Stoch Environ Res Risk Assess (2013) 27:877–897
DOI 10.1007/s00477-012-0613-x

ORIGINAL PAPER

# Parameter estimation of subsurface flow models using iterative regularized ensemble Kalman filter

**A. H. ELSheikh · C. C. Pain · F. Fang · J. L. M. A. Gomes · I. M. Navon**

Published online: 15 August 2012


**Abstract** A new parameter estimation algorithm based on ensemble Kalman filter (EnKF) is developed. The developed algorithm combined with the proposed problem parametrization offers an efficient parameter estimation method that converges using very small ensembles. The inverse problem is formulated as a sequential data integration problem. Gaussian process regression is used to integrate the prior knowledge (static data). The search space is further parameterized using Karhunen–Loève expansion to build a set of basis functions that spans the search space. Optimal weights of the reduced basis functions are estimated by an iterative regularized EnKF algorithm. The filter is converted to an optimization algorithm by using a pseudo time-stepping technique such that the model output matches the time dependent data. The EnKF Kalman gain matrix is regularized using truncated SVD to filter out noisy correlations. Numerical results show that the proposed algorithm is a promising approach for parameter estimation of subsurface flow models.

**Keywords** ensemble Kalman filter · inverse problems · regularization · Gaussian process regression · Karhunen–Loève expansion

A. H. ELSheikh (✉) · C. C. Pain · F. Fang · J. L. M. A. Gomes
Department of Earth Science and Engineering, Imperial College London, Prince Consort Road, London SW7 2BP, UK
e-mail: a.el-sheikh@imperial.ac.uk

I. M. Navon
Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, USA

## 1 Introduction

Inference of subsurface geological properties is essential for many fields. Accurate prediction of groundwater flow and the fate of subsurface contaminants is one example (McLaughlin and Townley 1996; Carrera et al 2005). The multiphase flow of hydrocarbons in an oil reservoir is another example where accurate predictions have large economic impact (Naevdal et al. 2005). Subsurface domains are generally heterogeneous and shows wide range of heterogeneities in many physical attributes such as permeability and porosity fields. In order to build high-fidelity subsurface flow models for environmental risk assessment and/or for future predictions, a large number of parameters have to be specified. These parameters are obtained through a parameter estimation step. However, the amount of available data to constrain the inverse problem is usually limited in both quantity and quality. This results in an ill-posed inverse problem that might admit many different solutions.

Two types of data are available to constrain subsurface flow models. Static data collected at well bores and dynamic data measured as a time series of observations at few locations in the model. In the context of model calibration, there are two difficulties to consider (Fu and Gomez-Hernandez 2009). The first is to build a model that produces realizations conforming to static data. The second problem is to sample from these realizations in order to build a posterior distribution conforming to dynamic production data. For the first problem, Geostatistical analysis is commonly used to generate a set of subsurface models assuming a certain correlation length between the samples. These models are good initial solutions for the inverse problem. For the second problem, different parameter estimation techniques can be applied. These techniques can be classified into Bayesian

methods based on Markov Chain Monte Carlo (MCMC) methods (Oliver et al. 1997; Ma et al. 2008; Fu and Gomez-Hernandez 2009), gradient based optimization methods (McLaughlin and Townley 1996; Carrera et al. 2005) and ensemble Kalman filter (EnKF) methods (Moradkhani et al. 2005; Naevdal et al. 2005).

EnKF is a parallel sequential Monte Carlo method (SMC) for data assimilation. EnKF was introduced by Evensen (1994) and since then have been used for subsurface model update (Moradkhani et al. 2005; Naevdal et al. 2005; Tong et al. 2010). Both model parameters (e.g. permeability and porosity) and state variables (e.g. phase saturation and pressure values) can be updated by EnKF. In EnKF a number of simulations are run in parallel and are sequentially updated based on their average response and the measured data. Standard implementation of EnKF methods incorporates time dependent data in an online fashion during the flow simulation as observations become available.

Different variants of ensemble filters have been proposed (Pham et al. 1998; Tippett et al. 2003; Ott et al. 2004). These methods differ in how the ensemble members are updated (i.e., the analysis step) and can be generally categorized into perturbation-based or deterministic filters (Sun et al. 2009). Perturbation-based ensemble filters add random noise to each observation and this added observation noise, becomes an extra source of inaccuracy. Deterministic ensemble filters apply linear algebraic transformations to produce analysis ensembles that match the desired sample mean and covariance. It was observed that deterministic ensemble filters are more robust than perturbation-based methods especially for small-sized ensembles (Tippett et al. 2003; Sun et al. 2009). The use of relatively small ensemble size fails to produce meaningful statistics about the complete distribution of the model state variables conditional on the available observations. Techniques to improve the covariance estimates include covariance localization (Gaspari and Cohn 1999; Houtekamer and Mitchell 2001; Tong et al. 2012) and local approximation of the state error covariance using Local Analysis (LA) (Anderson 2003; Ott et al. 2004). EnKF might underestimate the background error covariance due to sampling errors as well as due to the presence of model errors. This might lead to poor performance and in severe cases to filter divergence where the filter no longer responds to the observations. Covariance inflation algorithms address this issue by pushing ensemble members away from the ensemble mean (Anderson and Anderson 1999; Anderson 2001). In EnKF, the ensemble perturbations are assumed to be Gaussian but in comparison to Kalman filters, EnKF can handle some nonlinearities by using the nonlinear model in the ensemble forecast step. Several studies were carried out to extend EnKF to handle non-Gaussian estimation problems especially by adopting the method of anamorphosis (Bengtsson et al. 2003; Smith 2007; Simon and Bertino 2009; Zhou et al. 2011).

In this paper, a flexible parameter estimation algorithm is developed. The algorithm starts with a stochastic interpolation using Gaussian process regression (GPR) (Rasmussen and Williams 2005) to integrate prior knowledge about the unknown field. Following that, the search space is parameterized using a Karhunen–Loève (KL) dimension reduction technique (Kac and Siegert 1947; Loève 1948; Karhunen 1947). The parameter estimation problem is then solved by an iterative regularized EnKF algorithm on the reduced space. EnKF for parameter estimation uses a pseudo-time stepping technique and time dependant data are matched in a batch mode to evaluate the likelihood of the estimated parameters. This algorithm requires repeated flow simulations of the entire simulation time. A Kalman gain regularization based on truncated singular value decomposition (TSVD) (Hansen 1998) is used to filter out noisy correlations and to deal with the estimated covariance matrix rank deficiency. SVD is used in square root filters (Tippett et al. 2003) to generate new ensemble members that preserve the forecast covariance. However, TSVD is used in the proposed algorithm for regularization instead of the Bayesian regularization via the measurement error covariance matrix and standard covariance localization techniques. The resulting algorithm offers a flexible and efficient alternative to gradient based optimization techniques. It converges after a small number of iterations while using very small ensemble sizes.

The proposed algorithm presents several novelties that differentiate it from previously published work. First, EnKF is applied iteratively in a batch mode for parameter estimation. This is inspired by related methods for converting filters into optimization methods (Zhou et al. 2008; Wan and Van Der Merwe 2000; Zupanski et al. 2008). However, it is different from ensemble Kalman smoothers that operate on the state variables (Evensen and van Leeuwen 2000; Chen and Oliver 2012). The proposed algorithm exhibits some similarities with Maximum Likelihood Ensemble Filter (MLEF) (Zupanski et al. 2008) but the error covariance is not updated using an analysis step as in filtering methods. Instead, a random perturbation is applied to mimic a random stencil in a stochastic Newton like method. The perturbation magnitude is gradually reduced as the solution approaches the optimal solution. Second, the proposed algorithm utilizes GPR for static data integration instead of kriging (Chilès and Delfiner 1999). In kriging, models are usually fitted using a variogram which measures the dissimilarity between samples versus the separating distance. This fitting is commonly performed using a least square method. However, GPR with Gaussian measurement noise have analytically tractable integrals over the parameter space. This enables an efficient solution of the model comparison problem. The optimal correlation length can be evaluated efficiently by maximizing the logarithm of

the marginal likelihood. Thirdly, model reduction using KL expansion is applied at the start of the algorithm to parametrize the unknown field. This step is similar to KL used in Efendiev et al. (2005) and Dostert et al. (2009). However, in these studies pre-set correlation lengths were used. In the current work, the mean and covariance matrices are obtained by the static data integration step using GPR. The utilized KL-model reduction is different from the Dimension-Reduced Kalman Filter (Zhang et al. 2007) as we do not assimilate the state variables. For the Hydraulic conductivity we use a standard parameterization following (Ghanem and Spanos 1991) where the updated parameters are the weights of the eigen functions of the stochastic field. These are a set of random variables with a predefined prior of zero mean and a unit standard deviations. The eigen functions (KL basis) are evaluated once at the start of the algorithm. In the numerical testing we use very limited amount of dynamic data to constrain the subsurface flow models as it is the case for many practical problems. The efficiency of the proposed algorithm is evident in the size of ensembles used in the presented numerical testing and the total number of forward runs required to reduce the mismatch errors. These small ensembles enable extensive exploration of the parameter space for uncertainty quantification studies.

The organization of this paper is as follows: Sect. 2 presents two tools for parametrizing the search space, GPR for static data integration and KL-dimension reduction technique. Section 3 provides a simple description of the standard EnKF algorithm followed by a full description of EnKF method for parameter estimation and the TSVD regularization as it is used within the developed algorithm. Section 4 presents a brief formulation of the subsurface flow problem followed by an application of the proposed algorithm on several test problems. In Sect. 5, the relation between the proposed algorithm and other iterative EnKF methods is highlighted. Also, numerical results utilizing the Simultaneous Perturbation Stochastic Approximation (SPSA) (Sadegh and Spall 1998) for parameter estimation is presented for numerical comparison. The conclusions of the current work are drawn in Sect. 6.

## 2 Search space parameterization

In this section we present two parameterization techniques for the subsurface flow inverse problem. Trying to solve the inverse problem on a the simulation grid results in a very large search space. Further, it neglects any spatial correlations between the unknown field values. In subsurface flow problems, the unknown field is usually known at few points where well bores exits. Accounting for these point data is commonly denoted as a static data integration process. We utilized GPR for solving the static data integration problem. The output of the GPR is then subjected to a dimension reduction technique using KL expansion. The combined use of these two methods provide a consistent and straight forward method for parameterizing the inverse problem.

### 2.1 Gaussian process regression

In the context of regression problems, it is required to find a function that maps from the spatial coordinate $\mathbf{x}$ to a real value $y$. For example, $y$ could be the log-permeability field and it is required to find its value over the domain of interest. Formally, the input data $\mathbf{D}$ for the regression problem, is a set of data pairs of observation $\{(\mathbf{x}_i, y_i)|i = 1, \ldots, n\}$, where $n$ is the number of observations and $y_i$ is the target or collected data at the spatial position $\mathbf{x}_i \in \mathbb{R}^2$. The objective of the regression is to make predictions about new targets $\widetilde{y}$ given the corresponding input $\widetilde{\mathbf{x}}$. In addition to the input data set, one has to make additional assumptions about the distributions of the data points to get a well-posed problem.

A Gaussian process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution (Rasmussen and Williams 2005). If we assume a set of data points $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $y_i = y(\mathbf{x}_i)$ are samples form $\mathbf{y} = (y_1, \ldots, y_n)^T$, then GP is defined as

$$\mathbf{y} \sim \mathcal{N}(\mu, \mathbf{C}), \quad \mu \in \mathbb{R}^n, \mathbf{C} \in \mathbb{R}^{n \times n}, \tag{1}$$

where $\mu$ is the mean function and $\mathbf{C}$ is the covariance matrix. The covariance matrix is specified as $[\mathbf{C}]_{i,j} = cov(y_i, y_j) = \mathcal{C}(\mathbf{x}_i, \mathbf{x}_j)$, where $\mathcal{C}$ defines the covariance function. The covariance function specifies the similarity between two function values $y(\mathbf{x}_i)$ and $y(\mathbf{x}_j)$ based on their corresponding spatial vectors $\mathbf{x}_i$, $\mathbf{x}_j$. The covariance function can take many forms and one of widely used functions is the squared exponential function defined as

$$\mathcal{C}_{SE}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_c^2 \exp\left(-\frac{1}{2}\frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{l^2}\right) \tag{2}$$

where $\sigma_c$ is the signal variance and $l$ is a normalization length that defines the global smoothness of the function $y$. The set of covariance function parameters and the measurement noise variance $\sigma_n$ are known as the *hyperparameters* of the GP $\psi = \langle \sigma_c, l, \sigma_n \rangle$.

Making predictions using a GP is equivalent to estimating $p(\widetilde{y}|\widetilde{\mathbf{x}}, \mathbf{D})$,where $\widetilde{y}$ is the new function value at the location $\widetilde{\mathbf{x}}$. Using vector notations, the input data is defined as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ and $\mathbf{y} = [\, y_1, y_2, \ldots, y_n]$. Assuming a predefined mean and covariance functions for GP with the associated *hyperparameters*, the inference problem on the new data set $\widetilde{\mathbf{X}}$ is defined using the following distribution

$$\begin{bmatrix} \mathbf{y} \\ \widetilde{\mathbf{y}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m} \\ \widetilde{\mathbf{m}} \end{bmatrix}, \begin{bmatrix} (\mathbf{C}_{xx} + \sigma_n^2 \mathbf{I}) & \mathbf{C}_{\widetilde{xx}} \\ \mathbf{C}_{\widetilde{xx}} & \mathbf{C}_{\widetilde{xx}} \end{bmatrix}\right) \tag{3}$$

where, $\mathbf{m}$ is a vector of means corresponding to the input data vector $\mathbf{X}$, $\widetilde{\mathbf{m}}$ contains prior mean values for the new data points $\widetilde{\mathbf{X}}$, $\widetilde{\mathbf{y}}$ is a vector of the posterior means for each new data points, $\mathbf{C}_{xx}$ is the covariance matrix of the input data, $\mathbf{C}_{x\widetilde{x}}, \mathbf{C}_{\widetilde{x}x}$ are the two cross covariance matrices, $\mathbf{C}_{\widetilde{x}\widetilde{x}}$ is the prior covariance matrix for the new data points, and $\mathbf{I}$ is the identity matrix. The conditional distribution $p(\widetilde{y}|\widetilde{\mathbf{x}}, \mathbf{y}, \mathbf{X})$ is a normal distribution with the mean $\widetilde{\mu} = \widetilde{\mathbf{m}} + \mathbf{C}_{\widetilde{x}x}(\mathbf{C}_{xx} + \sigma_n^2\mathbf{I})^{-1}(\mathbf{y} - \mathbf{m})$ and covariance $\widetilde{\mathbf{C}} = \mathbf{C}_{\widetilde{x}\widetilde{x}} - \mathbf{C}_{\widetilde{x}x}(\mathbf{C}_{xx} + \sigma_n^2\mathbf{I})^{-1}\mathbf{C}_{x\widetilde{x}}$ (see Rasmussen and Williams 2005, Appendix A.2). Obtaining a realization from this distribution involves generating correlated Gaussian random numbers.

### 2.1.1 Covariance function specification

The covariance function $\mathcal{C}(\mathbf{x}_i, \mathbf{x}_j)$ controls the correlation and dependence of the function values $y(\mathbf{x}_i)$ and $y(\mathbf{x}_j)$ on the spatial input vectors $\mathbf{x}_i, \mathbf{x}_j$. The covariance function must be symmetric positive semi-definite function (i.e., $\alpha^T\mathbf{C}\alpha \geq 0$ for all $\alpha \in \mathbb{R}^n$,where $\alpha^T$ denotes the transpose of $\alpha$). In the current work, we use the Matérn covariance function (Rasmussen and Williams 2005) defined as

$$\mathcal{C}_{Matern}(r) = \frac{2^{1-v}}{\Gamma(v)}\left(\sqrt{2v}\frac{r}{l}\right)^{v} K_v\left(\sqrt{2v}\frac{r}{l}\right) \tag{4}$$

where $v$ is an order parameter, $\Gamma$ denotes the Gamma function, $K_v$ is the modified Bessel function of the second kind of order $v > 0, r = x - x'$ is the Euclidean distance between two points and $l$ is the correlation length.

Learning a GP model is the process of finding appropriate kernel for the problem at hand as well as the covariance function parameterization. This process falls in the class of model selection. Given a parametric covariance function, model selection tries to find the hyperparameters vector $\psi = \langle v, l, \sigma_n \rangle$ that maximizes the conditional evidence

$$\psi^* = \arg\max_{\psi} p(\mathbf{y}|\mathbf{X}, \psi) \tag{5}$$

If the elements of $\mathbf{y}$ are independent samples from the Gaussian process, the distribution $p(\mathbf{y}|\mathbf{X}, \psi)$ is a multivariate Gaussian density defined as (Rasmussen and Williams 2005)

$$p(y|X, \psi) = \left((2\pi)^{\frac{n}{2}}\mathbf{C}_y^{\frac{1}{2}}\right)^{-1} \times \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{m})^T\mathbf{C}_y^{-1}(\mathbf{y}-\mathbf{m})\right) \tag{6}$$

where $\mathbf{C}_y = \mathbf{C}_{xx} + \sigma_n^2\mathbf{I}, \sigma_n$ is the measurement noise variance. The logarithm of the marginal likelihood is simple to evaluate as

$$\ln p(\mathbf{y}|\mathbf{X}, \psi) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln \mathbf{C}_y - \frac{1}{2}(\mathbf{y} - \mathbf{m})^T\mathbf{C}_y^{-1}(\mathbf{y} - \mathbf{m}) \tag{7}$$

This value is also called the logarithm of the evidence and is maximized with respect to the hyperparameters to obtain an optimal set of parameters given the observed data (see Rasmussen and Williams 2005, Algorithm 5.1). The estimated optimal set of hyperparameters is called the Maximum Likelihood type II (ML-II) estimate. These parameters are found by solving a non-convex optimization problem using conjugate gradient optimization with random restarts (MacKay 1999). Selecting the optimal order parameter $v$ of the Matérn covariance function is a model choice problem which can be solved within a Bayesian framework (ELsheikh et al. 2012). In this paper, $v$ is preset to 3 as it produces realizations that are neither very smooth nor very noisy.

## 2.2 Karhunen–Loève dimension reduction

The KL expansion (Kac and Siegert 1947; Loève 1948; Karhunen 1947), is a classical method for Gaussian random vectors quantization. It is also known as proper orthogonal decomposition (POD) or principal component analysis (PCA) in the finite dimensional case. The result of the GPR is a real-valued random field $\mathcal{K}$ with mean $\mu(\mathbf{x})$ and a covariance function $\mathcal{C}(\mathbf{x}_1, \mathbf{x}_2)$. Let $\mathcal{K}(\mathbf{x}, \xi)$ be a function of the position vector $\mathbf{x}$ defined over the problem domain and $\xi$ belonging to space of random events. KL expansion provides a Fourier-like series form of $\mathcal{K}(\mathbf{x}, \xi)$ as

$$\mathcal{K}(\mathbf{x}, \xi) = \mu(\mathbf{x}) + \sum_{k=1}^{\infty}\sqrt{\lambda_k}\xi_k\psi_k(\mathbf{x}) \tag{8}$$

where $\xi_k$ is a set of random variables, $\lambda_k$ is a set of real constants and $\psi_k(\mathbf{x})$ are an orthonormal set of deterministic functions. The covariance function $\mathcal{C}$ is symmetric and positive semidefinite and has the spectral decomposition

$$\mathcal{C}(\mathbf{x_1}, \mathbf{x_2}) = \sum_{k=1}^{\infty}\lambda_k\psi_k(\mathbf{x_1})\psi_k(\mathbf{x_2}) \tag{9}$$

where $\lambda_k > 0$ are the eigenvalues, $\psi_k$ are the corresponding eigenvectors. The orthogonal basis functions $\psi_k(\mathbf{x})$ satisfy the following equation (Ghanem and Spanos 1991)

$$\int_{\Omega}\mathcal{C}(\mathbf{x_1}, \mathbf{x_2})\psi_k(\mathbf{x_2})d\mathbf{x_2} = \lambda_k\psi_k(\mathbf{x_1}), \quad k = 1, 2, \ldots \tag{10}$$

The basis functions $\psi_k$, are the eigenvectors of the covariance matrix and can be obtained by principal component analysis (PCA) or solving an eigenvalue

problem. The eigenvectors are orthogonal and are normalized as follows

$$\int_{\Omega} \psi_k(\mathbf{x})\psi_j(\mathbf{x})d\mathbf{x} = \delta_{kj} \tag{11}$$

where $\delta_{kj}$ is the Kronecker delta. The random variables $\xi_k$ are uncorrelated with zero mean and unit variance ($E[\xi_k] = 0, E[\xi_k\xi_j] = \delta_{kj}$). For the case where $\mathcal{K}$ is a GP, $\xi_k$ is an i.i.d sequence of normal random variables with zero mean and unit variance $\mathcal{N}(0,1)$, the general form for $\xi_n$ can be obtained from

$$\xi_k = \frac{1}{\lambda_k}\int (\mathcal{K}(\mathbf{x},\xi) - \mu(\mathbf{x}))\psi_k(\mathbf{x})d\mathbf{x} \tag{12}$$

KL expansion using the eigenvectors of the covariance kernel is optimal in minimizing the mean-square-error from a finite representation of the process (Ghanem and Spanos 1991). This property makes KL expansion an efficient method for model reduction by truncating the summation in Eq. 8 to a finite set of $n$-terms as

$$\mathcal{K}_n(\mathbf{x},\xi) = \mu(\mathbf{x}) + \sum_{k=1}^{n}\sqrt{\lambda_k}\xi_k\psi_k(\mathbf{x}) \tag{13}$$

Due to the orthogonality of the basis functions, the total variance (energy) of the truncated $\mathcal{K}$ converges to the complete version as $n$ tends to infinity.

$$\int E[\mathcal{K}_n(\mathbf{x},\xi) - \mu(\mathbf{x})]^2 d\mathbf{x} = \int \mathcal{C}_n(\mathbf{x},\mathbf{x})d\mathbf{x} = \sum_{k=1}^{n}\lambda_k \tag{14}$$

The summation of the eigenvalues represents the amount of variance explained by the structure associated to the corresponding eigenvectors. The logarithm of the permeability field can be parameterized using a limited number of eigenvectors as in Eq. 13. Different realizations can be generated for different values of $\xi_k$. The dynamic data integration problem is concerned with finding values of $\xi_k$ such that the measured production data matches the simulation results.

## 3 Ensemble Kalman filter

The EnKF is a parallel SMC for data assimilation. This method was introduced by Evensen (1994). EnKF relies on two steps: prediction and update. For a discrete time nonlinear system

$$\mathbf{x}_{t+1} = \mathcal{M}(\mathbf{x}_t) + w_t \tag{15}$$

$$\mathbf{y}_{t+1} = \mathcal{H}(\mathbf{x}_t) + r_t \tag{16}$$

where $\mathbf{x}$ is the state vector, $\mathbf{y}$ is the observation vector, $w_t$ and $r_t$ are zero-mean white noises with covariance matrices $\mathbf{W}$ and $\mathbf{R}$, respectively and $\mathcal{M}, \mathcal{H}$ are the parameter update and the measurement operator, respectively. For an ensemble of size $n$, at each time step, a set of realizations of the state vector $\mathbf{X}_t = [\mathbf{x}_t^1, \mathbf{x}_t^2, \ldots, \mathbf{x}_t^n]$ are generated and the corresponding measurements are $\mathbf{Y}_t = [\mathbf{y}_t^1, \mathbf{y}_t^2, \ldots, \mathbf{y}_t^n]$. The matrix $Y$ is of size $n \times p$ where $n$ is the ensemble size and $p$ is the number of observations. The state variables are updated using the following steps (Houtekamer and Mitchell 2005; Blum et al. 2008)

$$\mathbf{x}_i^f = \mathcal{M}(\mathbf{x}_i^a(t-1)) + w_i, \quad i = 1,\ldots,n \tag{17}$$

$$w_i \sim \mathcal{N}(0,\mathbf{W}) \tag{18}$$

$$\mathbf{K} = \mathbf{C}_{xy}(\mathbf{C}_{yy} + \mathbf{R})^{-1}, \tag{19}$$

$$y_i^o = y^o + r_i, \quad i = 1,\ldots,n, \tag{20}$$

$$r_i \sim \mathcal{N}(0,\mathbf{R}) \tag{21}$$

$$\mathbf{x}_i^a(t) = \mathbf{x}_i^f + \mathbf{K}(y_i^o - \mathcal{H}(\mathbf{x}_i^f)), \quad i = 1, \ldots, n \tag{22}$$

where the superscripts $a$ and $f$ are for the analysis and forecast steps respectively, $\mathbf{C}_{xy}$ is state-measurement cross covariance matrix, $\mathbf{C}_{yy}$ is the measurement covariance matrix, $\mathbf{K}$ is the Kalman gain matrix and $y^o$ are the unperturbed observations.

### 3.1 EnKF for parameter estimation

The objective of the current paper is to calibrate the model parameters (i.e. permeability field) to conform with the dynamic (production) data. The time stepping in data assimilation EnKF will be used as a pseudo time stepping to represent the iterative nature of the solution. Similar formulation was used for modifying filtering algorithms to solve optimization and parameter estimation problems (Zhou et al. 2008; Wan and Van Der Merwe 2000). The proposed method is also related to parameter estimation by extended Kalman filter as highlighted in (Navon 1998). However, the state variables are limited to only the model parameters. The unknown field (permeability in this study) will be assigned to the state vector $\mathbf{x}$ in Eq. 17 and the state update equation will have the form $\mathbf{x}_i^f = \mathbf{x}_i^a + \epsilon_k w_i$ where, $\epsilon_k$ is a scaling factor in the iteration $k$ and $w_i$ is a zero-mean white noise sampled from a Gaussian distribution $\mathcal{N}(0,1)$. In the current study, a scaling factor of the form $\epsilon_k = c/\log(k+1)$ is used, where $c$ is a user input constant (e.g. 0.01) and $k$ is the iteration number (Kushner 1987). The measurement equation 16, corresponds to the simulator output of the production data

$$\mathbf{y}_i^f = \mathcal{SIM}(\mathbf{x}_i^f), \quad i = 1,\ldots,n \tag{23}$$

where $\mathcal{SIM}$ represents the nonlinear operator defined by the numerical simulator. Given a set of $n$ realizations of the state parameters, the covariance between the different production data generated by the simulator can be calculated using the sample covariance matrix as

$$\mathbf{C}_{yy} = \frac{1}{n-1}\mathbf{Y}^f(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T)(\mathbf{Y}^f)^T \tag{24}$$

where $n$ is the ensemble size, $\mathbf{I}_n$ is an identity matrix of size $n$, $\mathbf{1}$ is a vector of $n$ ones and $\mathbf{Y}^f$ is the output matrix. Similarly, the cross covariance between the different realizations of the field $\mathbf{x}_i^f$ arranged as rows of $\mathbf{X}^f$ and the corresponding outputs $\mathbf{Y}^f$ is defined as

$$\mathbf{C}_{xy} = \frac{1}{n-1}\mathbf{X}^f\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)(\mathbf{Y}^f)^T \tag{25}$$

At the analysis step each ensemble member is updated using the Kalman gain relation

$$\mathbf{x}_i^a(t) = \mathbf{x}_i^f + \mathbf{K}(\mathbf{y}_{obs} - \mathbf{y}_i^f), \quad i = 1, \ldots, n \tag{26}$$

where $\mathbf{y}_{obs}$ is the observed data and $\mathbf{K}$ is the Kalman gain matrix defined as

$$\mathbf{K} = \mathbf{C}_{xy}(\mathbf{C}_{yy} + \mathbf{R})^{-1} \tag{27}$$

The Kalman gain equation can be thought of as a corrector step utilizing an approximate Hessian (Thacker 1989; Cohn 1997). In the current paper, the parameter space is modeled with reduced order basis and no clear distance function can be assumed for covariance localization methods (Gaspari and Cohn 1999; Houtekamer and Mitchell 2001). A standard regularization based on truncated SVD is proposed as a reliable general method for filtering spurious correlations in estimating the Kalman gain matrix.

### 3.2 Truncated SVD regularization

Regularization was developed to solve ill-posed problems of the form $\mathbf{A}\mathbf{z} = \mathbf{b}$ (Hansen 1998). The matrix $\mathbf{A}$ can be decomposed using singular value decomposition, to obtain a set of orthogonal basis functions satisfying

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{28}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, satisfying $\mathbf{U}\mathbf{U}^T = \mathbf{I}_k, \mathbf{V}^T\mathbf{V} = \mathbf{I}_k$ and $\mathbf{S}$ is a diagonal matrix with non-negative entries $\sigma_1 \geq \sigma_1 \geq \cdots \geq \sigma_k \geq 0$ corresponding to the singular values. The matrix $\mathbf{A}$ will have a condition number $cond(\mathbf{A}) = \sigma_1/\sigma_k$. Given the SVD decomposition the solution of the system is equal to

$$\mathbf{z} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^T\mathbf{b} = \sum_{i=1}^{k}\frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i \tag{29}$$

If $\mathbf{A}$ has some small singular values, the solution $\mathbf{z}$ will be dominated by the corresponding singular vectors $\mathbf{v}_i$. Regularization methods attempt to reduce the effects of the small singular values on the solution vector $\mathbf{z}$. This can be done by truncated SVD (TSVD) or by Tikhonov Regularization (Hansen 1998). In the TSVD, all terms corresponding to small singular values are truncated from the calculation of the solution vector. TSVD can be reduced to the following form (Hansen 1998)

$$\mathbf{z}_f = \sum_{i=1}^{k}\phi_i\frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i \tag{30}$$

where $\mathbf{z}_f$ is the filtered solution and $0 \leq \phi_i \leq 1$ is a filtering factor defined as

$$\phi_i \equiv \begin{cases} 1, & i = 1, 2, \ldots, t \\ 0, & i = t+1, t+2, \ldots, k \end{cases} \tag{31}$$

The parameter $t$ is the number of SVD components maintained in the regularized solution. In the current study, we retain a number of SVD components corresponding to 99 % of the total variance.

### 3.3 Regularization of the Kalman gain matrix

Regularizing the inverse of the matrix $(\mathbf{C}_{yy} + \mathbf{R})$ appearing in the Kalman gain matrix is essential in order that small singular values do not dominate the update equation. This is easily achieved by first performing SVD on this matrix, similar to Eq. 28, then regularizing the inverse using

$$(\widetilde{\mathbf{C}_{yy} + \mathbf{R}})^{-1} = \mathbf{V}\mathbf{S}^{+}\mathbf{U}^T \tag{32}$$

where $\mathbf{S}^+$ is the pseudoinverse of $\mathbf{S}$, which is formed by replacing every diagonal entry that is larger than a pre-specified threshold by its reciprocal and transposing the resulting matrix. The current application of TSVD for covariance regularization has some similarities with recent work by Dovera and Della Rossa (2011) but they were concerned with initial ensemble generation which is different from the current work. Recently, Sætrom and Omre (2011) used shrinkage based regularization techniques for regularizing the Kalman gain matrix within standard EnKF method for parameters and states estimation. The regularized Kalman gain matrix is defined as

$$\widetilde{\mathbf{K}} = \mathbf{C}_{xy}(\widetilde{\mathbf{C}_{yy} + \mathbf{R}})^{-1} \tag{33}$$

The TSVD regularization solves the problem of rank deficiency with the added cost of calculating SVD for the matrix $\mathbf{Y}$. However, the size of the observations vector is usually limited and efficient methods for SVD calculations can be used (Golub and Van Loan 1996).

### 3.4 Algorithmic details

We presented all the elements of a parameter estimation algorithm that is general and can be applied to any inverse problem that can be formulated as a nonlinear least square problem. Any numerical simulator can be viewed as a nonlinear function that takes an input parameter vector $\mathbf{x}_i$ and produces an output vector $\mathbf{y}_i = \mathcal{H}(\mathbf{x}_i)$. Given a set of

observations $\mathbf{y}_{osb}$, the Kalman filter estimates $\mathbf{x}_{est}$ obtained by iterative application of Eq. 26 can be viewed as the Maximum A Posteriori (MAP) estimator of the following objective function (Sørensen and Madsen 2004)

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_m)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_m) + \frac{1}{2}(\mathbf{y}_{osb}-\mathcal{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}_{osb}-\mathcal{H}(\mathbf{x})) \quad (34)$$

where $\mathcal{J}$ is the objective function, $\mathbf{B}$ the parameter covariance prior to the parameter estimation (Background error covariance), $\mathbf{R}$ is the output error covariance matrix, $\mathbf{x_m}$ is the mean of the parameters prior distribution. The derivation of the Kalman update Eq. 26 assumes that the parameters follows a Gaussian distribution. These equations can also be derived from the least square approach as the Best Linear Unbiased Estimator (BLUE) (Jazwinski 1970; Anderson and Moore 1979).

We want to reiterate that the proposed algorithm is simple and requires a limited number of input constants that need to be adjusted. Different forms of observation data can be included in the observation vector $\mathbf{y}_{obs}$ to account for any data than need to be assimilated. The outline of the Iterative regularized EnKF algorithm for parameter estimation is listed in Algorithm.

**Algorithm 1**: Parameter estimation using iterative regularized EnKF

**Input:** Total ensemble size $n_t$, initial perturbation constant $c$,
Maximum number of iteration $i_{max}$
**Initialization:**
Randomly initialize $n_t$ states into the states matrix $\mathbf{X}$
**for** $i = 1$ *to* $i_{max}$ **do**
  **– Perturb the ensemble:**
  Update $\epsilon_i$ based on the iteration number
  $\mathbf{x}_j^f \leftarrow \mathbf{x}_j^a + \epsilon_i\, w_j \qquad \forall j = 1,\ldots,n_t$
  **– Propagate the ensemble:**
  $\mathbf{y}_j^f \leftarrow \mathcal{SIM}(\mathbf{x}_j^f) \qquad \forall j = 1,\ldots,n_t$
  **– Update the ensemble:**
  Estimate: $\tilde{\mathbf{K}}$ using Eq. 33, and $\mathbf{y}_j^f, \mathbf{x}_j^f$
  $\mathbf{x}_j^a \leftarrow \mathbf{x}_j^f + \tilde{\mathbf{K}}(\mathbf{y}_{obs} - \mathbf{y}_j^f) \qquad \forall j = 1,\ldots,n_t$
**end**

This algorithm can be considered as an iterative Wiener filter (Hillery and Chin 1991) because we assume a non-stationary problem for the unknown parameters. Also it can be considered as a quasi-Newton type algorithm with a random stencil. The Kalman gain equation can be thought of as a corrector step utilizing an approximate Hessian (Thacker 1989; Cohn 1997; Zupanski 2005).

### 3.5 Different forms of perturbation

In the current algorithm, the background error covariance is not updated. Instead, a random perturbation is used where the perturbation magnitude was decreased as we approach the solution. In all our numerical testing we used a logarithmic rule proposed by Kushner (1987) for specifying $\epsilon_k$ as $c/\log(k+1)$, where $c$ is a user input and $k$ is the iteration number. However, other forms for decaying sequences can be used as proposed by Gelfand and Mitter (1991)

$$c_k = \frac{c}{\sqrt{(k+1)\log\log(k+1)}} \quad (35)$$

or by Fang et al. (1997)

$$c_k = \frac{c}{(k+1)^{\alpha/2}\log(k+1)}, \quad 0<\alpha<1 \quad (36)$$

## 4 Problem formulation and numerical testing

A two-phase immiscible flow in a heterogeneous porous subsurface region is considered. For clarity of exposition, gravity and capillary effects are neglected. However, the proposed 1 model calibration algorithm is independent of the selected physical mechanisms. The two phases will be referred to as water with the subscript $w$ for the aqueous phase and oil with the subscript $o$ for the non-aqueous phase. This subsurface flow problem is described by the mass conservation equation and Darcy's law

$$\nabla\cdot\mathbf{v}_t = q, \quad \mathbf{v}_t = -\mathbb{K}\lambda_t(S_w)\nabla p \text{ over } \Omega \quad (37)$$

where $\mathbf{v}_t$ is the total Darcy velocity of the engaging fluids, $q = Q_o/\rho_o + Q_w/\rho_w$ is the normalized source or sink term, $\mathbb{K}$ is the absolute permeability tensor, $S_w$ is the water saturation, $\lambda_t(S_w) = \lambda_w(S_w) + \lambda_o(S_w)$ is the total mobility and $p = p_o = p_w$ is the pressure. In which, $\rho_w, \rho_o$ are the water and oil fluid densities, respectively. These equations can be combined to produce the pressure equation

$$-\nabla\cdot(\mathbb{K}\lambda_t(S_w)\nabla p) = q \quad (38)$$

The pore space is assumed to be filled with fluids and thus the sum of the fluid saturations should add up to one (i.e., $S_o + S_w = 1$). Then, only the water saturation equations is solved

$$\phi\frac{\partial S_w}{\partial t} + \nabla\cdot(f(S_w)\mathbf{v}_t) = \frac{Q_w}{\rho_w} \quad (39)$$

where $\phi$ is the porosity, $f(S_w) = \lambda_w/\lambda_t$ is the fractional flow function. The relative mobilities are modeled using polynomial equations of the form

$$\lambda_w(S_w) = \frac{(S_{nw})^2}{\mu_w}, \quad \lambda_o(S_w) = \frac{(1-S_{nw})^2}{\mu_o}, \quad S_{nw} = \frac{S_w - S_{wc}}{1 - S_{or} - S_{wc}} \quad (40)$$

where $S_{wc}$ is the connate or irreducible water saturation, $S_{or}$ is the irreducible oil saturation and $\mu_w$, $\mu_o$ are the water and oil fluid viscosities, respectively. The pressure Eq. 38

is discretized using standard two-point flux approximation (TPFA) method and the saturation equation 39 is discretized using an implicit solver with standard Newton–Raphson iteration (Chen 2007). For simplicity, we limit the parameter estimation to the subsurface permeability map $\mathbb{K}$. We also assumed this permeability map as a lognormal random variable as it is usually heterogeneous and shows a high range of variability.

The proposed parameter estimation algorithm will be applied to three test cases. All three test cases simulate water flooding cases with no flow boundary conditions on the domain boundaries. The true permeability map is sampled at certain number of locations and these values are used to construct the GP. The Matérn covariance is used for all GPR analysis and the value of $v = 3$ is selected. Dynamic data is obtained by running the simulator on the reference permeability map and the resulting water-cut curve is considered to be true and replaces real calibration date in our testing. Each water-cut curve was sampled at 50 points and these samples were used for calculating the errors and the update equation. The measurement errors were set to a small value of $10^{-12}$ to test the algorithm convergence in an optimization setting.

### 4.1 Test case 1

In this test case, the model is based on a 2D regular grid of $41 \times 41$ blocks in the $x$ and $y$ directions, respectively. The size of each grid block is 10 meters in each direction and a unit thickness in the $z$ direction. The porosity is constant in all grid blocks and equals 0.2. The water viscosity $\mu_w$ is $0.3 \times 10^{-3}$ Pa s and the oil viscosity $\mu_o$ is set to $0.3 \times 10^{-3}$ Pa s. The irreducible water saturation and irreducible oil saturation are set as $S_{or} = S_{wc} = 0.2$. For this test case, an injector-producer pattern is used.

Figure 1a shows the logarithm of the reference permeability field (permeability units is Darcy $= 9.869233 \times 10^{-13} m^2$). On the reference field, the location of the injection well is plotted as a black dot and the production well is plotted as a white dot. This field was sampled at 15 points and the sampled values are used as an input to the GPR for static data integration. The correlation lengths are optimized using the ML-II estimator and Fig. 1b shows the mean log-permeability field obtained from the GP regression. The reference and the mean regression fields along with the $\pm 2$ standard deviations bounds are shown in Fig. 2, where the reference field is plotted in red and the interpolated mean field is shown in blue color. The first six scaled eigen modes of the GPR covariance matrix are shown in Fig. 3. The first 40 modes obtained by KL-expansion were retained and all higher order modes were truncated. Figure 4 shows the decay of the variance of each term versus the number of KL terms. For the current problem, 40 KL terms preserve 92 % of the total variance of the GPR results. The number of retained basis functions is usually selected heuristically as long as we are able to match the data (Efendiev et al. 2005; Dostert et al. 2009). However, this problem of selecting the optimal number of KL terms can be formulated as a model choice problem which can be solved within a Bayesian framework (Bulygina and Gupta 2010; ELsheikh et al. 2012). This is in

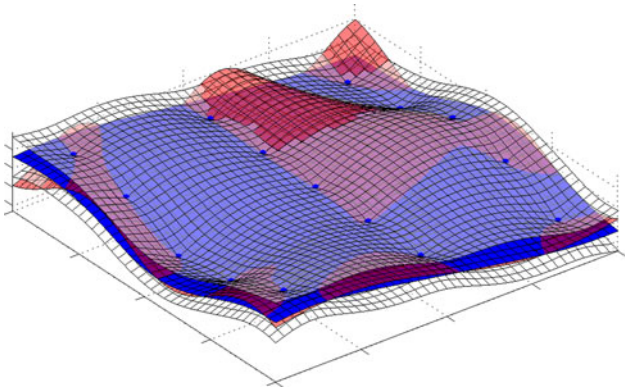

**Fig. 2** GP regression result along with the two standard deviations bounds for test case 1 (reference field in *red*, estimated mean in *blue* and two standard deviations bounds in *white*). (Color figure online)

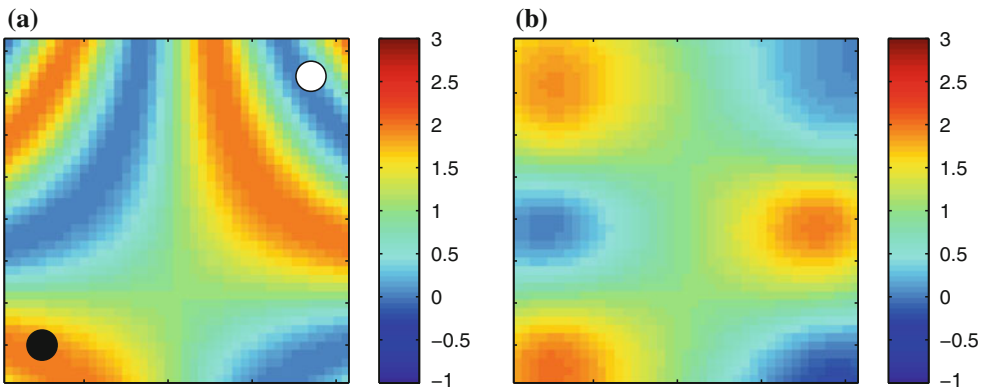


**Fig. 1** Log-permeability map for test case 1: **a** Reference and **b** GPR mean field

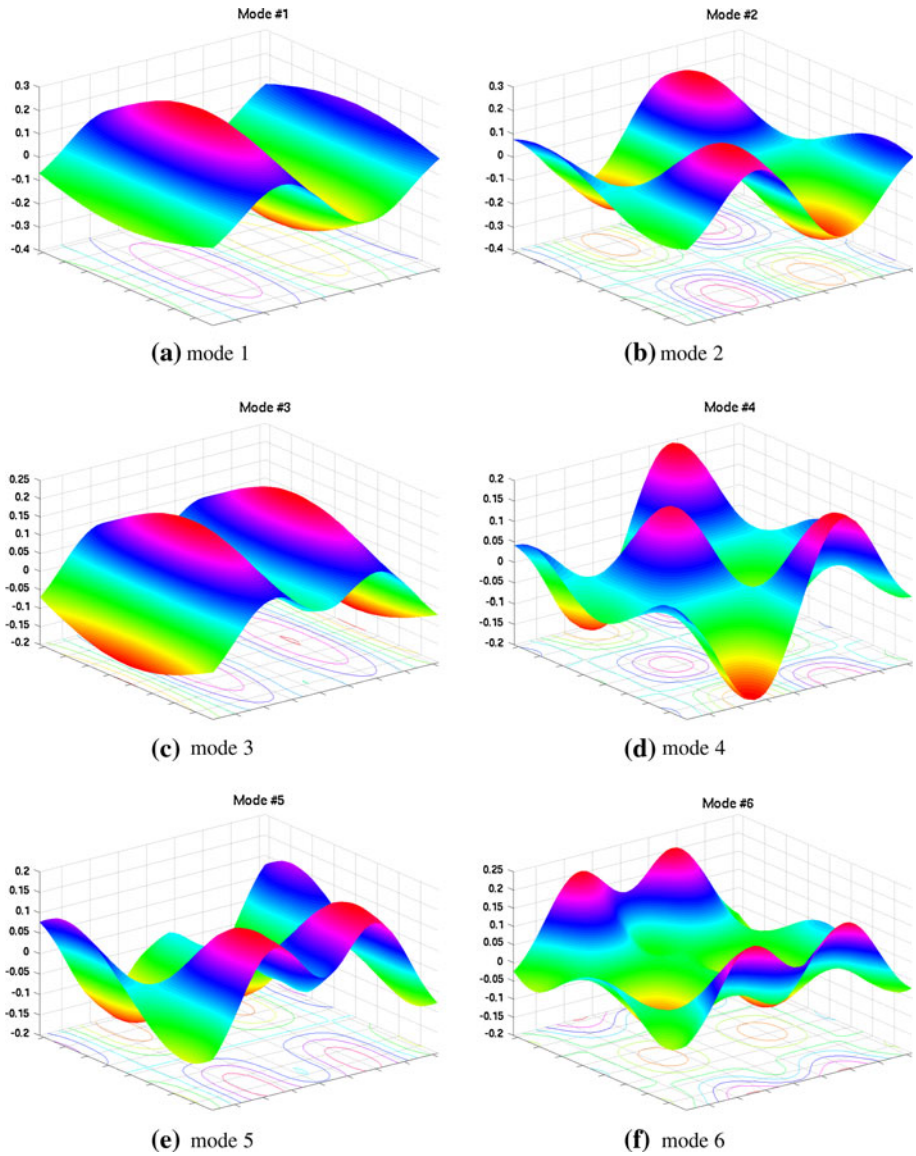


**Fig. 3** First six eigen modes obtained by KL-expansion of the GPR covariance matrix for test case 1

contrast to model reduction techniques where most of the energy has to be retained by the utilized KL terms (Fang et al. 2009).

Figure 5 shows the water fraction flow curve at the production cell for a set of ensembles with different sizes. The results are shown in terms of dimensionless time defined by the pore volume injected (PVI). It is noted that the water fraction flow curves converge to the reference data after 100 forward runs regardless of the ensemble size.

Figure 6 shows the optimized log-permeability field at the end of the EnKF iterations for two different ensemble sizes.

Figure 7 shows the root-mean-square error (RMSE) in the water fractional flow curve versus the number of EnKF forward runs. The ensemble of 5 members showed the best performance and did converge to the reference solution significantly faster than larger ensemble runs. This can be attributed to the smoothness of the problem. However, all ensembles have exhibited significant reduction in the RMSE with the increase of the number of forward runs. The difference in performance between small ensembles and large ensembles might suggest running smaller ensembles for the first few iterations followed by running larger ensemble for fine tuning the optimized parameters.

The results from this example show that the proposed method for integrating static data using GP regression along with the regularized EnKF for dynamic data integration is successful for conditioning the permeability maps to all measured and observed data. The ensemble regularization using TSVD is also successful in guiding the

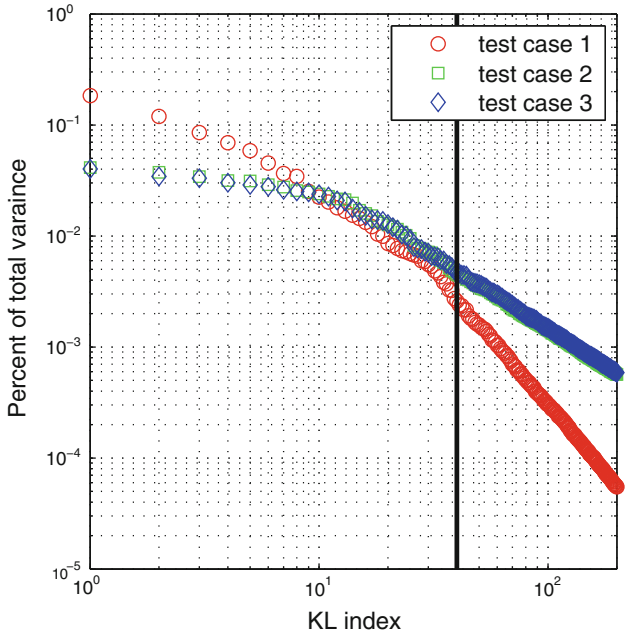


**Fig. 4** Variance decay rates of KL basis for test problem 1–3

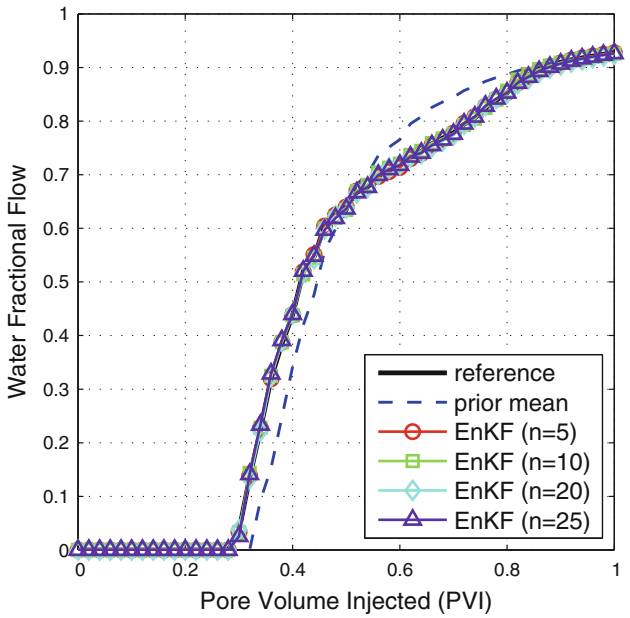


**Fig. 5** Reference, initial and optimized water fractional flow curves using different ensemble sizes for test case 1

optimization iteration from the mean GPR values to a solution that fits the production data.

### 4.2 Test case 2

This problem uses layer 20 of the Tenth SPE comparative test case (Christie and Blunt 1995) as the reference permeability field. The porosity is set to 0.2 over all grid blocks and the parameters $\mu_w = 0.3 \times 10^{-3}$ Pa s and $\mu_o = 3 \times 10^{-3}$ Pa s are used. The irreducible water and oil saturations are set as $S_{wc} = S_{or} = 0.2$. Figure 8a shows the logarithm of the reference permeability field (in Darcy $= 9.869233 \times 10^{-13}\text{m}^2$). We tested this problem under two wells pattern. The first wells pattern have one injector (black dot) and one producer (white dot) as plotted in Fig. 8a. The second wells pattern have two injectors and four produces located close the four corners of the domain as shown in Fig. 8b. The reference permeability field was sampled at 24 points and these values are used to constrain the GPR. The mean regression field obtained from the GPR is shown in Fig. 8c. The reference and the mean regression fields along with the $\pm$ 2 standard deviations bounds are shown in Fig. 9, where the reference field is plotted in red and the interpolated mean field is shown in blue color. The uncertainty in the GPR interpolated solution collapsed at the location of sampling points. KL-expansion was applied to reduce the search space and Fig. 4 shows the decay of the variance of each KL term versus the number of KL terms. A black vertical line corresponds to the cut off number of 40 retained KL terms. These terms preserved 63 % of the total variance of the GPR covariance matrix.

Figure 10a–c shows the optimized log-permeability map after running the EnKF using ensembles of 5,10 and 20 members. All runs were terminated after 100 forward runs regardless of the ensemble size. The corresponding optimized fractional flow curves are shown in Fig. 11. The water-cut curve fully matches the reference water cut curve after the optimization run. Figure 12 shows the RMSE in the fractional flow curve versus the number of EnKF iterations. A smooth convergence is observed for all ensemble sizes, with the small ensembles outperforming all other ensembles sizes in terms of error reduction. This is attributed to the success of the TSVD regularization in identifying the major search directions at each iteration. These search directions are adaptively updated at each iteration and smaller ensembles perform this step more frequently than larger ensembles. We also observe that as the ensemble approaches the solution, smaller ensembles fails in attaining further error reduction in comparison to larger ensembles. This is attributed to the approximate nature of the estimated gradients with small ensembles.

The regularized EnKF algorithm relies on the state update step (Eq. 15) that has a random component. This random component is scaled by a constant that vanishes with the number of EnKF iterations. However, the initial value of the scaling factor depends on the constant $c$ which is set to 0.01 in all our testing. In order to study the effect of $c$ on the convergence of the algorithm, two different runs were performed using different values of $c$. Figure 13 shows the RMSE of the water-cut curve for multiple runs

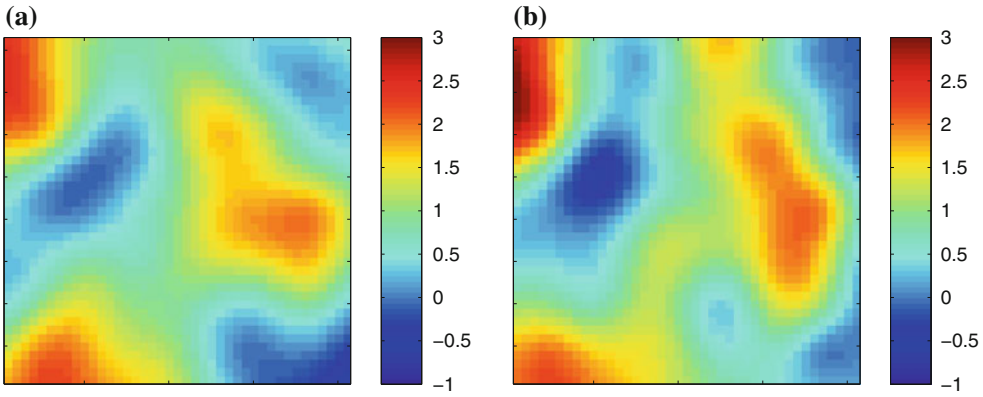


**Fig. 6** Optimized Log-permeability map for test case 1 with different ensemble sizes: **a** $n = 5$ and **b** $n = 10$

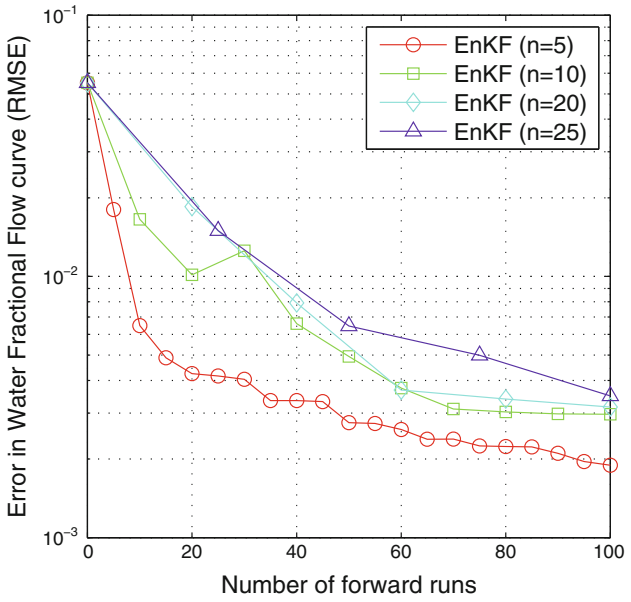


**Fig. 7** RMSE in water fractional flow curve versus the number of forward runs using different ensemble sizes for test case 1

with two different values for $c = 0.01, 0.04$. An ensemble of 10 members was used and all runs were initialized by the GPR mean. The convergence of the mean from different runs is observed for both cases. However, some runs showed local divergence. In order to increase the reliability of the algorithm, an adaptive step size might be applied within a line search strategy (Nocedal and Wright 2006).

Figure 14a–c shows the optimized log-permeability maps after running the EnKF using ensembles of 5,10 and 20 members for test case 2 under the wells pattern 2. All runs were terminated after 100 forward runs regardless of the ensemble size.

Figure 15 shows the corresponding RMSE in the fractional flow curve versus the number of forward runs. This wells pattern is considered more challenging as the integrated data is spatially distributed. Generally speaking all ensemble sizes managed to reduce the errors. However, it is observed that smaller ensembles performed initially better than larger ensembles for detecting major error reduction directions. Later in the iterative scheme, larger ensembles outperformed smaller ensembles as more accurate search direction were required.

### 4.3 Test case 3

This problem uses layer 80 of the Tenth SPE comparative test case as the reference permeability field. The porosity is set to 0.2 over all grid blocks and the water viscosity $\mu_w$ is set to $0.3 \times 10^{-3}$ Pa s and the oil viscosity $\mu_o$ is set to $3 \times 10^{-3}$ Pa s. The irreducible water and oil saturations are $S_{wc} = S_{or} = 0.2$. This test case has a channel along the length of the model as shown in Fig. 16a. Similar to test case 2, two wells pattern are tested. The first wells pattern have one injector (black dot) and one producer (white dot) as plotted in Fig. 16a. The second wells pattern have two injectors and four produces as shown in Fig. 16b. A set of 24 values of the permeability field are used for the GPR. The resulting mean field is shown in Fig. 16c. The reference and the mean regression fields along with the $\pm$ 2 standard deviations bounds are shown in Fig. 17. The GPR optimized correlation lengths are relatively short to accommodate the sharp features. Figure 4 shows the decay of the variance of each KL term versus the number of KL terms. Only 40 terms were retained and they preserved 62 % of the total variance of the GPR covariance matrix.

For the first injection pattern, the EnKF optimized log-permeability fields are plotted in Fig. 18a–c for ensembles of size 5,10 and 20 members, respectively. This figure shows that different ensembles have recovered different modes of the solution as the problem is ill-posed and might admit different solutions. The EnKF optimized water-cut curves are shown in Fig. 19 and a full agreement with the reference water-cut curve is observed. The regularized EnKF algorithm successfully solved the inverse problem and converged to the true curve regardless of the utilized ensemble size. Figure 20 shows RMSE in the water-cut curve versus the number of EnKF iterations and a clear

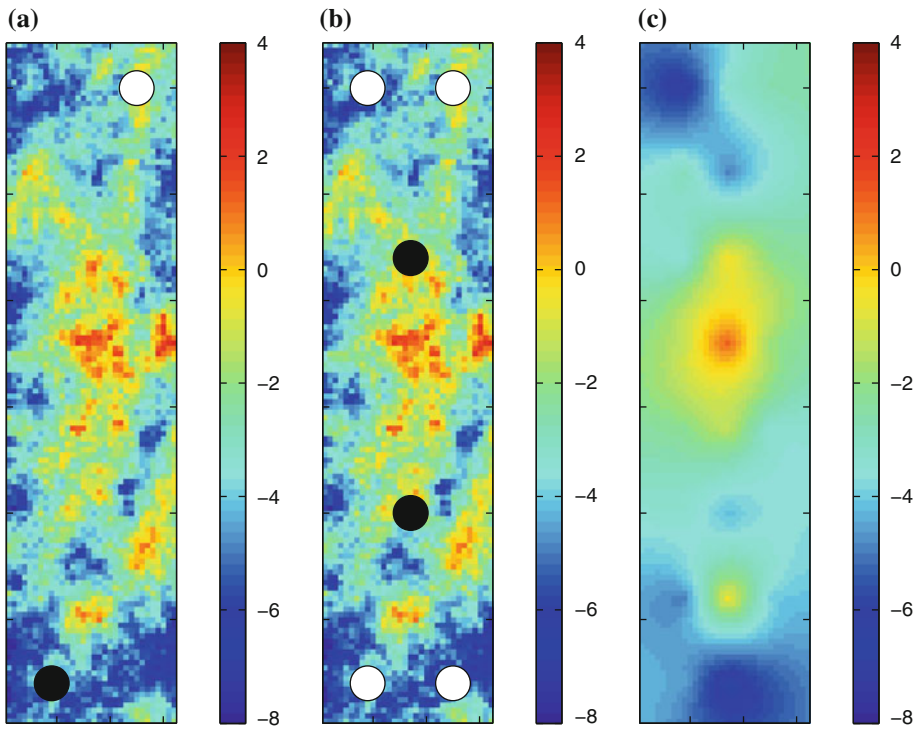


**Fig. 8** Log-permeability map for test case 2: **a** Reference with well locations for pattern 1, **b** Reference with well locations for pattern 2 and **c** GPR mean field (vertical to horizontal scale ratio is 2:1)

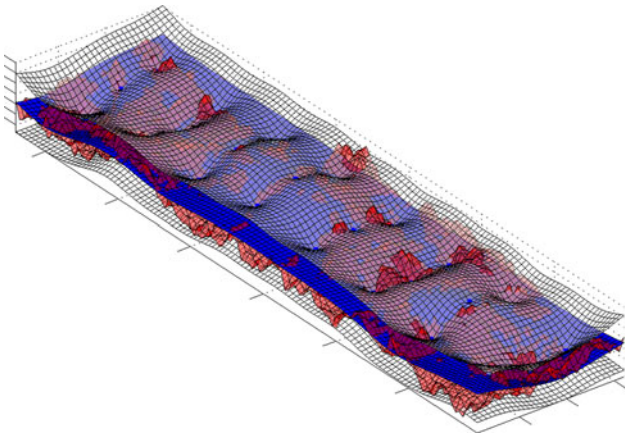

**Fig. 9** GP regression result along with the two standard deviations bounds for test case 2 (reference field in *red*, estimated mean in *blue* and two standard deviations bounds in *white*). (Color figure online)

convergence is observed after few iterations. The initial performance of the very small size ensemble with 5 members is very good. However, after 40 forward runs it fails in further reduction of the errors.

For the second injection pattern, the EnKF optimized log-permeability fields are plotted in Fig. 21a–c for ensembles of size 5,10 and 20 members, respectively.

These realization have less extreme values in comparison to those obtained for injection pattern 1. This might be attributed to the higher information content from four production wells that enables better model calibration. Figure 22 shows RMSE in the water-cut curve versus the number of EnKF iterations and a clear convergence is observed after few iterations. The initial performance of the very small size ensemble with 5 members is very good. However, larger ensembles of 10 members performed better after 20 forward runs. It is surprising that larger ensembles did not perform favourably. This can be attributed to the addition of the noise at each iteration and the magnitude of the error is related to iteration number and not to the number of forward runs.

### 4.4 Search Space Exploration

The proposed algorithm, similar to gradient based algorithms, might be attracted to a local minimum of the solution. A standard way to solve this problem is to restart the parameter estimation algorithm with different random initial values. In our setting, due to the limited amount of data, the inverse problem might admit many different solutions. Here, the objective is not to find the global minimum, instead we are interested in exploring the search space and recovering many different models that can be used for future forecasts. Problems 2 and 3 (wells pattern 1) were tested using a set of independent runs initialized by values that are different from the mean permeability maps obtained by GPR. The 40 parameters were initialized using random numbers following the Gaussian distribution $\mathcal{N}(0,1)$. These initial realizations conform to all static data even if they are different from the GPR mean.

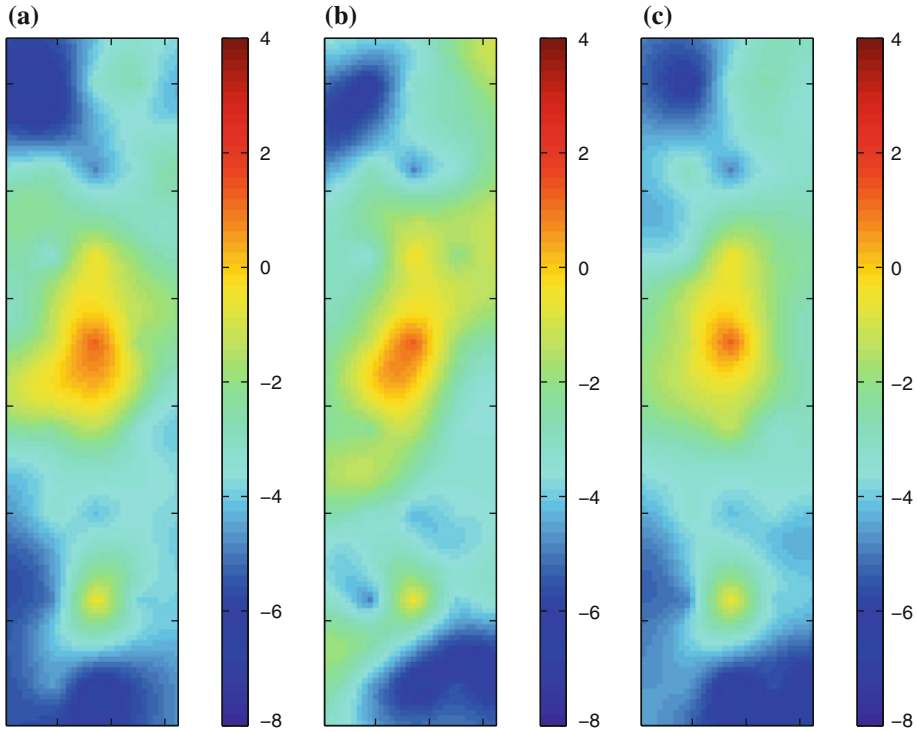


**Fig. 10** Log-permeability map for test case 2 with wells pattern 1: **a** EnKF optimized $n = 5$, **b** EnKF optimized $n = 10$ and **c** EnKF optimized $n = 20$ (vertical to horizontal scale ratio is 2:1)

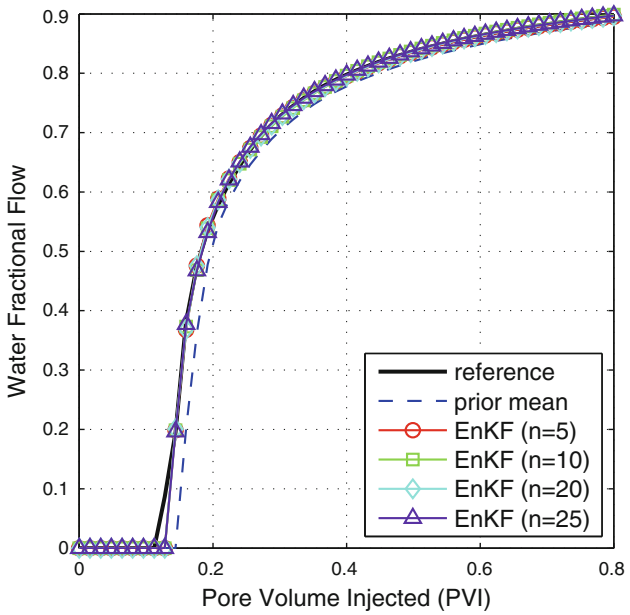


**Fig. 11** Reference, initial and optimized water fractional flow curves using different ensemble sizes for test case 2 with wells pattern 1

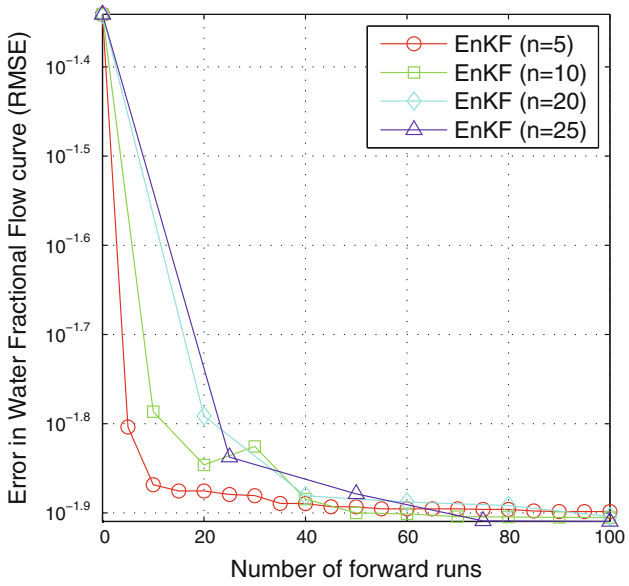


**Fig. 12** RMSE in water fractional flow curve versus the number of forward runs using different ensemble sizes for test case 2 with wells pattern 1

Figure 23 shows the initial ensemble water fractional flow curves versus the optimized curves for test case 2. Ensembles of 10 members were used and the optimization was done using the data up to PVI = 0.32. This is marked by a black vertical line in Fig. 23b. The rest of the fractional flow curve is an out-of-sample data that can be used effectively for future forecasts. Similarly, parallel simulations with random starting points were performed for test case 3 with calibration date up to PVI = 0.20. The results are shown in Fig. 24 in terms of water cut curves. The accuracy of the optimized ensembles is evident in the out-of-sample data.

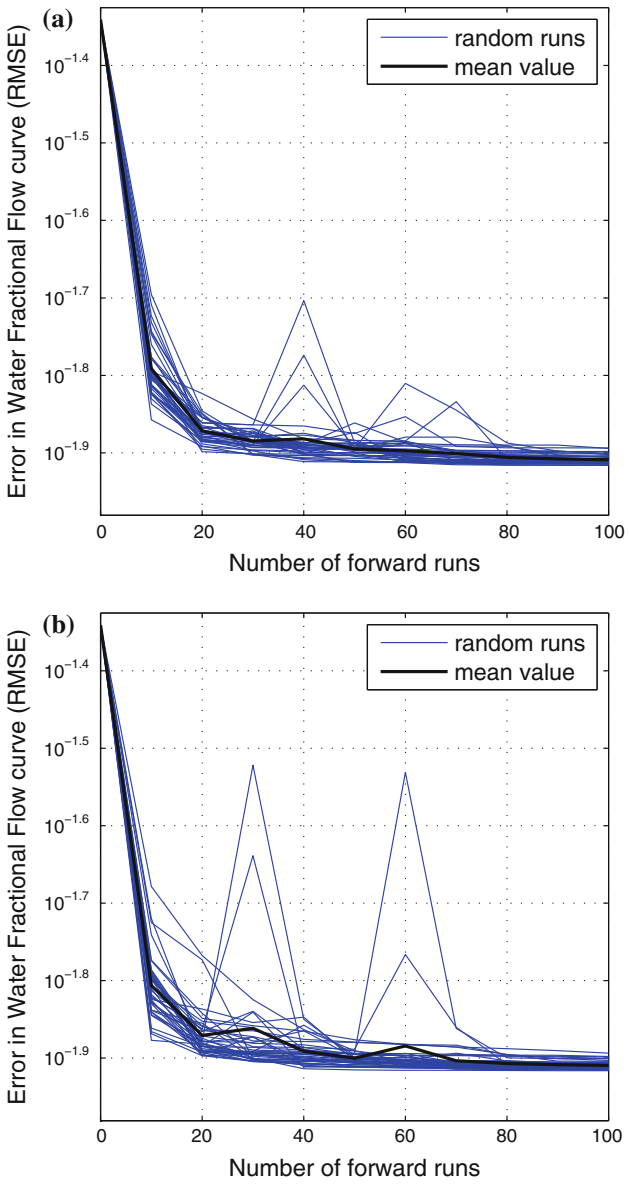


**Fig. 13** RMSE in water fractional flow curve versus the number of forward runs using different magnitudes of the random perturbation for test case 2 (wells pattern 1) with ensembles of 10 members, **a** $c = 0.01$ and **b** $c = 0.04$

## 5 Discussion and comparison

Recently, iterative ensemble based methods have attracted a large amount of research effort. Gu and Oliver (2007) introduced the ensemble randomized maximum likelihood (EnRML) method that is based on a formulation for nonlinear least square problems. The sensitivities were calculated using an ensemble based method. However, in high-dimensional problems the ensemble approximation of the sensitivity matrix is often poor (Chen and Oliver 2012). Li and Reynolds (2009) presented two iterative EnKF algorithms that relied on adjoint solutions. The first algorithm relied on estimating the sensitivities using an adjoint solution from the current data assimilation time back to time zero. The second algorithm relied on the adjoint solution over a single time step instead of all the way back to time zero. The presented results showed superior performance in comparison to standard EnKF Method.

Krymskaya et al. (2009) proposed a straight forward iteration of EnKF for both state and parameter estimation. This method was inspired by an outer loop iterative Kalman filter method (Jazwinski 1970). In this method, the mean of the estimated parameters at the end of the EnKF were used to initialize the ensemble for the next iteration. Interestingly, when re-running the filter, the mean estimator of initial guess was updated but the background error covariance was not changed during the iteration. Lorentzen and Naevdal (2011) presented an iterative EnKF method where an early stopping criteria was introduced. A likelihood function (similar to Eq. 34) was evaluated for each ensemble member and if the value of this likelihood function is higher or equal to value of the pervious iteration, the member is not updated. The iterations are stopped when no members are updated. The stopping criteria, provided a balance between the prior information and the overfitting of observations.

Sakov et al. (2012) presented an iterative EnKF algorithm for strongly nonlinear problems. The iterative EnKF algorithm has many similarities with EnRML (Gu and Oliver 2007). However, an ensemble square root filter was used within the iterative algorithm. The update equation have the form

$$\mathbf{x}_{i+1} = \mathbf{x}_i^f + \mathbf{K}(y_o - \mathcal{H}(\mathbf{x}_i^f)) + \mathbf{B}_{i+1}\mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_i) \qquad (41)$$

where $\mathbf{B}_{i+1}$ is the updated background error covariance. The last term tries to penalize the iterative procedure by including the prior information. As noted by the authors, the evaluation of the background error covariance $\mathbf{B}_{i+1}$ is technically difficult in the iterative process and a standard formula that rely on the linearity of the solution was used. Recently, Chen and Oliver (2012) presented an ensemble Kalman smoother based on EnRML where all the data are assimilated at once. The estimation of the sensitivities matrix based on sampling required large ensembles to provide reliable estimates.

All these iterative schemes have many commonalities and relied on almost the same assumptions. However, the difference between these methods are noticed in dealing with the following two questions: how to update uncertainties in nonlinear problems and how to regularize the estimated sensitivities. The first problem of updating uncertainties in nonlinear problems is challenging. In cases of large search space, the curse of dimensionality implies

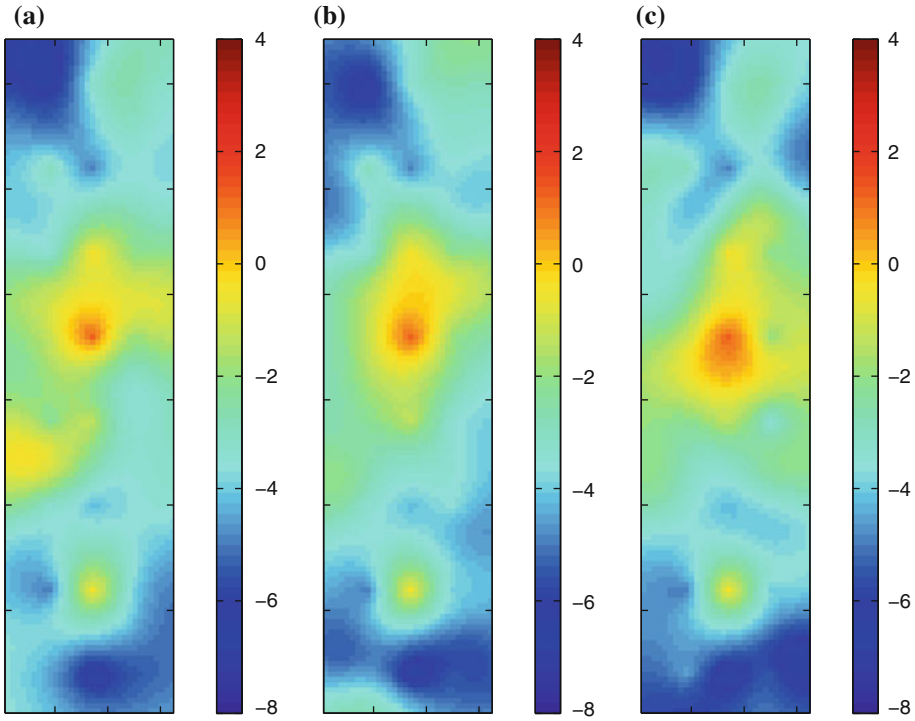


**Fig. 14** Log-permeability map for test case 2 with wells pattern 2: **a** EnKF optimized $n = 5$, **b** EnKF optimized $n = 10$ and **c** EnKF optimized $n = 20$ (vertical to horizontal scale ratio is 2:1)

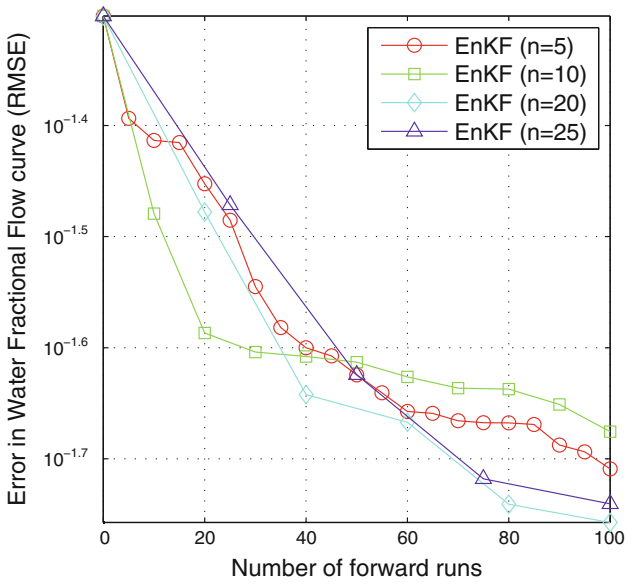


**Fig. 15** RMSE in water fractional flow curve versus the number of forward runs using different ensemble sizes for test case 2 with wells pattern 2

that the support of the posterior PDF is likely to by much smaller than the prior. This increases the probability of the case where none of the sampled parameters lies in that region. This might result in over estimation of the posterior PDF spread. On the other hand, Allen et al. (2006) highlighted that repeated use of the data will result in an underestimation of uncertainty or what is called data overfitting. The major question is how to balance the underestimation and the overestimation of uncertainties and how that affects the convergence of the solution. Different iterative algorithms tried to deal with this problem by early stopping of the iterative update (Lorentzen and Naevdal 2011) or by including the prior in the Kalman update equation (Gu and Oliver 2007; Sakov et al. 2012) or by not updating the error covariance in the iterative scheme (Krymskaya et al. 2009). The second problem is related to regularization of the Kalman gain matrix to avoid divergence. Generally speaking, including the prior in the objective function introduces some sort of regularization for the ill-posed problem as some preference is given to solutions that are close to the prior. However, regularization by TSVD penalizes the magnitude of the solution and gives a preference to solutions with minimum $L_2$ norm.

The proposed algorithm deals with the previously mentioned two questions in a different way. First, the proposed algorithm is a parameter estimation algorithm and does not try to update the error covariance. During initial numerical testing, we observed that updating the error covariance similar to the work presented by Sakov et al. (2012) have a negative effect on the convergence rates. We decided to use a standard technique utilized in many stochastic optimization methods where a gain sequence is formulated with a decaying magnitude to describe the random perturbations (Kushner 1987). If one is interested in evaluating the parameters uncertainties, Eq. 21 in (Zupanski et al. 2008) can be utilized at the converged solution.

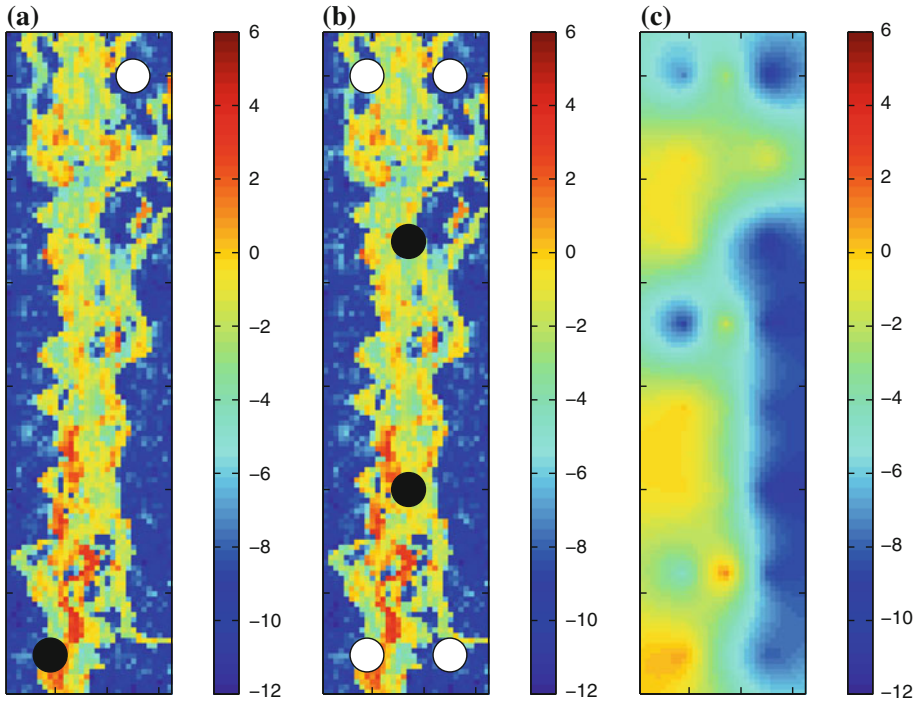


**Fig. 16** Log-permeability map for test case 3: **a** reference with well locations for pattern 1, **b** reference with well locations for pattern 2 and **c** GPR mean field (vertical to horizontal scale ratio is 2:1)

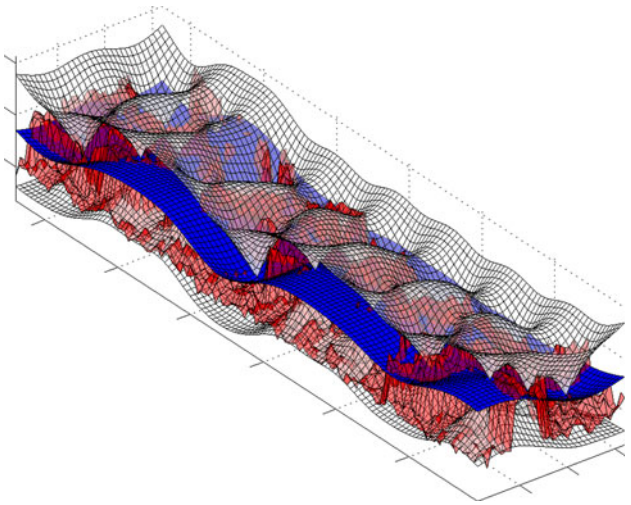

**Fig. 17** GP regression result along with the two standard deviations bounds for test case 2 (reference field in *red*, estimated mean in *blue* and two standard deviations bounds in *white*). (Color figure online)

As for regularization, we note that for the current parameterization using KL-expansion the parameters prior is centered around zero. This means that regularization using TSVD by minimizing the solution $L_2$ norm is equivalent to the Bayesian regularization by including the prior term. Once that is realized, only one type of regularization is applied and we decided to use TSVD as the filtering threshold is updated at each iteration. We notice the difference between our approach and the recent work of Chen and Oliver (2012) where Bayesian regularization was performed by including the prior in the update equation and then the sensitivities were evaluated from a rank deficient system after regularization. After that, the evaluation of the Kalman gain required the inverse of singular matrix that needed one more step of regularization. The repeated application of different regularization steps with different thresholds might result in excessive information loss. In the proposed algorithm, only one regularization operation is applied at each iteration.

### 5.1 Simultaneous Perturbation Stochastic Approximation

In this section, we compare the proposed algorithm to the SPSA method (Spall 2003). Both the proposed algorithm and the SPSA method are stochastic optimization algorithms that use a randomly selected stencil (Spall 2003) where all the unknowns are perturbed simultaneously to generate the stochastic gradient. SPSA relies on the following iterative scheme

$$\mathbf{x}_{k+1} = \mathbf{x}_k + a_k \hat{g}(\mathbf{x}_k) \tag{42}$$

where $\mathbf{x}_k$ is the parameters vector at iteration $k$, $a_k$ is the step size and $\hat{g}(\mathbf{x}_k)$ is the approximate gradient of the objective function. The two sided SPSA relies on the gradient calculated as

$$\hat{g}(\mathbf{x}_k) = \frac{\mathcal{J}(\mathbf{x}_k + c_k\Delta_k) - \mathcal{J}(\mathbf{x}_k - c_k\Delta_k)}{2c_k} \begin{bmatrix} \Delta_{k1}^{-1} \\ \Delta_{k2}^{-1} \\ \vdots \\ \Delta_{kp}^{-1} \end{bmatrix} \tag{43}$$

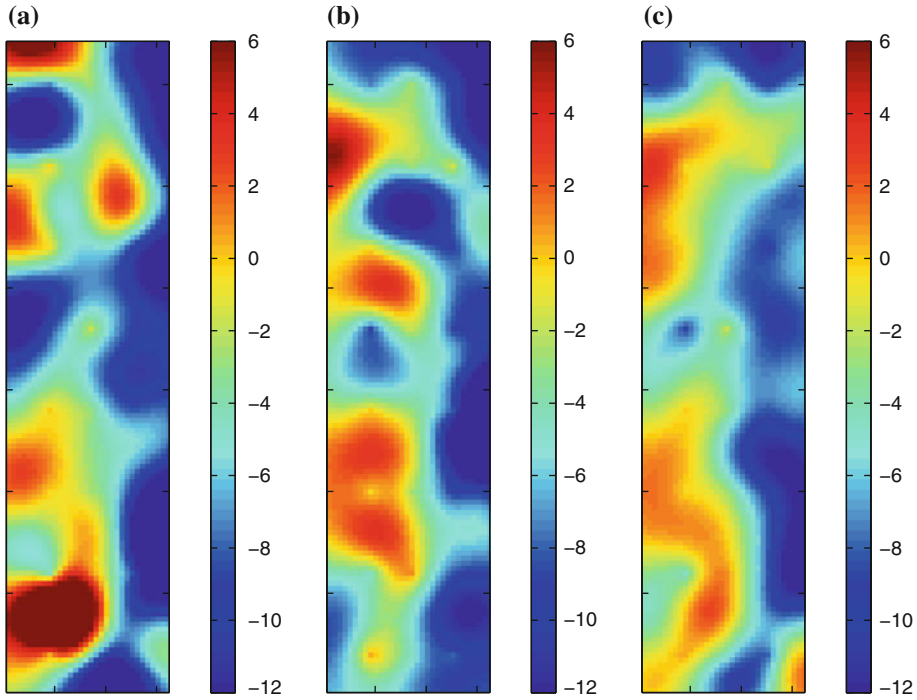


**Fig. 18** Log-permeability map for test case 3, wells pattern 1: **a** EnKF optimized $n = 5$, **b** EnKF optimized $n = 10$ and **c** EnKF optimized $n = 20$ (vertical to horizontal scale ratio is 2:1)

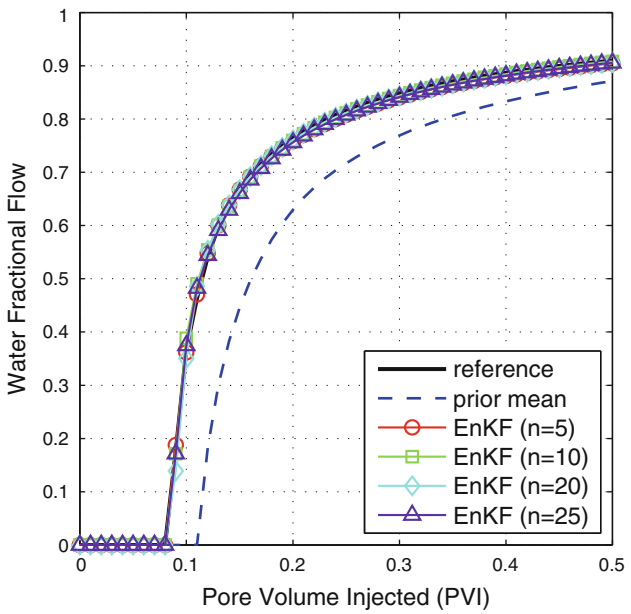


**Fig. 19** Reference, initial and optimized water fractional flow curves using different ensemble sizes for test case 3 with wells pattern 1

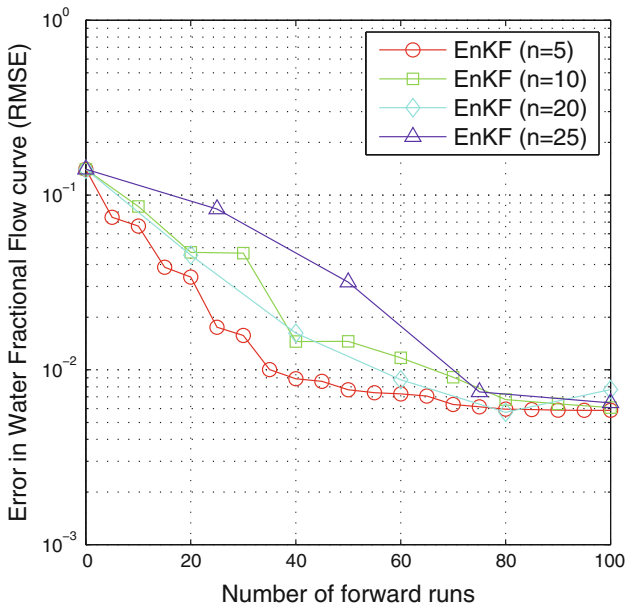


**Fig. 20** RMSE in water fractional flow curve versus the number of forward runs using different ensemble sizes for test case 3 with wells pattern 1

where $c_k$ is the perturbation magnitude and $\Delta_{ki}$ is $i$th component of the perturbation vector $\Delta_k$. The perturbation vector is generally sampled from a symmetric Bernoulli $\pm 1$ distribution (Spall 2003). The convergence of the SPSA iterations relies on the sequences $a_k$ and $c_k$, as they should go to 0 at the appropriate rates. The values of $\Delta_{ki}$ are independent and symmetrically distributed around 0. The SPSA gain sequences $a_k$ and $c_k$ are selected as (Spall 2003)

$$a_k = \frac{a}{(A+k+1)^{\alpha}}, \quad c_k = \frac{c}{(k+1)^{\gamma}} \tag{44}$$

where $k$ is the iteration number starting from $0, a, A, \alpha$ and $\gamma$ are a set of positive constants. Spall (2003) provided guidelines on how to pick these constants. For ensuring convergence of the algorithm the gain sequences have to satisfy the following conditions

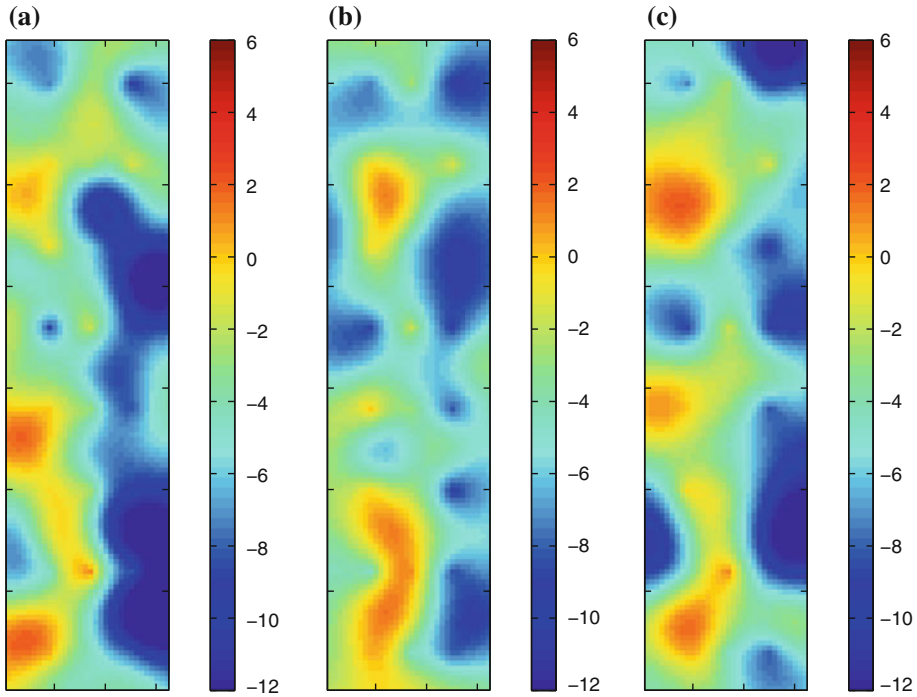


**Fig. 21** Log-permeability map for test case 3 with wells pattern 2: **a** EnKF optimized $n = 5$, **b** EnKF optimized $n = 10$ and **c** EnKF optimized $n = 20$ (vertical to horizontal scale ratio is 2:1)

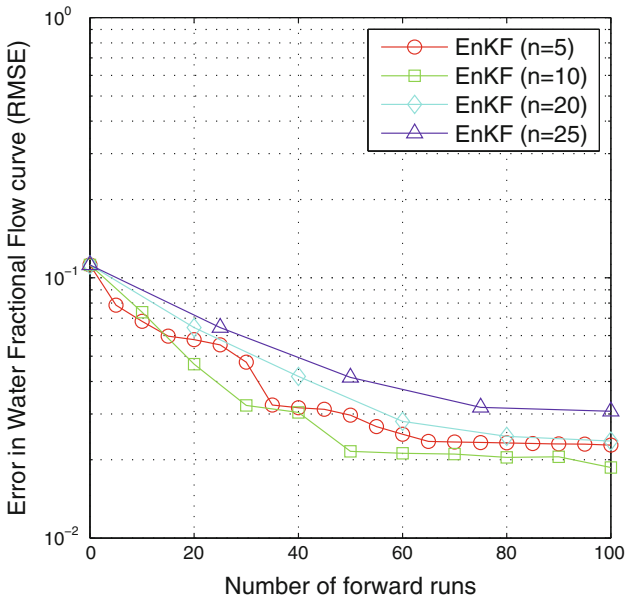


**Fig. 22** RMSE in water fractional flow curve versus the number of forward runs using different ensemble sizes for test case 3, wells pattern 2

$$a_k > 0, \quad c_k > 0, \quad a_k \to 0, \quad c_k \to 0, \quad \sum_{k=0}^{\infty} a_k = \infty,$$
$$\sum_{k=0}^{\infty} \left(\frac{a_k}{c_k}\right)^2 < \infty \tag{45}$$

Also the constants $\alpha$ and $\gamma$ should satisfy

$$\alpha - 2\gamma > 0, \quad 3\gamma - 0.5\alpha \geq 0 \tag{46}$$

The stability constant $A$ is recommended to be about 10 % of the allowed or expected number of search iterations (Spall 2003). The practical values for $\alpha$ and $\gamma$ are 0.602 and 0.101, respectively. These are the smallest values the satisfy the conditions in Eqs. 45 and 46.

In the current numerical study, we set $A$ to 10 which is about 10 % of the maximum 100 forward runs. The value of $a$ was adjusted to have a unit step size in the first iteration. We tested the algorithm with two different values of $c = 0.1$ and $c = 0.01$. Figure 25 shows the convergence rates of the SPSA method for problem 2 with wells pattern 1. Five different stochastic gradient were averaged at each iteration to produce an ensemble of 10 members. A total of 80 different runs were performed and both the individual convergence curves and their average are plotted. These results should be compared to results in Fig. 13. The proposed iterative regularized EnKF algorithm have higher convergence rates and clearly outperformed SPSA in reducing the RMSE in the first two iterations. Similar results for test problem 3 (wells pattern 1) are presented in Fig. 26. In comparison to the SPSA, the iterative regularized EnKF method produces signification error reduction in the first few iterations as shown in Fig. 20.

## 6 Conclusions

In this paper, a new parameter estimation method for subsurface flow models was presented. The proposed algorithm can be applied to any simulator and eliminates

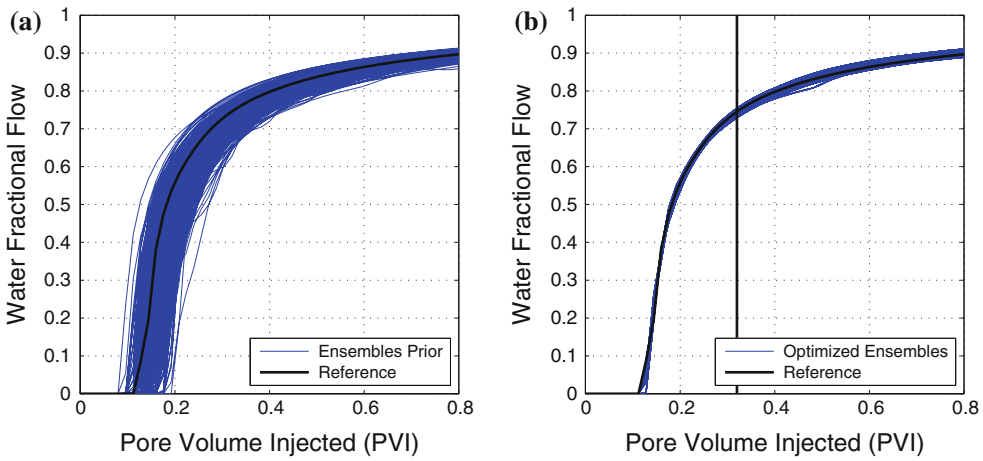


**Fig. 23** Results of a search space exploration study for test case 2, wells pattern 1. EnKF optimized versus initial ensemble fractional flow curves obtained using ensembles of 10 members, **a** Initial ensembles members and **b** Optimized ensembles members

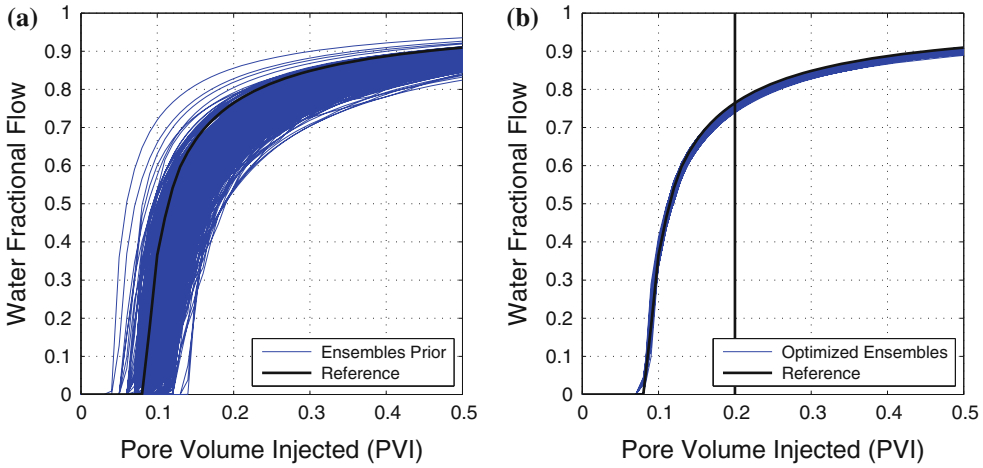


**Fig. 24** Results of a search space exploration study for test case 3, wells pattern 1. EnKF optimized versus initial ensemble fractional flow curves obtained using ensembles of 10 members, **a** initial ensembles members and **b** optimized ensembles members

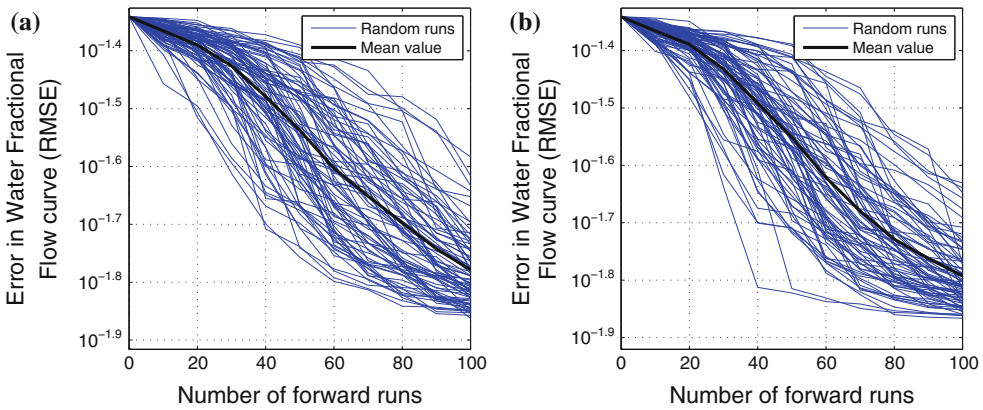


**Fig. 25** RMSE in water fractional flow curve versus the number of forward runs using SPSA with different magnitudes of the random perturbation for test case 2 (wells pattern 1) with ensembles of 10 members, **a** $c = 0.1$ and **b** $c = 0.01$

the need for expensive derivative evaluation required for gradient based algorithms. The algorithm relied on a novel combination of GPR, KL model reduction and TSVD regularized EnKF. GPR provided an easy method for incorporating static data into the model. Correlation lengths were obtained by maximizing the logarithm of the model evidence. KL expansion (aka. POD) was used as an effective dimension reduction tool. The use of GPR estimated mean field and covariance matrix for the KL dimension reduction eliminated the need for any pre-set parameters. This parameterization technique is essential for the smooth convergence of the inverse problem.

The inverse problem solution (dynamic data integration) was performed using an iterative regularized EnKF algorithm. EnKF was used in a batch mode where each time step corresponded to an iteration of a Newton like

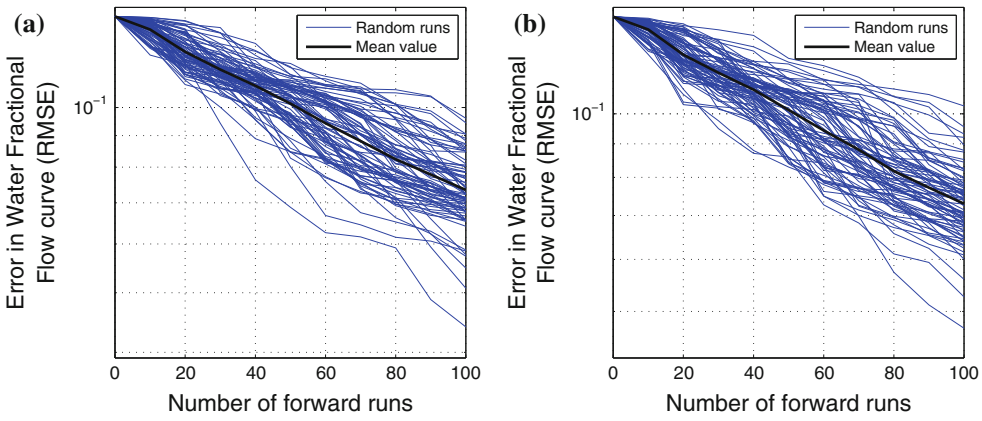


**Fig. 26** RMSE in water fractional flow curve versus the number of forward runs using SPSA with different magnitudes of the random perturbation for test case 3 (wells pattern 1) with ensembles of 10 members, **a** $c = 0.1$ and **b** $c = 0.01$

correction step. These iterations were repeated until convergence or a maximum number of steps was reached. The Kalman gain matrix was filtered using TSVD to eliminates spurious correlations. This method is automatic and is based on a standard regularization technique. Other methods based on covariance localization cannot be easily adopted as the de-correlation length is hard to define once the KL model reduction is applied.

The algorithms showed smooth convergence for very small ensembles of 5 or 10 members. The use of small ensembles enables extensive search space exploration as demonstrated in the examples. As a general optimization tool, this algorithm can get trapped at local minima similar to any unconstrained minimization gradient based algorithm. Also, the convergence of the algorithm will be dependent on step size. Here, we used a unit step size in all our numerical testing. Augmenting the algorithm with a line search step to automatically adjust the update step size might improve the convergence rates of the algorithm.

**Acknowledgments** The authors would like to thank the anonymous reviewers for their insightful and constructive comments which helped enhance this manuscript. Dr. A. H. Elsheikh and Dr. Jefferson L. M. A. Gomes carried this work as part of activities of the Qatar Carbonates and Carbon Storage Research Centre (QCCSRC). They gratefully acknowledge the funding of QCCSRC provided jointly by Qatar Petroleum, Shell, and the Qatar Science and Technology Park. Professor I. M. Navon acknowledges the support of NSF/CMG grant ATM-0931198.

## References

Allen M, Frame D, Kettleborough J, Stainforth D (2006) Model error in weather and climate forecasting. In: Palmer T, Hagedorn R (eds) Predictability of weather and climate. Cambridge University Press, Cambridge

Anderson JL (2001) An ensemble adjustment Kalman filter for data assimilation. Mon Weather Rev 129(12):2884–2903

Anderson JL (2003) A local least squares framework for ensemble filtering. Mon Weather Rev 131(4):634–642

Anderson JL, Anderson SL (1999) A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilation and forecasts. Mon Weather Rev 127:2741–2758

Anderson BDO, Moore JB (1979) Optimal filtering. Information and system sciences series. Prentice-Hall, Inc., Englewood Cliffs, NJ

Bengtsson T, Snyder C, Nychka D (2003) Toward a nonlinear ensemble filter for high-dimensional systems. J Geophys Res 108(D24):8775–8785

Blum J, Le Dimet FX, Navon IM (2008) Data assimilation for geophysical fluids. In: Ciarlet PG, Temam R, Tribbia J (eds) Computational methods for the atmosphere and the oceans. Handbook of numerical analysis, vol 14. Elsevier, Amsterdam, pp 385–442

Bulygina N, Gupta H (2010) How bayesian data assimilation can be used to estimate the mathematical structure of a model. Stoch Environ Res Risk Assess 24(6):925–937

Carrera J, Alcolea A, Medina A, Hidalgo J, Slooten LJ (2005) Inverse problem in hydrogeology. Hydrogeol J 13(1):206–222

Chen Z (2007) Reservoir simulation: mathematical techniques in oil recovery. Society for Industrial and Applied Mathematics, Philadelphia, PA

Chen Y, Oliver D (2012) Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math Geosci 44(1):1–26

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

Christie M, Blunt M (1995) Tenth SPE comparative solution project: a comparison of upscaling techniques. SPE Reserv Eval Eng 4:308–317

Cohn SE (1997) An introduction to estimation theory. J Meteorol Soc Jpn Ser II 75(1B):257–288

Dostert P, Efendiev Y, Mohanty B (2009) Efficient uncertainty quantification techniques in inverse problems for Richards' equation using coarse-scale simulation models. Adv Water Resour 32(3):329–339

Dovera L, Della Rossa E (2011) Improved initial ensemble generation coupled with ensemble square root filters and inflation to estimate uncertainty. Comput Geosci 1–17

Efendiev Y, Datta-Gupta A, Ginting V, Ma X, Mallick B (2005) An efficient two-stage markov chain monte carlo method for dynamic data integration. Water Resour Res 41(12)

ELsheikh AH, Jackson MD, Laforce TC (2012) Bayesian reservoir history matching considering model and parameter uncertainties. Math Geosci 44(5):515–543

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. J Geophys Res 99(C5):10,143–10,162

Evensen G, van Leeuwen PJ (2000) An ensemble Kalman smoother for nonlinear dynamics. Mon Weather Rev 128(6):1852–1867
Fang H, Gong G, Qian M (1997) Annealing of iterative stochastic schemes. SIAM J Control Optim 35(6):1886–1907
Fang F, Pain CC, Navon IM, Piggott MD, Gorman GJ, Allison PA, Goddard AJH (2009) Reduced-order modelling of an adaptive mesh ocean model. Int J Numer Methods Fluids 59(8):827–851
Fu J, Gomez-Hernandez J (2009) A blocking markov chain monte carlo method for inverse stochastic hydrogeological modeling. Math Geosci 41(2):105–128
Gaspari G, Cohn SE (1999) Construction of correlation functions in two and three dimensions. Q J R Meteorol Soc 125(554):723–757
Gelfand S, Mitter S (1991) Simulated annealing type algorithms for multivariate optimization. Algorithmica 6(1):419–436
Ghanem RG, Spanos PD (1991) Stochastic finite elements: a spectral approach. Springer-Verlag New York, Inc., New York
Golub G, Van Loan C (1996) Matrix computations, 3rd edn. Johns Hopkins University Press, Baltimore
Gu Y, Oliver DS (2007) An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. SPE J 12(4):438–446
Hansen C (1998) Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion. SIAM, Philadelphia
Hillery A, Chin R (1991) Iterative wiener filters for image restoration. IEEE Trans Signal Process 39(8):1892–1899
Houtekamer PL, Mitchell HL (2001) A sequential ensemble Kalman filter for atmospheric data assimilation. Mon Weather Rev 129(1):123–137
Houtekamer PL, Mitchell HL (2005) Ensemble Kalman filtering. Q J R Meteorol Soc 131(613):3269–3289
Jazwinski AH (1970) Stochastic processes and filtering theory. Academic Press, New York
Kac M, Siegert AJF (1947) An explicit representation of a stationary Gaussian process. Ann Math Stat 18:438–442
Karhunen K (1947) Über lineare Methoden in der Wahrscheinlichkeitsrechnung. Ann Acad Sci Fenn Ser A I Math Phys 1947(37):79
Krymskaya MV, Hanea RG, Verlaan M (2009) An iterative ensemble Kalman filter for reservoir engineering applications. Comput Geosci 13(2):235–244
Kushner HJ (1987) Asymptotic global behavior for stochastic approximation and diffusions with slowly decreasing noise effects: global minimization via monte carlo. SIAM J Appl Math 47(1):169–185
Li G, Reynolds AC (2009) Iterative ensemble Kalman filters for data assimilation. SPE J 14(3):496–505
Loève M (1948) Fonctions aléatoires de second order. In: Levy P (ed) Processus Stochastiques et Movement Brownien. Hermann, Paris
Lorentzen RJ, Naevdal G (2011) An iterative ensemble Kalman filter. IEEE Trans Autom Control 56(8):1990–1995
Ma X, Al-Harbi M, Datta-Gupta A, Efendiev Y (2008) An efficient two-stage sampling method for uncertainty quantification in history matching geological models. SPE J 13(1):77–87
MacKay DJC (1999) Comparison of approximate methods for handling hyperparameters. Neural Comput 11(5):1035–1068
McLaughlin D, Townley LR (1996) A reassessment of the groundwater inverse problem. Water Resour Res 32(5):1131–1161
Moradkhani H, Sorooshian S, Gupta HV, Houser PR (2005) Dual state–parameter estimation of hydrological models using ensemble Kalman filter. Adv Water Resour 28(2):135–147
Naevdal G, Johnsen L, Aanonsen S, Vefring E (2005) Reservoir monitoring and continuous model updating using ensemble Kalman filter. SPE J 10(1):66–74
Navon IM (1998) Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. Dyn Atmos Oceans 27(1–4):55–79
Nocedal J, Wright SJ (2006) Numerical Optimization, 2nd edn. Springer Verlag
Oliver D, Cunha L, Reynolds A (1997) Markov chain monte carlo methods for conditioning a permeability field to pressure data. Math Geol 29(1):61–91
Ott E, Hunt BR, Szunyogh I, Zimin AV, Kostelich EJ, Corazza M, Kalnay E, Patil DJ, Yorke JA (2004) A local ensemble Kalman filter for atmospheric data assimilation. Tellus A 56(5):415–428
Pham DT, Verron J, Roubaud MC (1998) A singular evolutive extended Kalman filter for data assimilation in oceanography. J Mar Syst 16(3–4):323–340
Rasmussen CE, Williams CKI (2005) Gaussian processes for machine learning (adaptive computation and machine learning). MIT Press, Cambridge
Sadegh P, Spall J (1998) Optimal random perturbations for stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Trans Autom Control 43(10):1480–1484
Sætrom J, Omre H (2011) Ensemble Kalman filtering with shrinkage regression techniques. Comput Geosci 15(2):271–292
Sakov P, Oliver DS, Bertino L (2012) An iterative enkf for strongly nonlinear systems. Mon Weather Rev 140(6):1988–2004
Simon E, Bertino L (2009) Application of the gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the north atlantic with the ENKF: a twin experiment. Ocean Sci 5(4):495–510
Smith KW (2007) Cluster ensemble Kalman filter. Tellus A 59(5): 749–757
Sørensen JVT, Madsen H (2004) Data assimilation in hydrodynamic modelling: on the treatment of non-linearity and bias. Stoch Environ Res Risk Assess 18(4):228–244
Spall J (2003) Introduction to stochastic search and optimization: estimation, simulation, and control. Wiley-Interscience series in discrete mathematics and optimization. Wiley-Interscience, Hoboken
Sun AY, Morris A, Mohanty S (2009) Comparison of deterministic ensemble Kalman filters for assimilating hydrogeological data. Adv Water Resour 32(2):280–292
Thacker WC (1989) The role of the hessian matrix in fitting models to measurements. J Geophys Res 94(C5):6177–6196
Tippett MK, Anderson JL, Bishop CH, Hamill TM, Whitaker JS (2003) Ensemble square root filters. Mon Weather Rev 131(7): 1485–1490
Tong J, Hu B, Yang J (2010) Using data assimilation method to calibrate a heterogeneous conductivity field conditioning on transient flow test data. Stoch Environ Res Risk Assess 24(8):1211–1223
Tong J, Hu B, Yang J (2012) Assimilating transient groundwater flow data via a localized ensemble Kalman filter to calibrate a heterogeneous conductivity field. Stoch Environ Res Risk Assess 26(3):467–478
Wan E, Van Der Merwe R (2000) The unscented Kalman filter for nonlinear estimation. In: Adaptive systems for signal processing, communications, and control symposium 2000. AS-SPCC. The IEEE 2000, Lake Louise, pp 153 –158
Zhang D, Lu Z, Chen Y (2007) Dynamic reservoir data assimilation with an efficient, dimension-reduced Kalman filter. SPE J 12(1):108–117
Zhou E, Fu M, Marcus S (2008) A particle filtering framework for randomized optimization algorithms. In: Simulation conference, 2008. WSC 2008, Winter, pp 647–654
Zhou H, Gomez-Hernandez JJ, Franssen HJH, Li L (2011) An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. Adv Water Resour 34(7):844–864
Zupanski M (2005) Maximum likelihood ensemble filter: theoretical aspects. Mon Weather Rev 133(6):1710–1726
Zupanski M, Navon IM, Zupanski D (2008) The maximum likelihood ensemble filter as a non-differentiable minimization algorithm. Q J R Meteorol Soc 134(633):1039–1050