds to the computational demands of the Full EnKF scheme. A discussion regarding the computational properties of the CV scheme when used to select the appropriate subspace dimension can be found in Sætrom and Omre [2011], where it is shown that the CV scheme will not lead to a significant increase in the computational demands if it is properly implemented.

## 4 Empirical study

To evaluate the performance of the EnKF updating schemes based on the Bayesian regression techniques presented in Sections 3.1 and 3.2, we consider a synthetic case study similar to the one used in Myrseth and Omre [2010b]. Here the unknown state vector of interest  $\mathbf{x}_k \in \mathbb{R}^{100 \times 1}$ , is defined for  $k = 0, \dots, 10$ , where  $x_{j,k}$  denotes the variable of interest at time step  $k$  and location  $j \in \{1, \dots, 100\}$ . The purpose of this study is to assimilate observed data,  $\mathbf{d}_k^o$  made at time steps  $0, \dots, 9$ , and predict at time step 10.

We consider a Gaussian prior at the initial time step, a linear forward function and a linear Gaussian likelihood model:

$$\begin{aligned}\mathbf{x}_0 &\sim \text{Gauss}_{n_x}(\mathbf{0}_{n_x,1}, \Sigma_{\mathbf{x}_0}) \\ \mathbf{x}_k &= \mathbf{A}_k \mathbf{x}_{k-1} \\ \mathbf{d}_k &= \mathbf{B} \mathbf{x}_k + \varepsilon_{\mathbf{d}_k}.\end{aligned}$$

Here an exponential covariance function defines  $\Sigma_{\mathbf{x}_0}$ , and  $\mathbf{A}_k$  defines a linear smoothing envelope moving from left to right in time over the grid domain. The error term of the likelihood model is Gaussian,  $\varepsilon_{\mathbf{d}_k} \sim \text{Gauss}_{n_d}(\mathbf{0}_{n_d,1}, \mathbf{I}_{n_d})$  and  $\mathbf{B} \in \mathbb{R}^{n_d \times n_x}$  is defined as a sparse matrix, so that  $d_i = \sum_{l=-1}^1 x_{j+l}$ ,  $i = 1, \dots, n_d$ , at  $n_d = 13$  different grid locations  $j \in \{1, \dots, 100\}$ . Further details can be found in Sætrom and Omre [2011].

We consider the following EnKF updating schemes:

- Standard EnKF: Estimated Kalman gain matrix computed based as  $\hat{\mathbf{K}}_{\text{Std}} = \hat{\Sigma}_{\mathbf{x}^u} \mathbf{B}^T (\mathbf{B} \hat{\Sigma}_{\mathbf{x}^u} \mathbf{B}^T + \Sigma_{\varepsilon_{\mathbf{d}}})^{-1}$ , using the correct error covariance matrix  $\Sigma_{\varepsilon_{\mathbf{d}}} = \mathbf{I}_{n_d}$ .
- CP-EnKF: Kalman gain matrix generated independently for each ensemble member based on the matrix- $t$  distribution in Eq. (12), obtained using a Conjugate Prior (CP) distribution on the mean and covariance as described in Section 3.1.
- JP-PLSR-EnKF: Kalman gain matrix generated independently for each ensemble member using the approximate matrix- $t$  distribution based on the non-informative Jeffreys' prior, described in Section 3.4. In addition,  $\hat{\mathbf{K}}$  is replaced with a shrinkage regression estimate based on PLSR, where the dimension of the

respective subspaces  $p$  and  $q$  are selected based on 10-fold CV, minimising the Predictive Error Sum of Squares (PRESS) [Hastie et al., 2009].

Two different ensemble sizes will be used,  $n_e = 20$  and  $n_e = 100$ . In addition, the initial ensembles are identical for all three updating schemes. For the CP-EnKF updating scheme we select the prior model hyperparameters  $\eta_x = \mu_x$ ,  $\xi = 0.0001$ ,  $\Psi_x = 10 \times \Sigma_x$  and  $\mathbf{v} = n_x + n_d + 3$ , with  $\Psi_{\mathbf{x},\mathbf{d}} = \Psi_x \mathbf{B}^T$  and  $\Psi_d = 10 \times \Sigma_d$ . Thus, preserving the structure of the reference model in  $\eta_y$  and  $\Psi_y$ , although the uncertainty regarding the size of  $\eta_y$  and  $\Psi_y$  is high.

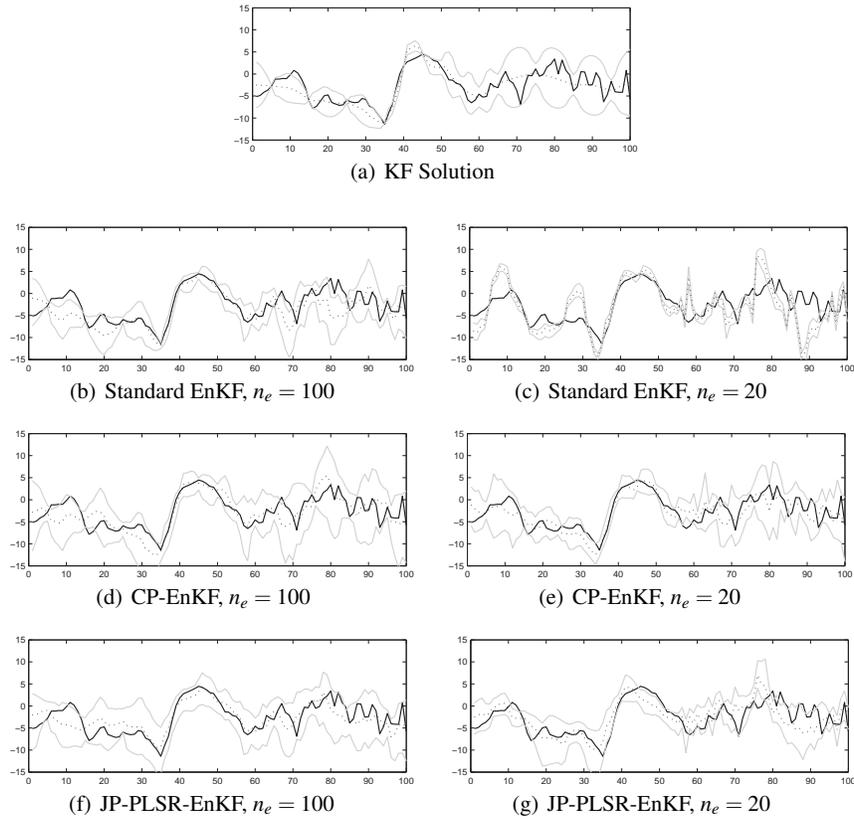
Because we consider a Gauss-linear model, the prediction mean  $E[\mathbf{x}_{10}^u]$ , and 95% prediction interval are analytically obtainable using the KF recursions, see Figure 2a. Notice that the reference state vector is centred around the prediction mean, with reduced uncertainty in the 95% prediction intervals in the neighbourhood of the observation sites. The results for the three different EnKF updating schemes are shown in Figures 2b through g.

As seen from Figure 2b, the Standard EnKF is able to produce reliable results for  $n_e = 100$ . Both the estimated posterior mean and prediction interval matches the analytical KF solution fairly well, albeit the ensemble average is not as smooth as the theoretical one. Moreover, there is a tendency to underestimate the prediction variance. For  $n_e = 20$ , the effects caused by increased estimation uncertainty and coupling of the ensemble members are more prominent. This results in an estimated posterior mean and prediction interval deviating dramatically from the KF solution, as expected from the results in Sætrum and Omre [2012].

The CP-EnKF updating scheme provides reasonable estimates of the posterior mean and prediction interval for  $n_e = 100$ , as seen in Figure 2. However, the bounds of the prediction interval is slightly increased at certain grid locations compared with the KF solution based on  $\Sigma_{\mathbf{x}^u|\mathbf{d}}$ . Because we are sampling the Kalman gain matrix, rather than using a common plug-in estimate, we should expect to see this behaviour for  $\hat{\Sigma}_{\mathbf{x}^c}$ , in light of the theoretical results presented in Sætrum and Omre [2012]. Because the matrix- $t$  distribution for the Kalman gain matrix in Eq. (12) requires the empirical covariance matrix as a model parameter, the ensemble members will be somewhat coupled. However, this coupling is reduced compared with the Standard EnKF updating scheme, leading to improved estimates of both the prediction mean and interval.

Both the estimated posterior mean and prediction interval resembles the analytical KF solution for the JP-PLSR-EnKF updating scheme with  $n_e = 100$ , as seen in Figure 2f. The smooth behaviour of the posterior mean,  $\mu_{\mathbf{x}^c}$  is present, with reduced uncertainty in the prediction variance around the observation sites. For  $n_e = 20$ , however, the ensemble average appears to be less smooth, and the prediction intervals are reduced. The results are comparable to the CP-EnKF results, and appears to be much better than the ones obtained with the Standard EnKF scheme.

To further quantify the performance of the three EnKF updating schemes, the case study is rerun 100 times with different initial ensembles. We consider the Root Mean Squared Error (RMSE) of the ensemble average to the analytical posterior mean and the percentage of the reference solution covered by the estimated prediction interval. Table 1 displays the results, together with the estimated RMSE and



**Fig. 2** Results obtained when running four different EnKF updating schemes on the linear case with two different ensemble sizes. The figure displays the reference  $\mathbf{x}_{10}^{\text{True}}$  (solid), the ensemble mean (dotted, black) and the estimated 90.5% ( $n_e = 20$ ) and 95% ( $n_e = 100$ ) confidence bounds of the prediction interval (solid, grey).

coverage computed when the initial ensemble is run through the forward model without conditioning on the observed data, referred to as the No Updating scheme. Ideally, the RMSE should be small and the Coverage should be close to 90.5% for  $n_e = 20$  and 95% for  $n_e = 100$ . The former represents accuracy while the latter is a measure for the reliability of the empirically estimated prediction intervals. For  $n_e = 100$ , the RMSE decreases significantly for all three updating schemes, relative to the RMSE of the No Updating scheme. The CP-EnKF updating scheme shows the largest improvement with a 62% decrease in the RMSE, although it is not found to be significantly better than the Standard EnKF updating scheme with a  $p$ -value of 0.92. The hypothesis test is based on the traditional two sample  $t$ -statistic with unequal variance [Casella and Berger, 2002]. The RMSE of the JP-PLSR-EnKF updating scheme is slightly larger compared with the RMSE of the other two methods.

**Table 1** Estimated Root Mean Squared Error (RMSE) of the posterior mean and coverage of the reference solution in the estimated 90.5% ( $n_e = 20$ ) and 95% ( $n_e = 100$ ) prediction intervals, plus-minus one standard deviation, for the linear case based on 100 different initial ensembles.

$n_e$	Scheme	RMSE	Coverage (%)
100	No Updating	$4.60 \pm 0.14$	$95.1 \pm 2.2$
100	Standard EnKF	$1.77 \pm 0.30$	$81.1 \pm 5.8$
100	CP-EnKF	$1.74 \pm 0.28$	$95.3 \pm 3.2$
100	JP-PLSR-EnKF	$1.91 \pm 0.25$	$88.4 \pm 3.8$
20	No Updating	$4.68 \pm 0.27$	$88.5 \pm 4.2$
20	Standard EnKF	$3.89 \pm 1.05$	$26.4 \pm 6.3$
20	CP-EnKF	$1.73 \pm 0.29$	$90.0 \pm 4.3$
20	JP-PLSR-EnKF	$3.02 \pm 0.52$	$70.7 \pm 9.4$

As expected from the results in Sætrum and Omre [2012], sampling the Kalman gain matrix, as done in the CP-EnKF and JP-PLSR-EnKF updating schemes, will increase the width of the prediction intervals, whilst it will be narrower for the Standard EnKF updating scheme. For this case with  $n_e = 100$  we find the coverage the CP-EnKF updating scheme and the No Updating scheme to be almost equal, and the value is close to 95% as it should be. Testing for equivalence between the coverage of the CP-EnKF and JP-PLSR-EnKF updating schemes, leads to a  $p$ -value of 0.007, and hence should be rejected at a 99% level of confidence.

For  $n_e = 20$ , the CP-EnKF updating scheme again leads to the largest improvement in the RMSE relative to the No Updating scheme with a 63 percent decrease. The JP-PLSR-EnKF updating scheme shows the second largest improvement with a 35 percent decrease, whilst the Standard EnKF updating scheme has a 17 percent decrease in the RMSE. We find all three updating schemes to give significantly different RMSE values at a 99.999% level of confidence.

The Standard EnKF updating scheme has a coverage of only 26.4%, which demonstrates a dramatic underestimation of the prediction intervals caused by the coupling of the ensemble members. Furthermore, we observe that the coverage of the CP-EnKF updating scheme is found to be significantly larger than the coverage of the Standard EnKF and JP-EnKF updating schemes. However, the coverage of the No Updating scheme is not found to be significantly different from the CP-EnKF updating scheme with a  $p$ -value of 0.59.

We obtain the best results for the linear problem using the CP-EnKF updating scheme, but the results appear to rely heavily on the selected prior hyperparameters. If the model is rerun with a poorly selected prior model, the CP-EnKF updating scheme provides terrible results, with a large increase in the RMSE and large decrease in the coverage. This is particularly true for  $n_e = 20$ , although we do not show these results here. For high dimensional reservoir evaluation problems, the task of selecting the prior hyperparameters is far from trivial. In addition, because of high computational and memory demands, the CP-EnKF updating scheme will require additional approximations, such as enforcing sparsity in the model parameters through localisation [Evensen, 2009]. This suggests that the JP-PLSR-EnKF updating scheme is the most robust alternative, although the CP-EnKF scheme with

a carefully chosen prior model performed best for the empirical evaluation considered here.

## 5 Conclusions

In this paper we have formulated alternative EnKF updating schemes based on classical results known from Bayesian regression. Rather than using a common estimate of the Kalman gain matrix in the update, each ensemble member is updated using a unique Kalman gain matrix generated from an analytically tractable matrix variate probability distribution. The suggested EnKF updating schemes based on Bayesian regression reduces the coupling between the updated ensemble members, which leads to improved predictions and associated prediction intervals.

Two alternative EnKF updating schemes are defined and empirically tested. The first one is based on a conjugate prior distribution on the unknown mean and covariance matrix (CP-EnKF). The second one uses a non-informative Jeffreys' prior distribution on a multivariate regression model combined with Partial Least Squares Regression (JP-PLSR-EnKF), leading to a singular matrix- $t$  distributed Kalman gain matrix defined in a reduced order space. The main objective of this additional dimension reduction is to improve the Kalman gain matrix estimate and to handle large dimensional problems with small ensemble sizes. Cross-validation is used to select the subspace dimension by evaluating the predictive strength of the regression model.

The two suggested schemes were tested on a synthetic model, inspired by reservoir evaluation problems, and compared with the results obtained using the Standard EnKF updating scheme. With a small ensemble size of 20, both the CP-EnKF and JP-PLSR-EnKF updating schemes performed significantly better than the Standard EnKF updating scheme. The coupling of the ensemble members in the Standard EnKF scheme caused the predictions and associated prediction intervals to be highly unreliable. Increasing the ensemble size to 100 improved the results for the Standard EnKF updating scheme and made it comparable to the results obtain with the alternative EnKF scheme based on Bayesian regression.

The CP-EnKF scheme appeared as the most reliable EnKF approach in the empirical study, closely followed by the JP-PLSR-EnKF scheme. The Standard EnKF scheme produced unreliable predictions and associated prediction intervals for small ensemble sizes relative to the number of observations. The CP-EnKF scheme is, contrary to the JP-PLSR-EnKF scheme, computationally demanding to perform and highly dependent on the specification of a representative prior model. Consequently, we recommend using the JP-PLSR-EnKF scheme for high dimensional spatiotemporal problems with a large number of observations.

**Acknowledgements** This work was funded by the Uncertainty in Reservoir Evaluation (URE) initiative at NTNU.

## Appendix 1

*Property 1 (Realisation from the Matrix  $t$ -distribution).* Let

$$\Sigma \sim W_u^{-1}(\Psi^{-1}, \mathbf{v} + u - 1)$$

independent of  $\mathbf{X} \sim \text{MatrixG}_{u,g}(\mathbf{0}_{u,g}, \Omega, \mathbf{I}_u)$ . Then

$$\mathbf{T} = \Sigma^{1/2}\mathbf{X} + \Gamma \sim \text{MatrixT}_{u,g}(\Gamma, \Omega, \Psi, \mathbf{v}). \quad (18)$$

*Proof.* By Definitions of the inverted Wishart and matrix Gaussian distributions [Gupta and Nagar, 2000]

$$f(\Sigma, \mathbf{X}) \propto |\Sigma|^{-(\mathbf{v}+u-1+u+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} (\Sigma^{-1}\Psi + \Omega^{-1}\mathbf{X}^T\mathbf{X}) \right\},$$

Following Gupta and Nagar [2000], let  $\mathbf{T} = \mathbf{g}(\mathbf{X}, \Sigma) = \Sigma^{1/2}\mathbf{X} + \Gamma$ . A transformation of the variables gives,

$$f(\Sigma, \mathbf{T}) = f(\Sigma, \mathbf{g}^{-1}(\mathbf{T}, \Sigma)) |\mathbf{J}_{\mathbf{xt}}|,$$

where by Deemer and Olkin [1951],

$$\begin{aligned} \mathbf{J}_{\mathbf{xt}} &= \frac{\partial \Sigma^{-1/2}(\mathbf{T} - \Gamma)}{\partial \mathbf{T}} \\ &= |\Sigma|^{-g/2}. \end{aligned}$$

Hence,

$$\begin{aligned} f(\mathbf{T}) &\propto \int |\Sigma|^{-(\mathbf{v}+u-1+u+1+g)/2} \exp \left\{ -\frac{1}{2} \text{tr} ([\Psi + (\mathbf{T} - \Gamma)\Omega^{-1}(\mathbf{T} - \Gamma)^T] \Sigma^{-1}) \right\} d\Sigma \\ &\propto |\Psi + (\mathbf{T} - \Gamma)\Omega^{-1}(\mathbf{T} - \Gamma)^T|^{-(\mathbf{v}+u+g-1)/2}, \end{aligned}$$

and the result follows by the general properties of the determinant and trace operators [Mardia et al., 1979].

*Property 2 (Fast sampling from the matrix- $t$  distribution).* Let  $\mathbf{U}_\Omega \in \mathbb{R}^{b \times b}$  and  $\mathbf{U}_\Sigma \in \mathbb{R}^{a \times a}$  be the upper triangular Cholesky factors of the positive definite matrices  $\Omega$  and  $\Sigma$  respectively. Further let

$$\mathbf{A} = \begin{bmatrix} u_{1,1} & z_{1,2} & z_{1,3} & \dots & z_{1,a-1} & z_{1,a} \\ 0 & u_{2,2} & z_{2,3} & \dots & z_{2,a-1} & z_{2,a} \\ 0 & 0 & u_{3,3} & \dots & z_{3,a-1} & z_{3,a} \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & u_{a-1,a-1} & z_{a-1,a} \\ 0 & 0 & \dots & 0 & 0 & u_{a,a} \end{bmatrix} \in \mathbb{R}^{a \times a}, \quad (19)$$

where  $u_{i,i} = \sqrt{v_i}$ , with  $v_i \stackrel{\text{i.i.d.}}{\sim} \chi^2(v + a - i)$ , and  $z_{i,j} \stackrel{\text{i.i.d.}}{\sim} \text{Gauss}_1(0, 1)$ , and let  $\mathbf{W} \in \mathbb{R}^{a \times b}$ , have elements  $w_{i,j} \stackrel{\text{i.i.d.}}{\sim} \text{Gauss}_1(0, 1)$ . Then

$$\mathbf{T} = \mathbf{U}_\Sigma^T \mathbf{A}^{-1} \mathbf{W} \mathbf{U}_\Omega + \mathbf{B} \sim \text{MatrixT}_{a,b}(\mathbf{B}, \Omega, \Sigma, \nu). \quad (20)$$

*Proof.* By Smith and Hocking [1972] and Jones [1985],  $\mathbf{A}^{-1} \mathbf{A}^{T^{-1}} \sim W_a^{-1}(\mathbf{I}_a, \nu + a - 1)$ . Thus, by Property 1,

$$\mathbf{X} = \mathbf{A}^{-1} \mathbf{W} \sim \text{MatrixT}_{a,b}(\mathbf{0}_{a,b}, \mathbf{I}_b, \mathbf{I}_a, \nu).$$

The result then follows by Theorem 4.3.8 in Gupta and Nagar [2000].

## Appendix 2

**Definition 1 (Singular matrix- $t$  distribution).** Let  $\mathbf{U}_\Sigma \in \mathbb{R}^{a \times p}$  be the matrix containing the  $p$  first eigenvectors of  $\Sigma \in \mathbb{R}^{a \times a}$  corresponding to the  $p$  non-zero eigenvalues given as the  $p$  elements of the diagonal matrix  $\Lambda_\Sigma \in \mathbb{R}^{p \times p}$ , with a similar notation for the matrix  $\Omega \in \mathbb{R}^{b \times b}$ ,  $\mathbf{U}_\Omega \in \mathbb{R}^{b \times q}$  and  $\Lambda_\Omega \in \mathbb{R}^{q \times q}$ . Assume  $\mathbf{Z} \sim \text{MatrixG}_{p,q}^{p,q}(\mathbf{0}_{p,q}, \Lambda_\Omega, \mathbf{I}_p)$  independent of  $\mathbf{V} \sim W_p^{-1}(\Lambda_\Sigma^{-1}, \nu + p - 1)$ , where the singular matrix Gaussian and inverted Wishart distributions are defined in Daz-Garca et al. [1997] and Bodnar and Okhrin [2008] respectively, then

$$\mathbf{T} = \mathbf{U}_\Sigma \mathbf{W}^{1/2} \mathbf{Z} \mathbf{U}_\Omega^T + \mathbf{M} \quad (21)$$

follows the *singular matrix- $t$*  distribution,  $\mathbf{T} \sim \text{MatrixT}_{a,b}^{p,q}(\mathbf{M}, \Omega, \Sigma, \nu)$ .

*Remark 1.* This definition is different from the one used in Díaz-García and Gutiérrez-Jáimez [2009], where they define the singular matrix- $t$  distribution through  $\mathbf{Y} = (\mathbf{W}^{1/2})^{-} \mathbf{X} + \mathbf{M}$ , where  $\mathbf{W} \sim W_a^{\tilde{p}}(\Sigma, \nu)$  independent of  $\mathbf{X} \sim \text{MatrixG}_{a,b}^{a,q}(\mathbf{0}_{a,b}, \Omega, \mathbf{I}_p)$ ,  $\tilde{p} = \min\{p, \nu\}$  and the superscript  $(-)$  denotes the Moore-Penrose inverse of a matrix [Strang, 1988].

*Property 3 (Probability density function of the singular matrix- $t$  distribution).* The pdf of  $\mathbf{T} \sim \text{MatrixT}_{a,b}^{p,q}(\mathbf{M}, \Omega, \Sigma, \nu)$  is

$$f(\mathbf{T}) = c(p, q, \nu) \frac{|\mathbf{I}_b + \Omega^{-}(\mathbf{T} - \mathbf{M})^T \Sigma^{-}(\mathbf{T} - \mathbf{M})|^{(\nu + p + q - 1)/2}}{|\Lambda_\Sigma|^{q/2} |\Lambda_\Omega|^{p/2}}, \quad (22)$$

where

$$c(p, q, \nu) = \frac{\Gamma_p(\nu + p + q - 1)}{\pi^{pq/2} \Gamma_p(\nu + p - 1)},$$

with

$$\Gamma_p(\nu) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma((\nu + 1 - i)/2), \quad (23)$$

where,  $\Gamma(\cdot)$  is the Gamma function, and the superscript  $(-)$  denotes the Moore-Penrose inverse of a matrix.

*Proof.* Using Eq. (21) and the results in Díaz-García [2007],

$$\mathbf{T}|\mathbf{W} \sim \text{Matrix}G_{a,b}^{p,q}(\mathbf{M}, \Omega, \mathbf{U}_\Sigma \mathbf{W} \mathbf{U}_\Sigma^T).$$

Hence, the joint pdf of  $\mathbf{T}$  and  $\mathbf{W}$  is:

$$f(\mathbf{T}, \mathbf{W}) = c_1 |\mathbf{W}|^{-(v+2p+q)/2} \exp \left\{ -\frac{1}{2} \text{tr}(\Theta^{-1} \mathbf{W}^{-1}) \right\}, \quad (24)$$

where

$$c_1 = \frac{(2\pi)^{-pq/2} |\Lambda_\Sigma|^{(v+p-1)/2}}{|\Lambda_\Omega|^{p/2} 2^{(v+p-1)p/2} \Gamma_p(v+p-1)}$$

and

$$\Theta^{-1} = \Lambda_\Sigma + \mathbf{U}_\Sigma^T (\mathbf{T} - \mathbf{M}) \Omega^{-1} (\mathbf{T} - \mathbf{M})^T \mathbf{U}_\Sigma.$$

The result follows by integrating Eq. (24) with respect to  $\mathbf{W}$  and using general properties of the determinant [Mardia et al., 1979].

## References

- S. I. Aanonsen, G. Nævdal, D. S. Oliver, A. C. Reynolds, and B. Valls. Ensemble Kalman filter in reservoir engineering - a review. *SPE Journal*, 14(3):393–412, 2009.
- J. L. Anderson. A local least squares framework for ensemble filtering. *Monthly Weather Review*, 131(4):634–642, 2003a.
- T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley, 3 edition, 2003b.
- T. Bodnar and Y. Okhrin. Properties of the singular, inverse and generalized inverse partitioned Wishart distributions. *Journal of Multivariate Analysis*, 99(10):2389–2405, 2008.
- G. E. P. Box and G. C. Tiao. *Bayesian inference in statistical analysis*. Wiley, 1992.
- L. D. Broemeling. *Bayesian Analysis of Linear Models*. Marcel Dekker, Inc, 1985.
- P. J. Brown. *Measurement, Regression, and Calibration*. Oxford Science Publications, 1993.
- P. J. Brown, T. Fearn, and M. Vannucci. The choice of variables in multivariate regression: A non-conjugate Bayesian decision theory approach. *Biometrika*, 86(3):635–648, 1999.
- G. Burgers, G. Evensen, and P. J. van Leeuwen. Analysis scheme in the Ensemble Kalman Filter. *Monthly weather review*, 1998.
- G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, 2002.

- A. P. Dawid. The infinite regress and its conjugate analysis. In *Bayesian Statistics 3*, pages 95–110. Oxford University Press, 1988.
- W. L. Deemer and I. Olkin. The Jacobians of certain matrix transformations useful in multivariate analysis. *Biometrika*, 38:345–367, 1951.
- J. A. Díaz-García. A note about measures and Jacobians of singular random matrices. *Journal of Multivariate Analysis*, 98:960–969, 2007.
- J. A. Díaz-García and R. Gutiérrez-Jáimez. Singular matrix and matrix variate  $t$  distributions. *Journal of Statistical Planning and Inference*, 139(7):2382–2387, 2009.
- J. A. Díaz-García and R. Gutiérrez-Jáimez. Distribution of the generalised inverse of a random matrix and its applications. *Journal of Statistical Planning and Inference*, 136:183–192, 2006.
- A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10:197–208, 2000.
- J. H. Drèze and J.-F. Richard. Bayesian analysis of simultaneous equation systems. In *Handbook of Econometrics*, pages 517–598. Elsevier, 1983.
- J. A. Daz-Garca, R. Gutierrez-Jimez, and K. V. Mardia. Wishart and pseudo-Wishart distributions and some applications to shape theory. *Journal of Multivariate Analysis*, 63(1):73–87, 1997.
- G. Evensen. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean mode. *J. Geophys. Res.*, 97(C11):17905–17924, 1992.
- G. Evensen. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003.
- G. Evensen. *Data assimilation the Ensemble Kalman Filter*. Springer, 2 edition, 2009.
- D. E. Farrer and R. R. Glauber. Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics*, 49(1):92–107, 1967.
- R. Furrer and T. Bengtsson. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis*, 98(2):227–255, 2007.
- A. Gelb. *Applied Optimal Estimation*. MIT Press Cambridge, 1974.
- G. Golub and C. van Loan. *Matrix Computations*. Johns Hopkins University Press, 1996.
- A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Chapman & Hall, 2000.
- T. Hastie, R. Tibshirani, and J. Freidman. *The Elements of Statistical Learning; Data Mining, Inference, and Prediction*. Springer, New York, 2 edition, 2009.
- P. L. Houtekamer and H. L. Mitchell. A sequential Ensemble Kalman Filter for atmospheric data assimilation. *Monthly Weather Review*, 129:123–137, 2001.
- P. L. Houtekamer and H. L. Mitchell. Data assimilation using an Ensemble Kalman Filter technique. *Monthly Weather Review*, 126:796–811, 1998.
- A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
- H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London (Ser. A)*, 186:453–461, 1946.
- M. C. Jones. Generating inverse wishart matrices. *Communications in Statistics - Simulation and Computation*, 14:511–514, 1985.

- R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, 82(Series D):35–45, 1960.
- C. L. Keppenne and M. Rienecker. Assimilation of temperature into an isopycnal ocean general circulation model using a parallel Ensemble Kalman Filter. *J. Marine. Sys.*, 40-41:363–380, 2003.
- J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, 2001.
- K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.
- H. Moradkhani, S. Sorooshian, H. V. Gupta, and P. R. Houser. Dual state parameter estimation of hydrological models using Ensemble Kalman Filter. *Advances in Water Resources*, 28:135–147, 2005.
- I. Myrseth and H. Omre. Hierarchical Ensemble Kalman Filter. *SPE Journal*, 15(2):569–580, 2010a.
- I. Myrseth and H. Omre. The Ensemble Kalman Filter and related filters. In L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. van Bloemen Waanders, and K. Willcox, editors, *Large-scale Inverse Problems and Quantification of Uncertainty*. John Wiley and Sons, 2010b.
- J. S. Press. *Applied multivariate analysis: using Bayesian and frequentist methods of inference*. Robert E. Krieger Publishing Company, 2 edition, 1982.
- S. Rännér, F. Lindgren, P. Gelandi, and S. Wold. A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm. *Journal of Chemometrics*, 8:111–125, 1994.
- C. P. Robert. *The bayesian choice*. Springer, 2 edition, 2001.
- W. Sacher and P. Bartello. Sampling errors in Ensemble Kalman filtering. Part I: Theory. *Monthly Weather Review*, pages 3035–3049, 2008.
- J. Sætrom and H. Omre. Ensemble Kalman filtering with shrinkage regression techniques. *Computational Geosciences*, 15(2):271–292, 2011.
- J. Sætrom and H. Omre. Uncertainty quantification in the Ensemble Kalman Filter. Paper submitted for publication, 2012.
- G. A. F. Seber and A. J. Lee. *Linear Regression Analysis*. Wiley, 2003.
- W. B. Smith and R. R. Hocking. Algorithm AS 53: Wishart variate generator. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 21(3):341–345, 1972.
- G. Strang. *Linear Algebra And Its Applications*. Thomson Learning, 1988.
- P. J. van Leeuwen. Comments on "Data assimilation using an Ensemble Kalman Filter technique. *Monthly Weather Review*, 127:1374–1377, 1999.
- M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, 2 edition, 1999.