DELFT UNIVERSITY OF TECHNOLOGY

REPORT 08-16

Comparison of the Ensemble Kalman Filter and a modified Representer Method for sensitivity to prior data

J.R. Rommelse, J.D. Jansen, A.W. Heemink, F. Wilschut and O. Kleptsova

ISSN 1389-6520

Reports of the Department of Applied Mathematical Analysis

Delft 2008

Comparison of the Ensemble Kalman Filter and a modified Representer Method for sensitivity to prior data

J.R.Rommelse [*] J.D. Jansen [†] A.W. Heemink [‡] F. Wilschut [§]
O. Kleptsova [¶]

July 14, 2008

Abstract

In closed-loop reservoir management, data assimilation algorithms or automated history matching methods can be used to support decision-making tools. Due to prior information, which the user has to supply to the data assimilation procedure, it can never be fully automatic. For example, a statistical description of the reservoir properties must be formulated, as well as the uncertainty of the measurements. In a synthetic numerical experiment these weighting factors are known; in a field application they are not. One could say that there are not only uncertainties in the data and the reservoir properties, but there is also "uncertainty in the uncertainty".

In this article the robustness of the Ensemble (Square Root) Kalman Filter and a gradient-based algorithm using Representer Expansions are compared with respect to prior input data. Some algorithm dependent settings are explored to try to make the filter reproduce the results of the Representer Method: the ensemble size, the initialization method and the Kalman update. The concept of assimilating data more then once with dampened weighting factors (added uncertainty) is introduced.

First the equations that underlie the Ensemble Filter and the Representer Method are given. Then numerical experiments are presented and two measures of quantifying the success of the methods are introduced. According to one such measure, the Representer Method performed better for all numerical examples considered. The parameters of the filter can be chosen such that the filter with the correct input data is equally successful as the Representer Method, using the second measure. When the methods are fed with the wrong prior input, the second measure also favours the Representer Method, so for the examples considered in this article the Representer Method is less sensitive to "wrong" user input that the Kalman Filter.

[*]Delft University of Technology & Alten Nederland, E-mail: J.R.Rommelse@tudelft.nl
[†]Delft University of Technology & Shell International Exploration & Production, E-mail: J.D.Jansen@TUDelft.nl
[‡]Delft University of Technology, E-mail: A.W.Heemink@TUDelft.nl
[§]TNO, E-mail: frank.wilschut@tno.nl
[¶]Delft University of Technology, E-mail: O.Kleptsova@TUDelft.nl

1 Introduction

Section 1.1 introduces the concept of running a reservoir simulator in strong (deterministic) or weak (stochastic) mode and its notation. The relationship between the Ensemble Filter and the Representer Method is discussed in section 1.2 as different approximations of Bayes rule.

1.1 Strong and weak constraint reservoir simulation

Running a reservoir simulator with incorrect model parameters, like permeability or porosity, causes a discrepancy between observations and measurement forecasts. There is an additional phenomenon that causes the discrepancy; the model is an approximation, so even if the parameters were known, then the model would still produce incorrect output. For simple numerical experiments these model errors can be explicitly set to zero. The reservoir simulator is then used deterministically. For more realistic experiments or field applications, the model errors can be modelled as additional parameters, which are sampled by a Monte Carlo method or estimated along with the other model parameters. In this case the reservoir simulator is used in stochastic mode. In variational or gradient-based data assimilation, the terms deterministic and stochastic mode are denoted by using the reservoir simulator as a strong or as a weak constraint respectively.

Running the reservoir simulator and predicting measurements can be written in state-space notation as

$$\mathbf{x}_i = \mathbf{M}\left(\mathbf{x}_{i-1}\right) \qquad , \qquad \mathbf{y} = \mathbf{H}\left(\mathbf{x}\right), \tag{1}$$

where $\mathbf{x}_i, i \in \{0, \cdots, n\}$ contains the state variables (typically pressures and saturations) at time t_i. The dependence of the forward model operator $\mathbf{M}$ on the model parameters, the model errors and the control parameters is omitted. In this article, running the reservoir simulator is denoted by

$$\mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right) = \mathbf{0}, \tag{2}$$

where the model parameters (for example permeabilities and porosities) are collected in the vector $\mathbf{a}$ and the model errors on interval $[t_{i-1}, t_i]$ are contained in the vector ε_i. In fact, the mass balance and Darcy equations [Aziz and Settari, 1979]

$$\frac{d}{dt}\left(\mathbf{f}_1\left(\mathbf{x}\right)\right) = \mathbf{f}_2\left(\mathbf{x}, \mathbf{a}\right), \tag{3}$$

are discretized in time using a the implicit Euler scheme, resulting in a deterministic reservoir simulator:

$$\mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \mathbf{0}\right) = \mathbf{f}_1\left(\mathbf{x}_i\right) - \left(t_i - t_{i-1}\right)\mathbf{f}_2\left(\mathbf{x}_i, \mathbf{a}\right) - \mathbf{f}_1\left(\mathbf{x}_{i-1}\right) = \mathbf{0}. \tag{4}$$

The presence of water and oil mass in the grid blocks is described by $\mathbf{f}_1$, while $\mathbf{f}_2$ models the flow through the grid block interfaces and the injection/production of fluids in the wells. The model errors are introduced as additional sources/sinks in all grid blocks. In other

words; after $\mathbf{x}_i$ has been solved from Eq. (4), the water and oil masses in the grid blocks have not correctly been predicted and must still be modified. The prediction gets worse as the time step $(t_i - t_{i-1})$ gets larger. Therefore the correction is modelled proportional to $(t_i - t_{i-1})$. If the additional sources become too strong, then unrealistically high pressures will be observed. If the additional sinks become too strong, then saturations outside $[0, 1]$ will occur. In this article, the additional sinks are non-linearly constrained by $\mathbf{f}_1$. The stochastic reservoir simulator has the form

$$\begin{aligned} &\mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right) \\ &= \mathbf{f}_1\left(\mathbf{x}_i\right) - \left(t_i - t_{i-1}\right)\mathbf{f}_2\left(\mathbf{x}_i, \mathbf{a}\right) - \mathbf{f}_1\left(\mathbf{x}_{i-1}\right) + \min\left\{\mathbf{f}_1\left(\mathbf{x}_i\right), \left(t_i - t_{i-1}\right)\varepsilon_i\right\} = \mathbf{0}. \end{aligned} \tag{5}$$

1.2 Bayesian data-assimilation

Reservoir simulation can be embedded in a stochastic or probabilistic framework. In that case the reservoir state variables (pressures and saturations in all grid blocks) do not have deterministic values, but are described by a multivariate probability distribution (or density) function (PDF). The stochastic nature of the state variables is caused by the uncertainty in the initial states, the uncertainty in the model parameters (permeability, porosity, etc. ...) and the fact that the reservoir simulator is imperfect (maybe gravity or capillary effects were not modelled or 3 components were modelled where 5 would have been more appropriate). The uncertainty in the measurements are caused by two effects; sensors try to monitor a stochastic quantity and are subject to influences that might damage them or otherwise corrupt the data.

Bayes theorem for continuous probability distributions [Bayes, 1763]

$$f\left(\mathbf{a}|\mathbf{y}\right) = \frac{f\left(\mathbf{y}|\mathbf{a}\right) f\left(\mathbf{a}\right)}{f\left(\mathbf{y}\right)} \qquad , \qquad f\left(\mathbf{a}|\mathbf{y}\right) \propto f\left(\mathbf{y}|\mathbf{a}\right) f\left(\mathbf{a}\right) = f\left(\mathbf{a}, \mathbf{y}\right), \tag{6}$$

states that the *posterior* density $f\left(\mathbf{a}|\mathbf{y}\right)$ (the probability of the model parameters given the data) is proportional to the *prior* density $f\left(\mathbf{a}\right)$ (the probability of the model parameters) multiplied by the *likelihood* $f\left(\mathbf{y}|\mathbf{a}\right)$ of the data given the model parameters. This basically means that the reservoir simulator and a measurement model must be used to construct the joint density of the parameters and the measurement forecasts $f\left(\mathbf{a}, \mathbf{y}\right)$. Then the domain is restricted by substituting the observations, which causes the integral over the new function no longer to be equal to 1. Renormalization then gives the posterior density. The hidden dependence of the measurement forecasts on the reservoir states $\mathbf{x}$ can also be explicitly written in Bayes theorem

$$f\left(\mathbf{a}, \mathbf{x}|\mathbf{y}\right) = \frac{f\left(\mathbf{x}, \mathbf{y}|\mathbf{a}\right) f\left(\mathbf{a}\right)}{f\left(\mathbf{y}\right)} \qquad , \qquad f\left(\mathbf{a}, \mathbf{x}|\mathbf{y}\right) \propto f\left(\mathbf{x}, \mathbf{y}|\mathbf{a}\right) f\left(\mathbf{a}\right) = f\left(\mathbf{a}, \mathbf{x}, \mathbf{y}\right). \tag{7}$$

For practical applications, it is hardly ever possible to compute Eq. (6) analytically, so approximations have to be made. For the special case where the prior $f\left(\mathbf{a}\right)$ is Gaussian

and the measurement forecasts linearly depend on the parameters (which requires that both the reservoir simulator and the sensor model are linear), the posterior $f(\mathbf{a}|\mathbf{y})$ is also Gaussian [Tarantola, 2005], which means that it suffices to only calculate the mean and covariance.

The Kalman Filter, section 2.1, exploits this property and sequentially updates the two statistical moments by performing time updates and measurement updates. In the non-linear case, the Ensemble Kalman Filter [Evensen, 2003] samples the prior and updates the samples until a sampled representation of the posterior is found. However, at every measurement update the density is again assumed to be Gaussian.

When the posterior is Gaussian, the logarithm is proportional to a weighted average of the model errors, the measurement errors and the parameter errors [Tarantola, 2005], [section 3.1]. This can be formulated as an objective that has to be minimized in order to find the mode of the posterior, which is equal to the mean for a Gaussian. This is done in variational methods. In the non-linear case, a mode of a probability distribution is found, but it is usually not clear how this distribution is related to the posterior. Only one mode is found if the posterior is multi-modal.

2 Ensemble Filter

2.1 EnKF

With increasing computer capacity, the Kalman filter [Kalman, 1960, Gelb, 1974] is becoming more popular for automated history matching. Implementing an Ensemble Kalman Filter (EnKF) [Evensen, 2003, Heemink et al., 2001] is easy and does not require modifications to existing reservoir simulators. With an Ensemble Kalman Smoother [Cohn et al., 1994], even time-lapse seismic data [Skjervheim et al., 2005] can be assimilated.

In the Ensemble Kalman Filter , the prior is represented by an ensemble of samples $\left\{\mathbf{a}^{(1)}, \cdots, \mathbf{a}^{(n)}\right\}$. This sampling can be done randomly or deterministically. In the absence of measurements, the samples are updated in the time update by the non-linear reservoir simulator

$$\mathbf{g}\left(\mathbf{x}_i^{(j)}, \mathbf{x}_{i-1}^{(j)}, \mathbf{a}^{(j)}, \varepsilon_i^{(j)}\right) = \mathbf{0}. \tag{8}$$

Whenever measurements are available, the measurement update is performed. Augmented ensemble members are created by concatenating the reservoir states, the measurement forecasts and (in case of parameter estimation) the parameters

$$\mathbf{l}_f^{(j)} = \begin{bmatrix} \mathbf{x}_i^{(j)} \\ \mathbf{y}_f^{(j)} \\ \mathbf{a}^{(j)} \end{bmatrix}, \tag{9}$$

where the measurement forecasts ensemble is obtained by running an ensemble of (non-linear) measurement models

$$\mathbf{y}_f^{(j)} = \mathbf{h}\left(\mathbf{x}_i^{(j)}\right). \tag{10}$$

The forecasted ensemble members are then split into a mean and a set of deviations

$$\mathbf{l}_f^{(j)} = \overline{\mathbf{l}_f} + \widetilde{\mathbf{l}}_f^{(j)}, \tag{11}$$

which are updated by

$$\mathbf{L}_f = \begin{bmatrix} \widetilde{\mathbf{l}}_f^{(1)} & \cdots & \widetilde{\mathbf{l}}_f^{(n)} \end{bmatrix} \quad , \quad \widetilde{\mathbf{L}} = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{L}_f, \tag{12}$$

$$\mathbf{K} = \mathbf{L}_f \widetilde{\mathbf{L}}^T \left(\widetilde{\mathbf{L}} \widetilde{\mathbf{L}}^T + (n-1)\, \mathbf{P}_y \right)^{-1}, \tag{13}$$

$$\overline{\mathbf{l}_a} = \overline{\mathbf{l}_f} + \mathbf{K} \left(\overline{\mathbf{y}}_o - \overline{\mathbf{y}}_f \right) \quad , \quad \mathbf{L}_a = \mathbf{L}_f + \mathbf{K} \left(\begin{bmatrix} \widetilde{\mathbf{y}}_o^{(1)} & \cdots & \widetilde{\mathbf{y}}_o^{(n)} \end{bmatrix} - \widetilde{\mathbf{L}} \right), \tag{14}$$

to form the analyzed mean and deviations that can be used to form the analyzed ensemble members. The Kalman gain matrix, Eq. (13), is built from the measurement forecasts autocovariance $\frac{\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^T}{n-1}$, the observations autocovariance $\mathbf{P}_y$ (which must be specified by the user) and the cross covariance of the measurement forecasts and the augmented forecasted ensemble members. The measurement errors $\widetilde{\mathbf{y}}_o^{(j)}$ are sampled using $\mathbf{P}_y$. The analyzed measurements are a weighted average of the measurement forecasts and the observations. The jump from the forecasted values to the analyzed values is linearly extrapolated to the reservoir states and the parameters using the cross covariances between them.

2.2 ESRKF

In the EnKF, the errors in the observations $\widetilde{\mathbf{y}}_o^{(j)}$ are sampled using $\mathbf{P}_y$. Without these errors, the uncertainty in the analyzed ensemble would be underestimated by

$$\mathbf{L}_a = \left(\mathbf{I} - \mathbf{K} \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix} \right) \mathbf{L}_f. \tag{15}$$

However, also deterministic solutions to this problem are available. In [Verlaan and Heemink, 1997], extra columns are added to Eq. (15). These columns are obtained as a square root $\mathbf{L}_y$ of $\mathbf{P}_y = \mathbf{L}_y \mathbf{L}_y^T$. After the augmentation, the number of columns is reduced to the original number by selecting the leading left-singular vectors. In the Ensemble Square Root Kalman Filter (ESRKF), Eq. (15) is used, but with a modified Kalman gain matrix which, in the one-dimensional case, can be written as

$$\widetilde{\mathbf{K}} = \frac{\mathbf{K}}{1 + \sqrt{\frac{(n-1)\mathbf{P}_y}{\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^T + (n-1)\mathbf{P}_y}}}. \tag{16}$$

For more dimensions, the modified Kalman gain matrix involves calculating the non-unique square roots of two matrices. Details can be found in [Whitaker and Hamill, 2002].

2.3 Duplicated measurements

As mentioned in section 2.1, the analyzed measurements are a weighted average of the measurement forecasts and the observations. The jump from the forecasted values to the analyzed values is linearly extrapolated to the reservoir states and the model parameters using the cross covariances between them. These are both linear operations which can cause problems when the measurement operator is non-linear. Furthermore, the weighting factors that are used for the averaging are the user-defined uncertainty in the observations and the covariance of the measurement forecasts. This second order statistical moment is not appropriate when the state variables are predicted by a non-linear simulator, even when it is obtained by a EnKF rather than an Extended Kalman Filter (EKF) [Evensen, 2003]. In fact, using a EnKF without extra modifications often causes saturation values outside $[0, 1]$ and unrealistic pressure and permeability estimates. A quick engineering trick is to apply a dampening factor to the jump from the forecasted values to the analyzed values, which is equivalent to adding extra uncertainty to the observations.

Since in practice it costs money to obtain data, it is a waste to discard data or to add extra uncertainty to it. Results in this article are obtained by duplicating measurements to assimilate them more than once, but with extra uncertainty. The extra uncertainty ensures that the jump from the forecasted values to the analyzed values, which is overestimated by the linear measurement update, is dampened. The duplication of measurements ensures that the jump is not dampened too much, so no valuable data are discarded. When a measurement is predicted with variance σ_f^2 and the uncertainty of the sensor is σ_o^2, then the measurement is assimilated n times with artificial uncertainty $\widetilde{\sigma}_o^2$ using two criteria:

1. The artificial sensor uncertainty $\widetilde{\sigma}_o^2$ is of the same magnitude as the measurement forecasts variance σ_f^2, so $\widetilde{\sigma}_o^2 \approx \sigma_f^2$.

2. The variance of the analyzed measurements is invariant under this method, so

$$\sigma_a^2 = \frac{\sigma_f^2 \sigma_o^2}{\sigma_f^2 + \sigma_o^2} = \frac{\sigma_f^2 \widetilde{\sigma}_o^2}{2^{n-1}\sigma_f^2 + \widetilde{\sigma}_o^2}. \tag{17}$$

Therefore, every measurement is used

$$n = 1 + \left\lceil \log_2 \left(\frac{\sigma_f^2 + \sigma_o^2}{2\sigma_o^2} \right) \right\rceil \tag{18}$$

times, with variance $\widetilde{\sigma}_o^2 = 2^{n-1}\sigma_o^2$. By construction, the variance of the analyzed measurements is identical to what it would be without duplicated measurements. However, the jump from the forecasted values to the analyzed values is not; the sum of a set of small jumps is less than one big jump.

3 Representer Method

The discrepancy between observed measurements and their model predicted antitheses can be formulated as an objective function that has to be minimized. This minimization process is regularized by penalizing the deviation of the estimated parameters from the prior parameters, and also the model errors can be added to the objective function, section 3.1. If the posterior distribution is Gaussian, then the mode (or mean) of the posterior minimizes the objective. If not, then the objective gets a different meaning, and the weighting factors might have to be changed accordingly; covariance matrices that are sufficient to represent Gaussians are no longer appropriate and the user has to supply an alternative. When the posterior is multi-modal, then the objective function has multiple local minima, so extra regularization is required. Calculating a gradient of the objective function with respect to the model parameters can be done efficiently using an adjoint reservoir simulator [Rommelse et al., 2007] in case the model errors are explicitly set to zero. Regularization is usually done by selecting a limited set of basis functions that are chosen as singular vectors of a root of the covariance matrix of the parameters. When model errors are taken into account, the reservoir simulator and its adjoint can no longer be used sequentially to produce a gradient. The Representer Method (RM) [Bennett and McIntosh, 1982, Eknes and Evensen, 1997, Bennett, 2002, Valstar et al., 2004, Baird and Dawson, 2005, Janssen et al., 2006, Przybysz et al., 2007] can then be used to solve the weak constraint minimization problem. Regularization is also included in the method, and the basis functions are data-driven, rather than specified by the user. Unfortunately the computational cost is proportional to the number of measurements times the computational cost of solving the strong constraint problem using the adjoint system.

The RM used in this article was introduced in [Rommelse et al., 2007] and differs from previous implementations. It produces a gradient that can be used in any minimization algorithm (with minor modifications); here Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) [Gao and Reynolds, 2006, Ulbrich, 2002, Rommelse et al., 2007] is used. The 1-1-relationship between isolated measurements in time and space, and the representer functions is abandoned, making the method computationally more attractive. Also, the concept of measurement representers is introduced to deal with non-linear measurement models and a new linearization is used that does no longer require the calculation of special correction terms.

3.1 Objective function and derivatives

The objective that has to be minimized is formulated as

$$J = \frac{1}{2}\left(\mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right) - \mathbf{m}\right)^T P_{\mathbf{m}}^{-1}\left(\mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right) - \mathbf{m}\right) + \tag{19}$$
$$+ \frac{1}{2}\left(\mathbf{a} - \mathbf{a}^{prior}\right)^T \mathbf{P}_{\mathbf{a}}^{-1}\left(\mathbf{a} - \mathbf{a}^{prior}\right) + \frac{1}{2}\sum_{i=1}^{n}\varepsilon_i^T P_{\varepsilon_i}^{-1}\varepsilon_i +$$
$$+ \sum_{i=1}^{n}\lambda_i^T \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right).$$

If the reservoir simulator $\mathbf{g}$ and the sensor model $\mathbf{h}$ were linear and the prior were Gaussian, then the weighting factors $\mathbf{P_a}$, $\mathbf{P}_{\varepsilon_i}$ and $\mathbf{P_m}$ could be chosen equal to the model parameter uncertainty, the model error uncertainty and the sensor uncertainty respectively, in order to find the same solution that a EnKF would find. Otherwise different factors might have to be chosen. The observations are contained in the vector $\mathbf{m}$. In general, the initial reservoir states need to be estimated along with the permeability, porosity and other parameters, $\mathbf{x}_0 = \mathbf{f}(\mathbf{a})$, but in this article they are assumed known.

The derivatives of Eq. (19) with respect to λ_i, ε_i, $\mathbf{x}_i$ and $\mathbf{a}$ are

$$\left(\frac{\partial J}{\partial \lambda_i}\right)^T = \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right), \tag{20}$$

$$\left(\frac{\partial J}{\partial \varepsilon_i}\right)^T = P_{\varepsilon_i}^{-1}\varepsilon_i + \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \varepsilon_i}\right)^T \lambda_i, \tag{21}$$

$$\left(\frac{\partial J}{\partial \mathbf{x}_i}\right)^T = \left(\frac{\partial \mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right)}{\partial \mathbf{x}_i}\right)^T P_{\mathbf{m}}^{-1}\left(\mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right) - \mathbf{m}\right) + \tag{22}$$
$$+ \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \mathbf{x}_i}\right)^T \lambda_i + \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_{i+1}, \mathbf{x}_i, \mathbf{a}, \varepsilon_{i+1}\right)}{\partial \mathbf{x}_i}\right)^T \lambda_{i+1},$$

$$\left(\frac{\partial J}{\partial \mathbf{a}}\right)^T = \mathbf{P}_{\mathbf{a}}^{-1}\left(\mathbf{a} - \mathbf{a}^{prior}\right) + \sum_{i=1}^{n}\left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \mathbf{a}}\right)^T \lambda_i. \tag{23}$$

For the strong constraint case, where ε_i is explicitly set to $\mathbf{0}$, the gradient of the objective function with respect to the model parameters, $\left(\frac{\partial J}{\partial \mathbf{a}}\right)^T$, can be calculated using Eq. (23), where the model states $\mathbf{x}_i$ and adjoint states λ_i follow from sequentially solving Eq. (20) and Eq. (22) with the left-hand sides, $\left(\frac{\partial J}{\partial \lambda_i}\right)^T$ and $\left(\frac{\partial J}{\partial \mathbf{x}_i}\right)^T$, set to zero. Eq. (21) does not need to be used; instead $\varepsilon_i = \mathbf{0}$ is used.

To interface with gradient-based minimization algorithms, a numerical routine must be implemented that evaluates J, Eq. (19), and $\left(\frac{\partial J}{\partial \mathbf{a}}\right)^T$, Eq. (20), Eq. (22) and Eq. (23). If orthogonal basis functions are chosen for the regularization and placed as columns in the matrix $\mathbf{Q}$, then $\mathbf{Q}\mathbf{Q}^T\left(\frac{\partial J}{\partial \mathbf{a}}\right)^T$ is the regularized gradient.

3.2 Local minimizer

In a stationary point of Eq. (19), all gradients are equal to zero, so

$$\mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right) = \mathbf{0}, \tag{24}$$

$$\varepsilon_i = -P_{\varepsilon_i}\left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \varepsilon_i}\right)^T \lambda_i, \tag{25}$$

$$\left(\frac{\partial \mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right)}{\partial \mathbf{x}_i}\right)^T P_{\mathbf{m}}^{-1}\left(\mathbf{m} - \mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right)\right) \tag{26}$$

$$= \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \mathbf{x}_i}\right)^T \lambda_i + \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_{i+1}, \mathbf{x}_i, \mathbf{a}, \varepsilon_{i+1}\right)}{\partial \mathbf{x}_i}\right)^T \lambda_{i+1},$$

$$\mathbf{a} = \mathbf{a}^{prior} - \mathbf{P}_{\mathbf{a}} \sum_{i=1}^{n} \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \mathbf{a}}\right)^T \lambda_i. \tag{27}$$

Calculating and regularizing a gradient as it was done in section 3.1 is no longer possible, since the forward model Eq. (24) and the adjoint model Eq. (26) are now coupled because the model errors ε_i^s are no longer equal to zero; they are related to the adjoint states λ_i^s by Eq. (25).

3.3 Representer expansions

Since it is not possible to directly obtain a gradient of the weak constraint minimization problem and then regularize it, the Representer Method was introduced to regularize Eq. (24) through Eq. (27) and then find a minimization scheme. Unlike in section 3.1, where the basis functions are chosen by the user, in the RM the regularization is done symbolically and then substituted into Eq. (24) through Eq. (27) to obtain the basis functions. In [Valstar et al., 2004, Przybysz et al., 2007] the number of basis functions is chosen equal to the number of measurements, such that

$$\mathbf{x}_i = \mathbf{x}_i^{forecast} + \mathbf{x}_i^{correction} + \mathbf{X}_i \mathbf{b}, \tag{28}$$

$$\lambda_i = \mathbf{\Lambda}_i \mathbf{b}, \tag{29}$$

$$\mathbf{a} = \mathbf{a}^{prior} + \mathbf{A}\mathbf{b}, \tag{30}$$

where the columns of $\mathbf{X}_i$, $\mathbf{\Lambda}_i$ and $\mathbf{A}$ are the basis functions, called the state representers, adjoint representers and parameter representers. The representer coefficients are contained in the vector $\mathbf{b}$.

In [Rommelse et al., 2007] a different regularization is introduced. The minimization algorithm is started with $\mathbf{a}^{init} = \mathbf{a}^{prior}$ and $\varepsilon_i^{init} = \lambda_i^{init} = \mathbf{0}$. Applying these initial conditions to Eq. (24) gives the initial system states $\mathbf{x}_i^{init}$. These are not initial in the sense

that they are defined at t_0, but they are initial in the minimization algorithm. Given the initial states, also initial measurements can be predicted, $\mathbf{h}\left(\mathbf{x}^{init}_{\{0,\cdots,n\}}\right)$. The causes why the variables move away from their prior are parameterized by $\mathbf{b}$. The 1-1-relationship between one such cause and an isolated measurement in space and time is abandoned. For computational purposes, it is interesting to assume that the number of parameters in the vector $\mathbf{b}$ is (much) smaller than the number of measurements. The deviations from the priors are decomposed as

$$\mathbf{x}_i - \mathbf{x}_i^{prior} = \mathbf{X}_i \mathbf{b}, \tag{31}$$

$$\lambda_i = \mathbf{\Lambda}_i \mathbf{b}, \tag{32}$$

$$\mathbf{a} - \mathbf{a}^{prior} = \mathbf{A}\mathbf{b}, \tag{33}$$

$$\mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right) - \mathbf{h}\left(\mathbf{x}^{prior}_{\{0,\cdots,n\}}\right) = \mathbf{H}\mathbf{b}. \tag{34}$$

The columns of $\mathbf{H}$ contain the measurement representers.

3.4 Representer equations

After substitution of Eq. (31), Eq. (32) and Eq. (33) into Eq. (24), Eq. (26) and Eq. (27) and some manipulation [Rommelse et al., 2007], the representer equations are obtained. These can be used to obtain a regularized gradient for the weak constraint minimization problem:

1. If the model errors ε_i have not yet been approximated by a previous iteration of the minimization algorithm, initialize them to zero. Given the model parameters $\mathbf{a}$ and the model errors ε_i, run the non-linear model Eq. (24). After the model has been run for the first time, save the prior states $\mathbf{x}_i^{prior}$ and forecast the prior measurements $\mathbf{h}\left(\mathbf{x}^{prior}_{\{0,\cdots,n\}}\right)$.

2. Choose $\mathbf{Q}$ [Rommelse et al., 2007].

3. Calculate the adjoint representers $\mathbf{\Lambda}_i$ from

$$\begin{aligned} &\left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \mathbf{x}_i}\right)^T \mathbf{\Lambda}_i \\ &= \left(\frac{\partial \mathbf{h}\left(\mathbf{x}_{\{0,\cdots,n\}}\right)}{\partial \mathbf{x}_i}\right)^T \mathbf{Q} - \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_{i+1}, \mathbf{x}_i, \mathbf{a}, \varepsilon_{i+1}\right)}{\partial \mathbf{x}_i}\right)^T \mathbf{\Lambda}_{i+1}. \end{aligned} \tag{35}$$

4. Calculate the parameter representers $\mathbf{A}$ from

$$\mathbf{A} = -\mathbf{P}_{\mathbf{a}} \sum_{i=1}^{n} \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i^s, \mathbf{x}_{i-1}^s, \mathbf{a}^s, \varepsilon_i^s\right)}{\partial \mathbf{a}^s}\right)^T \mathbf{\Lambda}_i. \tag{36}$$

5. Calculate the state representers $\mathbf{X}_i$ from

$$\begin{aligned}&\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \mathbf{x}_i}\mathbf{X}_i + \frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \mathbf{x}_{i-1}}\mathbf{X}_{i-1} + \frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \mathbf{a}}\mathbf{A} \\ &= \frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \varepsilon_i} P_{\varepsilon_i} \left(\frac{\partial \mathbf{g}\left(\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{a}, \varepsilon_i\right)}{\partial \varepsilon_i}\right)^T \mathbf{\Lambda}_i.\end{aligned} \tag{37}$$

6. Calculate the measurement representers $\mathbf{H}$ from

$$\mathbf{H} = \sum_{i=0}^{n} \frac{\partial \mathbf{h}\left(\mathbf{x}^s_{\{0,\cdots,n\}}\right)}{\partial \mathbf{x}^s_i}\mathbf{X}_i. \tag{38}$$

7. Calculate new representer coefficients $\mathbf{b}$ as the least-squares solution given by

$$\left(\mathbf{H} + P_{\mathbf{h}}\mathbf{Q}\right)\mathbf{b} = \mathbf{m} - \mathbf{h}\left(\mathbf{x}^{prior}_{\{0,\cdots,n\}}\right). \tag{39}$$

8. Calculate new adjoint states Eq. (32).

9. Calculate the model errors Eq. (25).

10. Calculate the regularized gradient of the weak constraint minimization problem

$$\left(\frac{\partial J}{\partial \mathbf{a}}\right)^T = \mathbf{a} - \mathbf{a}^{prior} - \mathbf{A}\mathbf{b}. \tag{40}$$

Comments:

- In the routine that calculates the gradient, also the model errors are updated. For minimization algorithms that calculate more than one gradient per iteration, there must be a mechanism in the gradient routine that can store the model errors as well as the temporary changes. The changes must be made definite whenever that is signaled by the minimization algorithm, for example after a successful line search. This requires a minor update to the minimization routine, meaning that third-party software libraries can only be used if the source code is also available.

- By setting the model errors equal to zero and ignoring steps 8 and 9, the RM can also be used to obtain a regularized gradient for the strong constraint minimization problem.

4 Numerical experiments

Experiments were done on a 2D 2phase waterflooding application with wells in a 5-spot configuration. Water is injected at a rate of one pore volume per year and the production wells are constrained to 0.25 pore volumes per year. The state of the reservoir is described by pressure and water saturation in all 21x21x1 grid blocks of 10x10x20 m. Capillary pressure is ignored, as well as gravity effects. Synthetic data is generated by picking one realization out of a database of 1000 realizations as the "true" permeability, Fig. (1), and running a reservoir simulation with model errors that are sampled as white noise, Fig. (2). The pressure and saturation responses in the wells after simulating with these model errors are shown in Fig. (3). This figure also shows 10 synthetic pressure measurements in all wells at 100 and 200 days of simulation. The other realizations from the database are used to construct a covariance matrix that is used in the objective function that has to be minimized.

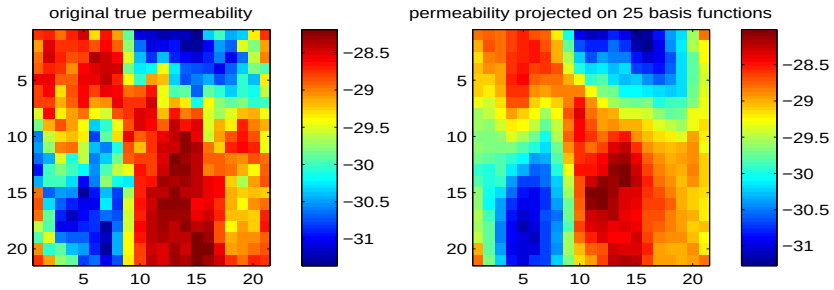

Figure 1: True permeability $[\ln(m^2)]$

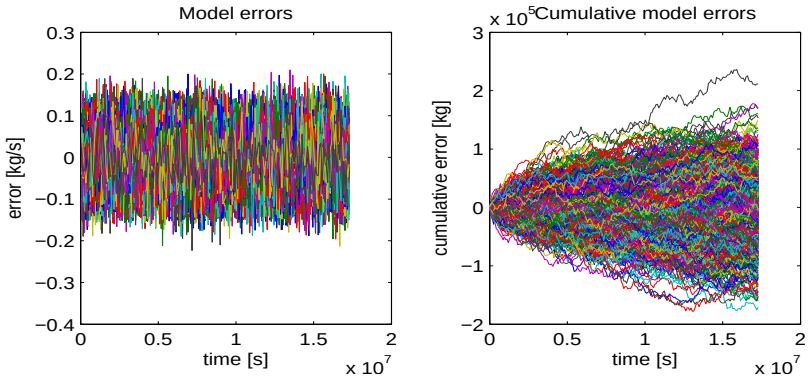

Figure 2: Synthetic true model errors

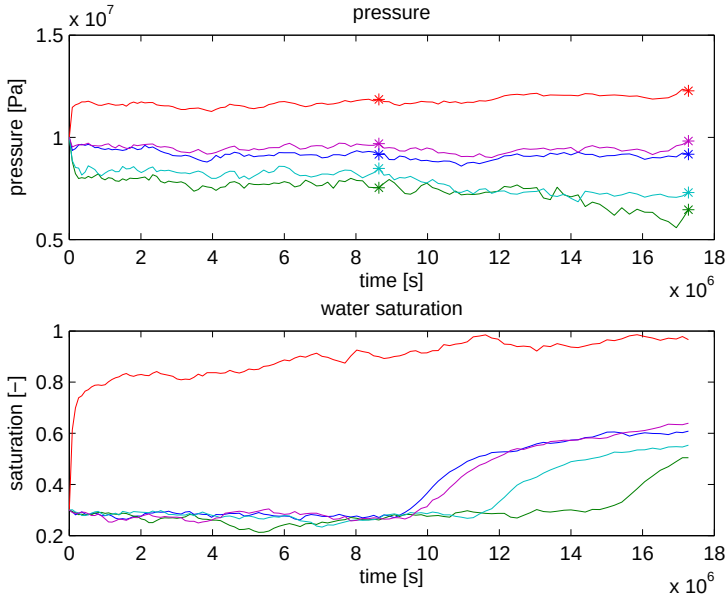

Figure 3: True pressure and saturation in the wells as functions of time. The top plot contains the 10 synthetic measurements at 100 and 200 days.

4.1 Correct prior and prior with exponentially decreasing correlation length

The filter is initialized by choosing an ensemble of permeability fields. For a variational method, like the RM, an initial permeability estimate must be specified as well as the uncertainty. These are extracted from the database as the mean and the sample covariance matrix. For an honest comparison of the methods, the same mean and covariance matrix can also be used to initialize the filter. In order to do so, samples must be drawn from the covariance matrix and added to the mean to obtain the ensemble members. This sampling can be done randomly, which is equivalent to picking ensemble members from the database, or deterministically, for example by selecting the principal components of a square root of the covariance matrix.

In field applications, specifying such a covariance matrix for the permeability is a rather arbitrary process. It could be said that there is "uncertainty in the uncertainty", meaning that it is not clear how the covariance matrix should be chosen. In this article, the effect of specifying the wrong covariance matrix is investigated. The "wrong" covariance matrix is obtained from the "correct" covariance matrix by preserving the magnitude of the variance, but imposing a different spatial correlation pattern. The covariance between the permeability values in two grid blocks i and j decreases exponentially as function of the distance between grid blocks i and j; the covariance is specified as

$$\widetilde{P}_{ij} = \sqrt{P_{ii}P_{jj}}e^{-\frac{dist(i,j)}{l}}, \tag{41}$$

where l is the correlation length; it is obtained by minimizing the square-sum-difference of $\widetilde{P}_{ij}$ and P_{ij}, resulting in a value of 72.43 $[m]$, Fig. (4)

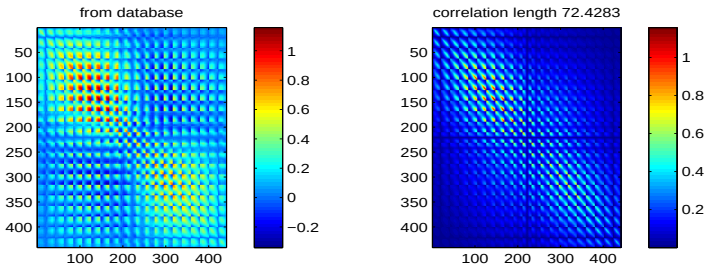

Figure 4: Correct prior covariance matrix and prior with exponentially decreasing covariance

4.2 Filter results

Some initial ensemble members are shown in Fig. (5). Fig. (6) shows the initial mean and covariance. The final estimates obtained with the correct prior covariance and 50 ensemble members of ESRKF are shown in Fig. (7). The estimated mean, the left plot of Fig. (8), resembles the true permeability, Fig. (1). A more quantitative comparison is given in section 4.4. The results using the prior with exponentially decreasing covariance are shown in Fig. (9) and Fig. (10).

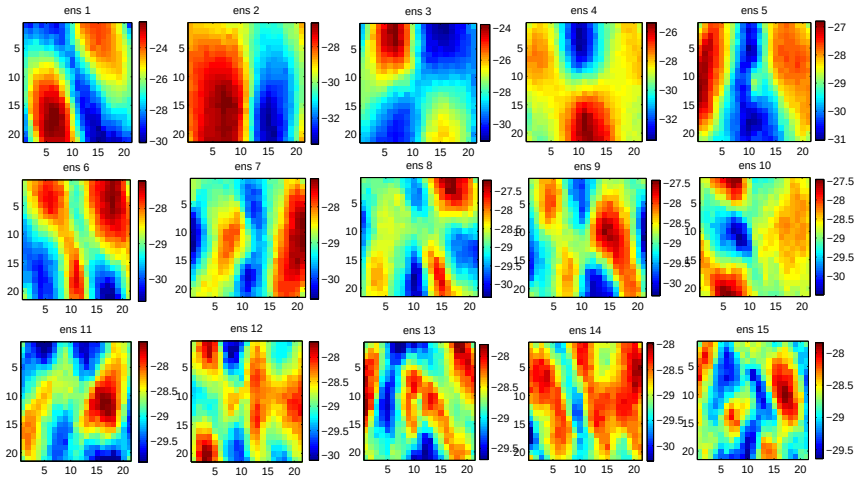

Figure 5: Some initial ensemble members

Fig. (11) shows the first four statistical moments of the ensemble after picking 100 realizations from the database. The covariance is a matrix, but only the diagonal elements

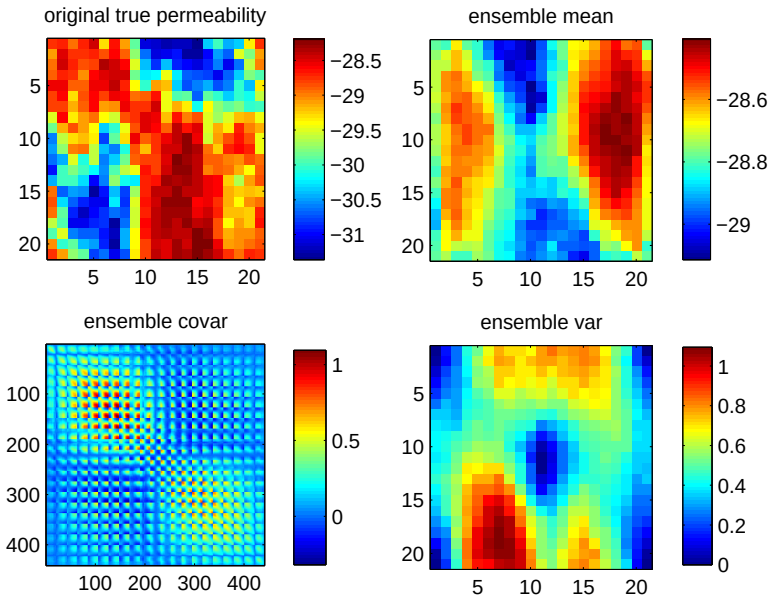

Figure 6: Synthetic true permeability, initial ensemble mean, covariance and variance

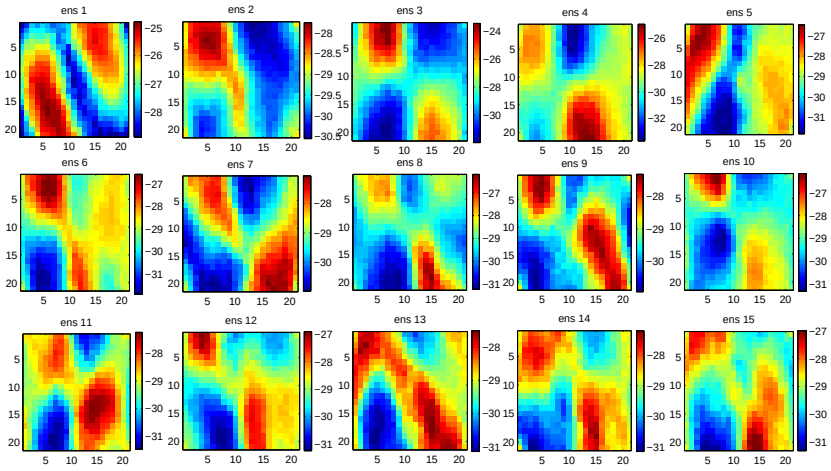

Figure 7: Some ensemble members, estimated with ESRKF using correct prior

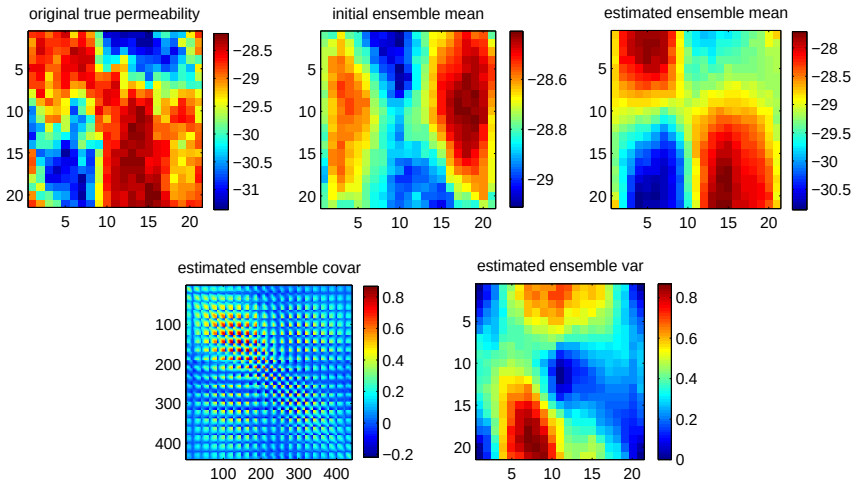

Figure 8: Synthetic true permeability, initial ensemble mean, and estimated ensemble mean, covariance and variance, estimated with ESRKF using correct prior

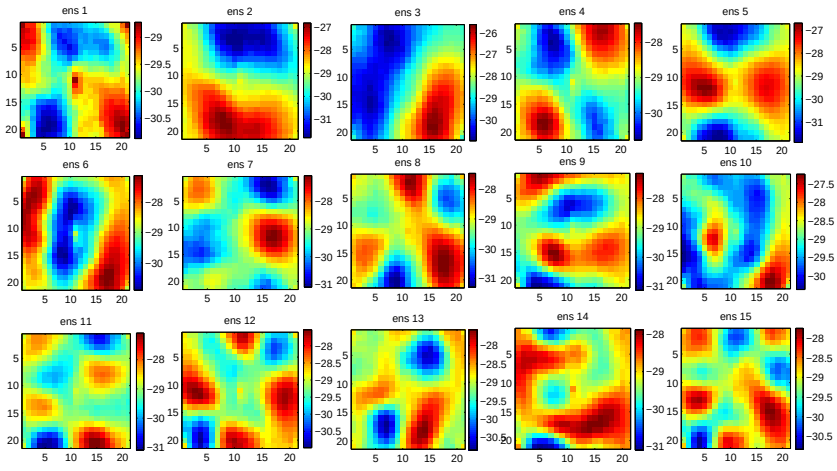

Figure 9: Some ensemble members, estimated with ESRKF using prior with exponentially decreasing covariance

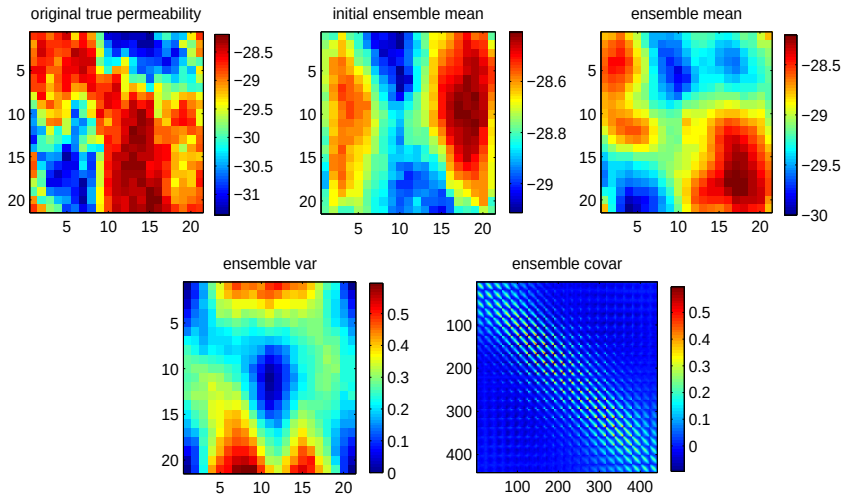

Figure 10: Synthetic true permeability, initial ensemble mean, and estimated ensemble mean, covariance and variance, estimated with ESRKF using exponential prior

are plotted. The third and fourth order moments are higher order tensors, but also only the "diagonals" are plotted. Fig. (13) shows what happens when a singular value decomposition is used to initialize the ensemble; the probability density that is represented by the ensemble is skewed and therefore it does not at all represent a Gaussian probability density. This negatively impacts the performance of the EnKF. A random orthogonal matrix can be used to de-skew the ensemble, preserving the first and second order statistical moments. This is shown in figures Fig. (12) and Fig. (14).

The randomness that is introduced by picking realizations from the database is suppressed for a large part by picking a large enough set of realizations, 100. If (11) were created again using 100 different realizations from the database, it looks quite similar, Fig. (15). This is not the case if the ensemble size is smaller.

4.3 Representer results

Permeability estimates of the RM using the correct prior and the prior with exponentially decreasing covariance are shown in Fig.(16) and Fig. (17). Fig. (18) and Fig. (20) show the parameter representers. The model errors that are reconstructed by the RM are shown in Fig. (22).

4.4 Alternative quantification of success

Unlike in a field application, in a synthetic experiment the reconstructed parameters can be compared to the ones that were used to synthesize the data. Consequently, the RMSE

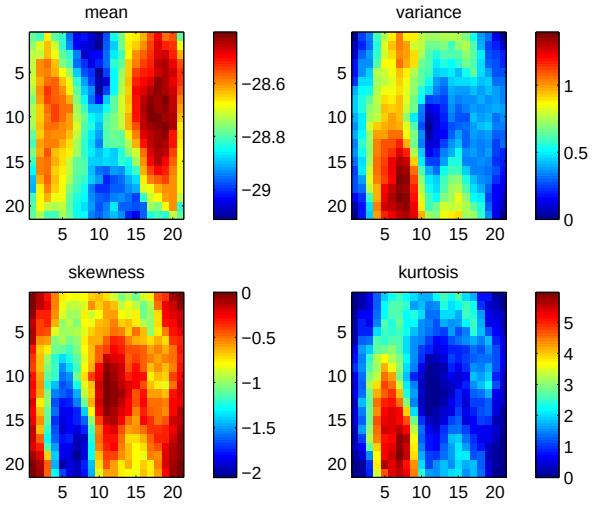

Figure 11: Four statistical moments of the ensemble. 100 members were picked from the database.

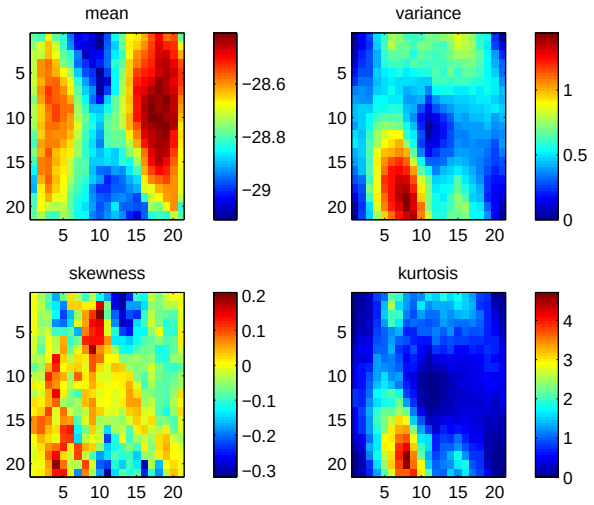

Figure 12: Four statistical moments of the ensemble. 100 members were picked from the database and then 100 random linear combinations were created.

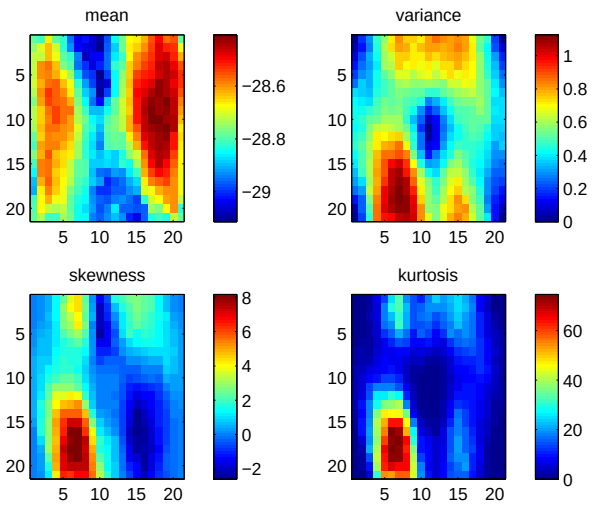

Figure 13: Four statistical moments of the ensemble. 100 leading singular vectors were calculated from the database.

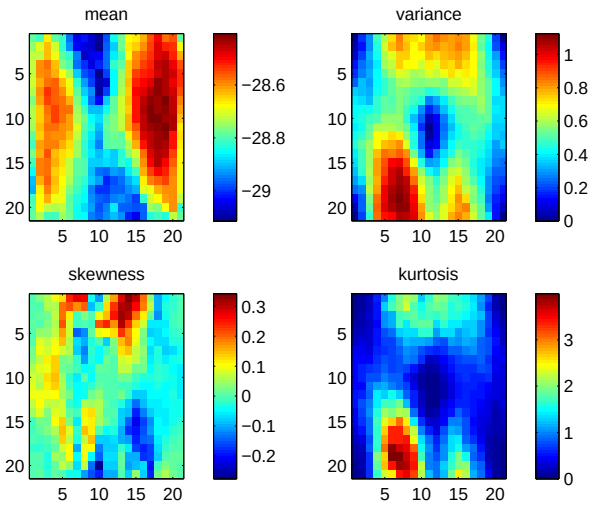

Figure 14: Four statistical moments of the ensemble. 100 leading singular vectors were calculated from the database and then 100 random linear combinations were created.

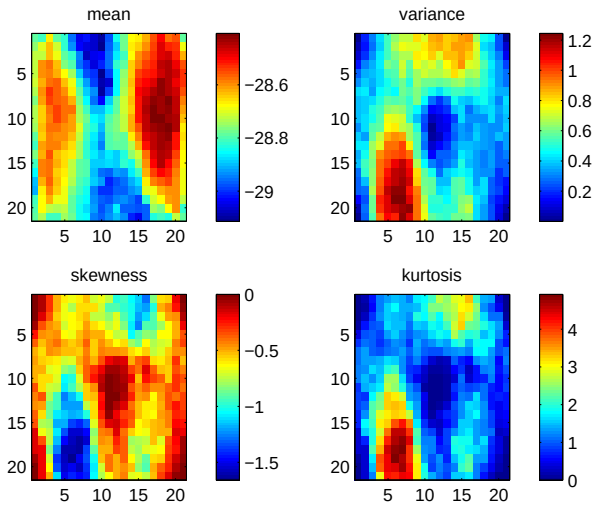

Figure 15: Four statistical moments of the ensemble. 100 members were picked again from the database, differently that Fig. (11).

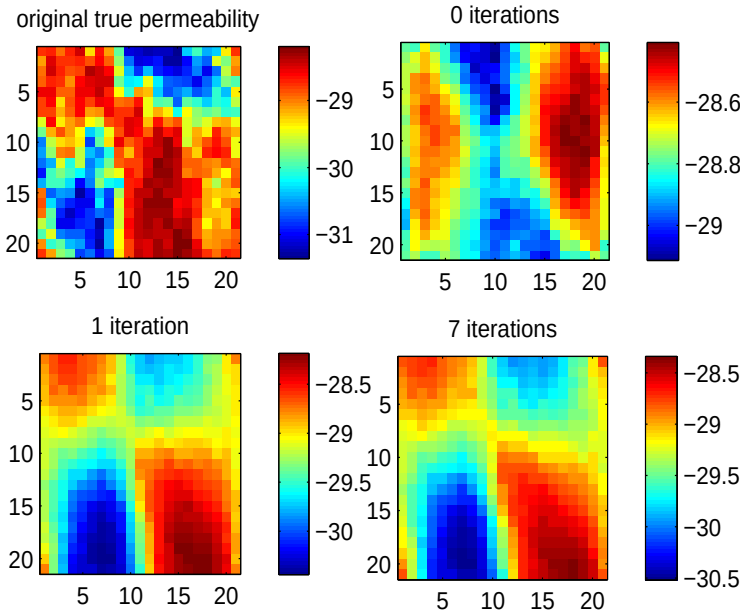

Figure 16: Permeability estimates of the RM using the correct prior

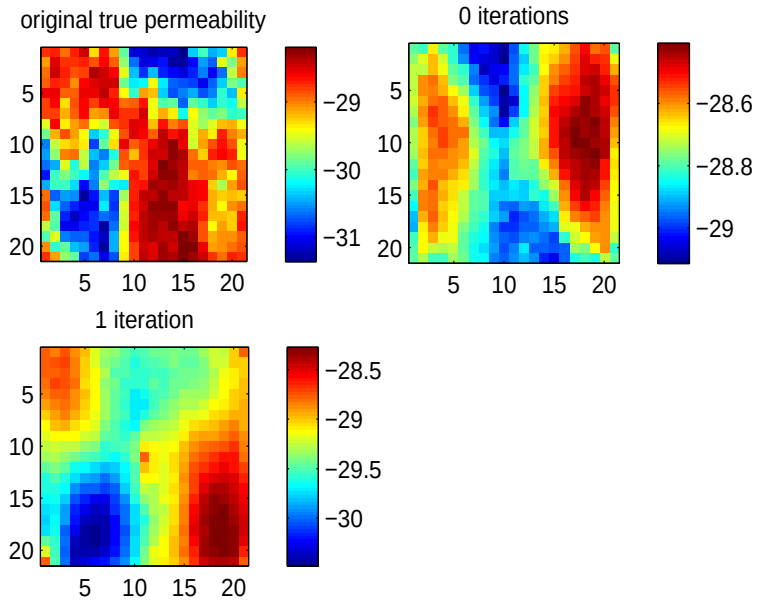

Figure 17: Permeability estimates of the RM using the prior with exponentially decreasing covariance

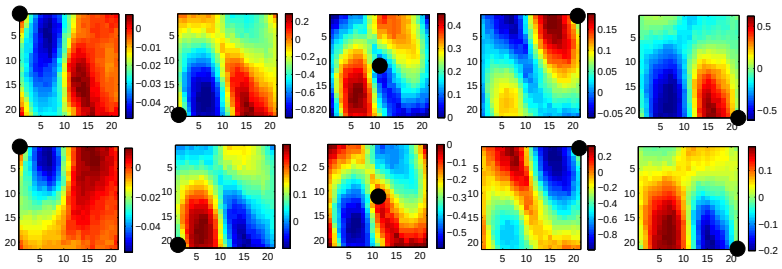

Figure 18: Parameter representers using correct prior. Different colums represent the deviation of the permeability estimates from the prior by different measurement locations (the wells). The locations are denoted by the black dots. Different rows represent different assimilation times.

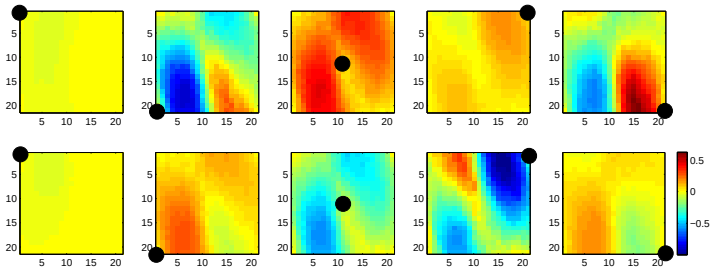

Figure 19: Parameter representers using correct prior, plotted on the same scale. Different colums represent the deviation of the permeability estimates from the prior by different measurement locations (the wells). The locations are denoted by the black dots. Different rows represent different assimilation times.

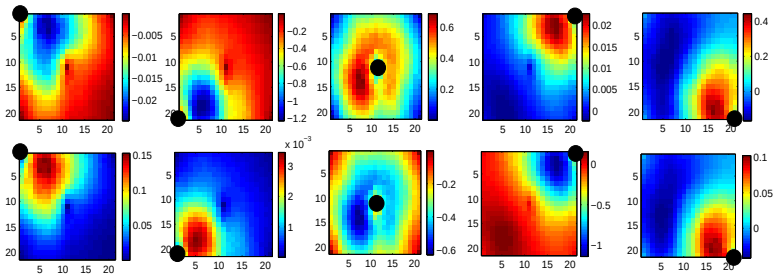

Figure 20: Parameter representers using exponential prior. Different colums represent the deviation of the permeability estimates from the prior by different measurement locations (the wells). The locations are denoted by the black dots. Different rows represent different assimilation times.

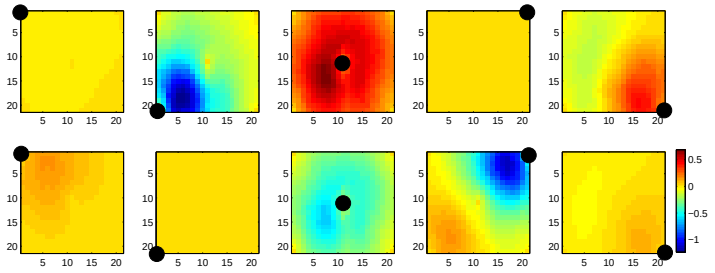

Figure 21: Parameter representers using exponential prior, plotted on the same scale. Different colums represent the deviation of the permeability estimates from the prior by different measurement locations (the wells). The locations are denoted by the black dots. Different rows represent different assimilation times.

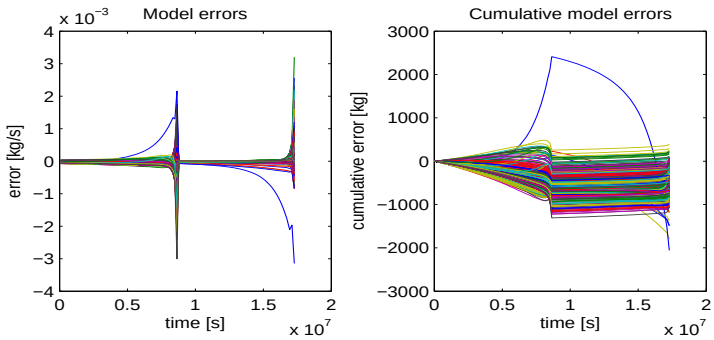

Figure 22: Model errors reconstructed by the RM using the correct prior

between all ensemble members and the truth can be calculated. A histogram is plotted in Fig. (23) for the case where the correct prior was used.

Another important quantification of the success of a data assimilation algorithm in reservoir management comes from looking at the water breakthrough curves in the wells. It is too late to react to water breakthrough in the production wells after it has been observed. A data assimilation method may therefore add considerable value if a reasonable prediction of the water breakthrough can be made (long) before it actually occurs. Only then the control strategy of the (smart) wells can be changed to delay the water breakthrough. Fig. (24) shows the water saturation curves for the synthetic truth in all wells. It also shows the same curves for one ensemble member of a EnKF. The curves are shifted in time to optimally fit the truth. The shift is a measure of how well the water breakthrough could be predicted. It is calculated as the integral average of the difference of the true and the predicted curve in the saturation domain:

$$\frac{1}{S_{final} - S_{init}} \int_{S_{init}}^{S_{final}} \left(t_{true}\left(S\right) - \left(t_{predicted} - t_{shift}\right)\right) dS = 0. \tag{42}$$

After t_{shift} has been calculated, the area between the true and the shifted predicted curve can be used to quantify the performance of the assimilation method to predict the shape of the water front as it breaks through:

$$A = \int_{S_{init}}^{S_{final}} \left| t_{true}\left(S\right) - \left(t_{predicted} - t_{shift}\right)\right| dS. \tag{43}$$

When t_{shift} is calculated for all ensemble members, a histogram can be plotted, Fig. (25). For comparison with the RM, where the ensemble consists of only one ensemble member, the mean of the ensemble of t_{shift} values is used to quantify the performance of the filter.

Some numbers:

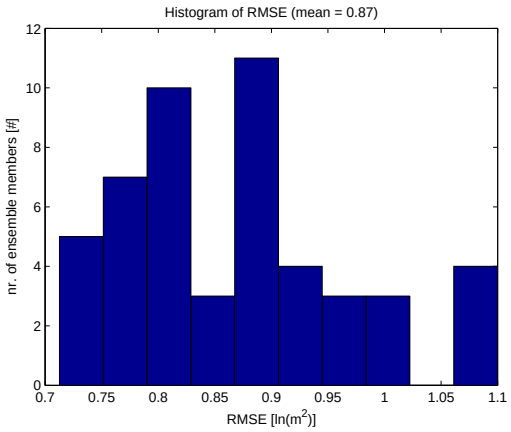

Figure 23: Root-mean-square difference between estimated permeability and the truth, for all ensemble members. 50 ensemble members were randomly picked from the database.

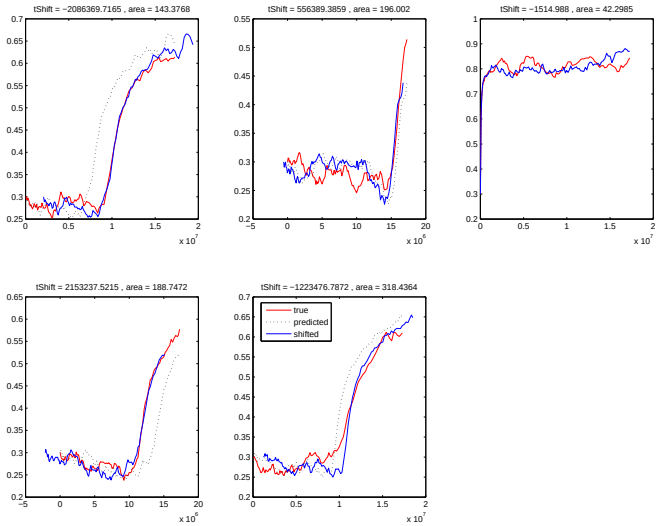

Figure 24: True, predicted and shifted $[s]$ water breakthrough curves of one ensemble member for all wells

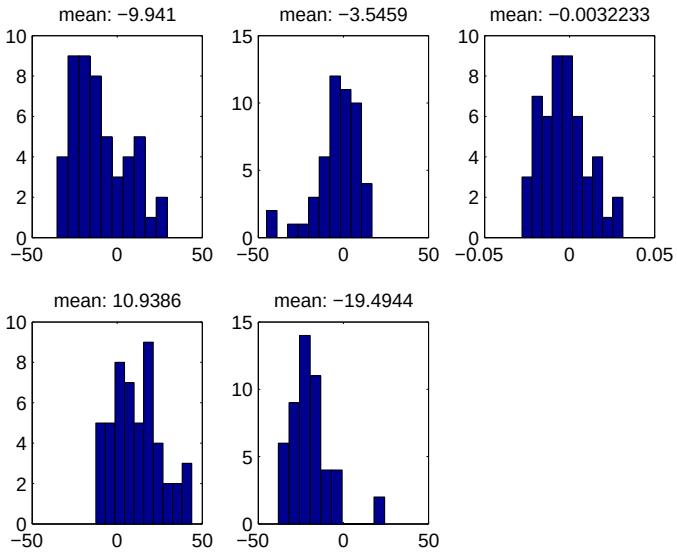

Figure 25: True, predicted and shifted [*days*] water breakthrough curves of all ensemble member for all wells

Forty experiments were done to gain insight in the effects of different prior, ensemble size, initialization method, de-skewing method and measurement update on the performance of the EnKF/ESRKF. In [Rommelse et al., 2007] the performance of the RM was investigated for different magnitudes of model errors, different line search algorithms, different minimization algorithms and compared to the results of adjoint-based strong constraint parameter estimation algorithms, but only the correct prior was used in the data assimilation process. The RM options that led to the best results of that research were used in this article and applied to a twin experiment where a prior with exponentially decreasing covariance was used. Fig. (26) visualizes the results of the forty-two experiments; its shows the cross plots of two different quality measures for all experiments. $\|\mathbf{t}_{Shift}\|$ is calculated as the 2-norm of the 5-dimensional vector with the means of Fig. (25). RMSE is the mean from Fig. (23). The five different plots represent results obtained with different ensemble sizes, or with different measurement updates, but they all contain the same RM results for comparison. In the ESRKF, measurements are assimilated sequentially using the duplication as described in section 2.3. In the EnKF, measurements at different times are assimilated sequentially, but measurements that were taken at the same time are also assimilated simultaneously and without duplication. Effects of randomness between the different experiments are reduced by storing and re-using the random numbers; this way "between-subject" variability is reduced and the effects of "within-subject" variation can be studied [van der Poel and Jansen, 2004]. The results of the ensemble methods are independent of the output of the random number generator if the ensemble size is chosen sufficiently large. For a smaller ensemble size (10, 25), results vary, but the effect of the

ensemble size decreases between 50 through 100.

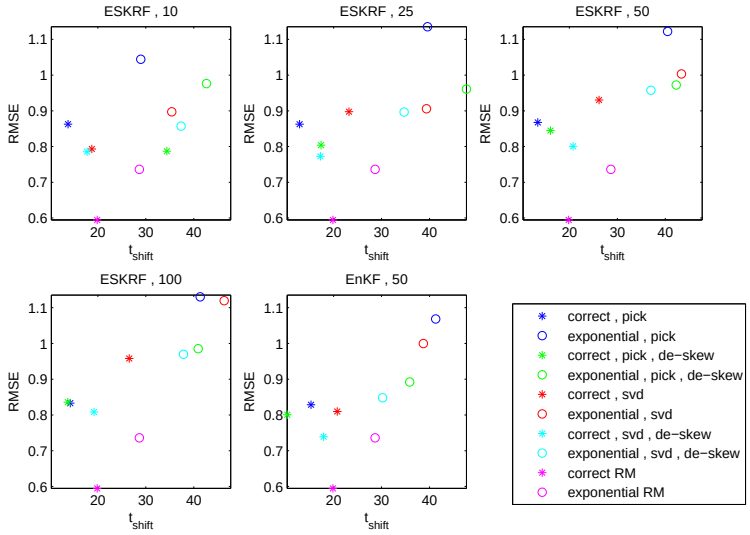

Figure 26: Crossplot of the two quality measures (RMSE and t_{shift}) for different ensemble sizes, EnKF or ESRKF, correct prior and prior with exponential decreasing covariance, ensemble randomly picked or obtained by svd and with or without de-skewing of the ensemble. All plots also contain the results of the Representer Method.

In terms of RMSE, the filter always performs worse than the RM; in terms of predicting the water breakthrough curves, the performance of the RM can be reached by a filter if the ensemble size is chosen sufficiently large and the correct prior is used. This is not true if a prior with exponentially decreasing covariance is used, so it can be said that the RM was less sensitive to using the "wrong" prior in this application than the filter. No matter what prior is used, the filter usually performs best if it is initialized by SVD and then de-skewed. If no de-skewing is applied, then randomly initializing the ensemble usually performs better than using SVD for the correct prior, but SVD performs better for the exponential prior. This can be explained as follows; randomly picking ensemble members from the database gives a rather symmetric ensemble, whereas the SVD ensemble is very skewed. For the prior with exponential decreasing covariance, no such database is available. Picking ensemble members involves picking columns from a Cholesky vector of the covariance matrix and results in a skewed ensemble as well.

It is worth mentioning that ESRKF with duplicated measurements performs reasonably better than EnKF in terms of RMSE and much better in terms of $\|t_{Shift}\|$ in the presence of no or little measurement errors. In the presence of moderate or large measurement errors, the results of EnKF or ESRKF are similar. The measurement update of EnKF makes a crude Gaussian assumption, whereas the version with duplicated measurements makes a set of smaller assumptions. The EnKF is a major improvement over the Extended

Kalman filter for dealing with nonlinearities in the reservoir model, but it is recommended to not use the EnKF for non-linear problems without a modified measurement update if the measurements are very accurate. Experiments to support this statement are not shown in this article, because the scope of this article is to compare the EnKF/ESRKF with the RM and not to show all details of the EnKF.

5 Conclusions

The RM always performed better than the filter, in terms of RMSE between estimated and true parameters. If the posterior probability of the parameters given the measurements is not symmetric or even multi-modal, then it makes sense that the RM performs better; in that case the probability that the true parameters lie close to the mode of the posterior is larger than the probability that the true parameters lie near the mean of the posterior. Also, the recursive filter introduces errors during every measurement update. These errors accumulate. An iterative algorithm has the opportunity in every iteration to reduce the errors that were made in the previous iteration.

When using the correct prior, predicting the water breakthrough can be done equally well by the filter as the RM if the ensemble size is chosen sufficiently large. This also shows that models that are not equally well history matched might produce future predictions equally well. This implies that a good history match may not guarantee good predictions. However, there is a high correlation between the RMSE and water breakthrough time error when results obtained from one method are compared with each other.

When using the prior with exponentially decreasing covariance, the filter does not perform as well as the RM, both in terms of RMSE and predicted water breakthrough. It can be said that the RM is less sensitive to using a "wrong" prior than a filter in this application.

The best filter results can be obtained if the ensemble is initialized with SVD and then de-skewed. Without de-skewing, SVD creates ensembles that are very skewed, causing poor results in estimating model parameters. Randomly initializing the ensemble for the case with the exponential prior involves picking columns from a Cholesky factor of the covariance matrix, which also results in a skewed ensemble.

References

[Aziz and Settari, 1979] Aziz, K. and Settari, A. (1979). *Petroleum reservoir simulation*. Chapman and Hall.

[Baird and Dawson, 2005] Baird, J. and Dawson, C. (2005). The representer method for data assimilation in single-phase darcy flow in porous media. *Computational Geosciences*, 9(4):247–271.

[Bayes, 1763] Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a letter to john canton, a. m. f. r. s. *Philosophical Transactions*, (53):370418.

[Bennett, 2002] Bennett, A. F. (2002). *Inverse modeling of the ocean and atmosphere*. Cambridge University Press.

[Bennett and McIntosh, 1982] Bennett, A. F. and McIntosh, P. C. (1982). Open ocean modeling as an inverse problem; tidal theory. *Journal of Physical Oceanography*, 12(10):1004 1018.

[Cohn et al., 1994] Cohn, S. E., Sivakumaran, N. S., and Todling, R. (1994). A fixed-lag kalman smoother for retrospective data assimilation. *Monthly Weather Review*, 122(12):28382867.

[Eknes and Evensen, 1997] Eknes, M. and Evensen, G. (1997). Parameter estimation solving a weak constraint formulation for an ekman flow model. *Journal of Geophysical Research*, 102(6C):12479 – 12491.

[Evensen, 2003] Evensen, G. (2003). The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53.

[Gao and Reynolds, 2006] Gao, G. and Reynolds, A. C. (2006). An improved implememntation of the lbfgs algorithm for automatic history matching. *SPE Journal*, 11(1):5 – 17.

[Gelb, 1974] Gelb, A. (1974). *Applied Optimal Estimation*. The MIT Press.

[Heemink et al., 2001] Heemink, A. W., Verlaan, M., and Segers, A. J. (2001). Variance reduced ensemble kalman filtering. *Monthly Weather Review*, 129(7):17181728.

[Janssen et al., 2006] Janssen, G. M. C. M., Valstar, J. R., and van der Zee, S. E. A. T. M. (2006). Inverse modeling of multimodal conductivity distributions. *Water Resources Research*, 42(3):Art. No. W03410.

[Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, pages 35–45.

[Przybysz et al., 2007] Przybysz, J. K., Hanea, R. G., Jansen, J. D., and Heemink, A. W. (2007). Application of the representer method for parameter estimation in numerical reservoir models. *Computational Geosciences*.

[Rommelse et al., 2007] Rommelse, J. R., Jansen, J. D., and Heemink, A. W. (2007). An efficient gradient-based parameter estimation algorithm using representer expansions. Technical Report 07-05, Delft University of Technology.

[Skjervheim et al., 2005] Skjervheim, J. A., Evensen, G., Aanonsen, S. I., Ruud, B. O., and Johansen, T. A. (2005). Incorporating 4d seismic data in reservoir simulation models using ensemble kalman filter. In *SPE Annual Technical Conference and Exhibition*, Dallas.

[Tarantola, 2005] Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia.

[Ulbrich, 2002] Ulbrich, S. (2002). Limited-memory bfgs-verfahren mit powell-wolfe schrittweitenregel.

[Valstar et al., 2004] Valstar, J. R., McLaughlin, D. B., te Stroet, C. B. M., and van Geer, F. C. (2004). A representer-based invserse method for groundwater flow and transport applications. *Water Resourses Research*, 40.

[van der Poel and Jansen, 2004] van der Poel, R. and Jansen, J. D. (2004). Probabilistic analysis of the value of a smart well for sequential production of a stacked reservoir. *Journal of Petroleum Science and Engineering*, 44:155 – 172.

[Verlaan and Heemink, 1997] Verlaan, M. and Heemink, A. W. (1997). Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrology and Hydraulics*, 11(5):349–368.

[Whitaker and Hamill, 2002] Whitaker, J. S. and Hamill, T. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130.