

Paper C

Effect of Large Number of Measurements on the Performance of EnKF Model Updating *

* Proceeding at 2008 *European Conference on the mathematics of oil recovery (ECMOR XI)*, Bergen, Norway, 2008.

Effect of Large Number of Measurements on the Performance of EnKF Model Updating

Abul Fahimuddin, Sigurd Ivar Aanonsen and Trond Mannseth
Center for Integrated Petroleum Research (CIPR),
University of Bergen, Norway

Abstract

The ensemble Kalman filter (EnKF) is a Monte Carlo method for data assimilation and assessment of uncertainties during reservoir characterization and performance forecasting. The method is based on a low-rank approximation to the system covariance matrix calculated from an ensemble which may be orders of magnitude smaller than the number of state variables. In practical applications, the ensemble size has to be kept relatively small. This may lead to poor approximation of the cross-covariance matrix, and sampling errors can result in spurious correlations and incorrect changes in the state variables. Also, since the rank of the covariance matrix cannot be larger than the number of ensemble members, the number of degrees of freedom may be too low when a large number of measurements are assimilated, such as with 4D seismic data. In this work, we have investigated the shortcomings of a straightforward EnKF implementation for small ensemble size, relative to a large number of measurements. This is done by considering a single update of a simple linear model and comparing the EnKF update to the traditional Kalman filter (or Kriging) solution, which in this case is exact. The quality of the EnKF update is assessed by considering the mean and variance of the updated state variable, as well as various error norms and the eigen-spectrum of the covariance matrix. Even for this simple model, spurious long-range correlation, ensemble collapse, etc. are clearly seen as the number of measurements increases for a given ensemble size. For a traditional implementation of EnKF, the ensemble size have to be much larger than the number of measurements to obtain an accurate solution, and the solution gets worse when the measurement uncertainty is reduced.

Introduction

The ensemble Kalman filter (EnKF) [3] has been increasingly popular in recent years for updating reservoir simulation models, and the method has been introduced to the petroleum engineering literature as a way to estimate uncertain reservoir properties and states [10, 1]. Because of computational limitations, the ensemble size needs to be small for practical field applications, and a straightforward approximation to the covariance matrix may lead to severe problems like spurious correlations (causing changes in the state variable in regions of no real influence) and rank deficiency/ensemble collapse. The latter problem is most significant for problems with large amounts of independent data, such as 4D seismic [1]. To avoid these problems, several approaches for covariance localization have been suggested, see e.g. [6, 7, 5, 4]. Skjervheim et al. [11] apply a simple local analysis to assimilate 4D seismic data with the EnKF.

Here, we investigate the accuracy of the traditional covariance estimate with a large number of measurements. No localization schemes are applied. We demonstrate that the ensemble size has to be much larger than the number of measurements in order to obtain an accurate solution, and that the problem becomes more severe when the measurement uncertainty decreases, indicating that some kind of localization may have to be applied more often than previously believed. First we give a brief presentation of the Kalman filter (KF) and ensemble Kalman filter (EnKF). Then we investigate, using a simple linear model, the relation between ensemble size, number of measurements, measurement uncertainty, and the accuracy of the estimated covariance and updated state variable. In our case, the KF update is the true reference case as for linear problems, KF is the best approximation. We compare the performance of the EnKF update with the KF update. An analysis of the eigenvalues and eigenvectors of the KF and EnKF covariance matrices has also been performed in order to investigate the spurious correlation effect.

In the last section we summarize our results and present some concluding remarks.

Kalman Filter and Ensemble Kalman Filter

In 1960 the Kalman filter was developed for estimating the state of a discrete-time controlled process governed by linear stochastic difference equations [8]. Let $\Theta^f \in \mathcal{R}^n$ represent the unobserved (forecast) state variable, and let $d \in \mathcal{R}^{N_d}$ denote a new set of observations. The data and the state are related by the state-to-observation operator, \mathbf{H} . Using the Kalman filter formula, the updated (analyzed) state is

$$\Theta^a = \Theta^f + \mathbf{K}(d - \mathbf{H}\Theta^f), \quad (1)$$

and the updated error covariance matrix is

$$\mathbf{C}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{C}^f. \quad (2)$$

Here \mathbf{I} is the identity matrix, \mathbf{C}^f is the prior model covariance matrix for Θ , and \mathbf{K} is the Kalman gain matrix given by

$$\mathbf{K} = \mathbf{C}^f \mathbf{H}(\mathbf{H}\mathbf{C}^f \mathbf{H}^T + \mathbf{C}_d)^{-1},$$

where \mathbf{C}_d is the measurement error covariance matrix.

The EnKF assumes a sample of size N_e from the forecast distribution. We denote the sample $\{\Theta_k^f\}$. Then, given a forecast ensemble member Θ_k^f , the EnKF algorithm generates a corresponding ensemble of synthetic data (by adding random perturbations, ϵ_k to the data, d , by drawing from a normal distribution with mean zero and covariance matrix \mathbf{C}_d), and updates Θ_k^f according to

$$\Theta_k^a = \Theta_k^f + \mathbf{K}(d + \epsilon_k - \mathbf{H}\Theta_k^f). \quad (3)$$

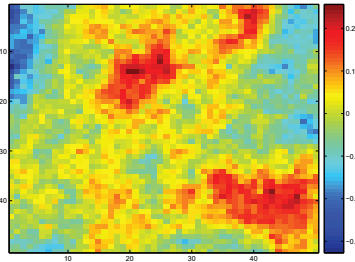


Figure 1: Reference field

Parameter Name	Value
System dimension, n	2500 (50 × 50 grid)
Model uncertainty, σ_m^2	0.03
Variogram model	Spherical
Correlation length	25 grid cells

Table 1: Input data

In EnKF a sample estimate of prior covariance, $\overline{\mathbf{C}^f}$ based on $\{\Theta_i^f\}$ is used, and the update step (3) is implemented by replacing \mathbf{K} by a sample Kalman gain matrix, \mathbf{K}_e . The approximation of the analyzed error covariance matrix with mean state $\overline{\Theta}^a$ can be expressed as

$$\overline{\mathbf{C}^a} = \frac{1}{(N_e - 1)} \{(\Theta_1^a - \overline{\Theta}^a) \dots (\Theta_{N_e}^a - \overline{\Theta}^a)\} \{(\Theta_1^a - \overline{\Theta}^a) \dots (\Theta_{N_e}^a - \overline{\Theta}^a)\}^T. \quad (4)$$

It is seen that the EnKF may be considered as a Gaussian simulation with Kriging update. The method has been extended to solve the combined parameter and state estimation problem by expanding the state vector to include also the unknown parameters to be estimated like porosity or log-permeability in petroleum applications, see e.g. Refs. [3, 1, 10].

If $\overline{\mathbf{C}^f}$ is a consistent estimator of \mathbf{C}^f , Slutsky's theorem can be used to show that the EnKF converges to Kalman filter as the ensemble size N_e approaches infinity, see e.g., [4]. However, for small N_e and large system dimension, $\overline{\mathbf{C}^f}$ will likely be a poor estimate of \mathbf{C}^f , and the influence of sample variability in \mathbf{C}^f will adversely affect the sample-based EnKF update and hence, spurious correlation effect may occur.

Description of Synthetic Experiments

We have performed a systematic study of the validity on the standard EnKF update by considering a single update of a very simple linear model. A “reference” model and a “forecast” model is taken as two realizations of a Gaussian random field with mean zero and covariance matrix $\mathbf{C}^f = \sigma_m^2 \times \mathbf{R}$, where σ_m^2 is model error variance. The correlation matrix, \mathbf{R} , is based on a spherical variogram. An example field is shown in Fig. 1. Then measurements are taken from the reference model and used to update the forecast model using both KF and EnKF. The problem then corresponds to a traditional, simple Kriging, and \mathbf{H} will contain only zero's and one's, with one's at measurement locations and zero elsewhere. The prior ensemble for EnKF is generated by adding realizations of the same Gaussian field to the forecast model. The measurement perturbations are generated from a Gaussian random field with mean zero and covariance matrix $\mathbf{C}^d = \sigma_d^2 \times \mathbf{R}$. To make sure that the measurements are consistent, the same correlation model was used as for the model covariance. All the measurements are located in the lower left corner of the grid for easier investigation of the long-range correlations. Input data are listed in Table 1.

In order to quantify the numerical findings, we have compared the updated states (equation no. 1 and 3) and updated covariance matrices (equation no. 2 and 4) for both

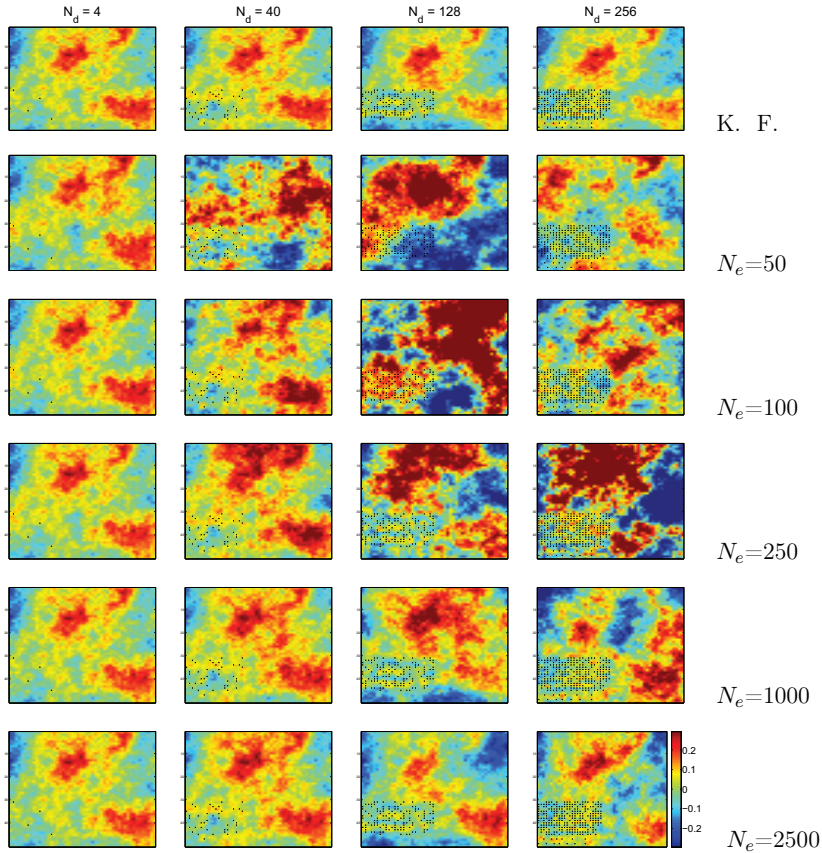


Figure 2: Updated fields vs. N_d and N_e . $\sigma_d = 0.001$. The same color bar is used in all plots

Kalman filtering and EnKF update using the root mean square (RMS) error norm

$$\|\Theta\|_{rms} = \sqrt{\sum_{i=1}^n (\bar{\Theta}^a(i) - \Theta^a(i))^2 / n}, \quad (5)$$

and the Frobenius matrix norm

$$\|\mathcal{C}\|_{frob} = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (\bar{\mathcal{C}}^a(ij) - \mathcal{C}^a(ij))^2 / n^2}. \quad (6)$$

Results

Fig. 2 delineates the degradation of the updated property fields with increased number of measurements. The measurement error is kept constant equal to 0.001. Each column of the figure show updated fields for a given value of N_d , where the first row shows the results with Kalman filtering, and the other rows show the variation with N_e . Measurements locations are marked with black dots. For $N_d = 4$, the updated field is reasonably

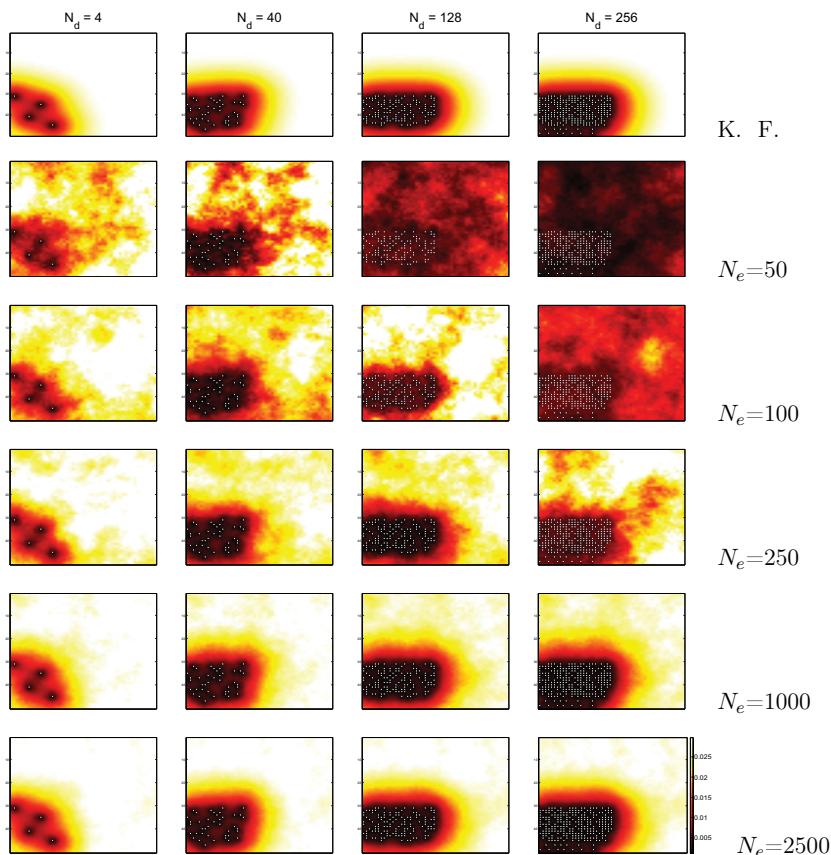


Figure 3: Posterior variances vs. N_d and N_e . $\sigma_d = 0.001$. The same color bar is used in all plots

accurate even with $N_e = 50$, which is considerably less than the “standard” ensemble size, $N_e = 100$. However, the scenario starts to change when N_d increases. For $N_d = 40$, we clearly see that the field far from the measurement locations has been incorrectly updated due to spurious correlations. In general, based on these and other cases not shown, we observe that the number of ensemble members has to be much larger (ten times or more) than the number of measurements to avoid the spurious correlations and obtain a good solution.

The posterior variances (given by the diagonal elements of the updated covariance) are plotted in Fig. 3, similarly. Considering the Kalman filter variance, we see that, as expected, there is very low uncertainty at the measurement locations. On the other hand, the variance approaches the model uncertainty, which in this case is equal to 0.03, far from the measurement locations. With EnKF, the estimated variance around the measurement locations seems OK, but far from the measurements, again a very large number of ensemble members are needed to obtain the correct variance. For smaller ensemble size, the estimated variances are typically too small.

The relative RMS norm, i.e., the norm of the difference between the EnKF and KF updates normalized with the KF update, is plotted versus the ensemble size, N_e , for an increasing number of measurements, N_d , in Fig. 4(a). Here the values for small N_e is

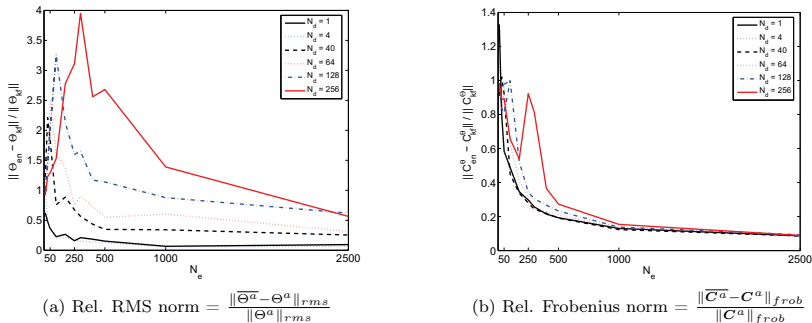


Figure 4: Relative Error Norms vs. ensemble size for different number of measurements, $\sigma_d^2 = 0.001$.

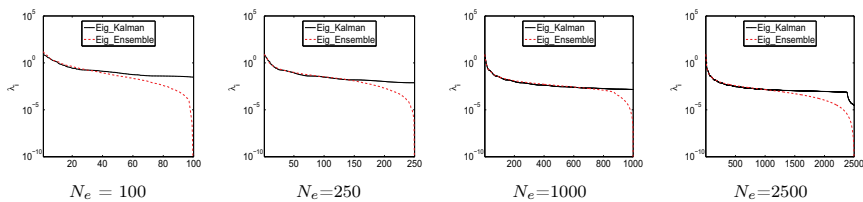


Figure 5: Eigen-spectrum of Covariance Matrices, $N_d = 128$

influenced by the random seed and should be disregarded. In accordance with Figs. 2 and 3, the RMS norm quickly approaches a very small value as we increase N_e when the number of measurements is low, e.g., $N_d = 1, 4$. In the EnKF literature there has been a considerable focus on the loss of rank which may occur when there are more measurements than ensemble members ($N_d \geq N_e$) [9, 3]. However, here we clearly see that the problem with spurious correlations and an inaccurate update will occur for a much lower number of measurements, and no clear transition in the accuracy is seen at $N_e = N_d$ in neither the updated fields, Fig. 2, nor the estimated variance, Fig. 3. When N_d increases, a very large ensemble size ($N_e \gg N_d$) is needed to obtain a low relative RMS value. For $N_d = 128$ and 256, the relative norm is still above 0.5 for an ensemble size of 2500.

Fig. 4(b) shows the relative Frobenius norm of the difference between the updated covariances matrix for Kalman filter and EnKF. Again, we see that a very large ensemble is needed to reach convergence. However, as opposed to the RMS plot, the curves are not sensitive to the number of measurements. The reason for this is not clear, but it indicates that the Frobenius norm is not a good measure on the accuracy of the covariance matrix estimate for the purpose of EnKF.

Fig. 5 shows the eigen-spectrum of the covariance matrix. The largest 15 eigenvalues are plotted in Fig. 6. Investigations have been performed with all the cases $N_d = 1, 4, 40, 64, 128, 256$, but we report here the case with $N_d = 128$ as all other cases show the similar pattern of results.

Fig. 5 clearly shows how the smaller eigenvalues deviates from the exact spectrum and drops to zero (from Eq. 4 it is clear that the rank cannot be larger than $N_e - 1$). The low rank approximation to the covariance matrix results in a reduction of variance for EnKF, insufficient projection of the variance in the direction of the trailing eigenvectors, and a reduction of the search space of the model parameters. This may lead to poor

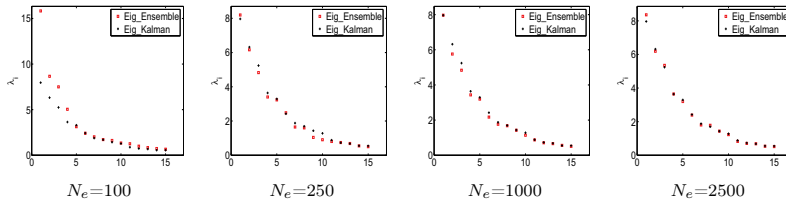


Figure 6: 15 largest eigenvalues, $N_d = 128$

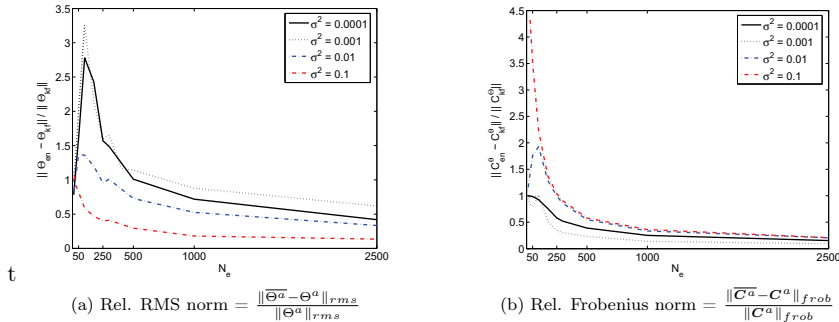


Figure 7: Relative Error Norms vs. ensemble size for different measurement error, $N_d = 128$.

parameter estimation and potentially filter divergence [2]. As expected, the approximate spectrum approaches the true spectrum when the ensemble size increase. The long-range correlations are determined by the largest eigenvector-eigenvalue pairs. It is interesting to note, however, that although the largest eigenvalues are well approximated for $N_e = 250$, spurious long-range correlations are still clearly seen in the corresponding plot of Figs. 2 and 3. Although not shown, we have also checked the eigenvectors, and also the largest eigenvectors seem to be well approximated in this case.

Fig. 7 shows the relative RMS error and the associated relative Frobenius matrix norm of the updated covariance matrix vs. ensemble size for different value of the measurement error. We report here only the case with $N_d = 128$ as the other cases show almost similar trends in result. As expected, the error is larger for smaller σ_d^2 , and again a very large ensemble is needed to reach a convergence of the RMS. Note that the behavior of the Frobenius norm is opposite to the RMS plot, i.e., the smaller the σ_d^2 , the larger the Frobenius norm, again indicating that this norm is not a good measure of the accuracy of the covariance matrix with respect to EnKF update.

Fig. 8 is intended to show the ensemble collapse effect with small measurement uncertainty. The top row of the each of the sub-plots shows the ensemble of property estimates in a row of grid cells far from the measurement locations; the bottom row shows estimates in a row of grid cells through the measurement locations. Also we have plotted in each case the true Kalman filter update (red line) along with the ensemble mean (green line). We see that as the measurement error is reduced (for a given ensemble size), all the ensemble members tend to collapse into a single solution with almost no spread, i.e., no random variation at all. This situation may be avoided by increasing the ensemble size. However, we see that by increasing σ_d^2 , we may retain the variability of the ensemble members and still obtain a reasonably good estimate of the mean. This shows that ensemble collapse can be avoided if we have large enough σ_d^2 . Although the estimated variance will be incorrect, the mean may still be OK.

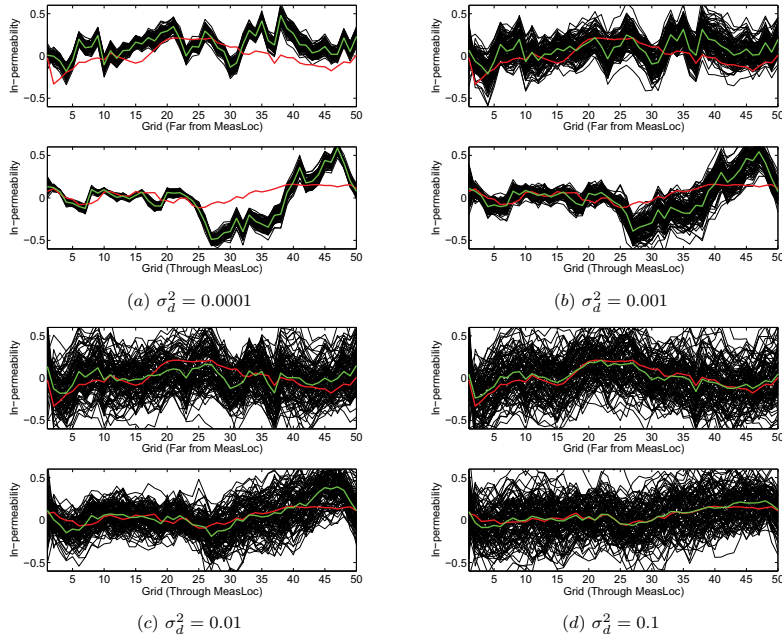


Figure 8: Ensemble Members on Grids: $N_d = 128$, $N_e = 100$, Exp. 2

Concluding Remarks

We have performed a series of numerical experiments with a simple, linear model in order to investigate the validity of the straightforward EnKF algorithm with an increased number of measurements. All the experiments clearly demonstrated that EnKF solution gets worse with increased number of measurements and $N_e \gg N_d$ is required for an accurate solution. Spurious, long range correlations are clearly seen when the number of measurements increases. Also ensemble collapse is demonstrated in cases with small measurement uncertainty.

References

- [1] AANONSEN, S. I., NÆVDAL, G., OLIVER, D. S., REYNOLDS, A. C., AND VALLÈS, B., *Review paper: Ensemble kalman filter to petroleum engineering*. Submitted to SPE Journal, 2008.
- [2] DEVEGOWDA, D., ARROYO, E., DATTA-GUPTA, A., *Efficient and robust reservoir model updating using ensemble kalman filter with sensitivity based covariance localization*, SPE 106144 presented at the SPE Reservoir Simulation Symposium, Houston, TX, (2007).
- [3] EVENSEN, G., *Data Assimilation: The Ensemble Kalman Filter*. Springer, Berlin, 2007.
- [4] FURRER, R., BENGTTSON, T., *Estimation of high-dimensional prior and posterior covariance matrices in kalman filter variants*, Journal of Multivariate Analysis 98(2007), (2006).

- [5] GASPARI, G. AND COHN, S. E., *Construction of correlation functions in two and three dimensions*, Quarterly Journal of the Royal Meteorological Society, 125 (1999), pp. 723–757.
- [6] HOUTEKAMER, P. L. AND MITCHELL, H. L., *Data assimilation using an ensemble Kalman filter technique*, Monthly Weather Review, 126 (1998), pp. 796–811.
- [7] HOUTEKAMER, P. L., MITCHELL, H. L., *A sequential ensemble kalman filter for atmospheric data assimilation*, Monthly Weather Review, 129(1), (2001).
- [8] KALMAN, R. E., *A New Approach to Linear Filtering and Prediction Problems*, Transactions of the ASME-Journal of Basic Engineering, 82(Series D), (1960).
- [9] J. D. KEPERT, *On ensemble representation of the observation-error covariance in the Ensemble Kalman Filter*, Ocean Dynamics, 54 (2004), pp. 561–569.
- [10] NAEVDAL, G., JOHNSEN L. M., AANONSEN, S. I., VERFRING, E. H., *Reservoir Monitoring and Continuous Model Updating Using Ensemble Kalman Filter*, SPE Journal 10, (2005).
- [11] SKJERVHEIM, J.-A., EVENSEN, G., AANONSEN, S. I., RUUD, B., JOHANSEN, T. A., *Incorporating 4D Seismic Data in Reservoir Simulation Models Using Ensemble Kalman Filter*, SPE Annual Technical Conference and Exhibition, Dallas, Texas, USA, (2005).

