

Noname manuscript No. (will be inserted by the editor)
--

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

Received: date / Accepted: date

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 and without any tuning parameters. The inverse problem is formulated as a sequential data integration problem. Gaussian Process Regression (GRP) 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 ensemble Kalman filter 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
Department of Earth Science and Engineering,
Imperial College London,
Prince Consort Road, London, SW7 2BP, UK
Tel.: +44 20 759 47325
E-mail: a.el-sheikh@imperial.ac.uk

C.C. Pain · F. Fang · J.L.M.A. Gomes
Department of Earth Science and Engineering,
Imperial College London, SW7 2BP, 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; Fu and Gomez-Hernandez, 2008). 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 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, 2008, 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, 2008, 2009), gradient based optimization methods (McLaughlin and Townley, 1996; Carrera et al, 2005) and ensemble Kalman filter methods (Moradkhani et al, 2005; Naevdal et al, 2005; Chen and Zhang, 2006).

Ensemble Kalman Filter (EnKF) is a parallel sequential Monte Carlo method (SMC) for data assimilation. EnKF was introduced by Evensen (Evensen, 1994) and since then have been used for subsurface model update (Moradkhani et al, 2005; Naevdal et al, 2005; Chen and Zhang, 2006). 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 variations of ensemble filters have been proposed (Pham et al, 1998; Bishop et al, 2001; Ott et al, 2004; Tippett et al, 2003). 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, 2009b). 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, 2009b). The use of relatively small ensemble size might results in overshooting of the estimated parameters due to sampling errors in the estimated covariance matrices. Techniques to solve this problem include covariance localization (Gaspari and Cohn, 1999; Houtekamer and Mitchell, 2001), local approximation of the state error covariance using Local Analysis (LA) (Anderson, 2003; Ott et al, 2004; Sakov and Bertino, 2011) and covariance inflation (Anderson and Anderson, 1999; Anderson, 2001). EnKF algorithms are limited to Gaussian system as they rely on the first two moments of the ensemble statistics. Several studies were carried out to extend EnKF to handle non-Gaussian estimation problems (Bengtsson et al, 2003; Smith, 2007; Sun et al, 2009a; 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 Siebert, 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 small number of iterations while using very small ensemble sizes.

The proposed algorithm have 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). The proposed algorithm have 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 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; Dostert et al, 2009; Zeng and Zhang, 2010). 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. 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 as it is an order of magnitude lower than any previously published results. These small ensembles enable extensive exploration of the parameter space for uncertainty quantification for subsurface flow models.

The organization of this paper is as follows: section 2 presents two tools for parametrizing the search space, Gaussian process regression 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. Section 5 utilizes the EnKF parameter estimation algorithm for an uncertainty quantification study to produce a diverse ensemble of models for future forecast. The conclusions of the current work are drawn in section 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 exists. Accounting for these point data is commonly denoted as a static data integration process. We utilized Gaussian process regression for solving the static data integration problem. The output of the GPR is then subjected to a dimension reduction technique using Karhunen-Loève expansion. The combined used 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 $\{(\mathbf{x}_i, y_i) | i = 1, \dots, n\}$, 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 \tilde{y} given the corresponding input $\tilde{\mathbf{x}}$. 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 $\mathbf{y} = (y_1, \dots, y_n)^T$, then GP is defined as

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

where $\boldsymbol{\mu}$ is the mean function and \mathbf{C} is the covariance matrix. The covariance matrix is specified as $[\mathbf{C}]_{i,j} = \text{cov}(y_i, y_j) = \mathcal{C}(\mathbf{x}_i, \mathbf{x}_j)$, 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

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

where σ_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 σ_n are known as the *hyperparameters* of the Gaussian process $\psi = \langle \sigma_c, l, \sigma_n \rangle$.

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

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

where, \mathbf{m} is a vector of means corresponding to the input data vector \mathbf{X} , $\tilde{\mathbf{m}}$ contains prior mean values for the new data points $\tilde{\mathbf{X}}$, $\tilde{\mathbf{y}}$ is a vector of the posterior means for each new data points, \mathbf{C}_{xx} is the covariance matrix of the input data, $\mathbf{C}_{x\tilde{x}}, \mathbf{C}_{\tilde{x}x}$ are the two cross covariance matrices, $\mathbf{C}_{\tilde{x}\tilde{x}}$ is the prior covariance matrix for the new data points, and \mathbf{I} is the identity matrix. The conditional distribution $p(\tilde{y} | \tilde{\mathbf{x}}, \mathbf{y}, \mathbf{X})$ is a normal distribution with the mean $\tilde{\boldsymbol{\mu}} = \tilde{\mathbf{m}} + \mathbf{C}_{\tilde{x}x}(\mathbf{C}_{xx} + \sigma_n^2 \mathbf{I})^{-1}(\mathbf{y} - \mathbf{m})$ and covariance $\tilde{\mathbf{C}} = \mathbf{C}_{\tilde{x}\tilde{x}} - \mathbf{C}_{\tilde{x}x}(\mathbf{C}_{xx} + \sigma_n^2 \mathbf{I})^{-1}\mathbf{C}_{x\tilde{x}}$ (Appendix A.2 in (Rasmussen and Williams, 2005)). 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 α^T denotes the transpose of α). In the current work, we use the Matérn covariance function (Rasmussen and Williams, 2005) defined as

$$\mathcal{C}_{\text{Matern}}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{r}{l} \right)^\nu J_\nu \left(\sqrt{2\nu} \frac{r}{l} \right) \quad (4)$$

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

Learning a Gaussian process 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 $\psi = \langle \nu, l, \sigma_n \rangle$ that maximizes the conditional evidence

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

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)

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

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

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

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). In this paper, the order parameter ν of the Matérn covariance function was set to 3.

2.2 Karhunen-Loève dimension reduction

The Karhunen-Loève expansion (KL) (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 (Berkooz et al, 1993). The results 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 ξ belonging to space of random events. Karhunen-Loève expansion provides a Fourier-like series form of $\mathcal{K}(\mathbf{x}, \xi)$ as

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

where ξ_k is a set of random variables, λ_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:

$$\mathcal{C}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{k=1}^{\infty} \lambda_k \psi_k(\mathbf{x}_1) \psi_k(\mathbf{x}_2) \quad (9)$$

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

$$\int_{\Omega} \mathcal{C}(\mathbf{x}_1, \mathbf{x}_2) \psi_k(\mathbf{x}_2) d\mathbf{x}_2 = \lambda_k \psi_k(\mathbf{x}_1), \quad k = 1, 2, \dots \quad (10)$$

The basis functions ψ_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

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

where δ_{kj} is the Kronecker delta. The random variables ξ_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 Gaussian process, ξ_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 ξ_n can be obtained from

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

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

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

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.

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

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 ξ_k . The dynamic data integration problem is concerned with finding values of ξ_k such that the measured production data matches the simulation results.

3 Ensemble Kalman filter (EnKF)

The Ensemble Kalman Filter (EnKF) is a parallel sequential Monte Carlo method (SMC) for data assimilation. This method was introduced by Evensen (Evensen, 1994). Ensemble Kalman Filter relies on two steps: prediction and update (Gillijns et al, 2006). For a discrete time nonlinear system:

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

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

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 $\mathbf{X}_t = [\mathbf{x}_t^1, \mathbf{x}_t^2, \dots, \mathbf{x}_t^n]$ are generated and the corresponding measurements are $\mathbf{Y}_t = [\mathbf{y}_t^1, \mathbf{y}_t^2, \dots, \mathbf{y}_t^n]$. The matrix \mathbf{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):

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

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

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

$$\mathbf{y}_i^o = \mathbf{y}_i^f + r_i, \quad i = 1, \dots, n, \quad (20)$$

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

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

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 the overbar denotes the ensemble mean.

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. 15 and the state update equation will have the form $\mathbf{x}_i^f = \mathbf{x}_i^a + \epsilon_k w_i$ where, ϵ_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. The measurement Eq. 16, corresponds to the simulator output of the production data:

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

where \mathcal{SLM} 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:

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

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:

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

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

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

where d_{obs} is the observed data and \mathbf{K} is the Kalman gain matrix defined as:

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

The measurement error covariance matrix \mathbf{R} will be eliminated from the Kalman gain matrix as the measurement errors are set to zero for the optimization algorithm. The Kalman gain equation can be thought of as a corrector step utilizing an approximate Hessian based on the Kalman gain

matrix (Sivia and Skilling, 2006). 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 the covariance 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 SVD, to obtain a set of orthogonal basis functions satisfying

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (28)$$

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 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k \geq 0$ corresponding to the singular values. The matrix \mathbf{A} will have a condition number $\text{cond}(\mathbf{A}) = \sigma_1/\sigma_k$. Given the SVD decomposition the solution of the system is equal to

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

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) (Golub, G. H. and Kahan, W., 1965; Hanson, 1971) or by Tikhonov Regularization (Tikhonov and Arsenin, 1977). 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 et al, 2006):

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

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

$$\phi_i \equiv \begin{cases} 1, & i = 1, 2, \dots, t \\ 0, & i = t + 1, t + 2, \dots, k \end{cases} \quad (31)$$

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

The Kalman gain matrix has two components, the cross covariance and the output covariance matrix. Stabilizing the inverse of the output covariance matrix $\widetilde{\mathbf{C}}_{yy}$ is essential such that small singular values do not dominate the solution. The output matrix \mathbf{Y} will be regularized using TSVD as follows

$$\widetilde{\mathbf{Y}} = (\mathbf{U}_y(\Phi\mathbf{S}_y)\mathbf{V}_y^T) \quad (32)$$

where Φ is a diagonal matrix, with diagonal values corresponds to the regularization parameters ϕ . Using this regularized (filtered) matrix, the regularized covariance matrix $\widetilde{\mathbf{C}}_{yy}$ can be estimated. The inverse of the covariance matrix $\widetilde{\mathbf{C}}_{yy}$ can be efficiently evaluated as:

$$(\widetilde{\mathbf{C}}_{yy})^{-1} = (n-1) \left(\mathbf{U}_y(\Phi(\mathbf{S}_y)^{-2})\mathbf{U}_y^T \right) \quad (33)$$

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. The regularized Kalman gain matrix is defined as:

$$\widetilde{\mathbf{K}} = \mathbf{C}_{xy}(\widetilde{\mathbf{C}}_{yy})^{-1} \quad (34)$$

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).

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 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

$$\nabla \cdot \mathbf{v}_t = q, \quad \mathbf{v}_t = -\mathbb{K}\lambda_t(S_w)\nabla p \quad \text{over } \Omega \quad (35)$$

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, ρ_w, ρ_o are the water and oil fluid densities, respectively. These equations can be combined to produce the pressure equation

$$-\nabla \cdot (\mathbb{K}\lambda_t(S_w)\nabla p) = q \quad (36)$$

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

$$\phi \frac{\partial S_w}{\partial t} + \nabla \cdot (f(S_w) \mathbf{v}_t) = \frac{Q_w}{\rho_w} \quad (37)$$

where ϕ 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

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

$$S_{nw} = \frac{S_w - S_{wc}}{1 - S_{or} - S_{wc}} \quad (38)$$

where S_{wc} is the connate or irreducible water saturation, S_{or} is the irreducible oil saturation and μ_w, μ_o are the water and oil fluid viscosities, respectively. The pressure Eq. 36 is discretized using standard two-point flux approximation (TPFA) method and the saturation Eq. 37 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 with an injector-producer pattern. The injection of water is done at the first grid block (south-west) and the production well is located at the last grid block (north-east). No flow boundary conditions is applied on the domain boundaries. The true permeability map is sampled at certain number of locations and these values are used to construct the Gaussian process. The Matérn covariance is used for all GPR analysis and the value of $\nu = 3$ 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 observation in our testing.

4.1 Test case 1

In this test case, the model is based on a 2D regular grid of 41×41 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 μ_w is 0.3×10^{-3} Pa.s and the oil viscosity μ_o is set to 0.3×10^{-3} Pa.s. The irreducible water saturation and irreducible oil saturation are set as $S_{or} = S_{wc} = 0.2$.

Fig. 1(a) shows the logarithm of the reference permeability field (permeability units is Darcy = $9.869233 \times$

$10^{-13}m^2$). 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. 1(b) shows the mean log-permeability field obtained from the GP regression. The mean values along with the ± 2 standard deviations bounds are shown in Fig. 2, where the reference field is plotted in red and the interpolated mean value 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. This limits the search space to only 40 dimensions. This represents a significant reduction of search-space size in comparison to direct parameter estimation techniques where the size of the search space corresponds to the number of grid blocks in the model.

Fig. 4 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 converges to the reference data after 100 forward runs regardless of the ensemble size. Fig. 5 shows the optimized log-permeability field at the end of the EnKF iterations for three different ensemble sizes. This figure shows that different ensembles have recovered different modes of the solution as the problem is ill-posed and might admit different solutions. Fig. 6 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 reference solution significantly faster than larger ensemble runs. This can be attributed to the smoothness of the problem. However, all ensembles have shown 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 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

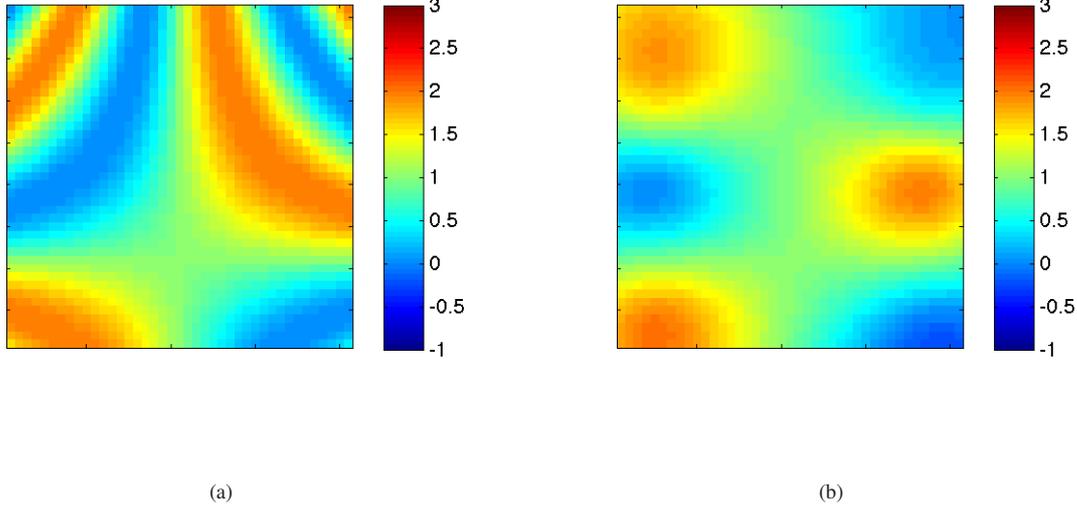


Fig. 1 Log-permeability map for test case 1: (a) Reference and (b) GPR mean value.

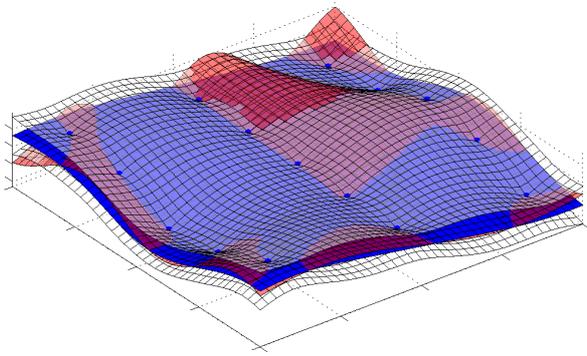


Fig. 2 GP regression result along with the two standard deviations bounds for test case 1 (reference field in red, estimated mean in blue and two standard deviations bounds in white).

$\mu_o = 3 \times 10^{-3}$ Pa.s are used. The irreducible water and oil saturations are set as $S_{wc} = S_{or} = 0.2$. Fig. 7(a) shows the logarithm of the reference permeability field (in Darcy = $9.869233 \times 10^{-13} m^2$). This field is sampled at 24 points and these values are used to constrain the Gaussian process regression. The mean regression results obtained from the GPR is shown in Fig. 7(b). The KL-expansion was applied to reduce the search space to the first 40 eigen modes as they retain most of the variance of GPR solution.

Fig. 7 parts (c),(d) and (e) show 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. 8. The water-cut curve fully matches the reference water cut curve after the optimization run. Fig. 9 shows the RMSE in the fractional flow curve versus the number of EnKF iterations. A smooth convergence is observed for all ensemble sizes, however the very small ensemble of 5 members shows some instability in error reduction. This instability can be eliminated by augmenting the algorithm with a line search step. The ensembles of 10 member performed much better than larger ensembles. 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.

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 . Fig. 10 shows the RMSE of the water-cut curve for multiple runs 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

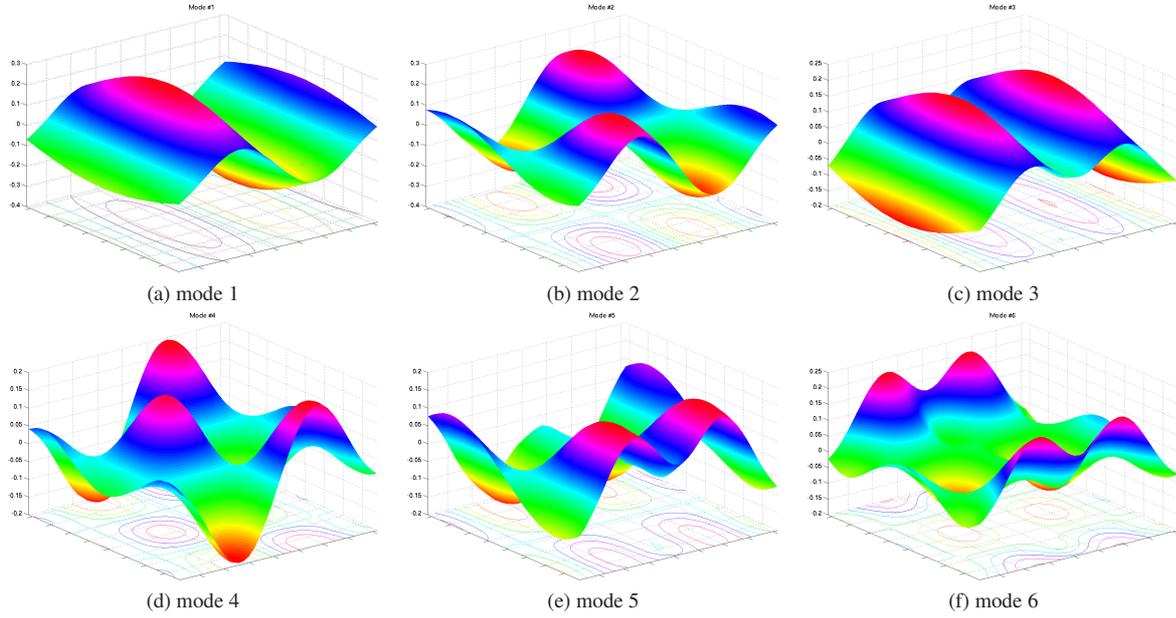


Fig. 3 First six eigen Modes obtained by KL-expansion of the GPR covariance matrix for test case 1.

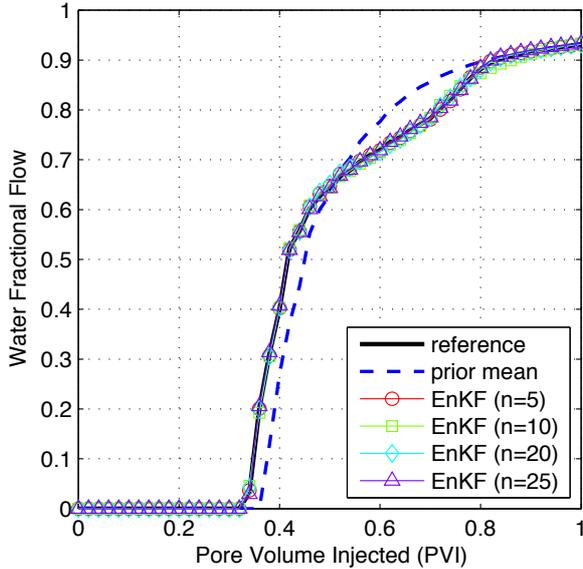


Fig. 4 Reference, initial and optimized water fractional flow curves using different ensemble sizes for test case 1.

step size might be applied within a line search strategy using Wolfe or Goldstein condition (Nocedal and Wright, 2006).

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 μ_w

is set to 0.3×10^{-3} Pa.s and the oil viscosity μ_o is set to 3×10^{-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. 11(a). A set of 24 values of the permeability field are used for the GPR. The resulting mean field is shown in Fig. 11(b) and the EnKF optimized log-permeability fields are plotted in Fig. 11(c),(d) and (e) for ensembles of size 5, 10 and 20 members, respectively.

The EnKF optimized water-cut curves are shown in Fig. 12 and shows a full agreement with the reference water-cut curve. The regularized EnKF algorithm successfully solved the inverse problem and converged to the true curve regardless of the utilized ensemble size. Fig. 13 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, after 20 forward runs it fails in further reduction of the errors. This can be attributed to the addition of the noise at each update step and that amount of noise needs to be reduced at a higher rate for small ensembles. It is also observed that ensembles of 10 members did show faster convergence rates than ensembles with more members.

5 Uncertainty quantification

In this section we use the iterative regularized EnKF algorithm for an uncertainty quantification study. The proposed algorithm, similar to gradient based algorithms, might be at-

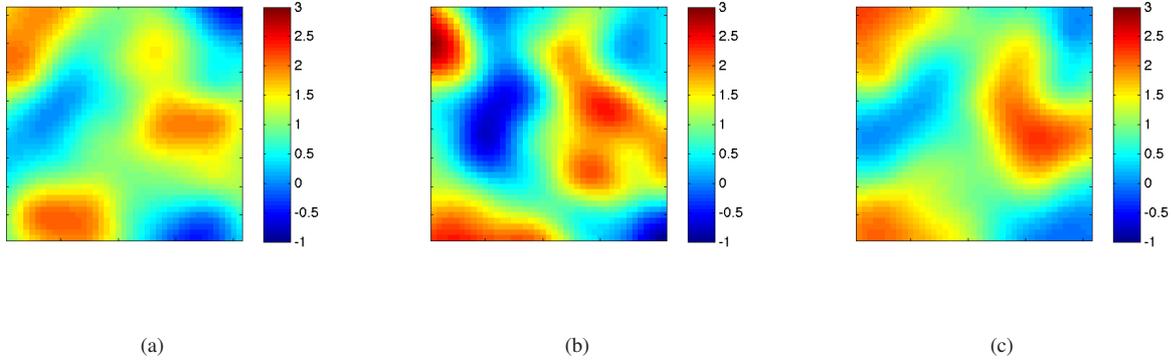


Fig. 5 Optimized Log-permeability map for test case 1 with different ensemble sizes: (a) $n = 5$, (b) $n = 10$ and (c) $n = 20$.

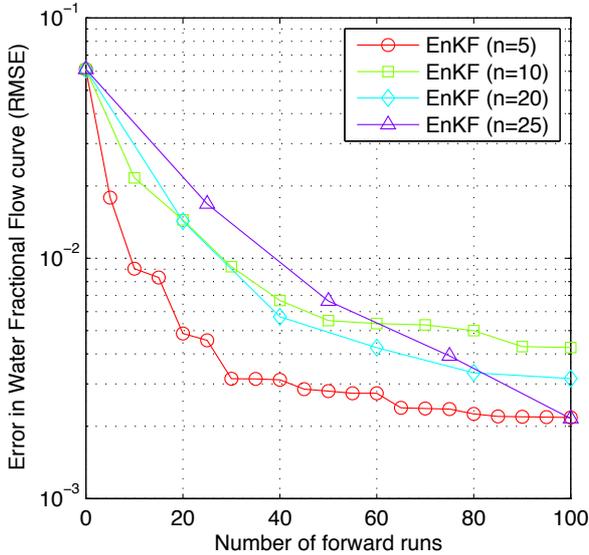


Fig. 6 RMSE in water fractional flow curve versus the number of forward runs using different ensemble sizes for test case 1.

tracted 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 subsequent uncertainty quantification and future forecasts. Problems 2 and 3 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 conforms to all static data even if it is different than the GPR mean.

Fig. 14 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. 14(b). The rest of the fractional flow curve is an out-of-sample data that can be used effectively for future forecasts. All plotted curves represent the ensemble mean response and each curve represents a different run of the algorithm. Similarly, parallel simulations with random starting points were performed for test case 3. The results are shown in Fig. 15 in terms of water cut curves. The accuracy of the optimized ensembles is evident in the out-of-sample data. The ensemble predictions via the estimated sample mean and associated variance can be used for uncertainty quantification as a simple Monte Carlo method. It can be also used as an input for a non-intrusive stochastic collocation method (Xiu, 2009).

6 Conclusions

In this paper, a new parameter estimation for subsurface flow models was presented. This algorithm can be applied to any simulator and eliminates the need for expensive derivative evaluation required for gradient based algorithms. The algorithm relies on a novel combination of GPR, KL model reduction and TSVD regularized EnKF. Gaussian process regression provided an easy method for incorporating static data into the model. Correlation lengths are obtain by maximizing the logarithm of the model evidence. KL expansion (aka. POD) is used as an effective dimension reduction tool. The use of GPR estimated mean field and covariance matrix for the KL dimension reduction eliminates the need for pre-set parameters. These parameterization techniques are essential for the smooth convergence of the inverse problem.

The inverse problem solution (dynamic data integration) was performed using an iterative regularized EnKF algo-

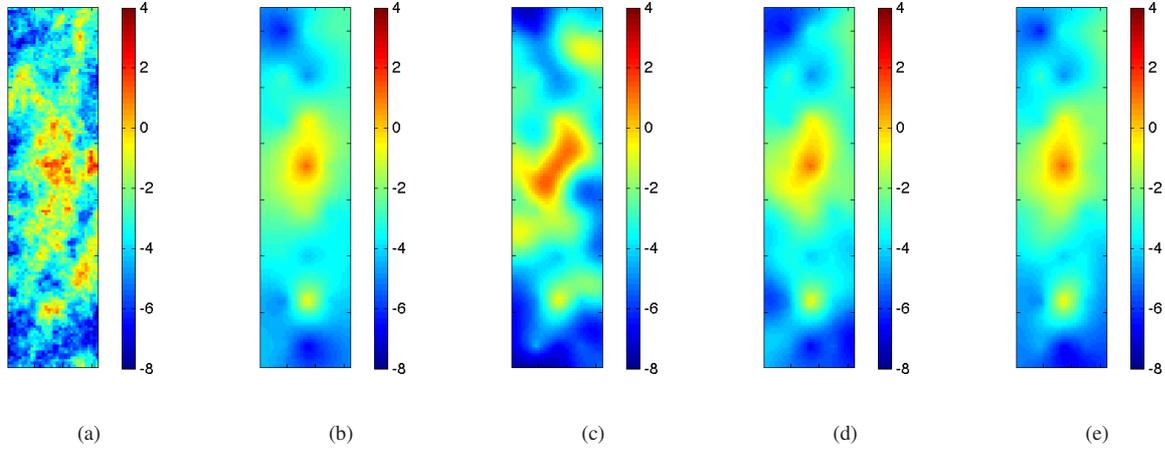


Fig. 7 Log-permeability map for test case 2: (a) Reference, (b) GPR mean value, (c) EnKF optimized $n = 5$, (d) EnKF optimized $n = 10$ and (e) EnKF optimized $n = 20$.

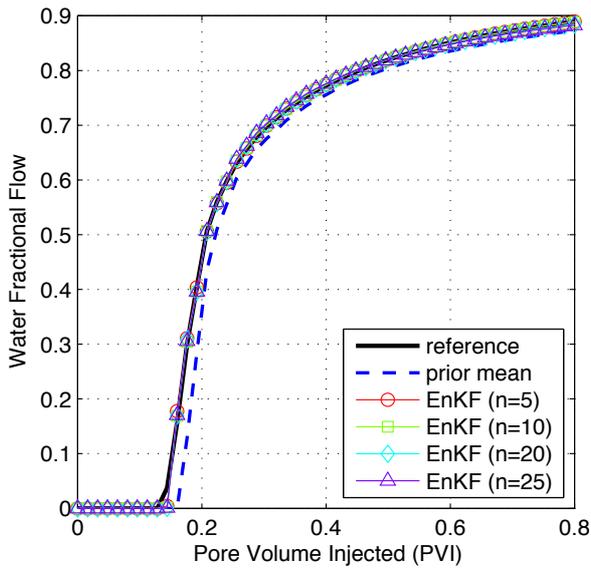


Fig. 8 Reference, initial and optimized water fractional flow curves using different ensemble sizes for test case 2.

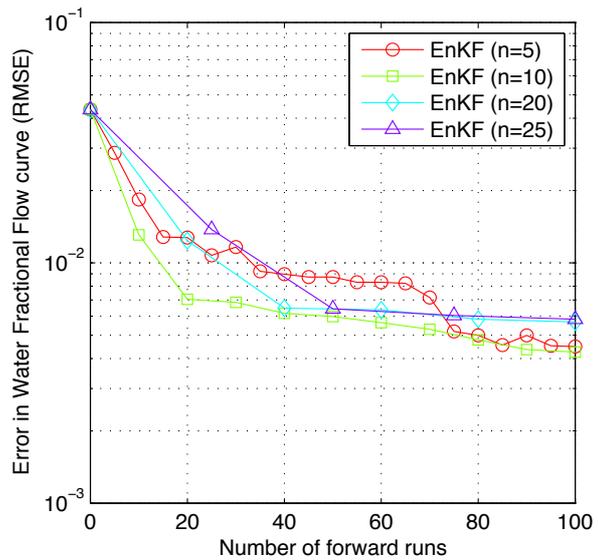


Fig. 9 RMSE in water fractional flow curve versus the number of forward runs using different ensemble sizes for test case 2.

rithm. The proposed EnKF algorithm is used in a batch mode where each time step corresponds to an iteration of a Newton like optimization algorithm. These iterations are repeated until convergence or a maximum number of steps is reached. The Kalman gain matrix was filtered using TSVD to eliminate spurious correlations. This method is automatic and is based upon a standard and reliable 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. Regularization based on re-sampling methods as introduced in

(Zhang and Oliver, 2010) might not be applicable for small ensembles.

The algorithms showed smooth convergence for very small ensembles of 5 or 10 members. These ensembles are an order of magnitude smaller than those used in related published work. The use of small ensembles enables extensive search space exploration for uncertainty quantification studies as demonstrated in the examples. As a general optimization tool, this algorithm can get trapped at local minima similar to any 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. Aug-

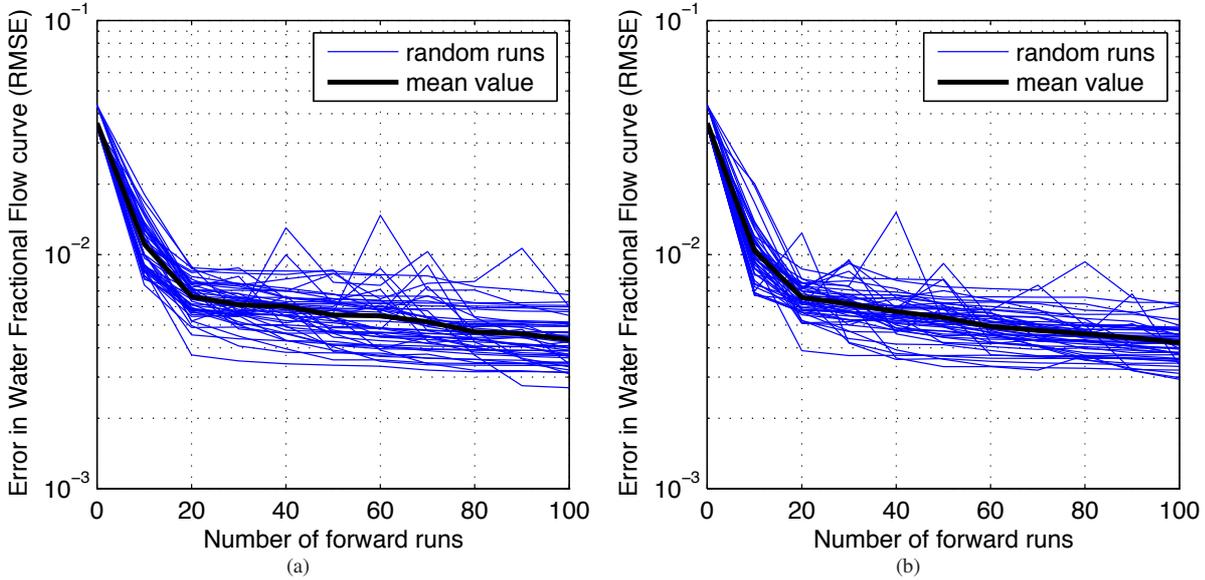


Fig. 10 RMSE in water fractional flow curve versus the number of forward runs using different magnitudes of the random perturbation for test case 2 with ensembles of 10 members, (a) $c = 0.01$ and (b) $c = 0.04$

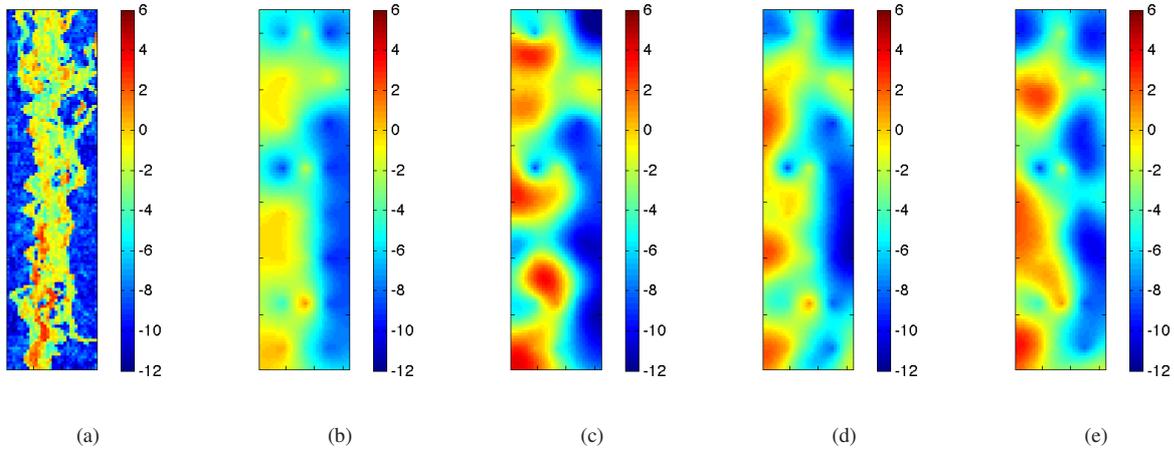


Fig. 11 Log-permeability map for test case 3: (a) Reference, (b) GPR mean value, (c) EnKF optimized $n = 5$, (d) EnKF optimized $n = 10$ and (e) EnKF optimized $n = 20$.

menting the algorithm with a line search step to automatically adjust the update step size can solve this problem.

Acknowledgements Professor I. M. Navon acknowledges the support of NSF/CMG grant ATM-0931198.

References

- Anderson JL, Anderson SL (1999) A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilation and forecasts. *Monthly Weather Review* 127:2741–2758
- Anderson JL (2001) An ensemble adjustment kalman filter for data assimilation. *Monthly Weather Review* 129(12):2884–2903
- Anderson JL (2003) A local least squares framework for ensemble filtering. *Monthly Weather Review* 131(4):634–642
- Bengtsson T, Snyder C, Nychka D (2003) Toward a nonlinear ensemble filter for high-dimensional systems. *J Geophys Res* 108(D24)
- Berkooz G, Holmes P, Lumley JL (1993) The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics* 25(1):539–575

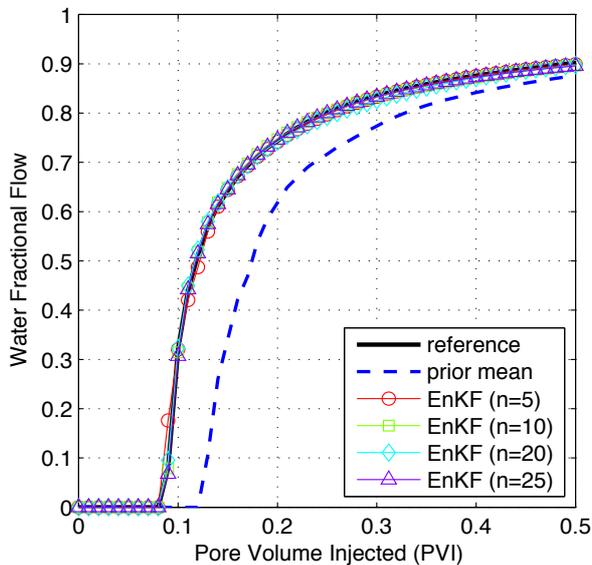


Fig. 12 Reference, initial and optimized water fractional flow curves using different ensemble sizes for test case 3.

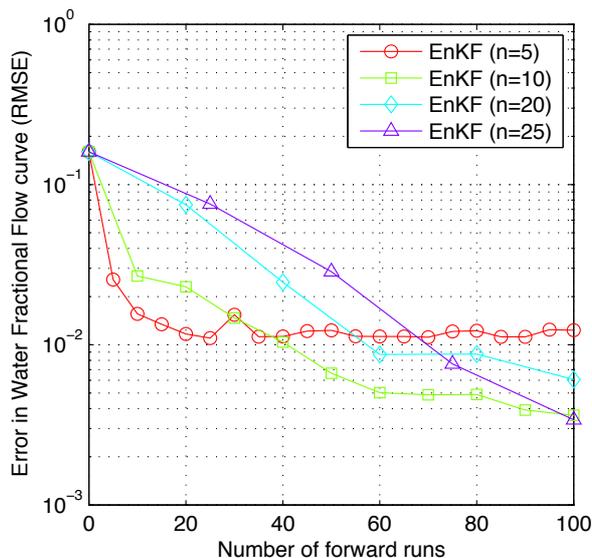


Fig. 13 RMSE in water fractional flow curve versus the number of forward runs using different ensemble sizes for test case 3.

Bishop CH, Etherton BJ, Majumdar SJ (2001) Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly Weather Review* 129(3):420–436

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

Chen Y, Zhang D (2006) Data assimilation for transient flow in geologic formations via ensemble kalman filter. *Advances in Water Resources* 29(8):1107–1122

Chen Z (2007) *Reservoir Simulation: Mathematical Techniques in Oil Recovery* (CBMS-NSF Regional Conference Series in Applied Mathematics). Society for Industrial and Applied Mathematics, Philadelphia, PA, USA

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 Reser Eval Eng* 4:308–317

Dostert P, Efendiev Y, Mohanty B (2009) Efficient uncertainty quantification techniques in inverse problems for richards' equation using coarse-scale simulation models. *Advances in Water Resources* 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. *Computational Geosciences* pp 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)

Evensen G, van Leeuwen PJ (2000) An ensemble kalman smoother for nonlinear dynamics. *Monthly Weather Review* 128(6):1852–1867

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

Fu J, Gomez-Hernandez JJ (2008) Preserving spatial structure for inverse stochastic simulation using blocking markov chain monte carlo method. *Inverse Problems in Science and Engineering* 16(7):865–884

Fu J, Gomez-Hernandez J (2009) A blocking markov chain monte carlo method for inverse stochastic hydrogeological modeling. *Mathematical Geosciences* 41:105–128

Gaspari G, Cohn SE (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society* 125(554):723–757

Ghanem RG, Spanos PD (1991) *Stochastic finite elements: a spectral approach*. Springer-Verlag New York, Inc., New York, NY, USA

Gillijns S, Mendoza OB, Chandrasekar J, De Moor BLR, Bernstein DS, Ridley A (2006) What is the ensemble kalman filter and how well does it work? 2006 American Control Conference pp 4448–4453

Golub, G H, Kahan, W (1965) Calculating the singular values and the pseudo inverse of a matrix. *SIAM J Numer Anal* 2:205–224

Golub G, Van Loan C (1996) *Matrix Computations*, 3rd edn. Johns Hopkins University Press

Hansen C (1998) *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, Philadelphia

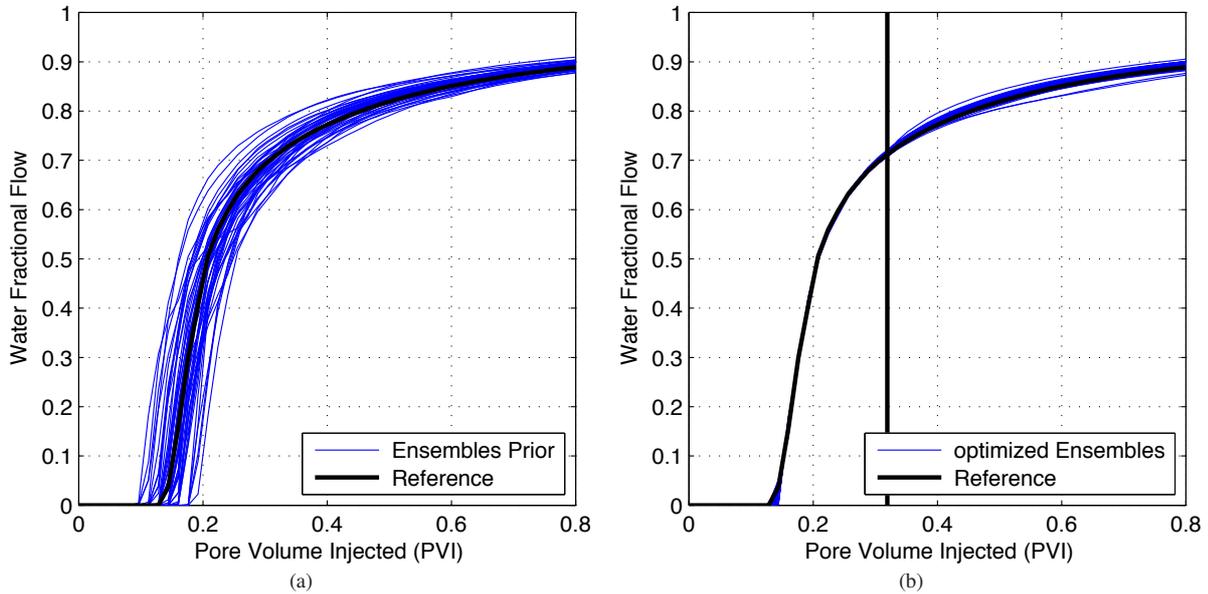


Fig. 14 Results of a UQ study for test case 2. EnKF optimized versus initial ensemble fractional flow curves obtained using ensembles of 10 members, (a) Initial ensembles means and (b) Optimized ensembles means.

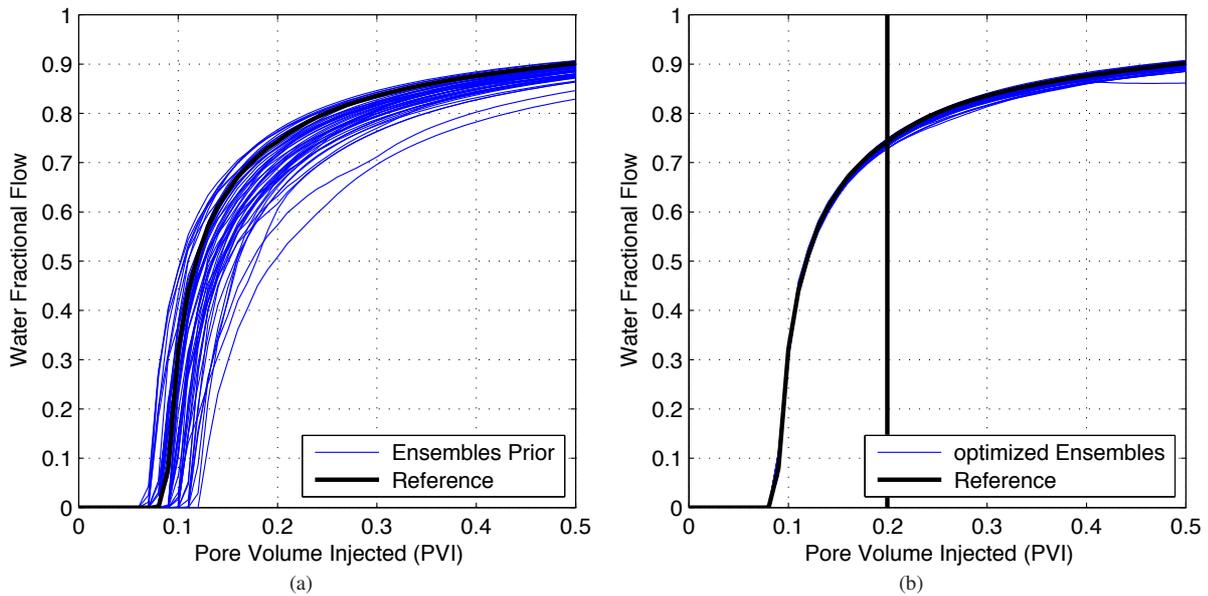


Fig. 15 Results of a UQ study for test case 3. EnKF optimized versus initial ensemble fractional flow curves obtained using ensembles of 10 members, (a) Initial ensembles means and (b) Optimized ensembles means.

Hansen PC, Nagy JG, O’Leary DP (2006) *Deblurring Images: Matrices, Spectra, and Filtering (Fundamentals of Algorithms)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA

Hanson RJ (1971) A numerical method for solving Fredholm integral equations of the first kind using singular values. *SIAM Journal on Numerical Analysis* 8(3):616–622

Houtekamer PL, Mitchell HL (2001) A sequential ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review* 129(1):123–137

Houtekamer PL, Mitchell HL (2005) Ensemble kalman filtering. *Quarterly Journal of the Royal Meteorological Society* 131(613):3269–3289, DOI 10.1256/qj.05.135

Kac M, Siebert AJF (1947) An explicit representation of a stationary Gaussian process. *Ann Math Statistics* 18:438–

- 442
- Karhunen K (1947) Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann Acad Sci Fennicae Ser A I Math-Phys* 1947(37):79
- Loève M (1948) Fonctions aléatoires de second order. In: Levy, P., Ed., *Processus Stochastiques et Movement Brownien*, Hermann, Paris
- MacKay DJC (1999) Comparison of approximate methods for handling hyperparameters. *Neural Comput* 11:1035–1068, DOI 10.1162/089976699300016331
- 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 Journal* 13(1):77–87
- McLaughlin D, Townley LR (1996) A reassessment of the groundwater inverse problem. *Water Resources Research* 32(5):1131–1161
- Moradkhani H, Sorooshian S, Gupta HV, Houser PR (2005) Dual state–parameter estimation of hydrological models using ensemble kalman filter. *Advances in Water Resources* 28(2):135–147
- Naevdal G, Johnsen L, Aanonsen S, Vefring E (2005) Reservoir monitoring and continuous model updating using ensemble kalman filter. *SPE Journal* 10:66–74
- Navon I (1998) Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dynamics of Atmospheres and Oceans* 27(14):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. *Mathematical Geology* 29:61–91
- 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. *Journal of Marine Systems* 16(34):323–340
- Rasmussen CE, Williams CKI (2005) *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press
- Sakov P, Bertino L (2011) Relation between two common localisation methods for the enkf. *Computational Geosciences* 15:225–237
- Sivia D, Skilling J (2006) *Data analysis: a Bayesian tutorial*. Oxford science publications, Oxford University Press
- Smith KW (2007) Cluster ensemble kalman filter. *Tellus A* 59(5):749–757
- Sun AY, Morris AP, Mohanty S (2009a) Sequential updating of multimodal hydrogeologic parameter fields using localization and clustering techniques. *Water Resources Research* 45(W07424):1131–1161
- Sun AY, Morris A, Mohanty S (2009b) Comparison of deterministic ensemble kalman filters for assimilating hydrogeological data. *Advances in Water Resources* 32(2):280–292
- Tikhonov A, Arsenin V (1977) *Solutions of ill-posed problems*. Scripta series in mathematics, Winston
- Tippett MK, Anderson JL, Bishop CH, Hamill TM, Whitaker JS (2003) Ensemble square root filters. *Monthly Weather Review* 131(7):1485–1490
- 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*, pp 153–158, DOI 10.1109/ASSPCC.2000.882463
- Xiu D (2009) Fast numerical methods for stochastic computations: a review. *Communications in Computational Physics* 5:242–272
- Zeng L, Zhang D (2010) A stochastic collocation based kalman filter for data assimilation. *Computational Geosciences* 14:721–744
- Zhang Y, Oliver D (2010) Improving the ensemble estimate of the kalman gain by bootstrap sampling. *Mathematical Geosciences* 42:327–345
- 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. *Advances in Water Resources* 34(7):844–864
- Zupanski M, Navon IM, Zupanski D (2008) The maximum likelihood ensemble filter as a non-differentiable minimization algorithm. *Quarterly Journal of the Royal Meteorological Society* 134(633):1039–1050