Environmental Science   College of Arts and Sciences

2015

The Compressed State Kalman Filter for Nonlinear State Estimation: Application to Large-Scale Reservoir Monitoring

J Y. Li

Amalia Kokkinaki
University of San Francisco, akokkinaki@usfca.edu

H Ghorbanidehno
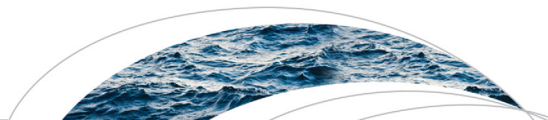
E F. Darve

P K. Kitanidis

Follow this and additional works at: http://repository.usfca.edu/envs

Part of the Environmental Monitoring Commons

Recommended Citation

Li, J. Y., A. Kokkinaki, H. Ghorbanidehno, E. F. Darve, and P. K. Kitanidis (2015), The compressed state Kalman filter for nonlinear state estimation: Application to large-scale reservoir monitoring, Water Resour. Res., 51, 9942–9963, http://dx.doi.org/10.1002/2015WR017203.

The compressed state Kalman filter for nonlinear state estimation: Application to large-scale reservoir monitoring

Judith Yue Li[1], Amalia Kokkinaki[1], Hojat Ghorbanidehno[2], Eric F. Darve[2,3], and Peter K. Kitanidis[1,3]

[1]Department of Civil and Environmental Engineering, Stanford University, Stanford, California, USA, [2]Department of Mechanical Engineering, Stanford University, Stanford, California, USA, [3]Institute for Computational and Mathematical Engineering, Jen-Hsun Huang Engineering Center, Stanford University, Stanford, California, USA

Abstract Reservoir monitoring aims to provide snapshots of reservoir conditions and their uncertainties to assist operation management and risk analysis. These snapshots may contain millions of state variables, e.g., pressures and saturations, which can be estimated by assimilating data in real time using the Kalman filter (KF). However, the KF has a computational cost that scales quadratically with the number of unknowns, m, due to the cost of computing and storing the covariance and Jacobian matrices, along with their products. The compressed state Kalman filter (CSKF) adapts the KF for solving large-scale monitoring problems. The CSKF uses N preselected orthogonal bases to compute an accurate rank-N approximation of the covariance that is close to the optimal spectral approximation given by SVD. The CSKF has a computational cost that scales linearly in m and uses an efficient matrix-free approach that propagates uncertainties using $N + 1$ forward model evaluations, where $N \ll m$. Here we present a generalized CSKF algorithm for nonlinear state estimation problems such as CO_2 monitoring. For simultaneous estimation of multiple types of state variables, the algorithm allows selecting bases that represent the variability of each state type. Through synthetic numerical experiments of CO_2 monitoring, we show that the CSKF can reproduce the Kalman gain accurately even for large compression ratios (m/N). For a given computational cost, the CSKF uses a robust and flexible compression scheme that gives more reliable uncertainty estimates than the ensemble Kalman filter, which may display loss of ensemble variability leading to suboptimal uncertainty estimates.

1. Introduction

Reservoir monitoring is an essential part of many field operations such as aquifer recharge [*Nenna et al.*, 2011], groundwater contamination identification [*McLaughlin et al.*, 1993], and CO_2 storage [*Arogunmati and Harris*, 2012]. Mathematically, the monitoring problem can be formulated as a state estimation problem, in which state variables including pressure, phase saturation, and other time-varying quantities are sequentially updated using data collected at discrete time steps. This sequential approach is particularly useful for real-time control applications, where management decisions need to be made based on newly acquired data and changes in reservoir conditions [*Yeh*, 1986]. Each time new data are assimilated, an inverse problem is solved. The unknowns at a given time and their respective uncertainties are inferred by combining a physical model and the actual observations at that time [*Tarantola*, 2005].

One of the challenges of such inverse problems for applications in reservoir monitoring is the high-dimensional state space that results from the discretization of a heterogeneous parameter field, combined with the nonlinear multiphase flow equations governing the evolution of the state variables. This high dimensionality translates into significant computational costs that often surpass the power of even modern computers. Methods that are able to handle such high-dimensional problems are necessary in cases where fine resolution of the estimated field is important, such as for CO_2 leakage detection.

The Kalman filter [*Kalman*, 1960] and its nonlinear version, the extended Kalman filter (EKF) [*Gelb*, 1974], provide a Bayesian framework to continuously assimilate observed data in the predictions of an uncertain forward simulator. In the case of modeling multiphase flow using reservoir simulators, data assimilation can be used to correct errors due to assumed initial conditions, boundary conditions, or input parameters [*Liu and Gupta*, 2007]. For a reservoir model of typical size, i.e., 10^4 to 10^6 unknowns after discretization, the original implementation of KF is computationally intractable primarily because of the requirement to store and

update a large covariance matrix of size equal to m^2 (m is the number of the unknowns). In addition, because of the nonlinearity of multiphase flow physics, at every assimilation step, the model has to be linearized around the current estimates using either expensive numerical differentiation techniques or adjoint solvers, which are often difficult to implement and require code modifications of the forward model. Therefore, another important contributor to the computational burden is the cost of computing and storing the Jacobian matrices and their products with other matrices.

One approach to address the computational challenge of such large systems is to use a low-rank approximation of the covariance matrix. Examples of low-rank KF algorithms include the ensemble Kalman filter (EnKF) [*Evensen*, 1994], the singular evolutive extended Kalman (SEEK) filter [*Tuan Pham et al.*, 1998], and the compressed state Kalman filter (CSKF) [*Kitanidis*, 2015]. Such methods approximate the estimation error covariance matrix by a reduced rank matrix, which is often justified in applications [*Oliver and Chen*, 2011]. Low-rank methods can provide significant computational speedup compared to the original KF algorithm. Fast linear algebra techniques for fast matrix-vector multiplication can further improve the computational efficiency of low-rank methods [*Kitanidis*, 2015]. The computational cost of such methods is determined by the size of the reduced rank covariance matrix, which also determines the number of forward runs of the simulator required. For all low-rank methods, the rank must be chosen such that the error subspace captures most of the variability in the unknown state, otherwise rank deficiency can lead to underestimation of uncertainty and result in inaccurate estimates of the unknown states and their uncertainties. This trade-off between computational efficiency and estimation accuracy is the critical point of low-rank methods. Methodologies that provide robust control of this trade-off are necessary for efficient and reliable estimation.

A critical requirement for achieving reliable estimation is the ability of a low-rank filter to reproduce the optimal Kalman gain. The Kalman gain is a quantity that determines how state variables should be adjusted in order to match the observations. Each element of the Kalman gain matrix can be thought of as a "linear regression coefficient" between the state variable being estimated and a set of observations [*Anderson*, 2003]. Inaccurate low-rank approximations of the covariance can result in inaccurate Kalman gain estimation, which may lead to unphysical state estimates and degradation of the filtering accuracy as more data are being assimilated. One well-known case of this problem is when the EnKF is used with insufficient ensemble size, resulting in deteriorated Kalman gain estimates [*Zhou et al.*, 2006; *Zhang and Oliver*, 2010]. This may lead to a fast reduction in the ensemble variability and often a severe underestimation in the uncertainty, a phenomenon that is referred to as "ensemble collapse" [*Lorenc*, 2003; *Bengtsson et al.*, 2008]. Increasing the ensemble size can improve the quality of Kalman gain and provide more robust error variance estimates [*Reichle et al.*, 2002], yet at a rather slow rate, i.e., $\mathcal{O}(1/\sqrt{N})$ [*Furrer and Bengtsson*, 2007], and at an increased computational cost. In such cases, additional steps that introduce additional tuning parameters, such as localization [*Houtekamer and Mitchell*, 1998; *Anderson*, 2012; *Zhang and Oliver*, 2010], shrinkage regression [*Sætrom and Omre*, 2011], covariance tapering [*Furrer and Bengtsson*, 2007], or covariance inflation [*Anderson*, 2001] have to be used to reduce the effect of sampling errors.

In this paper, we present the generalized compressed state Kalman filter (CSKF) that can be applied for both linear and nonlinear large-scale state estimation problems. The core idea behind the method's computational efficiency, as proposed in *Kitanidis* [2015], is to provide an accurate low-rank representation of the covariance matrix and thereby the optimal Kalman gain, and to avoid explicit calculation of the Jacobian by linearizing the nonlinear equations using an efficient matrix-free approach. Several modifications to the method of *Kitanidis* [2015] are presented that allow application of CSKF to more general cases such as CO_2 monitoring, where there is nonlinearity in both the model and the observation process, and also allowing the simultaneous estimation of multiple types of state variables that depend on each other. The paper begins with an overview of the extended Kalman filter and strategies to compress the covariance effectively. Then, we illustrate how the CSKF reduces the computational cost for a single type of state variable, and for multiple types of state variables by using a compression scheme that allows a different compression ratio for each type of state variable. The algorithm is validated and analyzed using synthetic CO_2 monitoring examples.

2. Methods

In this section, we formulate the state estimation problem in the context of reservoir monitoring. We first summarize the solution given by the extended Kalman filter for the nonlinear estimation problem in order to establish notation and then we present the generalized CSKF algorithm.

2.1. State-Space Model and Extended Kalman Filter

Consider a state-space model that describes a nonlinear dynamic system,

$$x_{k+1} = \phi(x_k) + w_k, \quad x \in \mathbb{R}^{m \times 1} \tag{1}$$

and the measurement relation to the state variables:

$$y_{k+1} = h(x_{k+1}) + v_k, \quad y \in \mathbb{R}^{n \times 1} \tag{2}$$

where m denotes the number of unknowns and n denotes the number of measurements. In the context of reservoir monitoring, the unknown state variables x_k (e.g., saturation and pressure) refer to the current reservoir condition at time instance k, evolving according to the discretized multiphase flow equation $\phi(\cdot)$. The term w_k represents the model errors due to incorrect initial, boundary conditions or parameters in equation (1). The observations at time k, y_k, (e.g., flux and pressure) are related to the reservoir state x_k through equation (2), where the term v_k represents measurement errors.

The operational assumption is that the unknown initial state x_0 is Gaussian with mean the best linearized estimate $\hat{x}_0$ and covariance Σ_0, and x_0, w_k, and v_k are mutually independent white (uncorrelated in time) Gaussian processes, i.e.,

$$x_0 \sim N(\hat{x}_0, \Sigma_0) \tag{3}$$

$$w_k \sim N(0, Q) \tag{4}$$

$$v_k \sim N(0, R) \tag{5}$$

with Q and R representing the model error and measurement error covariance matrices, respectively.

For the above state-space models, the extended Kalman filter (EKF) (see textbooks such as *Anderson and Moore* [1979]), i.e., the nonlinear version of the Kalman filter, gives the posterior mean $\hat{x}_{k+1|k+1}$ and covariance $\Sigma_{k+1|k+1}$ conditioned on the observations y_{k+1}. The subscript notation of $x_{i|j}$ denotes the estimate at time i conditioned on the measurements up to time step j. The EKF consists of a forecast step, which uses the forward model to predict the state at time $k + 1$, and an analysis step that uses the observations obtained at time $k + 1$ to improve the corresponding states predicted at the forecast step (Algorithm 1). The posterior mean $\hat{x}_{k+1|k+1}$ in the analysis step of Algorithm 1 is obtained by minimizing the linearized approximation of the objective function:

$$J(x) = ||y_{k+1} - h(x)||_{R^{-1}} + ||x - \phi(\hat{x}_{k|k})||_{\hat{\Sigma}_{k+1|k}^{-1}} \tag{6}$$

where the notation $||x||_P$ denotes the product of $x^T P x$. The computational challenge in the above minimization is that it involves expensive matrix operations with the covariance matrix Σ, of size $m \times m$, and the computation of Jacobian matrices $\Phi = \frac{\partial \phi}{\partial x}|_{x=\hat{x}_{k|k}}$ and $H = \frac{\partial h}{\partial x}|_{x=\hat{x}_{k+1|k}}$, which are computationally prohibitive for large systems. The next section outlines how these operations can be made more efficiently.

2.2. Covariance Compression

Covariance compression can be used to reduce the cost of storing and updating large covariance matrices. For many data assimilation problems, the covariance matrix, $\Sigma_{k+1|k+1}$, can be approximated adequately with a low-rank matrix [*Pham et al.*, 1998; *Kitanidis*, 2015; *Tippett et al.*, 2003]. Consider that the $m \times m$ state error covariance matrix Σ can be approximated using a matrix $\hat{\Sigma}$ of rank N ($N \ll m$) through the following factorization:

Algorithm 1. Extended Kalman filter algorithm (EKF)

Forecast:

Forecast state	$\hat{x}_{k+1|k} = \phi(\hat{x}_{k|k})$
Compute Jacobian	$\Phi = \frac{\partial \phi}{\partial x}|_{\hat{x}_{k|k}}$
Forecast covariance matrix	$\hat{\Sigma}_{k+1|k} = \Phi \hat{\Sigma}_{k|k} \Phi^T + Q$

Analysis:

Compute Jacobian	$H = \frac{\partial h}{\partial x}|_{\hat{x}_{k+1|k}}$
Kalman gain	$K = \hat{\Sigma}_{k+1|k} H^T \left(H \hat{\Sigma}_{k+1|k} H^T + R \right)^{-1}$
Update state	$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K(y_{k+1} - h(\hat{x}_{k+1|k}))$
Posteriori covariance	$\hat{\Sigma}_{k+1|k+1} = \hat{\Sigma}_{k+1|k} - KH\hat{\Sigma}_{k+1|k}$

$$\Sigma \simeq \hat{\Sigma} = ACA^T \tag{7}$$

where $A \in \mathcal{R}^{m \times N}$ satisfies $A^T A = I$, and the inner matrix $C \in \mathbb{R}^{N \times N}$ is a symmetric and positive definite matrix. The columns of A are N orthonormal vectors that form the basis spanning a subspace of the full m-dimensional space. As suggested by *Kitanidis* [2015], in our method,

the basis A is kept constant through the filtering process so that only the small matrix C is updated. Also, both Σ and Q are compressed using the same orthogonal basis, i.e., $\hat{Q} = AVA^T$.

To obtain the low-rank approximation of the covariance efficiently, A can be preselected following one of several possible approaches. For example, A can be the first N eigenvectors of a representative covariance matrix Z. The standard approach to compute the full factorization $Z = ACA^T$ is through direct eigen-decomposition, which has computational complexity of $\mathcal{O}(m^3)$. However, since only the first N dominant eigenvectors are desired, they can be obtained using the randomized low-rank approximation at a much lower cost of $\mathcal{O}(m^2N)$ [*Halko et al.*, 2011]. The computational cost is further reduced to $\mathcal{O}(mN)$ when the latter method is combined with fast linear algebraic techniques [*Lee and Kitanidis*, 2014]. An $\mathcal{O}(m)$ randomized low-rank algorithm used in this article is described in the supporting information S2.

Provided that the problem is discretized on a regular grid, a reasonable orthonormal basis A can be computed more efficiently using the discrete cosine transform (DCT) [*Strang*, 1999]. The DCT is well known as an effective image compression technique [*Ahmed et al.*, 1974]. Recently, the DCT has also found applications in reservoir modeling for representing complex geologic structures using a handful of coefficients [*Jafarpour et al.*, 2009a, 2009b]. In one dimension, the elements of the DCT matrix $A \in \mathcal{R}^{m \times N}$ are defined by

$$A(k, 1) = \frac{1}{\sqrt{m}}, \quad k = 1, \ldots, m \tag{8}$$

$$A(k, j) = \sqrt{\frac{2}{m}} \cos \left[\frac{\pi(2k-1)(j-1)}{2m} \right] \tag{9}$$

$$k = 1, \ldots, m, j = 2, \ldots, N \tag{10}$$

where m is the number of grid blocks for a given discretization, and N is the number of orthonormal vectors. Any spatial patterns can be represented by combining low-frequency and high-frequency DCT bases. For spatial fields characterized by smooth covariance kernels whose eigenspectrum drops rapidly, their variability can be captured using only the low-frequency DCT basis, i.e., N may be small and the compression therefore can be considerable.

Generally, the optimal choice of rank N for the estimation covariance compression depends on the eigenspectrum of the posterior covariance and the level of accuracy we are expecting from filtering. Even though the posterior covariance is not known a priori, we can choose N based on the smoothness of the problem and an assumed covariance that represents this smoothness reasonably well. Alternatively, the basis can be obtained from the training images [*Jafarpour and McLaughlin*, 2008].

Once A and N are known, the compressed state covariance matrix $C = A^T(\Sigma A)$, and the compressed model error covariance matrix $V = A^T(QA)$ can be computed efficiently as follows: assume that the initial state covariance matrix Σ_0 and the model error covariance matrix Q take the form of covariance kernels, examples of which are exponential, or Gaussian covariance functions that are frequently encountered in kriging and state estimation problems. Then, for the selected orthonormal basis $A \in \mathcal{R}^{m \times N}$, each of the matrix-vector products $\Sigma_0 A$ and QA can be computed with $\mathcal{O}(N^2m)$ computational cost with the aid of the hierarchical matrix approach [*Saibaba et al.*, 2012; *Saibaba and Kitanidis*, 2012; *Ambikasaran et al.*, 2013a,b; *Li et al.*, 2014; *Aminfar et al.*, 2016]. Each remaining product has an $\mathcal{O}(Nm)$ computational cost, resulting in a total of $(N + N^2)m$ operations.

The CSKF algorithm proposed in *Kitanidis* [2015] is based on the idea of using covariance compression to reduce the computational cost of EKF. In this paper, we present the generalized CSKF for solving nonlinear state estimation problems with a reduced computational cost. Compared to the algorithm of *Kitanidis* [2015] which is applicable for a linear measurement operator H and a single-state variable, the implementation presented here is for applications where both the model and measurement operators are nonlinear (see section 2.3) and for applications where we want to estimate multiple unknown states that depend on each other (see section 2.4), e.g., estimation of the reservoir saturation and pressure given real-time production data.

2.3. CSKF for Single-Variable Nonlinear State Estimation

Let us denote the products of the Jacobians and the eigenvectors A as $A^\Phi = \Phi A$ and $A^H = HA$. Assume both A^Φ and A^H are computed and stored. In the forecast step of each data assimilation cycle, we obtain the best state estimate given by the forward model and the previous state

$$\hat{x}_{k+1|k}=\phi(\hat{x}_{k|k}) \tag{11}$$

and its compressed covariance

$$C_{k+1|k}=(A^T A^\Phi)C_{k|k}(A^T A^\Phi)^T+V \tag{12}$$

with $\mathcal{O}(m)$ operations. Note that by compressing the model error covariance matrix Q using the selected basis A, i.e., $Q = AV\,A^T$, a matrix-free approach can be used to compute the otherwise expensive matrix-matrix products QH^T. This is necessary for cases with a nonlinear measurement operator $h(\cdot)$, as opposed to a linear $h(\cdot)$ case like the one discussed in *Kitanidis* [2015].

In the analysis step, the state at time step $k + 1$ is updated using measurements at time step $k + 1$:

$$x_{k+1|k+1}=x_{k+1|k}+K(y_{k+1}-h(x_{k+1|k})) \tag{13}$$

The Kalman gain, K, which is used to compute the correction made to the prior state $x_{k+1|k}$, is given by the following procedure:

First, compute the $\mathcal{X}$ by solving the system of linear equations

$$(A^H C_{k+1|k}A^{HT}+R)\mathcal{X}=A^H C_{k+1|k} \tag{14}$$

where $\Sigma_{HH}=A^H C_{k+1|k}A^{HT}+R$ is the innovation covariance and can be computed with $\mathcal{O}(m)$ instead of $\mathcal{O}(m^2)$ operations. Then, the Kalman gain is given by

$$K=A\mathcal{X}^T \tag{15}$$

By substituting equation (15) into equation (13), it can be shown that the correction being made to $x_{k+1|k}$ is restricted in the subspace spanned by the orthogonal bases included in A. A more detailed derivation of equations (12)–(15) can be found in Appendix A. If the innovation covariance matrix Σ_{HH} of size $n \times n$ is ill conditioned or singular, i.e., Σ_{HH}^{-1} does not exist [*Evensen*, 2007], the truncated singular value decomposition (TSVD) may be used to compute the pseudoinverse of Σ_{HH} in order to solve equation (14). For data of different magnitudes, scaling of Σ_{HH} is required before the TSVD to avoid loss of useful information [*Wang et al.*, 2010] (Appendix B).

The updated compressed covariance at time $k + 1$ is given by

$$C_{k+1|k+1}=(I_m-\mathcal{X}^T A^H)C_{k+1|k} \tag{16}$$

The full posterior covariance matrix is given by $\Sigma_{k+1|k+1}=AC_{k+1|k+1}A^T$. However, using the above sequence of operations, $\Sigma_{k+1|k+1}$ is never computed explicitly, thereby reducing the storage and computation costs dramatically.

Finally, note that in equations (12)–(16), instead of computing the Jacobian matrices $\Phi=\frac{\partial\phi}{\partial x}\big|_{x_{k|k}}$ and $H=\frac{\partial h}{\partial x}\big|_{x_{k+1|k}}$ directly, only the matrix-vector products of the Jacobian matrices A^Φ ($=\Phi A$) and A^H ($=HA$) are required. This matrix-vector product can be efficiently obtained using a matrix-free approach [*Lee and Kitanidis*, 2014] like the finite difference method for calculating matrix-vector products,

$$A_{:,i}^\Phi=\frac{\phi(x+\delta||x||A_{:,i})-\phi(x)}{\delta||x||}\bigg|_{x_{k|k}} \tag{17}$$

$$A_{:,i}^H=\frac{h(x+\delta||x||A_{:,i})-h(x)}{\delta||x||}\bigg|_{x_{k+1|k}} \tag{18}$$

where $A_{:,i}$ is the ith column of A, a dimensionless orthonormal vector, i.e., $||A_{:,i}||=1$. Using this approach, the computation of A^Φ and A^H requires $N + 1$ evaluations of the forward model $f(\cdot)$ and $h(\cdot)$. The approximation error can be reduced if a higher-order finite difference scheme (e.g., central difference) is used, at the cost of increased computational burden. The selection of δ depends on the machine precision and the sensitivity. A reasonable choice is $\delta=\sqrt{\epsilon}$, where ϵ is the reading precision of the output $f(x)$ and $h(x)$.

2.4. CSKF for Multiple-State Variables

In certain applications, we often want to update more than one type of uncertain state variable or model parameter simultaneously using information from the observations. In this section, we show how the CSKF can be extended to estimate two or more state variables that depend on each other (e.g., pressure and saturation for multiphase flow), with a different orthogonal basis for each type of variable. Including multiple states increases the computational cost of filtering, as there are more unknowns and respective covariances to be estimated.

The multiple-state estimation problem can be solved using a block compression scheme of the covariance. Consider an augmented state vector with two unknown states (e.g., pressure and saturation): $x_k = [p_k, s_k]^T$, whose covariance matrix has a block structure, with blocks on the diagonal representing the covariance of each variable and the blocks off the diagonal representing the cross-covariances (equation (19)):

$$\Sigma = \begin{bmatrix} \Sigma_{pp} & \Sigma_{ps} \\ \Sigma_{ps}^T & \Sigma_{ss} \end{bmatrix} \tag{19}$$

where Σ_{ij} is the covariance matrix for state variable i ($i = j$) or the cross-covariance for state variable i and j ($i \neq j$). A reasonable initialization choice for the state covariance is $\Sigma_{ii} = K(h|l, \sigma)$, where $K(h)$ is a covariance function of distance h (e.g., Gaussian and exponential), which is parameterized with standard deviation σ and correlation length l.

Their low-rank approximation of the form $\Sigma_{ij} = A_i C_{ij} A_j^T$ can be obtained following the description in section 2.2, where A_i is a preselected dimensionless orthogonal basis for each type of unknown i, and the inner matrix C_{ij} is the compressed covariance or cross-covariance that accounts for the units. The factorized state error covariance matrix has a block low-rank structure

$$\Sigma = \begin{bmatrix} A_p C_{pp} A_p^T & A_p C_{ps} A_s^T \\ A_s C_{ps}^T A_p^T & A_s C_{ss} A_s^T \end{bmatrix} \tag{20}$$

Now we write Σ in the factorized form, i.e., $\Sigma = ACA^T$, where

$$A = \begin{bmatrix} A_p & 0 \\ 0 & A_s \end{bmatrix} \tag{21}$$

$$C = \begin{bmatrix} C_{pp} & C_{ps} \\ C_{ps}^T & C_{ss} \end{bmatrix} \tag{22}$$

The factorization of model error covariance $Q = AVA^T$ can be obtained following the same procedure.

Similarly, the terms A^Φ and A^H for the augmented state can be obtained by computing individual parts for each variable first and then assembling them together as follows,

$$A^\Phi = [A_p^\Phi, A_s^\Phi] = \begin{bmatrix} \dfrac{\partial \phi_{k+1}}{\partial p_{k|k}} A_p & \dfrac{\partial \phi_{k+1}}{\partial s_{k|k}} A_s \end{bmatrix} \tag{23}$$

$$A^H = [A_p^H, A_s^H] = \begin{bmatrix} \dfrac{\partial h_{k+1}}{\partial p_{k+1|k}} A_p & \dfrac{\partial h_{k+1}}{\partial s_{k+1|k}} A_s \end{bmatrix} \tag{24}$$

where ϕ_{k+1} denotes $x_{k+1|k} = \phi(x_{k|k})$, and h_{k+1} denotes $y_{k+1} = h(x_{k+1|k})$. Equations (23) and (24) consist of a total of $N = \text{rank}(A_p) + \text{rank}(A_s)$ calls of forward model f_{k+1} and h_{k+1}. With the newly defined C, A^H, and A^Φ, the CSKF algorithm described in equations (1–18) can be applied to the augmented state for an arbitrary number of state variables. A summary of the generalized CSKF algorithm for multiple-state estimation is given by Algorithm 2.

2.5. Computational Cost

We compare the CSKF with two EnKF variants: a standard EnKF and a more typical EnKF variant that includes localization (termed EnKF+LOC hereafter). Localization is performed directly on the Kalman gain using the bootstrap-based method proposed by *Zhang and Oliver* [2010]. For both EnKF variants, an

Algorithm 2. The nonlinear compressed state Kalman filter algorithm (CSKF)

Initialization	Initialize A, C, and V so that $\Sigma = ACA^T$ and $Q = AVA^T$ for augmented state estimation:
	$A = \begin{bmatrix} A_p & 0 \\ 0 & A_s \end{bmatrix}$, $\quad C = \begin{bmatrix} C_{pp} & C_{ps} \\ C_{ps}^T & C_{ss} \end{bmatrix}$ and $V = \begin{bmatrix} V_{pp} & V_{ps} \\ V_{ps}^T & V_{ss} \end{bmatrix}$
Forecast state	$x_{k+1|k} = f_{k+1}(x_{k|k})$
Compute A^Φ	$A^\Phi_{:,j} = \frac{\phi(x+\delta\|x\|A_{:,j}) - \phi(x)}{\delta\|x\|}\big|_{x_{k|k}}$
	$A^\Phi = [A^\Phi_p, A^\Phi_s]$ for augmented state estimation
Forecast compressed covariance	$C_{k+1|k} = (A^T A^\Phi)C_{k|k}(A^T A^\Phi)^T + V$
Compute A^H	$A^H_{:,j} = \frac{h(x+\delta\|x\|A_{:,j}) - h(x)}{\delta\|x\|}\big|_{x_{k+1|k}}$
	$A^H = [A^H_p, A^H_s]$ for augmented state estimation
Kalman gain	1. Form $\Sigma_{HH} = A^H C_{k+1|k} A^{HT} + R$
	2. Solve system $\Sigma_{HH}\mathcal{X} = A^H C_{k+1|k}$ for $\mathcal{X}$
	3. Compute Kalman gain $K = A\mathcal{X}^T$
Update state	$x_{k+1|k+1} = x_{k+1|k} + K(d_{k+1} - h_{k+1}(x_{k+1|k}))$
Update compressed covariance	$C_{k+1|k+1} = (I_m - \mathcal{X}^T A^H)C_{k+1|k}$

improved sampling scheme was used, to ensure consistency between the initial covariances of the EnKF and the CSKF (see supporting information S1.2 for details).

Table 1 summarizes the computational cost of EKF, CSKF, and EnKF. It is shown that the CSKF and EnKF with the same rank-N approximation of the covariance matrix have roughly the same computational cost, which consists of $N+1$ evaluations of the forward model and $\mathcal{O}(m)$ operations to update mean and covariance, while EKF requires at least $\mathcal{O}(m^2)$ operations for Jacobian computation and covariance propagation. Here the sampling required by EnKF is done efficiently with the aid of fast linear algebra techniques (supporting information S1.2). Note that the computational cost of EnKF+LOC is roughly the same as EnKF, as the bootstrap sampling does not require additional forward simulations (supporting information S1.3).

Table 1. Comparison of Computational Cost of EKF, CSKF, and EnKF ($m \gg n > N$)

Equation	$\mathcal{O}$(Operations)
EKF Initialization	
Construct initial covariance $\Sigma_{0|0}$	m^2
CSKF Initialization	
1. Randomized SVD + $\mathcal{H}^2$-matrix	$N^2m + Nm$
2. DCT + $\mathcal{H}^2$-matrix	$m\log m + Nm + N^2m$
EnKF Initialization	
Randomized SVD + $\mathcal{H}^2$-matrix	$N^2m + Nm$
Generate samples (supporting information S1.2)	N^2m
EKF Forecast	
Jacobian and state forecast	$m^2 + 1$ evaluations of $f(\cdot)$
Covariance matrix forecast	$m^3 + m^2$
CSKF Forecast	
Compute A^Φ and forecast mean	$N+1$ evaluations of $f(\cdot)$
Forecast compressed covariance	$N^2m + N^3$
EnKF Forecast	
Forecast ensemble	$N+1$ evaluations of $f(\cdot)$
Sample from Q (supporting information S1.2)	N^2m
EKF Update	
Compute H	nm evaluations of $h(\cdot)$
Compute Kalman gain K	$nm^2 + n^2m + n^3$
Update mean and covariance	$nm^2 + nm$
CSKF Update	
Compute A^H	$N+1$ evaluations of $h(\cdot)$
Compute Kalman gain K	$Nnm + Nn^3 + N^2n$
Update mean and compressed covariance	$nm + N^2n + N^3$
EnKF Update	
Simulate data ensemble	$N+1$ evaluations of $h(\cdot)$
Compute Kalman gain	$Nnm + n^3 + Nn^2$
Update ensemble	Nnm

3. Numerical Example

Validation of the CSKF for a linear diffusion problem has been presented in *Kitanidis* [2015]. Here we address the additional challenges involved in applying the CSKF to a nonlinear data assimilation problem. For this purpose, we use a scenario of injecting supercritical CO_2 in a deep geological formation for storage, in which case the states, including CO_2 pressure and saturation, evolve according to a model describing multiphase flow and transport in porous media, i.e., the mass conservation equations for two fluid phases, accompanied with corresponding constitutive relationships [*Pruess and Spycher*, 2007].

Multiphase flow and transport models often require a large number of grid blocks to characterize the reservoir states (e.g., pressure and saturation). The
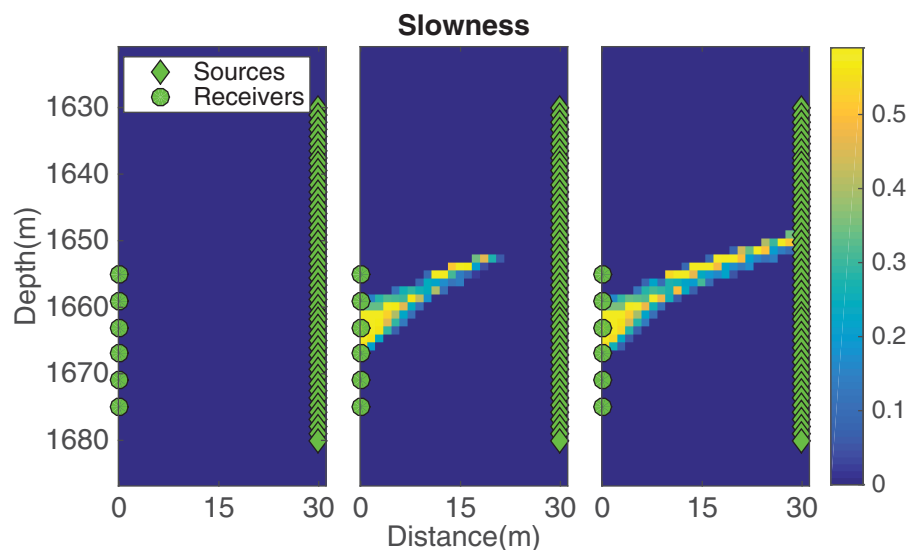
Figure 1. Case A: seismic survey geometry and the true slowness before the injection, 30 and 120 h after CO$_2$ injection.

resulting high state-space dimension prevents the use of standard KF or EKF for state estimation due to computational demands. In addition, with the number of unknowns being considerably higher than the number of observations, the inverse problem becomes severely underdetermined. To investigate the performance of CSKF for such underdetermined data assimilation problems, we use two 2-D CO$_2$ monitoring benchmark examples. In both examples, CO$_2$ injection is simulated using TOUGH2 [*Pruess*, 1991], a multiphase multicomponent reservoir simulator, combined with the module ECO2N [*Pruess and Spycher*, 2007] that models the physicochemical process of CO$_2$ displacing brine water in a deep aquifer. The TOUGH2-ECO2N code has been validated and is well documented in many carbon sequestration modeling studies [*Doughty*, 2010; *Pruess and Nordbotten*, 2011; *Zhou et al.*, 2010].

3.1. Case A: Validation

Before applying the proposed algorithm to a nonlinear case, let us first consider a case in which the true state is simulated using TOUGH2, but for the filtering a simple linear transition model, i.e., the random walk model, is used to approximate the nonlinear dynamics. For such a linear state-space model, it is possible to evaluate the full KF and use the KF estimates (i.e., the optimal estimate) to validate the generalized CSKF algorithm. Furthermore, we will show the convergence of the CSKF to the KF as the rank N increases.

This case study is based on a seismic CO$_2$ monitoring example, Frio II, following *Daley et al.* [2011] and *Li et al.* [2014]. Figure 1 shows the design of the Frio II experiment and the changes in seismic slowness (inverse of velocity) due to CO$_2$ injection. To map the velocity reduction zone due to CO$_2$ injection, 288 seismic travel time measurements are collected every 3 h using six seismic sources deployed at the injection well and 48 receivers deployed at the monitoring well [see *Li et al.*, 2014].

The objective is to continuously track a CO$_2$ plume with an unknown injection rate and unknown injection location, by estimating the changes in slowness x given the most up-to-date measurements of travel time delay y. Here we assume a linear time-invariant measurement operator $y=h(x)=Hx$ and a random walk transition equation $x_{k+1}=\phi(x_k)=x_k$, an approximation of the true transition equation that is used in cases where the true transition equation may be unavailable or too expensive to evaluate. Approximating the transition equation as a random walk model is reasonable for cases where measurements are collected frequently [*Nenna et al.*, 2011].

The data assimilation problem is solved using KF, CSKF, EnKF, and EnKF+LOC assuming the same initial conditions and regularization parameters. For the standard EnKF, in which no localization or covariance filtering is applied (supporting information S1.1 and S1.2), the difference from CSKF in the solution is attributed solely to how the covariance matrix is represented. For EnKF+LOC, there is the added difference due
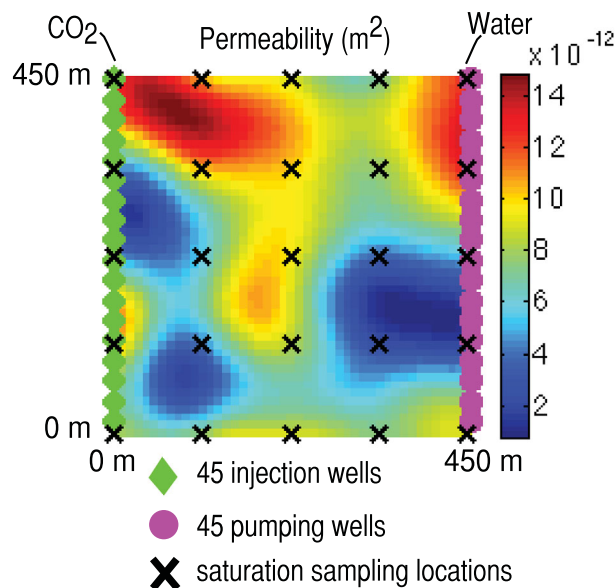
Figure 2. Case B: experiment design for the CO_2 flooding synthetic experiment. No flow boundary is assumed at both the top and the bottom.

to localization (supporting information S1.3), which aims to reduce spurious correlations due to sampling errors.

Assuming there is no CO_2 present in the field initially, the initial state is assumed to have zero mean and the error covariance Σ is assumed to be zero. To capture the sharp changes in the slowness developed during the evolution of the CO_2 plume, the model error covariance Q is represented through the kernel

$$K_1 = \sigma^2 \exp[-(h/l)^{0.5}] \qquad (25)$$

with $\sigma^2 = 1.14 \times 10^{-4}$, $l = 900$ m. h is the separation distance between two state variables, σ^2 is the variance, and l is a characteristic length parameter. The covariance matrix is compressed using the first N eigenvectors of K_1 generated by SVD. This kernel contains a large portion of high-frequency components and hence can capture the sharp CO_2 front.

3.2. Case B: Application

The second case used to evaluate the CSKF corresponds to a 450 m×450 m×10 m horizontal reservoir defined on a 45 × 45 × 1 grid shown in Figure 2. The domain has a no-flux boundary on the north and south boundary except at the wells. Forty-five vertical injection wells deployed on the left boundary inject CO_2 at a constant rate of 0.01 kg/s, and 45 vertical extraction wells deployed on the right boundary extract CO_2 with a constant bottom hole pressure of 206 bar. The true synthetic case is simulated with the permeability field shown in Figure 2 generated from a Gaussian distribution with a correlation length of 200 m.

The quantities we are interested to estimate are the CO_2 saturation and pressure every 10 days, assuming the permeabilities are known. To avoid generating saturation estimates that are outside the appropriate range, the saturation estimated here is the transformed saturation s_T in equation (26) as suggested in *Jafarpour et al.* [2009b]. The transformed saturation s_T is obtained by projecting the saturation from the [0,1] domain to the infinite domain through the forward transform:

$$s_T = \mathrm{erf}^{-1}(2s - 1), \quad s \in (0, 1) \qquad (26)$$

The back transform is defined as

$$s = 0.5(\mathrm{erf}(s_T) + 1), \quad s_T \in (-\infty, +\infty) \qquad (27)$$

which converts the transformed saturation to the actual saturation that can be accepted as an input to the reservoir simulator TOUGH2.

In terms of measurements, a total of 115 observations are collected every 10 days, consisting of 45 pressure measurements at the injection wells, 45 water flux
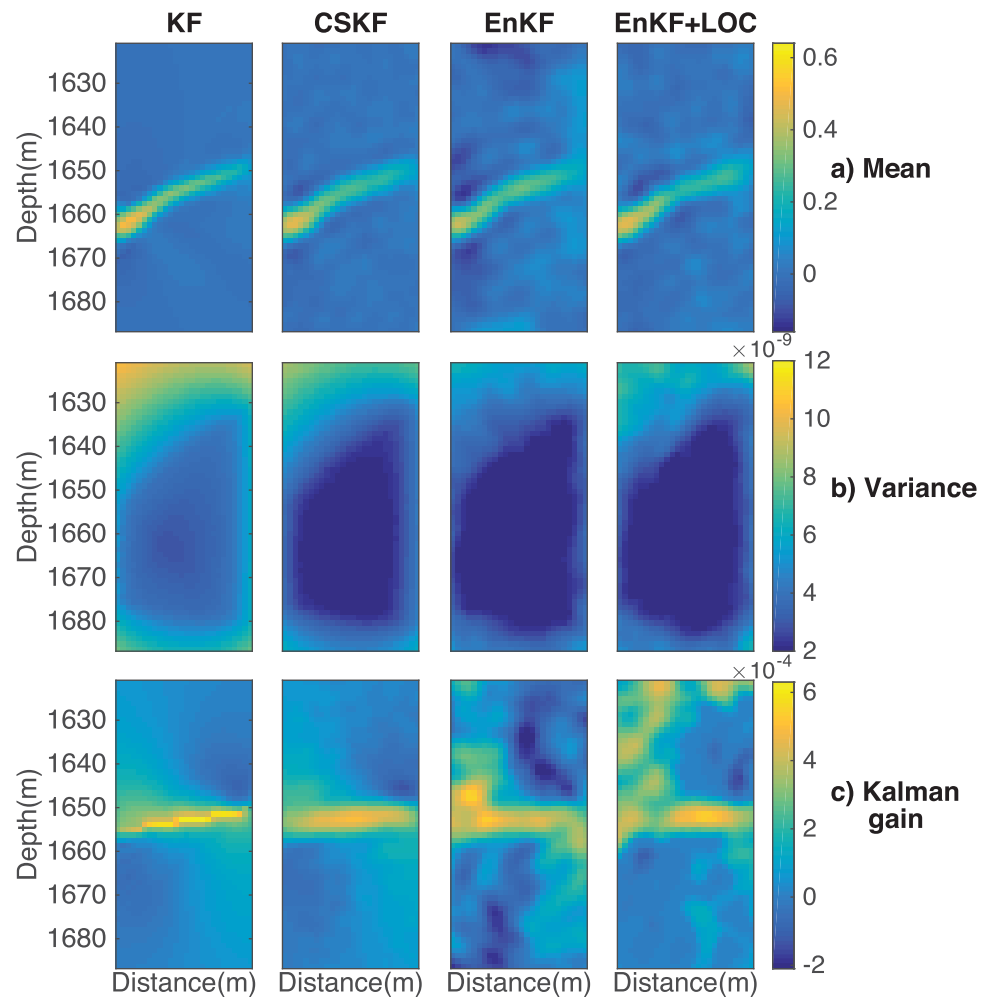
Table 2. Simulation and Data Assimilation Parameters for Case B

Forward Simulation Parameters	
Phases	CO_2/brine
Simulation time	5 × 10 days
Grid system	45 × 45 × 1
Cell dimensions	10 m × 10 m × 10 m
Rock porosity	0.2 (constant)
Permeability	Heterogeneous (Figure 2)
Number of injection well	45
Number of extraction well	45
Injection well constraints	Injection rate (0.05 kg/s)
Extraction well constraints	Pressure (200 bar)
Initial CO_2 saturation	Transformed Gaussian $N(-1^a, K2)$
Initial pressure	200 bar
Observation Error	
Water flux STD	0.01 kg/s
Pressure STD	0.05 bar
Saturation STD	0.02^a
Data Assimilation Parameters	
Initial pressure	200 bar(constant)
Initial saturation	-1^a
Initial saturation STD	1^a (case B1)/0.1^a (case B2)
Model error (saturation STD)	0.1^a
Model error (pressure STD)	0.005 bar

^{a}Value is for the transformed saturation.

Figure 3. Solution at the final assimilation step (fifth day) given by KF, CSKF with 300 bases, EnKF and EnKF+LOC with 301 realizations (equivalent to rank $N = 300$). (a) Posterior mean; (b) posterior variance; and (c) Kalman gain corresponding to the twentieth observation (i.e., the twentieth column of the Kalman gain matrix).

measurements at the extraction wells, and 25 saturation measurements sampled at 25 selected locations to track the CO_2 plume. The observation noise and other parameters used for the forward simulation are summarized in Table 2.

Unlike in case A, here we consider the case where prior to injection there was CO_2 left from a previous injection event. The true initial saturation field is generated from equation (27), where s_T is a Gaussian field with mean -1 and covariance as

$$K_2 = \sigma^2 \exp[-(h/l)^2] \tag{28}$$

where $\sigma = 1$ and $l = 100$ m. Note that the nonlinear back transformation (equation (27)) does not preserve the Gaussianity of the original random field, i.e., the initial saturation field s_0 is not a Gaussian field.

Since in a realistic application, we most likely would have no knowledge of whether CO_2 is present in the reservoir before the injection, in the filtering we assume that no CO_2 is present. The 115 observations of flux, pressure, and saturation along are used to correct 2025 gridblock reservoir states forecasted by TOUGH2 using this wrong initial saturation. The heterogeneous permeability field, the boundary conditions, the initial state error covariance Σ_0, and the observation error covariance R are assumed to be known perfectly.
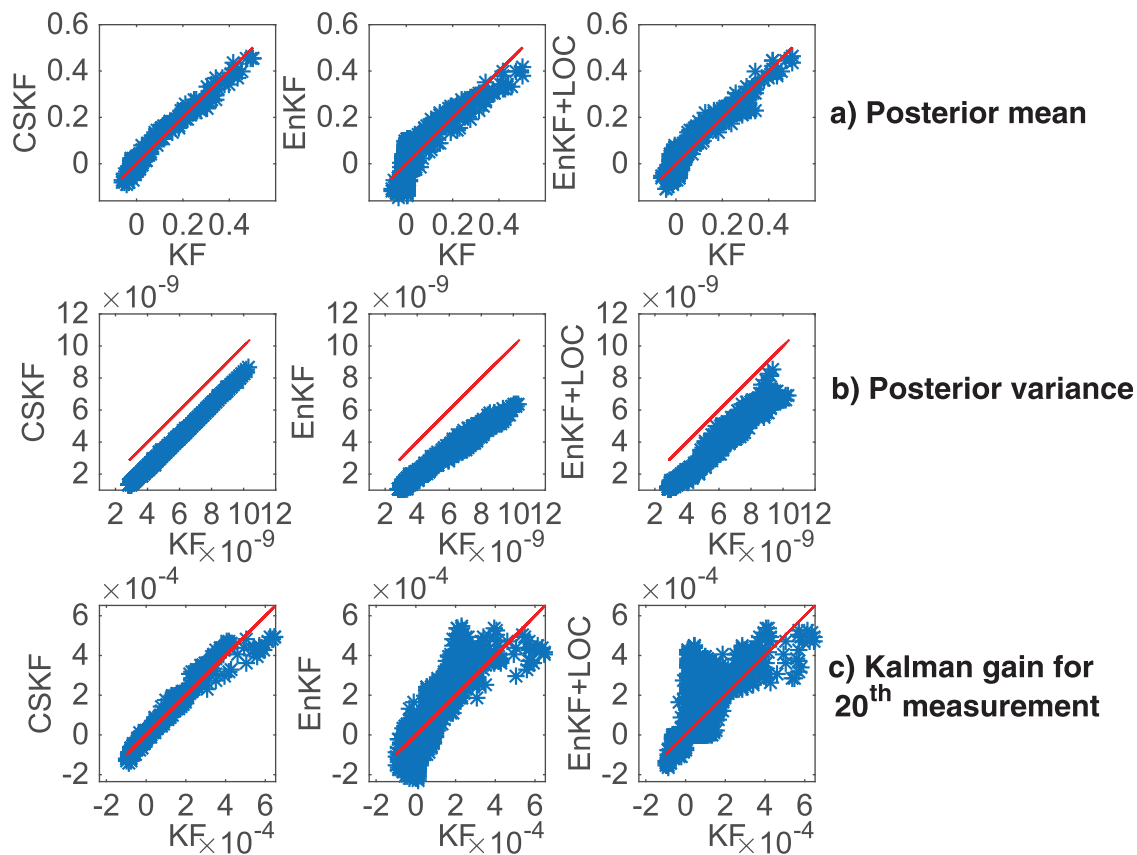
Figure 4. Case A: comparison of CSKF, EnKF, and EnKF+LOC to KF for (a) the posterior mean, (b) the posterior variance, and (c) the Kalman gain corresponding to the twentieth measurement.

We will present two variants of case B to demonstrate the abilities of the CSKF:

1. Case B1, single unknown variable: estimate CO_2 saturation distribution in space and assume the CO_2 pressure is completely known at each location over time. Flux and saturation measurements are used.
2. Case B2, multiple unknown variables: estimate CO_2 saturation and pressure simultaneously using water flux, saturation, and pressure measurements.

4. Results and Analysis

4.1. Case A: Validation

Figure 3 presents the posterior mean, variance and Kalman gain at the final step (the fifth day) given by KF, CSKF, EnKF, and EnKF+LOC. CSKF, EnKF, and EnKF+LOC adopt the same N-rank approximation of the model covariance matrix Q defined in equation (25) using randomized SVD and minimum second-order exact sampling (see supporting information S1.2). This means we are using an improved version of the EnKF. For a valid comparison of all methods, the initial mean and covariance are the same. Given the same rank $N = 300$, i.e., the same compression ratio ($m/N = 3245/300$), all methods give a good estimate of the posterior mean (Figure 3a), similar to KF and capture the spatial pattern of the posterior variance (Figure 3b). As expected, the images given by CSKF, EnKF, and EnKF+LOC are smoother compared to KF; however, the EnKF solution contains several nonzero values in the areas away from the CO_2 plume, where the values should be equal to 0. All methods slightly underestimate the magnitude of the variance. The smoothness and underestimation in variance result from using a low-rank approximation of the covariance matrix in all methods. The unphysical values in the posterior mean given by the EnKF are due to sampling errors. Localization improves the mean and variance given by EnKF by reducing spurious values, as expected.

The correction made to the state at each location is the weighted sum of the data residuals, where the weights are provided by the Kalman gain (equation (13)). Therefore, the accuracy of Kalman gain is related
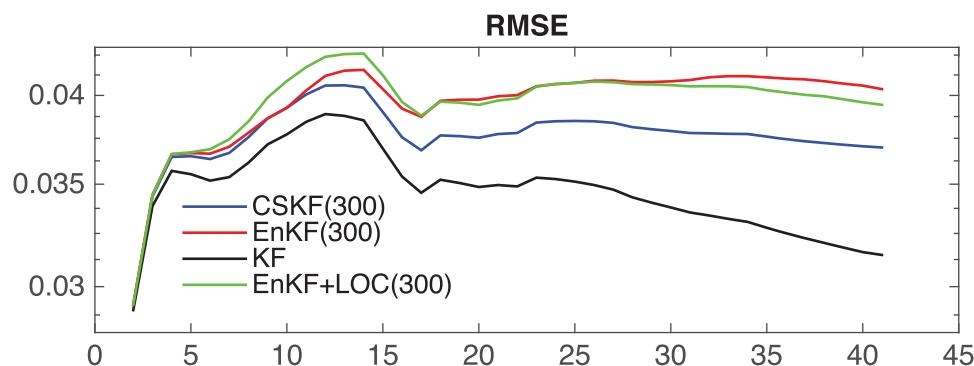
Figure 5. RMSE of state given by CSKF, EnKF, and EnKF+LOC and KF for 41 assimilation steps. Number in parentheses indicates number of bases used.

to estimation accuracy of the mean. Each column of the Kalman gain matrix contains the sensitivity of each state location to a particular measurement. Figure 3c shows the estimated Kalman gain computed by each method corresponding to the twentieth observation, by rearranging the twentieth column of Kalman gain matrix at the final assimilation step into an image. Warm colors in Figure 3c indicate regions with high sensitivity to the twentieth observation, suggesting which state variables to adjust in order to fit the particular observation. It is shown that the CSKF with 300 bases accurately estimates the optimal Kalman gain given by KF (note the smoothing due to low-rank approximation), while the EnKF produces spurious correlations. Such spurious correlations may prompt the EnKF to adjust a region that should not be corrected. Figure 3c shows that the localization used here indeed filters out some spurious Kalman gain values (values below -2×10^{-4}).

A pixel to pixel comparison of the posterior mean, variance, and the Kalman gain shown in Figure 3 is given in Figure 4, where CSKF, EnKF, and EnKF+LOC (y axis) are compared to KF (x axis). Whether the estimates fall above or below the red line indicates an overestimation or underestimation, respectively. Overall, all three methods accurately reproduce the posterior mean, although EnKF tends to underestimate the peak values and overestimate the zeros values. All methods underestimate the posterior variance because of the low-rank approximation, with the CSKF displaying a uniform deviation from KF that is smaller than the respective deviations from the EnKF and EnKF+LOC. Figure 4c shows that the CSKF gives the most reliable Kalman gain estimates among all the methods tested, while localization introduces some spurious high-value Kalman gain entries.

Next, we evaluate the KF, CSKF, EnKF, and EnKF+LOC for a fixed rank of $N = 300$. Figure 5 plots the estimation accuracy as given by the root-mean-square error (RMSE) over time for each of the four methods. The RMSE is calculated using the difference between the true and estimated state for each method. For the calculation of the RMSE, we used only the state variables in the area covered by the straight rays, in order to exclude the effect of errors in the part of the domain that is not affected by the data. The RMSE profile shows that as the CO_2 plume expands with time, the error increases, until about the fifteenth time step, when the plume and the associated RMSE stabilize for all low-rank methods. Because of the low-rank approximation, all low-rank methods start diverging from the KF early on and show an increasing error over time. However, the CSKF estimate is consistently closer to the optimal estimate given by KF compared to the EnKF methods. We found that the seed used for the EnKF methods affects the relative performance of EnKF and EnKF+LOC. Interestingly, the difference in the RMSE between EnKF and EnKF+LOC is not significant despite the clear differences in the spatial distribution (Figure 4a). The same was observed for the total variance, where all low-rank methods underestimated the total variance compared to the KF, but the difference between the EnKF and EnKF+LOC was not significant.

Next, we examine the convergence rate of each method with an increasing rank N. As the rank N and the ensemble size increase, both the CSKF and the standard EnKF should converge to the KF. The EnKF with localization, however, is not expected to display such convergence due to postprocessing made in the Kalman gain. The following metrics are used here to quantify the errors of CSKF, EnKF, and EnKF+LOC relative to KF, i.e.,

1. The total prior variance (trace of the prior covariance) at step 0,

$$SD1 = \frac{trace(\hat{\Sigma}_{1|0}) - trace(\Sigma_{KF})}{trace(\Sigma_{KF})} \tag{29}$$

2. The total posterior variance (trace of the posterior covariance) at tenth time step,

$$SD2 = \frac{trace(\hat{\Sigma}_{10|10}) - trace(\Sigma_{KF})}{trace(\Sigma_{KF})} \tag{30}$$

3. The Frobenius norm of the posterior covariance at tenth time step,

$$SD3 = \frac{||\hat{\Sigma}_{10|10} - \Sigma_{KF}||_F}{||\Sigma_{KF}||_F} \tag{31}$$

4. The Frobenius norm of the Kalman gain at tenth time step,

$$SD4 = \frac{||K - K_{KF}||_F}{||K_{KF}||_F} \tag{32}$$

Figure 6 shows how the four performance metrics change with an increasing rank N for both CSKF and EnKF. The plot of SD1 (Figure 6a) shows that the initial rank-N approximation of the initial covariance matrix Σ given by all methods is exactly the same, as we generated EnKF realizations using the minimum second-order exact sampling (see supporting information S1.2). Figure 6a (SD1) shows that for both methods a rank of $N = 100$ results in 5% underestimation in the total prior variance, which is reduced to 3% by increasing the rank to $N = 500$. However, as shown in Figure 6b (SD2), after assimilating the same 10 data sets, the underestimation increases to around 30% for $N = 100$. The underestimation by EnKF is bigger than by CSKF, especially for a small rank, e.g., $N = 50$, in which case the EnKF underestimates the posterior variance of KF by 50%. Variance underestimation is often observed when a small ensemble is used to assimilate a large amount of data. To reduce this effect and improve filter stability, EnKF applications typically use a larger ensemble size or localization techniques. In our implementation of EnKF+LOC, the variance was indeed boosted.

Figures 6c and 6d show that the CSKF reproduces the posterior covariance and the Kalman gain more accurately than the EnKF given the same rank N, as indicated by a smaller value of SD3 and SD4. For example, the SD4 value for CSKF with $N = 100$ is less than 1%, while for EnKF the value is around 10%, which suggests that the CSKF with 100 bases captures 99% of the Frobenius norm of the Kalman gain, while EnKF captures less than 90%. Most importantly, these results illustrate that the CSKF can achieve the same accuracy with a smaller rank compared to EnKF. In this specific case, the CSKF with $N = 50$ gives a relative error of 0.01 in the posterior covariance, while to reach the same accuracy EnKF needs at least $N = 300$ (Figure 6c). This is consistent with the theoretical basis of the CSKF, which performs a low-rank approximation based on singular value decomposition; as a result, the CSKF has spectral convergence to the KF, rather than statistical convergence, which is the case for the EnKF.

The rank N and the compression ratio that can be used in low-rank methods are closely related to the characteristics of the problem considered (i.e., smoothness) and the corresponding covariance structures. Generally, low-frequency components of the covariance spectrum are sufficient to capture only main features of the error covariance; finer details can be obtained by adding more high-frequency components. The slower the eigenspectrum decays, the more high-frequency components it has, and the larger rank N is required for a reasonably accurate low-rank approximation. For example, the initial prior covariance used in case A has an eigenspectrum that decays slowly compared to a more typical smooth case where the state is not expected to have sharp features (Figure 7). It can be seen that for the same index (200), in the smooth case (i.e., case B), the eigenvalues drop by 15 orders of magnitude, while for the nonsmooth case (i.e., case A), they drop by 5 orders of magnitude. In order to capture the sharp edge of the CO_2 front of case A, in
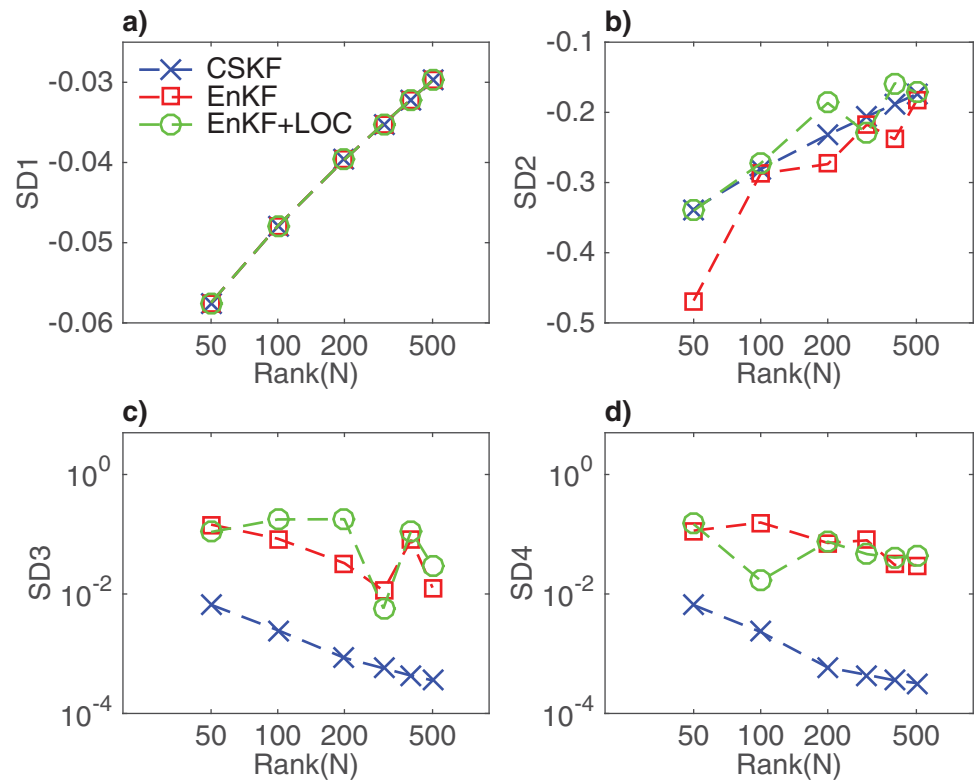
Figure 6. Convergence analysis for case A. (a) SD1: semilog plot of difference in total prior variance ($trace(\Sigma_{0|0})$, negative values indicate underestimation in variance); (b) SD2: semilog plot of difference in the total posterior variance ($trace(\Sigma_{10|10})$) at tenth time step; (c) SD3: log-log plot of errors in the Frobenius norm of the posterior covariance ($\Sigma_{10|10}$); and (d) SD4: log-log plot of errors in the Frobenius norm of Kalman gain K. All metrics measure errors relative to KF.

addition to using a nonsmooth kernel, the rank N used was also larger than what is typically used in the EnKF literature (300 versus typically used 100).

4.2. Case B1: Nonlinear, Single-Variable Estimation

In case B1, CO_2 saturation is estimated every 10 days using 45 pressure measurements at the injection well, 45 water mass flow rate measured at the extraction well, and 25 uniformly distributed saturation measurements to cover the region away from the wells, assuming a uniform initial saturation cl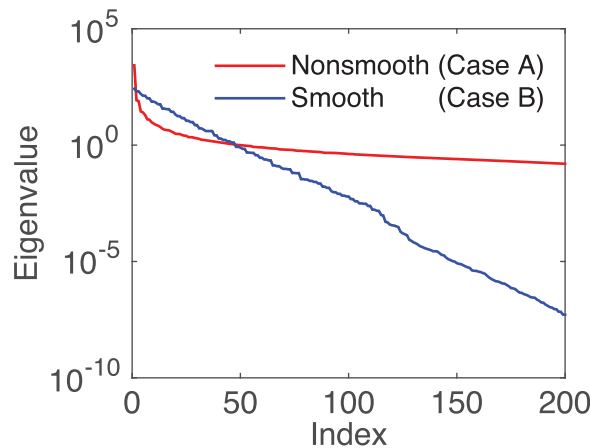ose to zero ($s = 0.07$ or $s_T = -1$). The true saturation field is a nonzero transformed Gaussian field. Snapshots of the true CO_2 saturation simulated using TOUGH2 are shown every 10 days until 50 days in Figure 8a. For this case, we used CSKF with a rank of 50, which corresponds to a compression ratio of $m/N = 2025/50$. The first 50 DCT bases reproduce 99% of the variability of the initial covariance (assumed Gaussian, equation (28)) and are therefore expected to provide accurate estimates. The same data assimilation problem is solved using EnKF and EnKF+LOC with an initial ensemble consisting of 51 realizations sampled from the same Gaussian distribution using the minimum second-order exact sampling approach (supporting information
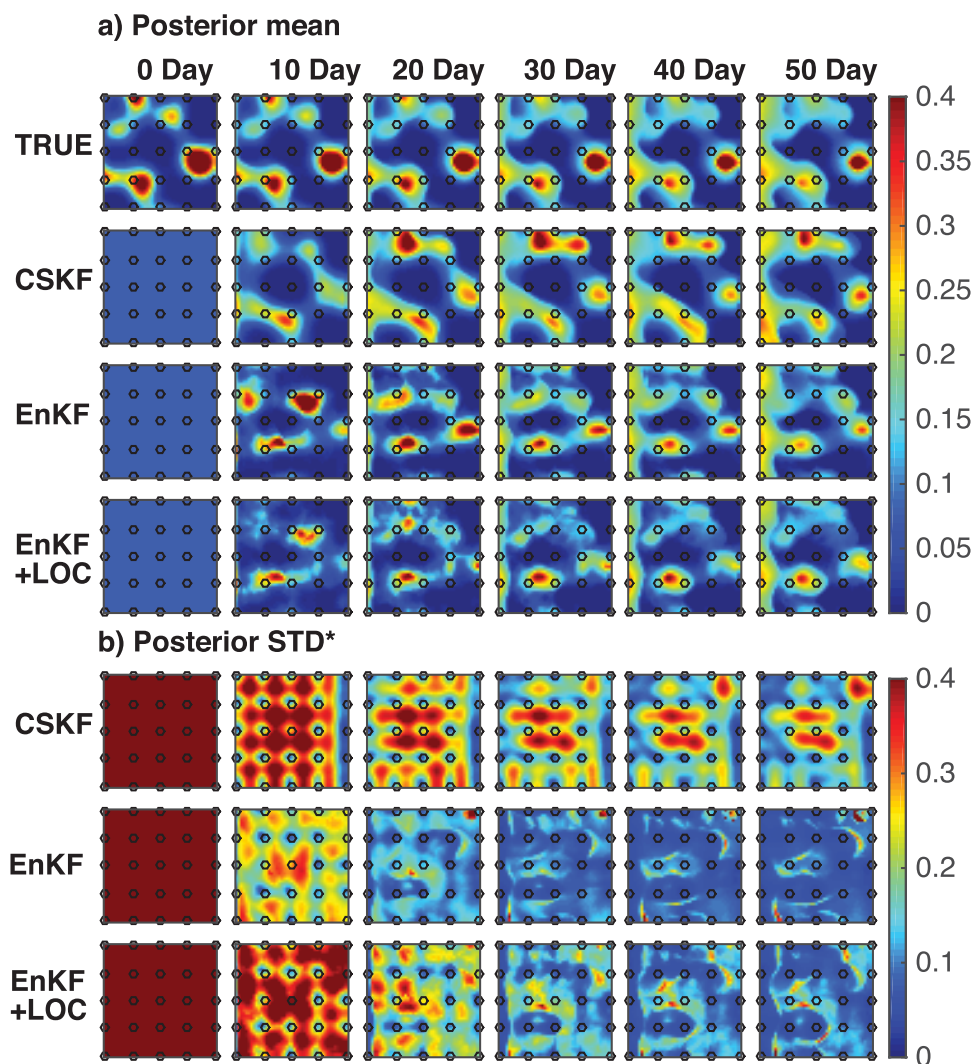

Figure 7. Eigenspectrum of the initial prior covariance Σ_0 used in case A (nonsmooth, equation (25)) and case B (smooth, equation (28)).

Figure 8. Case B1 solution given by CSKF, EnKF, and EnKF+LOC with $N = 50$: (a) true CO_2 saturation and its posterior mean given by each method. (b) Posterior standard deviation (*measured in transformed saturation domain). The sampling locations of saturation are marked by circles.

S1.2). The mean of the 51 realizations is used as the initial guess of the saturation. For a valid comparison, the initial rank-N-approximated covariances given by CSKF, EnKF, and EnKF+LOC are kept the same.

The conditional mean and standard deviation given by CSKF, EnKF, and EnKF+LOC for case B1 are shown in Figures 8a and 8b, respectively. All methods are able to capture the most pronounced features of the CO_2 plume, i.e., the three high saturation spots. As mentioned previously, the corrections made by CSKF are a linear combination of the 50 low-frequency DCT bases and are therefore expected to provide images that are smoother than the true saturation. Similarly, the corrections made by EnKF and EnKF+LOC are expected to be smooth as they are generated from realizations that are sampled from the rank-N approximation of the covariance which only contains low-frequency components. Spurious corrections can be identified in the EnKF solution starting from the first time step (i.e., tenth day), which are partially corrected through localization as more data are assimilated (see EnKF+LOC results). Overall, all methods result in similar estimation at the final assimilation step, even though intermediate results were different.

Figure 8b compares the conditional standard deviation (STD) given by CSKF, EnKF, and EnKF+LOC. The saturation standard deviation given by CSKF indicates a low uncertainty at measurement locations, in particular at the 25 sampling locations of saturation and on the right boundary where the water flux is measured, as expected. Moreover, over time, the uncertainty decreases gradually in the area where the CO_2 flows. In
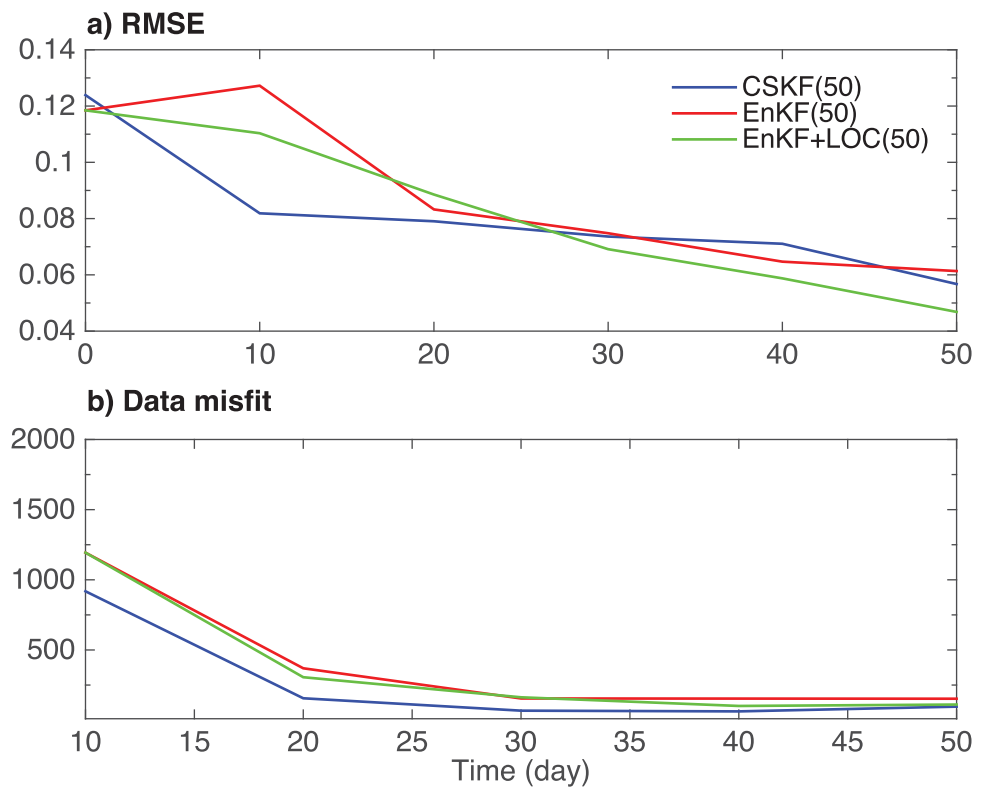
Figure 9. Case B1. (a) SD5: RMSE error in CO_2 saturation. (b) SD6: data misfit over time.

contrast, EnKF experiences a significant reduction in uncertainty after 10 days. This reduction is observed everywhere except at a few locations that have a high STD. This is consistent with our observation in case A where for a small N, EnKF tends to show a fast decay in the posterior variance after a few data assimilation steps (Figure 6b). Standard deviations given by EnKF+LOC (third row of Figure 6b) show that localization boosts the STD of EnKF, which displays a pattern similar to CSKF in the first few steps. However, as more data are assimilated, EnKF+LOC displays random features that are not consistent with the location of the measurements. Even though for this nonlinear case, the results cannot be compared to the true variance because of the prohibitive computational cost of KF, the CSKF gives uncertainty estimates that behave as predicted, indicating that the method is less prone to low-rank approximation errors than the EnKF. Localization does boost the variance of EnKF as expected, but it does not significantly alter its spatial distribution.

Figures 9a and 9b plot the relative estimation error of the saturation field and a measure of data misfit against time, as given by equations (33) and (34), respectively:

$$SD5 = ||s - s_{true}|| / ||s_{true}|| \tag{33}$$

$$SD6 = (y_{true} - h(s))^T R^{-1} (y_{true} - h(s)) \tag{34}$$

Figure 9a shows that the estimation accuracy and data misfit of CSKF, EnKF, and EnKF+LOC are improved progressively as more data are assimilated. The CSKF gives a consistently smaller data misfit (Figure 9b) among all methods. As shown by the EnKF+LOC results, localization improves the estimation accuracy of EnKF. Overall, these results from case B1 suggest that all three methods can give a good estimate of CO_2 saturation using a high compression ratio $m/N = 2025/50$, but CSKF gives more useful uncertainty estimates.

4.3. Case B2: Nonlinear Multiple Variables
In case B2, both saturations and pressures are estimated simultaneously (Figure 10). For this case, we compared CSKF to the standard EnKF only, as it was found from case B1, that localization improves the
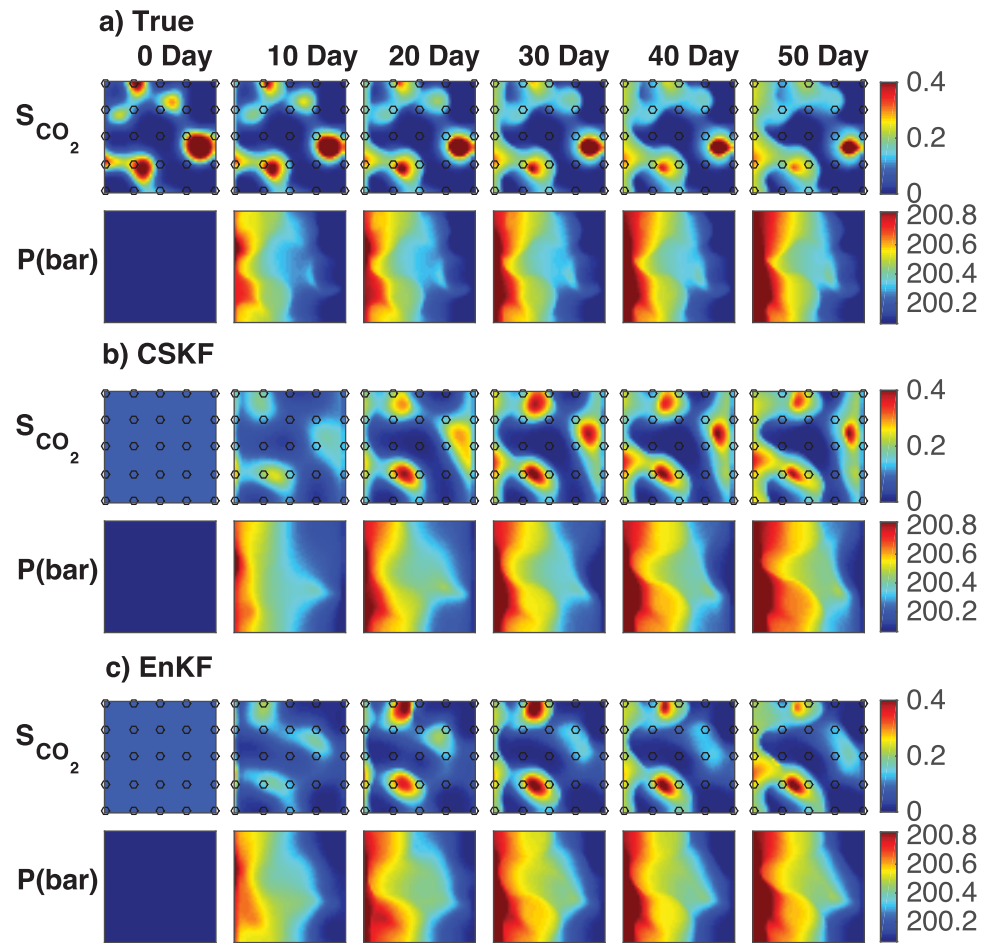
Figure 10. Case B2: true and estimated saturation and pressure every 10 days. (a) True saturations and pressures, (b) CSKF estimates ($N = 50$, 40 DCT basis for saturation and 10 DCT basis for pressure) and (c) EnKF estimates ($N = 50$, 51 realizations). The sampling locations of saturation are marked using circles.

magnitude but not the spatial structure of the uncertainty estimates. Note that in comparison to case B1, case B2 uses an augmented covariance matrix, which is compressed differently by CSKF (equation (20)) and EnKF (equation (5) in supporting information 1.1). CSKF adopts a flexible block low-rank representation that allows to compress saturation and pressure using different bases. This allows us to take advantage of the fact that the pressure field changes smoothly in respond to the injection due to its diffusion dominated nature compared to the CO_2 saturation that has a hyperbolic nature, and changes abruptly over time and space. Based on this observation, we use more bases to capture the changes in saturation and fewer bases for the pressure. Figure 10 shows the CSKF results with 40 bases compressing the saturation covariance and 10 bases compressing the pressure covariance, which are compared to the EnKF results with 51 realizations such that the computational cost is the same between the two methods. A high compression ratio ($m/N = 4050/50$) is selected here in order to avoid evaluations of the expensive forward simulations. Both methods require a total of $N + 1 = 51$ calls of the forward model TOUGH2. Starting with the same initial guess, both CSKF and EnKF can track the evolving CO_2 saturation and pressure with reasonable accuracy and they give similar conditional mean with EnKF performing slightly better by the final step (the fiftieth day), consistent with case B1. It can be observed that the CSKF estimates the pressure accurately even with a small number of bases ($m/N = 2025/10$). In comparison, saturation is more difficult to capture ($m/N = 2025/40$). Using the same number of basis for both variables would have been a waste of computational resources. This is avoided with the CSKF covariance compression scheme, which makes it possible to allocate more resources to estimate saturation instead of pressure given a limited computational budget.

Figure 11a shows the Kalman gain that represents the sensitivity of the saturation state estimate to the fourth saturation measurement, located on the left boundary at each step. The saturation measured at a
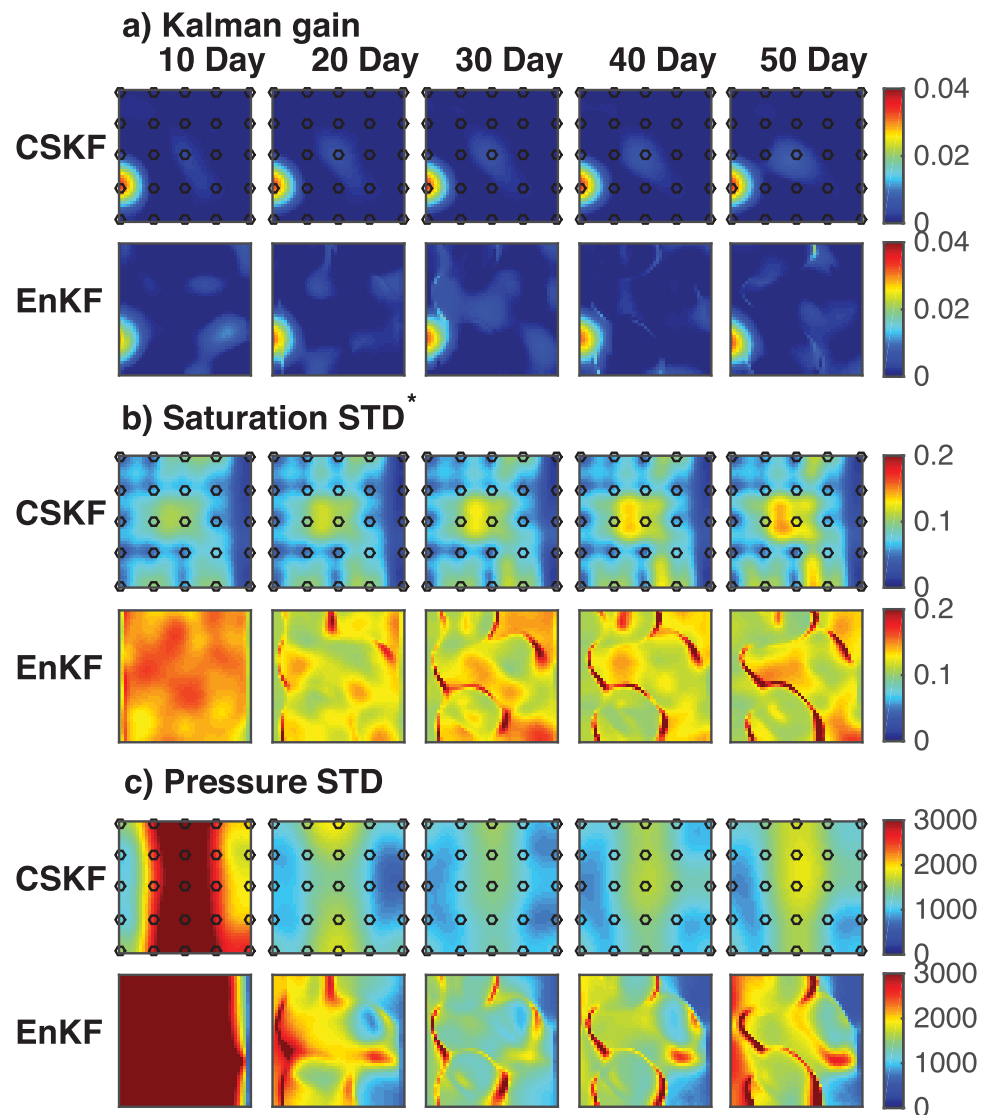
Figure 11. Case B2: for each method and every 10 days: (a) Kalman gain corresponding to fourth saturation measurement on the left boundary. (b) Standard deviation of saturation (*measured in the transformed saturation domain). (c) Standard deviation of pressure.

point directly constraints the saturation value at this location and those within a small distance and should have no effect on the saturation values beyond a certain distance. Both CSKF and EnKF capture the localized high sensitivity zone around the saturation measurement location. However, spurious correlations are observed for both methods, an effect that results from low-rank approximation, which is more pronounced in EnKF than in CSKF (same with case A). This leads to deteriorated estimates of uncertainty, as shown in Figures 11b and 11c, where the standard deviation of pressure and saturation are plotted every 10 days. The saturation standard deviation given by CSKF has similar characteristics compared to case B1 (Figure 8b) and the pressure standard deviation indicates lower uncertainty around the monitoring wells located on the left and right boundary. However, the EnKF with the same rank N cannot produce meaningful uncertainty estimates.

5. Conclusions

In this paper, we present the generalized compressed state Kalman filter (CSKF), a new efficient algorithm for large-scale nonlinear state estimation problems. The CSKF reduces the computational requirements of

the state estimation significantly, by decreasing the number of evaluations of the forward model and by taking advantage of fast linear algebra techniques. Furthermore, it is straightforward to implement as it can be combined with any forward model in a black-box fashion, with no need for explicit Jacobian computations. Our method belongs to the category of low-rank approximation methods for the Kalman filter, featuring a smart compression scheme appropriate to the characteristics of the covariance, and offers an alternative to ensemble-based low-rank methods.

The CSKF was first validated for a simple linear case and its performance was compared to the KF and the EnKF with and without localization. It was shown that the CSKF accurately reproduced the Kalman gain and state estimates of KF with a small number of bases and gave high quality uncertainty quantification. Compared to the EnKF, for this simple case, the CSKF provides slightly better results for the mean. The covariance was underestimated by both CSKF and EnKF, more so by EnKF. Adding localization to EnKF can boost the variance but does not necessarily improve the quality of uncertainty analysis. We also performed a convergence analysis and investigated the effect of the compression ratio (i.e., the number of unknowns versus the rank) on the results; our results indicate that the lower the rank (and therefore the higher the compression), the higher the chance for overcorrection, an effect appears to be more pronounced for the EnKF than for the CSKF.

The CSKF was then evaluated for a CO_2 monitoring case, which is a more complex, nonlinear problem governed by multiphase physics. The CSKF was shown to perform equally well to the EnKF in terms of the estimated mean. However, large discrepancies were observed between the two methods in the computed posterior variance. In particular, the CSKF provided robust Kalman gain and uncertainty estimates that did not degrade as more data were assimilated, as opposed to drastic variance reductions predicted by the EnKF indicating potential ensemble collapse and inaccurate Kalman gain calculations. Localization was able to boost the variance, however, it introduced bias into the Kalman gain and the uncertainty estimate with spatial structures that were not consistent with the measurement locations and had spurious features. This finding was corroborated by the results of a second CO_2 monitoring case, where two different types of state variables were estimated through state augmentation. Overall, for the specific cases investigated here, it appears that for the same compression ratio, and therefore the same computational cost, the CSKF gives equally good results for the mean estimate compared to EnKF, but more reliable uncertainty estimates, being less subject to spurious correlations and excessive variance reduction.

With the spectrum-based compression scheme of CSKF, significant compression and computational savings are possible. This was also shown in our convergence analysis, which showed that the CSKF requires smaller rank, i.e., fewer forward simulations to reach the same accuracy of EnKF. Furthermore, the CSKF does not require further adjustments or optimization, while localization is necessary for EnKF. In the latter case, the decision of how to localize involves ad hoc choices and tuning, which makes the method less robust than CSKF. The CSKF only requires the decision on the type and the number of bases used, which can be made based on the variability and smoothness inherent to the physical problem.

Adding to the efficiency of our method is its ability to adapt the compression scheme to cases where more than one type of state variables are estimated simultaneously; the method allows more compression for smoothly varying variables like pressure and allocates more computational resources to less smooth state variables, like CO_2 saturation in our case. Therefore, an optimal choice of compression is possible based on an understanding of the physics of each application, as the appropriate rank and potential computational benefits depend on the nature and characteristics of each specific problem.

The computational efficiency of the CSKF has important implications for problems with large numbers of unknowns, such as large-scale field studies that may extend over several kilometers and where the unknowns may be in excess of 10^6 after discretization. In such cases, the major limitation of inversion techniques is that due to cost/time constraints, domains are coarsely discretized to reduce the number of unknowns, leading to inaccurate estimation with low resolution, in addition to increased errors in forward simulations. The reduction in computational cost of data assimilation using the CSKF allows for finer discretization, enabling the delineation of smaller-scale features.

From a CO_2 storage and a contaminant hydrogeology perspective, the computational efficiency of our method and its potential to detect fine-scale features are of great importance. Detecting such features at a

reasonable computational cost could enable prevention of high-risk events, such as CO_2 leakage through fractures, preferential flow of contaminants through high permeability lenses, contaminant accumulation in low permeability formations, as well as better estimation of reservoir and aquifer capacity, for CO_2 storage, managed aquifer recharge and other near-surface hydrogeological applications.

The CSKF provides a reliable and robust method to control the trade-off between the accuracy and the computational cost of low-rank Kalman filtering for nonlinear estimation problems and is a practical alternative to conventional computationally intensive inversion techniques and ensemble-based compression techniques. Despite the mathematical sophistication of the method, the method as presented here is well suited for black-box use, where the user can apply the method for data assimilation without specialized knowledge of inversion and statistics, and for any application, as long as the forward model is available. Software is currently under development to make the algorithm available to the research community for further comparison and evaluation for other applications. The method can be further extended for combined parameter and state estimation, topics that will be pursued in future work.

Appendix A: Derivation of the Analysis Scheme of CSKF

Assume the prior covariance

$$\Sigma_{k+1|k} = A^\Phi C_{k|k} A^{\Phi T} + AV A^T \tag{A1}$$

can be factorized as $AC_{k+1|k}A^T$. Then, instead of computing $\Sigma_{k+1|k}$ explicitly, we only need to compute its compressed form by

$$C_{k+1|k} = (A^T A^\Phi) C_{k|k} (A^T A^\Phi)^T + V \tag{A2}$$

with $\mathcal{O}(m)$ operations.

In the analysis step, the state at time step $k + 1$ is updated using measurements at time step $k + 1$:

$$x_{k+1|k+1} = x_{k+1|k} + K(y_{k+1} - h(x_{k+1|k})) \tag{A3}$$

The Kalman gain, K, which is used to compute the correction made to the prior state $x_{k+1|k}$, is given by

$$\Sigma_{HH} K^T = \Sigma_{Hx} \tag{A4}$$

where $\Sigma_{HH} = H\Sigma_{k+1|k}H^T + R$ and $\Sigma_{Hx} = H\Sigma_{k+1|k}$. Use the relationship $\Sigma_{k+1|k} = AC_{k+1|k}A^T$, they can be computed with $\mathcal{O}(m)$ instead of $\mathcal{O}(m^2)$ operations through

$$\Sigma_{HH} = A^H C_{k+1|k} A^{HT} + R \tag{A5}$$

$$\Sigma_{Hx} = A^H C_{k+1|k} A^T \tag{A6}$$

As suggested in *Kitanidis* [2015], K can be computed efficiently by solving the reduced system

$$\Sigma_{HH}(K^T A) = \Sigma_{Hx} A \tag{A7}$$

Note that $\Sigma_{Hx}A = A^H C_{k+1|k}$. First, $\mathcal{X} = K^T A$ is computed by solving the system

$$\mathcal{A}\mathcal{X} = A^H C_{k+1|k} \tag{A8}$$

Then, the Kalman gain is given by

$$K = A\mathcal{X}^T \tag{A9}$$

Appendix B: Scaling of Innovation Covariance

The innovation covariance $H\hat{\Sigma}H^T + R$ is used to compute Kalman gain,

$$K = \hat{\Sigma}H^T (H\hat{\Sigma}H^T + R)^{-1} \tag{B1}$$

When the innovation covariance matrix is rank deficient, its inverse can be computed using truncated SVD. However, when observation contains different units, the innovation matrix has to be scaled following

$$K = \hat{\Sigma}H^T U^{-1T}\left(U^{-1}H\hat{\Sigma}H^T U^{-1T} + I\right)^{\dagger} U^{-1} \tag{B2}$$

before computing its pseudoinverse. The derivation is based on $R = UU^T$. The innovation matrix here is dimensionless.

References

Ahmed, N., T. Natarajan, and K. Rao (1974), Discrete cosine transform, *IEEE Trans. Comput.*, *C-23*(1), 90–93, doi:10.1109/T-C.1974.223784.
Ambikasaran, S., J. Li, E. Darve, and P. K. Kitanidis (2013a), Large-scale stochastic linear inversion using hierarchical matrices, *Comput. Geosci.*, *17*(6), 913–927, doi:10.1007/s10596-013-9364-0.
Ambikasaran, S., A. K. Saibaba, E. Darve, and P. K. Kitanidis (2013b), *Fast Algorithms for Bayesian Inversion, The IMA Volumes in Mathematics and Its Applications*, vol. 156, Springer, New York, doi: 10.1007/978-1-4614-7434-0_5.
Aminfar, A., S. Ambikasaran, and E. Darve (2016), A fast block low-rank dense solver with applications to finite-element matrices, *J. Comput. Phys.*, *304*, 170–188.
Anderson, B., and J. Moore (1979), *Optimal Filtering*, vol. 11, Prentice Hall, Englewood Cliffs, N. J.
Anderson, J. L. (2001), An ensemble adjustment Kalman filter for data assimilation, *Mon. Weather Rev.*, *129*(12), 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.
Anderson, J. L. (2003), A local least squares framework for ensemble filtering, *Mon. Weather Rev.*, *131*(4), 634–642, doi:10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.
Anderson, J. L. (2012), Localization and sampling error correction in ensemble Kalman filter data assimilation, *Mon. Weather Rev.*, *140*(7), 2359–2371, doi:10.1175/MWR-D-11-00013.1.
Arogunmati, A., and J. M. Harris (2012), An effective crosswell seismic traveltime-estimation approach for quasi-continuous reservoir monitoring, *Geophysics*, *77*(2), M17–M26, doi:10.1190/geo2011-0197.1.
Bengtsson, T., P. Bickel, B. Li (2008), Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, in *Probability and Statistics: Essays in Honor of David A. Freedman*, pp. 316–334, Inst. of Math. Stat., doi:10.1214/193940307000000518.
Daley, T., J. Ajo-Franklin, and C. Doughty (2011), Constraining the reservoir model of an injected CO_2 plume with crosswell CASSM at the Frio-II brine pilot, *Int. J. Greenhouse Gas Control*, *5*(4), 1022–1030, doi:10.1016/j.ijggc.2011.03.002.
Doughty, C. (2010), Investigation of CO_2 plume behavior for a large scale pilot test of geologic carbon storage in a saline formation, *Transp. Porous Media*, *82*(1), 49–76, doi:10.1007/s11242-009-9396-z.
Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162, doi:10.1029/94JC00572.
Evensen, G. (2007), *Data Assimilation: The Ensemble Kalman Filter*, Springer, Berlin, doi:10.1007/978-3-642-03711-5.
Furrer, R., and T. Bengtsson (2007), Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants, *J. Multivariate Anal.*, *98*(2), 227–255.
Gelb, A. (1974), *Applied Optimal Estimation*, MIT Press, Cambridge, Mass.
Halko, N., P. Martinsson, and J. Tropp (2011), Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, *SIAM Rev.*, *53*(2), 217–288, doi:10.1137/090771806.
Houtekamer, P., and H. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, *126*(3), 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
Jafarpour, B., and D. B. McLaughlin (2008), History matching with an ensemble Kalman filter and discrete cosine parameterization, *Comput. Geosci.*, *12*(2), 227–244, doi:10.1007/s10596-008-9080-3.
Jafarpour, B., and D. B. McLaughlin (2009a), Reservoir characterization with the discrete cosine transform, *SPE J.*, *14*(01), 182–201, doi:10.2118/106453-PA.
Jafarpour, B., and D. B. McLaughlin (2009b), Estimating channelized-reservoir permeabilities with the ensemble Kalman filter: The importance of ensemble design, *SPE J.*, *14*(02), 374–388, doi:10.2118/108941-PA.
Kalman, R. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, *82*(1), 35–45, doi:10.1109/9780470544334.ch9.
Kitanidis, P. K. (2015), Compressed state Kalman filter for large systems, *Adv. Water Resour.*, *76*, 120–126, doi:10.1016/j.advwatres.2014.12.010.
Lee, J., and P. K. Kitanidis (2014), Large-scale hydraulic tomography and joint inversion of head and tracer data using the principal component geostatistical approach (PCGA), *Water Resour. Res.*, *50*, 5410–5427, doi:10.1002/2014WR015483.
Li, J. Y., S. Ambikasaran, E. F. Darve, and P. K. Kitanidis (2014), A Kalman filter powered by $\mathcal{H}^2$ matrices for quasi continuous data assimilation problems, *Water Resour. Res.*, *50*, 3734–3749, doi:10.1002/2013WR014607.
Liu, Y., and H. V. Gupta (2007), Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resour. Res.*, *43*, W07401, doi:10.1029/2006WR005756.
Lorenc, A. C. (2003), The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var, *Q. J. R. Meteorol. Soc.*, *129*(595), 3183–3203, doi:10.1256/qj.02.132.[10.1256/qj.02.132]
McLaughlin, D., L. B. Reid, S.-G. Li, and J. Hyman (1993), A stochastic method for characterizing ground-water contamination, *Ground Water*, *31*(2), 237–249, doi:10.1111/j.1745-6584.1993.tb01816.x.
Nenna, V., A. Pidlisecky, and R. Knight (2011), Application of an extended Kalman filter approach to inversion of time-lapse electrical resistivity imaging data for monitoring recharge, *Water Resour. Res.*, *47*, W10525, doi:10.1029/2010WR010120.
Oliver, D. S., and Y. Chen (2011), Recent progress on reservoir history matching: A review, *Comput. Geosci.*, *15*(1), 185–221, doi:10.1007/s10596-010-9194-2.
Pham, D. T., J. Verron, and M. C. Roubaud (1998), A singular evolutive extended Kalman filter for data assimilation in oceanography, *J. Mar. Syst.*, *16*(3–4), 323–340, doi:10.1016/S0924-7963(97)00109-7.
Pruess, K. (1991), TOUGH2: A general-purpose numerical simulator for multiphase fluid and heat flow, Lawrence Berkeley Lab, Berkeley, Calif., doi:10.2172/5212064.

Pruess, K., and J. Nordbotten (2011), Numerical simulation studies of the long term evolution of a CO_2 plume in a saline aquifer with a sloping caprock, *Transp. Porous Media*, *90*, 135–151, doi:10.1007/s11242-011-9729-6.

Pruess, K., and N. Spycher (2007), ECO2N—A fluid property module for the TOUGH2 code for studies of CO_2 storage in saline aquifers, *Energy Convers. Manage.*, *48*(6), 1761–1767, doi:10.1016/j.enconman.2007.01.016.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, *130*, 103–114, doi:10.1175/1520-0493(2002)130 < 0103:HDAWTE>2.0.CO;2.

Sætrom, J., and H. Omre (2011), Ensemble Kalman filtering with shrinkage regression techniques, *Comput. Geosci.*, *15*(2), 271–292, doi:10.1007/s10596-010-9222-2.

Saibaba, A., S. Ambikasaran, J. Li, P. Kitanidis, and E. Darve (2012), Application of hierarchical matrices to linear inverse problems in geostatistics, *Oil Gas Sci. Technol.*, *67*(5), 857–875, doi:10.2516/ogst/2012064.

Saibaba, A. K., and P. K. Kitanidis (2012), Efficient methods for large-scale linear inversion using a geostatistical approach, *Water Resour. Res.*, *48*, W05522, doi:10.1029/2011WR011778.

Strang, G. (1999), The discrete cosine transform, *SIAM Rev.*, *41*(1), 135–147, doi:10.1137/S0036144598336745.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, Pa., doi:10.1137/1.9780898717921.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker (2003), Ensemble square root filters, *Mon. Weather Rev.*, *131*(7), 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.

Tuan Pham, D., J. Verron, and M. Christine Roubaud (1998), A singular evolutive extended Kalman filter for data assimilation in oceanography, *J. Mar. Syst.*, *16*(3–4), 323–340, doi:10.1016/S0924-7963(97)00109-7.

Wang, Y., G. Li, A. C. Reynolds, et al. (2010), Estimation of depths of fluid contacts by history matching using iterative ensemble-Kalman smoothers, *SPE J.*, *15*(02), 509–525, doi:10.2118/119056-PA.

Yeh, W. W.-G. (1986), Review of parameter identification procedures in groundwater hydrology: The inverse problem, *Water Resour. Res.*, *22*, 95–108, doi:10.1029/WR022i002p00095.

Zhang, Y., and D. S. Oliver (2010), Improving the ensemble estimate of the Kalman gain by bootstrap sampling, *Math. Geosci.*, *42*(3), 327–345.

Zhou, Q., J. T. Birkholzer, E. Mehnert, Y.-F. Lin, and K. Zhang (2010), Modeling basin-and plume-scale processes of CO_2 storage for full-scale deployment, *Ground Water*, *48*(4), 494–514, doi:10.1111/j.1745-6584.2009.00657.x.

Zhou, Y., D. McLaughlin, and D. Entekhabi (2006), Assessing the performance of the ensemble Kalman filter for land surface data assimilation, *Mon. Weather Rev.*, *134*(8), 2128–2142, doi:10.1175/MWR3153.1.