2017

Smoothing-based Compressed State Kalman Filter for Joint State-parameter Estimation: Applications in Reservoir Characterization and CO2 Storage Monitoring

Y. J. Li

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

E F. Darve

P K. Kitanidis

Recommended Citation
Li, Y. J., A. Kokkinaki, E. F. Darve, and P. K. Kitanidis (2017), Smoothing-based compressed state Kalman filter for joint state-parameter estimation: Applications in reservoir characterization and CO2 storage monitoring, Water Resour. Res., 53, 7190–7207, doi:10.1002/2016WR020168.

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2016WR020168

Key Points:
- A Bayesian consistent solution to physical inconsistency issues in large-scale nonlinear combined state-parameter estimation problems
- Reduce overshootings in estimating hyperbolic-type states with sharp changes where prediction uncertainty is from unknown model parameters
- A nonensemble approach that leverages covariance compression and matrix-free approach for reduced computational cost

Supporting Information:
- Supporting Information S1

Correspondence to:
Y. J. Li,
yuel@alumni.stanford.edu

Citation:
Li, Y. J., A. Kokkinaki, E. F. Darve, and P. K. Kitanidis (2017), Smoothing-based compressed state Kalman filter for joint state-parameter estimation: Applications in reservoir characterization and CO_2 storage monitoring, *Water Resour. Res.*, *53*, 7190–7207, doi:10.1002/2016WR020168.

Received 24 NOV 2016
Accepted 8 JUN 2017
Accepted article online 16 JUN 2017
Published online 19 AUG 2017

Smoothing-based compressed state Kalman filter for joint state-parameter estimation: Applications in reservoir characterization and CO_2 storage monitoring

Y. J. Li[1], Amalia Kokkinaki[2], Eric F. Darve[3,4], and Peter K. Kitanidis[1,3]

[1]Department of Civil and Environmental Engineering, Stanford University, Stanford, California, USA, [2]Department of Environmental Science, University of San Francisco, San Francisco, California, USA, [3]Institute for Computational and Mathematical Engineering, Jen-Hsun Huang Engineering Center, Stanford University, Stanford, California, USA, [4]Department of Mechanical Engineering, Stanford University, Stanford, California, USA

Abstract The operation of most engineered hydrogeological systems relies on simulating physical processes using numerical models with uncertain parameters and initial conditions. Predictions by such uncertain models can be greatly improved by Kalman-filter techniques that sequentially assimilate monitoring data. Each assimilation constitutes a nonlinear optimization, which is solved by linearizing an objective function about the model prediction and applying a linear correction to this prediction. However, if model parameters and initial conditions are uncertain, the optimization problem becomes strongly nonlinear and a linear correction may yield unphysical results. In this paper, we investigate the utility of one-step ahead smoothing, a variant of the traditional filtering process, to eliminate nonphysical results and reduce estimation artifacts caused by nonlinearities. We present the smoothing-based compressed state Kalman filter (sCSKF), an algorithm that combines one step ahead smoothing, in which current observations are used to correct the state and parameters one step back in time, with a nonensemble covariance compression scheme, that reduces the computational cost by efficiently exploring the high-dimensional state and parameter space. Numerical experiments show that when model parameters are uncertain and the states exhibit hyperbolic behavior with sharp fronts, as in CO_2 storage applications, one-step ahead smoothing reduces overshooting errors and, by design, gives physically consistent state and parameter estimates. We compared sCSKF with commonly used data assimilation methods and showed that for the same computational cost, combining one step ahead smoothing and nonensemble compression is advantageous for real-time characterization and monitoring of large-scale hydrogeological systems with sharp moving fronts.

Plain Language Summary Geologic CO_2 storage is a promising technology to reduce the CO_2 in the atmosphere by injecting them into the deep saline reservoir for permanent storage. To assure safe operations and effective containment of CO_2, numerical models are developed to accurately predict the CO_2 behaviors underground in order to make informed decisions, such as adjusting the volume and rate of injection to prevent fracturing the surrounding rock. However, because of our limited knowledge about the reservoir properties, often the numerical model is highly uncertain. Statistical techniques like Kalman filtering use sensor data to reduce the prediction uncertainty in the numerical model by correcting the unknown reservoir properties recursively in time when data becomes available. The amount of correction is determined by solving an optimization problem. However, it is computationally intractable to find feasible solutions to such problems if reservoir properties to be estimated are high dimensional. Moreover, when the optimization problem is nonlinear, Kalman-type approaches can give unphysical results. By improving the way information is extracted from the sensor data, we present a new Kalman-type approach that can solve this optimization problem with better accuracy and reduced uncertainty.

1. Introduction

The major uncertainty in predicting fluid flow in the subsurface with numerical models arises from the heterogeneity in geologic parameters, which is inherently difficult to observe directly. This has implications for

a wide range of applications, like groundwater storage [e.g., *Zhou et al*., 2010], groundwater contamination [e.g., *Snodgrass and Kitanidis*, 1997], and storage of hazardous materials in the subsurface [e.g., *Shapiro*, 1996]. One specific application in the latter category is the storage of carbon dioxide (CO_2) in subsurface formations as a climate change mitigation measure. The permanence, and thereby the safety of such storage is determined by the geologic properties of the utilized formations, with failures increasing the risks for leakage and groundwater contamination. To reduce these risks, continuous monitoring of the spatial distribution of the injected CO_2 is required.

The location of injected CO_2 can be estimated using data assimilation, a collection of powerful statistical techniques that estimate the state of a system using direct or indirect measurements, such as information collected at wells in the case of CO_2 injection. Although data assimilation techniques are traditionally used for state estimation, they can be adapted by using state augmentation, such that the unknowns include both the state variables and model parameters [e.g., *Naevdal et al*., 2003]. In the CO_2 storage case, this would allow the estimation of CO_2 saturation in space and simultaneous estimation of the heterogeneous rock properties (e.g., permeability).

Direct application of conventional data assimilation techniques, like the extended Kalman filter (EKF), with state augmentation for joint state-parameter estimation is challenging. First, the EKF is computationally prohibitive for large-scale systems because its implementation operates directly on the full-order covariance matrix, and the resulting computational cost scales with the squared number of unknowns. For a three-dimensional reservoir that is coarsely discretized, the unknowns parameters may be in excess of 10^6, and computations, which involve numerous runs of reservoir simulators, could take days or even months depending on site-specific conditions. This problem has been handled with low-rank filters such as the ensemble Kalman filter (EnKF) [*Evensen*, 1994; *Burgers et al*., 1998] and the compressed state Kalman filter (CSKF) [*Kitanidis*, 2015; *Li et al*., 2015]. Ensemble-based methods have been shown to suffer from filter inbreeding [*Hendricks Franssen and Kinzelbach*, 2008] and variance reduction problem [*Lorenc*, 2003]. Most importantly, in both ensemble methods like the EnKF, and in nonensemble methods, like the CSKF, that follow the conventional prediction-correction sequence, the extension to combined state and parameter estimation (e.g., permeability and saturation) can result in physical inconsistencies between the estimated states and parameters [*Wen and Chen*, 2006].

Physical inconsistencies are usually manifested as state predictions that are not physically possible for the corresponding estimated parameter field. Such inconsistencies between estimated parameters and states are caused by errors in an approximate linear correction in the state variables that occurs during the updating step. An important practical implication of physical inconsistencies in estimated states and parameters is that the filter becomes unstable; such cases have been reported using the joint EnKF [e.g., *Moradkhani et al*., 2005; *Wen and Chen*, 2007]. Unphysical linear corrections, also referred to as overshooting, are often attributed to strong nonlinearities in the relationship between measurements and unknowns. The problem is exacerbated when model parameters and initial conditions are poorly known, in which case significant linearization errors are expected, since Kalman filter methods were designed and are expected to work for problems that, at least locally, vary linearly. Linearization errors can be reduced through iterative approaches, where linearization points are improved successively by repetitively conditioning the state variables on the same observations. This, however, increases the computational cost, as several iterations may be needed until convergence is achieved [*Bar-Shalom and Li*, 1993; *Kitanidis*, 1995].

Noniterative methods that address the physical inconsistency problem often add a heuristic step to the conventional prediction-correction filtering process, that involves generating the state variables from the physical model as the final step of the estimation. For example, *Wen and Chen* [2006] included a third step in the prediction-correction Kalman filter recurrence, the verification step, where the updated parameters were used as input for an additional simulation to avoid nonphysical updates in the state variables. *Gu and Oliver* [2007] used an approach, where again, forward simulations were restarted after permeabilities were updated to obtain the final state estimates to reduce physical inconsistencies. This procedure is often applied even for iterative approaches to avoid overshooting [*Gu and Oliver*, 2007; *Man et al*., 2016]. However, such fixes result in algorithms that are no longer consistent with the Bayesian framework [*Hendricks Franssen and Kinzelbach*, 2008].

Dual KF approaches, designed as two interactive filters that update state and model parameters separately, have been introduced to reduce the physical inconsistency. A dual extended Kalman filter was applied to neural networks models to estimate signals (states) and weights (parameters) separately by *Wan and Nelson* [2001]. A dual EnKF was applied for joint state-parameter estimation by first updating the parameters and then applying the EnKF again with the updated parameters to obtain the final state estimation [*Moradkhani et al*., 2005]. A similar two-step updating approach was used by *Hendricks Franssen and Kinzelbach* [2008] and then by *Nowak* [2009] where the final state was updated via simulation with the updated parameter field instead of filtering to ensure physically consistent state and parameter estimates. Such approaches resolve the physical inconsistency problems that arise from the joint updating of states and parameters, however, deviate from the formal Bayesian filtering framework.

Recently, *Desbouvries et al*. [2011], followed by *Gharamti et al*. [2015] showed that the previously heuristically implemented dual filtering concept can be derived within the Bayesian framework by introducing one-step-ahead smoothing in the probabilistic formulation. Compared to the standard Kalman Filter sequence, in which model prediction is followed by parameter updates, the one-step ahead smoothing algorithm adopts a smoothing step that simultaneously updates states at the previous time step and parameters and then propagates states to the next time step using a physical model. This smoothing-based filtering approach for joint estimation of states and parameters avoids physical inconsistencies while also being Bayesian consistent [*Desbouvries et al*., 2011]. However, the smoothing-based filters developed so far are based on ensemble representations of the covariance matrix. An evaluation of one-step ahead smoothing in a setting unaffected by sampling and approximation errors has not been presented in the literature, resulting in a limited understanding of the method and its utility in data assimilation.

This paper focuses on the relevance of one-step ahead smoothing for applications of hydrogeological interest, specifically for systems with unknown model parameters and states exhibiting sharp moving fronts. In hydrogeological problems, sharp fronts can be observed in the transport of immiscible fluids through the heterogeneous porous rocks. We modify and combine one step ahead smoothing with our efficient, nonensemble-based filtering algorithm, compressed state Kalman filter (CSKF) [*Kitanidis*, 2015; *Li et al*., 2015], to conduct joint state and parameter estimation for a CO_2 injection application. Our results suggest that the smoothing approach is more appropriate than the conventional filtering approach for reservoir applications, where the model prediction errors stem from uncertain model parameters and where the state variable exhibits sharp fronts and is physically bounded. The new algorithm, termed smoothing based CSKF (sCSKF) is compared to alternative filters with the same computational cost, namely the CSKF, the ensemble Kalman filter (EnKF), and a two-step iterative Kalman filter for nonlinear large-scale state-parameter estimation.

The paper is organized as follows: section 2 reviews two alternative information processing sequences for Kalman-type data assimilation techniques and shows how sCSKF is derived. Section 3 compares the filter performance with and without smoothing for a one-dimensional numerical benchmark and sheds light on how smoothing affects the performance of the filter with a focus on physical consistency and the reliability of the uncertainty estimates. In section 4, we apply the one-step-ahead smoothing-based CSKF (sCSKF) for a synthetic application in characterizing and monitoring a heterogeneous reservoir during CO_2 injection.

2. Theory and Methods

2.1. Two Paths for Bayesian Recursion

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

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

and the measurement operator:

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

with m denoting the number of unknowns and n is the number of measurements, x_k is the uncertain state at time k, and α is the vector of uncertain parameters. In the context of reservoir monitoring, equation (1) describes the multiphase flow process where $f(\cdot)$ is the forward model, and equation (2) relates the observations collected at the wells (e.g., flux and pressure), to the changes in the reservoir states (e.g., saturation, pressure) through the measurement function $h(\cdot)$. The terms w_k and v_k represent the model and

measurement noise at time k, respectively. To jointly estimate both state and parameters, state augmentation can be used, combining x_k and α into an augmented unknown state $X_k = [x_k, \alpha]^T$.

To estimate the augmented state X_k based on measurements y_k, there are two ways in which information can be processed:

$$\textbf{P-path} : X_{k-1|k-1} \xrightarrow{Predict} X_{k|k-1} \xrightarrow[y_k]{Correct} X_{k|k}$$

$$\textbf{S-path} : X_{k-1|k-1} \xrightarrow[y_k]{Smoothing} X_{k-1|k} \xrightarrow{Predict} X_{k|k}$$

where the subscript of $X_{k|k-1}$ denotes state at time step k conditioned on the data up to time step $k-1$, and the notation y_k below the arrow indicates conditioning on observation. *Desbouvries et al*. [2011] refers to the two alternative processing sequences as the prediction-based P-path and the smoothing-based S-path, respectively, and we will adopt the same notation. As indicated from the graph, the P-path first obtains a prediction of the current state through a forward model and then performs a correction to the predicted state $X_{k|k-1}$ by conditioning on the observation y_k. The S-path on the other hand, first uses the observations y_k to correct the initial condition X_{k-1}, and then predicts the current state X_k is by the forward model using the improved initial condition $X_{k-1|k}$.

The solution to the joint state-parameter filtering problem is given by finding the state X_k that maximizes the posterior distribution $p(X_k|y_{0:k})$. We can express the unknown target posterior distribution $p(X_k|y_{0:k})$ at time k in terms of the known posterior distribution $p(X_{k-1}|y_{0:k-1})$ at time $k-1$ by first applying the law of total probability:

$$p(X_k|y_{0:k}) = \int p(X_k, X_{k-1}|y_{0:k}) dX_{k-1} \tag{3}$$

where $p(X_k, X_{k-1}|y_{0:k})$ denotes that X_k and X_{k-1} are jointly conditioned on the data history between the initial step 0 and the current step k. Bayes rule and the conditional independence properties of the hidden Markov chain are used recursively to obtain [*Ho and Lee*, 1964]:

$$p(X_k|y_{0:k}) \propto \int p(X_k, y_k|X_{k-1}) p(X_{k-1}|y_{0:k-1}) dX_{k-1} \tag{4}$$

Then, given the posterior density $p(X_{k-1}|y_{0:k-1})$ at time $k-1$, the posterior density $p(X_k|y_{0:k})$ can be computed by expressing the transition probability density function $p(X_k, y_k|X_{k-1})$ explicitly in terms of known information.

In calculating $p(X_k, y_k|X_{k-1})$, the observation y_k can be processed in two different orders. The P-path, which is the conventional filtering process, is based on factorizing the fundamental transition density $p(X_k, y_k|X_{k-1})$ by:

$$p(X_k, y_k|X_{k-1}) = p(y_k|X_k) p(X_k|X_{k-1}) \tag{5}$$

Substituting equation (5) into equation (4), we obtain the posterior density $p(X_k|y_{0:k})$ as a function of known quantities i.e.,

$$p(X_k|y_{0:k}) \propto p(y_k|X_k) \int p(X_k|X_{k-1}) p(X_{k-1}|y_{0:k-1}) dX_{k-1} \tag{6}$$

In the P-path, equation (6) is evaluated in two steps: a prediction step that first integrates over X_{k-1} to compute the predictive density (equation (7)) and then a correction step that multiplies the result with the likelihood $p(y_k|X_k)$ to obtain the posterior density (equation (8)):

$$p(X_k|y_{0:k-1}) \propto \int p(X_k|X_{k-1}) p(X_{k-1}|y_{0:k-1}) dX_{k-1} \tag{7}$$

$$p(X_k|y_{0:k}) \propto p(y_k|X_k) p(X_k|y_{0:k-1}) \tag{8}$$

The S-path is based on factorizing the fundamental transition density by (cf., equation (5)):

$$p(X_k, y_k|X_{k-1}) = p(X_k|X_{k-1}, y_k)p(y_k|X_{k-1}) \tag{9}$$

Substituting equation (9) into equation (4), the posterior density takes the following form:

$$p(X_k|y_{0:k}) \propto \int p(X_k|X_{k-1}, y_k)p(y_k|X_{k-1})p(X_{k-1}|y_{0:k-1})dX_{k-1} \tag{10}$$

equation (10) is evaluated in two steps: the second and third term of the integral are evaluated in a correction or smoothing step which uses the measurements y_k to correct the initial conditions X_{k-1}, i.e.,

$$p(X_{k-1}|y_{0:k}) \propto p(y_k|X_{k-1})p(X_{k-1}|y_{0:k-1}) \tag{11}$$

and then a prediction step integrates over X_{k-1} to compute the posterior density:

$$p(X_k|y_{0:k}) \propto \int p(X_k|X_{k-1}, y_k)p(X_{k-1}|y_{0:k})dX_{k-1} \tag{12}$$

where $p(X_k|X_{k-1}, y_k) = \frac{p(y_k|X_k)p(X_k|X_{k-1})}{p(y_k|X_{k-1})}$, which are all known quantities at a given data assimilation step. It should be noted that a filter using the S-path is different from a smoother. While a smoother uses both future and past observations to update the current state, the S-path only uses the most recent observation to characterize the state in real time.

In the context of hydrogeological data assimilation, equation (12) can be further simplified. For cases like subsurface transport predictions where the model prediction errors are dominated by model parameter errors (e.g., soil permeability), the dynamic noise w is comparably small, i.e., the dynamic equation (1) is often replaced by $X_{k+1} = f(X_k, \alpha)$, which is deterministic. Even for cases where the dynamic noise, which often stems from the wrong model formulation, plays a significant role, it is often difficult to quantify and can become another error source if not quantified correctly. In such cases, the dynamic noise is considered unknown a priori and can be incorporated as part of the unknown initial condition, i.e., $X_{k+1} = f(X_k, \alpha, w_k)$. For both cases, the likelihood $p(y_k|X_k)$ equals the likelihood $p(y_k|X_{k-1})$, hence $p(X_k|X_{k-1}, y_k) = \frac{p(y_k|X_k)p(X_k|X_{k-1})}{p(y_k|X_{k-1})} = p(X_k|X_{k-1})$. Similar assumptions are made by *Gharamti et al.* [2015] [see *Gharamti et al.*, 2015, equation (16)].

With these assumptions, equation (12) can be simplified to:

$$p(X_k|y_{0:k}) = \int p(X_k|X_{k-1})p(X_{k-1}|y_{0:k})dX_{k-1} \tag{13}$$

Note that the above derivation is general and does not make any assumptions about the probabilistic distribution. Although the information is processed in a different order in the P and the S-path, both paths give the posterior density $p(X_k|y_{0:k})$ by updating the probabilistic density of state X from time $k-1$ to time k using only data from time k, y_k. Therefore, both paths are designed for real-time estimation. For linear problems and when the probability densities are perfectly known, the two paths give exactly the same posterior distribution (see supporting information Text S1).

2.2. EKF Versus Smoothing-Based EKF

If the probability density functions involved in equations (5–13) can be reasonably approximated by Gaussian densities and if the physical models f and h are linear, then a closed-form analytical solution to the state estimation problem can be derived from the equations presented in the previous section. These sequences, originally derived for linear problems, can be extended to nonlinear problems, by applying linearization techniques. When the information is processed according to the P-path, the derivation results in the extended Kalman filter (EKF) [*Anderson and Moore*, 1979].

Here, using the S-path and applying the simplifications introduced in equation (13) and linear Gaussian assumptions, we derive the smoothing-based EKF (sEKF) (Algorithm 1). Rather than first obtaining a prediction, the sEKF first improves the initial condition $\bar{X}_{k-1|k-1}$ using the new observation (smoothing step, equation (11)) and then integrates the improved initial condition $\bar{X}_{k-1|k}$ to obtain a posterior mean estimate $f(\bar{X}_{k|k})$ (prediction step).

As mentioned in the previous section, sEKF is different than the filter previously proposed by *Desbouvries et al.* [2011] in that it specifically applies to state space models with negligible dynamic noise term w_k compared to the parameter uncertainty, an assumption that is justifiable in a wide range of hydrogeophysical estimation problems. The sEKF algorithm presented in this work, compared to the version derived in *Desbouvries et al.* [2011], does not employ a second correction step after prediction and therefore guarantees the physical consistency of states and parameters while requiring fewer evaluations of the forward model.

Both the EKF (P-path) and the sEKF (S-path) are least square estimators that linearize the nonlinear functions $f(x)$ about the reference point $\bar{X}_{k-1}$ (equation (14)) using Taylor series expansion:

$$x_k = f_k(X_{k-1}) \approx f_k(\bar{X}_{k-1}) + F_k(X_{k-1} - \bar{X}_{k-1}) + \mathcal{O}([X_{k-1} - \bar{X}_{k-1}]^2) \tag{14}$$

in which F_k is the Jacobian matrix $\frac{df_k}{dX_{k-1}}\Big|_{X_{k-1}=\bar{X}_{k-1}}$. These approximations generate second-order linearization errors. The smaller the linearization errors are, the more accurate is the linear Gaussian approximation. The linearization errors may become significant for models that are highly nonlinear and where the uncertainty of X_{k-1} is particularly high. For example, when $f(x)$ is a hyperbolic function, i.e., x exhibits sharp changes, the linear approximation by equation (14) will yield large errors at these fronts. The errors will become more significant when the reference point $\bar{X}_{k-1}$ is away from the truth, e.g., when the parameters or the initial conditions are poorly known.

The major difference between sEKF and EKF in terms of the linearization errors is the reference point used for the linearization. EKF computes the Jacobian F_k about the initial state and parameters ($\bar{X}_{k-1|k-1}$), while sEKF computes the Jacobian about the improved initial conditions and parameters ($\bar{X}_{k-1|k}$) from the smoothing step. By constraining the optimization solution using the information in current measurements one step ahead, sEKF gives model predictions that are both more accurate, and by design physically consistent with the parameters.

Algorithm 1 Smoothing-based EKF (sEKF) algorithm

Smoothing step:	
Forecast state	$y_k^s = h(f(\bar{X}_{k-1\|k-1}))$
Compute Jacobian	$H^* = \frac{\partial h(f(x))}{\partial x}\Big\|_{\bar{X}_{k-1\|k-1}}$
Kalman gain	$K = \Sigma_{k-1\|k-1}(H^*)^T(R + H^*\Sigma_{k-1\|k-1}(H^*)^T)^{-1}$
Update state	$\bar{X}_{k-1\|k} = \bar{X}_{k-1\|k-1} + K(y_k - y_k^s)$
Update covariance	$\Sigma_{k-1\|k} = (I - KH^*)\Sigma_{k-1\|k-1}$
Prediction step:	
Compute Jacobian	$F = \frac{\partial f}{\partial x}\Big\|_{\bar{X}_{k-1\|k}}$
Forecast state	$\bar{X}_k^f = f(\bar{X}_{k-1\|k})$
Forecast covariance	$\Sigma_{k\|k} = F\Sigma_{k\|k}F^T$

2.3. The sCSKF for Joint State and Parameter Estimation

Our motivation in evaluating sEKF is to investigate its utility for monitoring CO_2 injection in heterogeneous domains. In such applications the computational cost of running an expensive reservoir simulator numerous times becomes the computational bottleneck. In this section, we will show how covariance compression and the matrix-free approach adopted by the Compressed State Kalman Filter (CSKF) [*Kitanidis*, 2015; *Li et al.*, 2015] can be used to reduce the computational cost of sEKF (Algorithm 1), specifically by reducing the cost associated with constructing and updating large matrices. We refer the readers to supporting information Text S2 for a detailed description of the CSKF algorithm. The new CSKF variant, termed smoothing based Compressed State Kalman Filter (sCSKF), can handle cases with large number of unknowns, and is designed for joint state-parameter estimation.

To demonstrate the algorithm, we will consider an augmented state vector with two unknown states (e.g., pressure and saturation) and one unknown parameter (e.g., permeability): $X_k = [p_k, s_k, \alpha_k]^T$. The augmented covariance matrix has a block structure, with diagonal blocks representing the covariance of each variable, and the off-diagonal blocks representing the cross covariances:

$$\Sigma_0 = \begin{bmatrix} \Sigma_{pp} & \Sigma_{ps} & \Sigma_{p\alpha} \\ \Sigma_{ps}^T & \Sigma_{ss} & \Sigma_{s\alpha} \\ \Sigma_{p\alpha}^T & \Sigma_{s\alpha}^T & \Sigma_{\alpha\alpha} \end{bmatrix} \tag{15}$$

Factorization of equation (15) can be efficiently done following the methods described by *Li et al*. [2015] such that the computational costs scale linearly with m, the number of unknowns and state dimension by using a low-rank approximation of the state error covariance for each type of variable *i* with a constant basis *A*.

$$\Sigma_{ij} = A_i C_{ij} A_j^T \tag{16}$$

where A_i is a preselected dimensionless orthogonal basis for each type of unknown *i*, and the inner matrix C_{ii} is the compressed covariance for unknown state *i* and C_{ij} the compressed cross covariance for state *i* and *j*. The factorized state error covariance matrix has a block low-rank structure

$$\Sigma = \begin{bmatrix} A_p C_{pp} A_p^T & A_p C_{ps} A_s^T & A_p C_{p\alpha} A_\alpha^T \\ A_s C_{ps}^T A_p^T & A_s C_{ss} A_s^T & A_s C_{s\alpha} A_\alpha^T \\ A_\alpha C_{p\alpha}^T A_p^T & A_\alpha C_{s\alpha}^T A_s^T & A_\alpha C_{\alpha\alpha} A_\alpha^T \end{bmatrix} \tag{17}$$

Then, Σ can be written in the factorized form, i.e., $\Sigma = ACA^T$, where

$$A = \begin{bmatrix} A_p & 0 & 0 \\ 0 & A_s & 0 \\ 0 & 0 & A_\alpha \end{bmatrix} \tag{18}$$

$$C = \begin{bmatrix} C_{pp} & C_{ps} & C_{p\alpha} \\ C_{ps}^T & C_{ss} & C_{s\alpha} \\ C_{p\alpha}^T & C_{s\alpha}^T & C_{\alpha\alpha} \end{bmatrix} \tag{19}$$

Given the low-rank factorization form of the covariance, matrix-vector products of the Jacobian matrix and the column of A, i.e., $AH = A_{:,i}^{H^*}$ and $AF = A_{:,i}^F$ can be calculated efficiently using a matrix-free approach like finite differences. The matrix-vector product A^{H^*} and A^F for the augmented state can be obtained by computing individual parts for each variable first and then assembling them as follows:

$$A^{H^*} = [A_p^{H^*}, A_s^{H^*}, A_\alpha^{H^*}] = \left[\frac{\partial \phi}{\partial p_{k|k}} A_p \quad \frac{\partial \phi}{\partial s_{k|k}} A_s \quad \frac{\partial \phi}{\partial \alpha_{k|k}} A_\alpha \right] \tag{20}$$

$$A^F = [A_p^F, A_s^F, A_\alpha^F] = \left[\frac{\partial f}{\partial p_{k|k+1}} A_p \quad \frac{\partial f}{\partial s_{k|k+1}} A_s \quad \frac{\partial f}{\partial \alpha_{k|k+1}} A_\alpha \right] \tag{21}$$

equations (20) and (21) consist of a total of $N = \text{rank}(A_p) + \text{rank}(A_s) + \text{rank}(A_\alpha)$ calls of forward model $h(\cdot)$ and $2N$ calls of forward model $f(\cdot)$. Replacing the matrix vector products in Algorithm 1 with the compressed state covariance *C* and matrix-vector product A^{H^*} and A^Φ, the sCSKF is derived. Algorithm 2 shows the steps of the sCSKF algorithm, as well as the computational costs for each step. The sCSKF algorithm combines one-step ahead smoothing for a system with no model error and covariance compression. The full covariance Σ is replaced by its compressed counterpart *C*. Compared to the sEKF, sCSKF requires much fewer forward model runs, specifically $N + 1$ evaluations of $h(x)$ and $2\times(N+1)$ evaluations of $f(x)$, where *N* is the effective rank of the covariance and the number of basis functions used (typically $N \sim 100$) is far less than the number of unknowns (e.g., $m \sim 10^6$), i.e., $N \ll m$. Most importantly, the covariance compression is performed in an optimal way with eigenvalue decomposition, and is not affected by sampling errors, as ensemble-based methods are.

Algorithm 2 Smoothing-based CSKF algorithm for negligible dynamic noise

	Smoothing step	Computational cost
Forecast	$y_k^s = h(f(\bar{X}_{k-1\|k-1}))$	1f + 1h
Jacobian matrix product	$A^{H^s}_{:,i} = \frac{h(f(x+\delta\|\|x\|\|A_{:,i}))-h(f(x))}{\delta\|\|x\|\|}\Big\|_{\bar{X}_{k-1\|k-1}}$	$N(h+f)$
Kalman gain	solve for χ from $(R+A^{H^s}C_{k\|k}A^{H^s T})^T\chi = A^{H^s}C^T_{k\|k}$	
	$K = A\chi^T$	$2mnN+n^3$
Update augmented state	$\bar{X}_{k-1\|k} = \bar{X}_{k-1\|k-1} + K(y_k - y_k^s)$	
	$C_{k-1\|k} = (I-\chi^T A^{H^s})C_{k-1\|k-1}$	$mnN+N^2m$
	Prediction step	
Forecast	$\bar{X}_k^p = f(\bar{X}_{k-1\|k})$	$1f$
Jacobian	$A^F_{:,i} = \frac{f(x+\delta\|\|x\|\|A_{:,i})-f(x)}{\delta\|\|x\|\|}\Big\|_{\bar{X}_{k-1\|k}}$	Nf
Forecast state	$\bar{X}_{k\|k} = \bar{X}_k^p$	
Forecast covariance	$C_{k\|k} = A^T A^F C (A^T A^F)^T$	mN^2
Full computational cost	$(2N+2)f+(N+1)h+\mathcal{O}(mN(n+N))$	

3. 1-D Problem: Validation

In this section, we will first evaluate one-step ahead smoothing of sEKF and compare results to EKF for a simple 1-D nonlinear system. The purpose of this rather simplified comparison is to evaluate the impact of one step ahead smoothing for a case small enough that can be run without covariance compression, and for nonlinear models where nonlinearities can be directly controlled. Although previous studies [e.g., *Desbouvries et al.*, 2011; *Gharamti et al.*, 2015] have found one step ahead smoothing to be beneficial in several test cases, it has not been clarified whether this is true for all nonlinear systems or only for those tested, and the final findings were affected by low-rank or sampling errors.

In the following, EKF and sEKF are compared; the only difference between the two methods is the order in which the observation is processed and no other approximations are performed except linearization. Therefore, the differences observed are due to linearization errors in the filtering process. For linear forward models, the two methods are strictly equivalent, as shown in the supporting information Text S1.

We will focus the analysis on a specific type of nonlinear model: one in which the state exhibits hyperbolic behavior and sharp abrupt changes and that has bounded values. One such model is the following [*Kitagawa*, 1991]:

$$x_{k+1} = f_k(x_k, \alpha) = f_k(x_k) + \alpha \tag{22}$$

$$y_k = g(x_k) + v_k \tag{23}$$

with

$$f_k(x_k) = 0.5x_k + 25x_k/(1+x_k^2) + 8\cos(1.2(k+1)) \tag{24}$$

$$g(x_k) = x_k^2/20 \tag{25}$$

α as unknown additive model parameter in $f_k(x_k, \alpha)$. The model parameter α and the observation noise v_k are i.i.d., mutually independent and independent of the initial condition x_0, with $\alpha \sim \mathcal{N}(0.1, 0.5)$, $v_k \sim \mathcal{N}(0, R)$ and $x_0 \sim \mathcal{N}(0.5, 0.5)$. The dynamic model $f(x)$ is characterized by sharp changes near the origin $x = 0$ and linear behavior away from the origin (Figure 1). The state values vary between -20 and 20 over time, forming a bounded oscillating time series (Figure 2). Because of the dramatic change near $x = 0$, the predictive density $p(x_1|x_0)$ is bimodal when x_0 is close to the origin. In simple terms, the model prediction using erroneous parameters will either undershoot or overshoot the sharp front. The bimodality also extends to the filtered density $p(x_1|y_1)$. Such characteristics can be found in hyperbolic-type variables that change sharply near the front. For example, the concentration in advection-dominated groundwater transport or saturation in multiphase flow, also exhibit bimodal behavior even if the initial concentration and the conductivity are generated from a Gaussian distribution. In the following section, we will focus on such a problem.

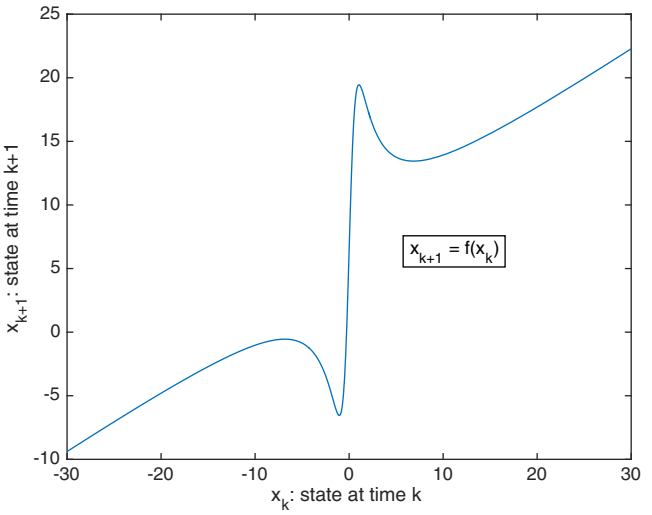

Figure 1. Plot of the state transition from time k to $k+1$ governed by the state transition equation (equation (22)).

The 1-D data assimilation problem was run with EKF and sEKF for joint parameter and state estimation. Figure 2 shows the state estimates, numerical bias, confidence intervals, and parameter estimates given by EKF and sEKF. These results are obtained by averaging of 300 independent simulations with different random initialization of initial condition and observation noise. The averaging step ensures that the difference in the final solution comes from the algorithm and not from the random noise.

Figure 2 shows that sEKF gives state estimates that are closer to the true states than EKF, with less numerical bias especially in the first few steps. While the estimation accuracy in the state variables for EKF improves in the following steps, the parameter estimation becomes less accurate than the estimates by sEKF (bottom row of Figure 2). The inability of EKF to provide accurate states and parameters is likely related to the fact that EKF approximates the predictive and filtered density as unimodal Gaussian while the actual density is bimodal when x is close to 0. In such cases, sEKF performs better than EKF, as sEKF uses the future observation to single out the location of the initial state during the smoothing step.

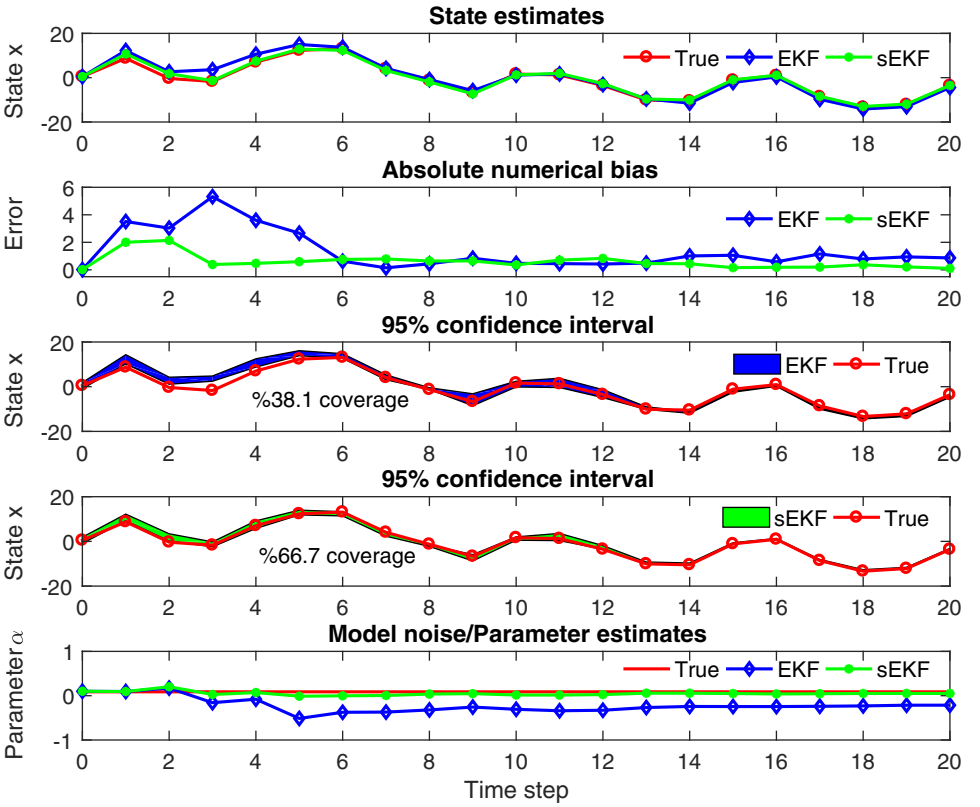

Figure 2. State estimates, absolute numerical bias, 95% confidence interval, and parameter estimates given by EKF and sEKF for the 1-D state-parameter estimation problem.

A good indicator of the reliability of the uncertainty estimates is the percentage of true values that falls into the 95% confidence interval predicted by each method (Figure 2). As shown in Figure 2, the 95% confidence interval given by sEKF is closer to the true state than EKF and captures 66.7% of the true value over the simulated time span, which is more than the 38.1% coverage given by EKF. The fact that neither method can fully capture 95% of the true values is most likely due to linearization errors.

Overall, it can be observed that the sEKF gives state and parameter estimates with less bias and more reliable uncertainty estimates than the EKF in the benchmark evaluated here, where the state variable exhibits hyperbolic behavior and a sharp front. The improvement is more significant near the sharp front where the state changes dramatically. The presence of that front creates a bimodal predictive density which the EKF approximates using a single Gaussian density resulting in overshooting. In contrast, sEKF does not suffer from overshooting as it uses measurements to constrain the predictive density. Interestingly, these findings are not true for parabolic dynamics. The same experiments were run for $f(x)=ax+bx^2$ and it was found that sEKF is not consistently better than the EKF (results not shown). In parabolic systems that vary smoothly, the predictive density is indeed approximately Gaussian, and therefore the EKF does not experience overshooting as pronounced as for hyperbolic models.

4. 2-D CO_2 Monitoring and Characterization

In this section, we evaluate one step ahead smoothing for estimating the CO_2 distribution after injection into a heterogeneous reservoir. The synthetic case has 2025 unknowns. Running the EKF and sEKF for this problem is computationally intractable. Therefore, we employ sCSKF instead, where covariance compression reduces the computational cost significantly. The problem examined is consistent with the assumptions of sCSKF: most of the uncertainty is due to unknown parameters and not by inherent model errors, and CO_2 exhibits sharp fronts as it travels as nonwetting phase in a water-saturated domain.

The sCSKF is compared with its nonsmoothing version (CSKF), as well as EnKF, using a 2-D reservoir monitoring example where the reservoir states are governed by nonlinear multiphase flow and transport equations. Sensor data are assimilated sequentially to reduce the prediction uncertainty in the reservoir simulation due to the unknown model parameters. We choose to benchmark sCSKF against the popular EnKF because (a) both are designed to solve large nonlinear data assimilation problems and is therefore useful to assess their relative performance for the same problem, and (b) to make the argument, via benchmark results, that a compression-based method with smoothing can alleviate the problems related to sampling error and unphysical corrections faced by EnKF. In our analysis, we choose to implement the EnKF version with improved sampling [*Evensen*, 2004] (see supporting information Text S3). The purpose is to provide a fair comparison between EnKF and sCSKF by ensuring that both methods start with the same initial covariance, to avoid tuning parameters for covariance inflation and localization and to avoid the comparison with basic EnKF versions that have known issues. The sCSKF is also compared against a two-step iterative CSKF to evaluate whether iterative improvement of linearization points can achieve the same results given the same computational effort. Information on the iterative CSKF used can be found in supporting information Text S4.

4.1. Experiment Setting

The synthetic CO_2 flooding example is designed following *Li et al.* [2015]. As shown in Figure 3, a 450 m$\times$ 450 m$\times$10 m reservoir is simulated on a 45 $\times$ 45 $\times$ 1 grid with no-flux conditions on the north and south boundary. CO_2 is injected in the 2-D reservoir at a constant rate of 0.01 kg/s through 45 injection wells located on the left boundary, while 45 wells on the right boundary extract water at a constant bottom hole pressure of 206 bar. The true synthetic case is simulated with the permeability field shown in Figure 3 generated as a realization from a distribution with Gaussian covariance function $\Sigma_{\alpha\alpha}=\sigma_\alpha^2\exp\left(-h^2/l^2\right)$ and a correlation length of 200 m. Snapshots of the simulated CO_2 saturation and pressure are also shown in Figure 3. The forward simulation is performed using the parallel version of TOUGH2 [*Pruess*, 1991], a multiphase and multicomponent reservoir simulator, with the module ECO2N [*Pruess and Spycher*, 2007] used to simulate the fluid properties of the brine and the CO_2 systems.

The objective of this numerical benchmark is to monitor CO_2 saturation and pressure with time and simultaneously estimate the unknown permeability of the heterogeneous formation given measurements collected

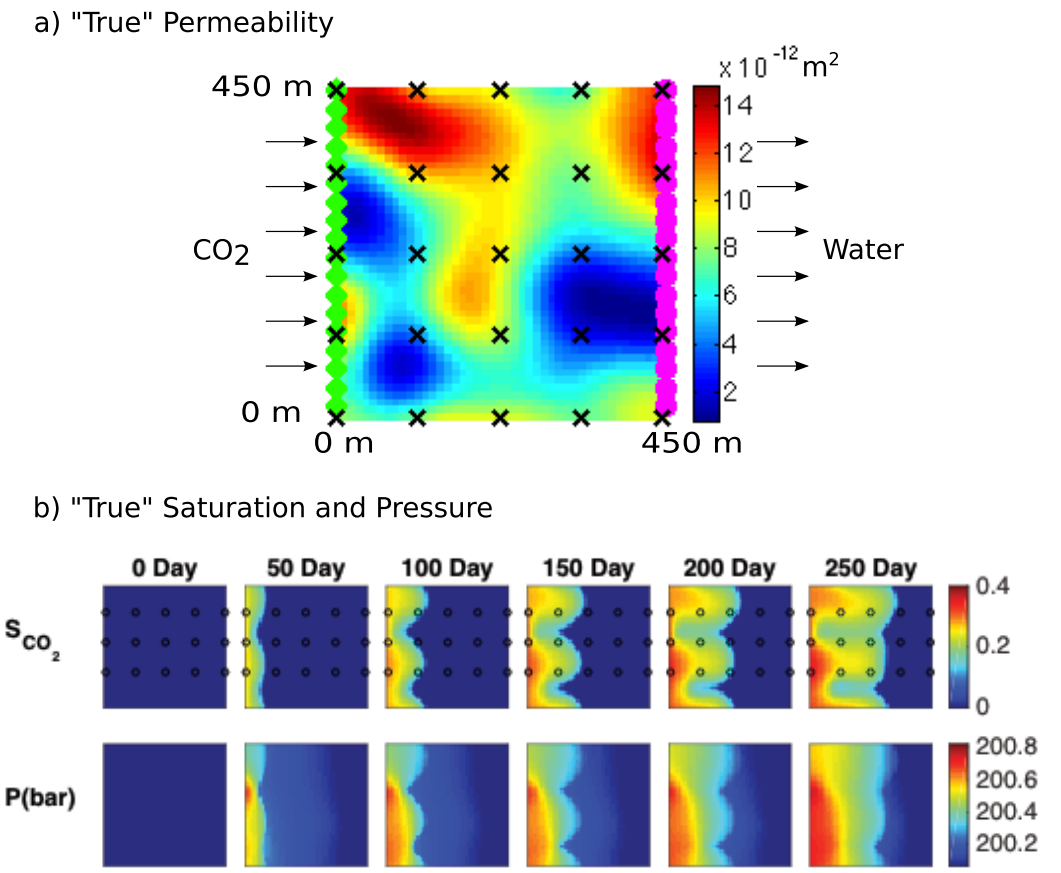

Figure 3. (a) True permeability and CO_2 monitoring experiment setting (the injection wells are marked by green diamonds on the left boundary, the extraction wells are marked by the pink diamonds on the right boundary and the saturation sampling locations are marked by the x markers) (b) True CO_2 pressures and saturations are shown every 50 days until 250 days after CO_2 injection.

every 50 days. The measurements include 45 pressure values at the injection wells, 45 water flux measurements at the extraction wells, and CO_2 saturation measurements sampled at 15 locations uniformly distributed within the domain. The observation noise and parameters used for the forward simulation are summarized in Table 1. To constrain the value of estimated CO_2 saturation s in the range $[0,1]$ and permeability α in $(0,+\infty)$, we truncate the saturation values that are out of range and apply log transformation to permeability before filtering, i.e., $\alpha_T=\log(\alpha), \alpha \in (0,+\infty)$ and back transformation before running the reservoir simulation i.e., $\alpha=\exp(\alpha_T), \alpha_T \in (-\infty,+\infty)$. The initial conditions for pressure and saturation are 200 bar and zero, respectively; they are assumed to be perfectly known, implying that their initial covariances are zero, i.e., $P_{ss}=P_{pp}=0$.

Table 1. Simulation and Data Assimilation Parameters for Reservoir Monitoring Experiment

Forward Simulation Parameters	
Phases	CO_2/brine
Simulation time	5 × 50 days
Grid system	45×45×1
Cell dimensions	10 m×10 m×10 m
Rock porosity	0.2 (constant)
Permeability	Heterogeneous (Figure 3)
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	Zero
Initial pressure	200 bar
Observation error	
Water flux STD	0.008 kg/s
Pressure STD	0.05 bar
Saturation STD	0.01
Data assimilation parameters	
Initial pressure mean	200 bar(constant)
Initial saturation mean	Zero
Initial permeability mean	2 darcy (case A)/ heterogeneous (case B)
Initial permeability variance	0.5 (log transformed)

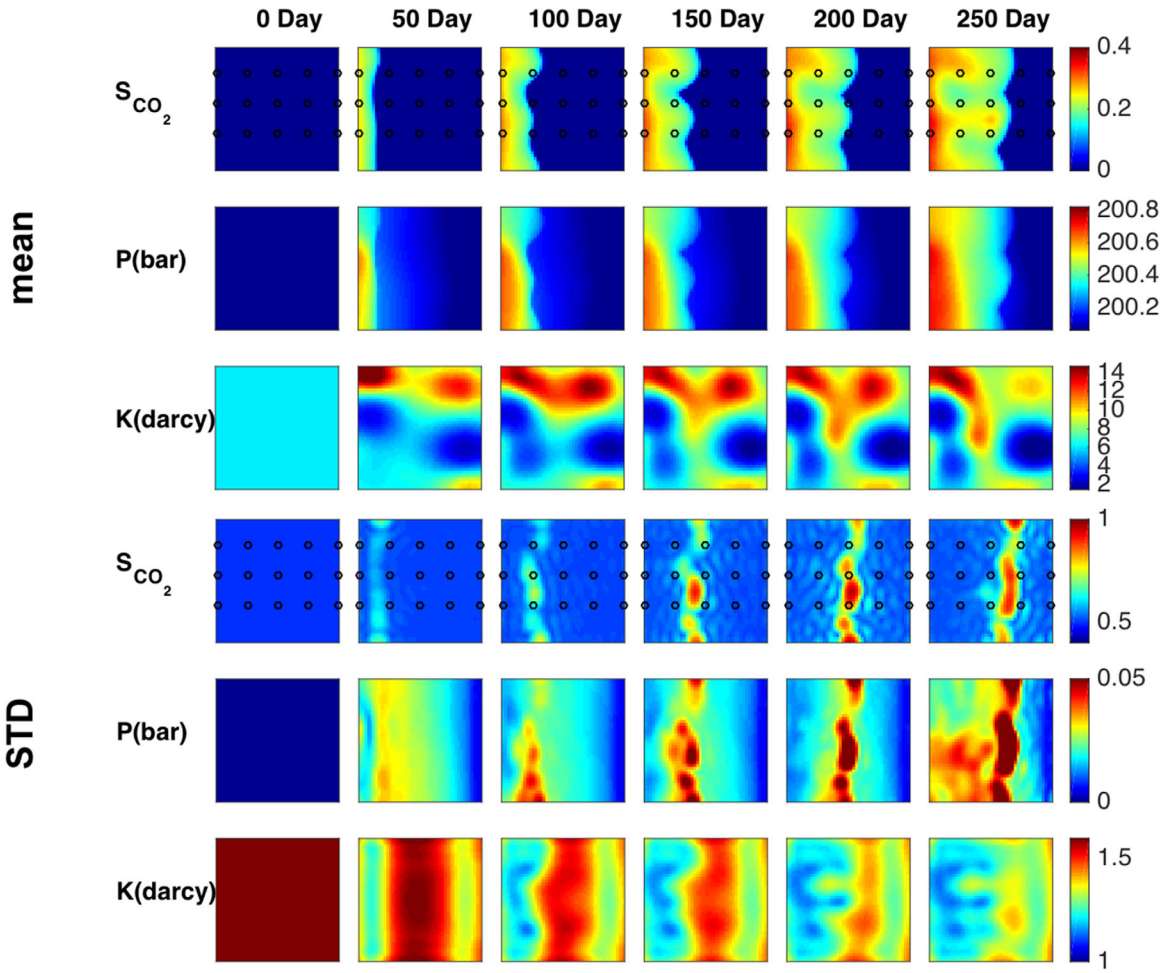

Figure 4. Posterior mean and standard deviations of estimated states and parameters for case A given by sCSKF by assimilating observations every 50th day.

The true covariance function for the permeability field is also known . To achieve computational efficiency while preserving the estimation accuracy, the covariance matrix is compressed using 100 DCT bases, which gives an approximation error of $1.8e-2$ and represents 99% of the variability in the covariance matrix.

Two test cases are designed to evaluate the filter performance under different degrees of parameter uncertainty. Both test cases start with the same initial guess for pressure and saturation. The two cases are only different in the initial guess for permeability; in case A (Figures 4 and 5), the initial guess for the permeability is assumed to be a constant homogeneous field with a value of 2 darcy, and in case B (Figure 6) it is assumed to be a close (but not exact) estimate of the true permeability field. In other words, case B provides a better initial guess for the parameters than case A, and therefore the nonlinearity of the problem is weaker by design compared to case A.

A complete list of the inversion parameters can be found in Table 1.

4.2. Results

For case A, where the data assimilation problem is affected by stronger nonlinearity, the estimated saturation, pressure and permeability estimated by sCSKF every 50 days are shown in Figure 4. The filter is able to accurately resolve the high and low permeability areas, track the CO_2 front, and the accumulation of pressure, using just 100 DCT basis functions per variable ($N = 100$), which corresponds to a compression ratio of $2025/100 \approx 20$. The resolution of the permeability is improved over time as more observations are assimilated. Pressure is resolved at a higher accuracy than CO_2 saturation. This is due to the fact that the pressure evolves parabolically and changes smoothly, such that $N = 100$ provides more than enough bases to reconstruct the smooth field with a small low rank approximation error. In contrast, the saturation displays a

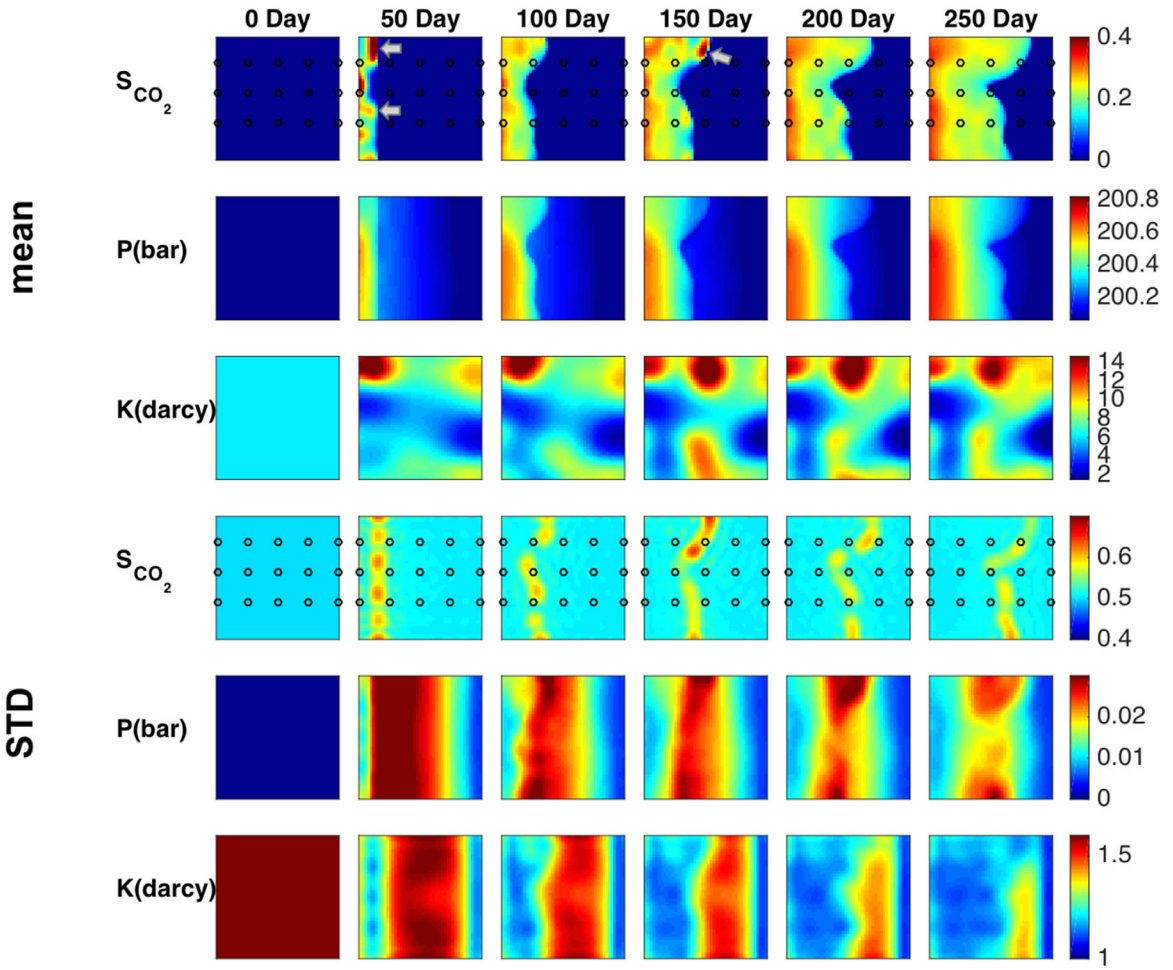

Figure 5. Posterior mean and standard deviations of unknown states and parameters for case A given by CSKF by assimilating observations every 50th day. Overshoots in saturations are marked by arrows. Overshoots due to linearization errors are noted with arrows and can be seen in the saturation posterior mean (first row). Comparison to Figure 4 (sCSKF) shows the beneficial effects of smoothing.

hyperbolic nature with a sharp front moving at different speeds due to permeability heterogeneity, hence requiring more bases to resolve the spatial variability. Although the number of basis functions are chosen to be the same in this test case, sCSKF allows representing each state variable using different number of basis functions. Estimation of nonsmooth variables like the saturation would benefit from including higher-frequency components (larger N), albeit at higher computational cost.

CSKF, the variant without one step ahead smoothing, was applied to the same case A to evaluate the reduction in accuracy when one step ahead smoothing is not used. Figure 5 shows the posterior mean given by CSKF for case A with observations assimilated every 50th day. As the initial guess of the parameter is far from the truth, after assimilating the observation on day 50, it can be observed that CSKF has estimated the saturation field inaccurately, significantly overestimating the saturations near the CO_2 front (pointed by the arrows, first row of Figure 5). This is a manifestation of overshooting that can happen close to sharp interfaces due to inaccurate linear corrections at a location where the forward model exhibits strong nonlinear, hyperbolic behavior. The same phenomenon can be observed in the permeability estimates of CSKF, where the highest permeable zone is inaccurately estimated to be located at the top middle of the reservoir starting at day 150.

While CSKF estimates for saturation and permeability are visibly less accurate than those of sCSKF, for pressure estimates there is no significant difference between the two filters. This contrast between pressure estimates that exhibit parabolic behavior, and saturation estimates, that exhibit hyperbolic behavior, corroborates our previous finding that overshooting and linearization errors mostly affect variables that are subject to dominantly hyperbolic conservation laws.

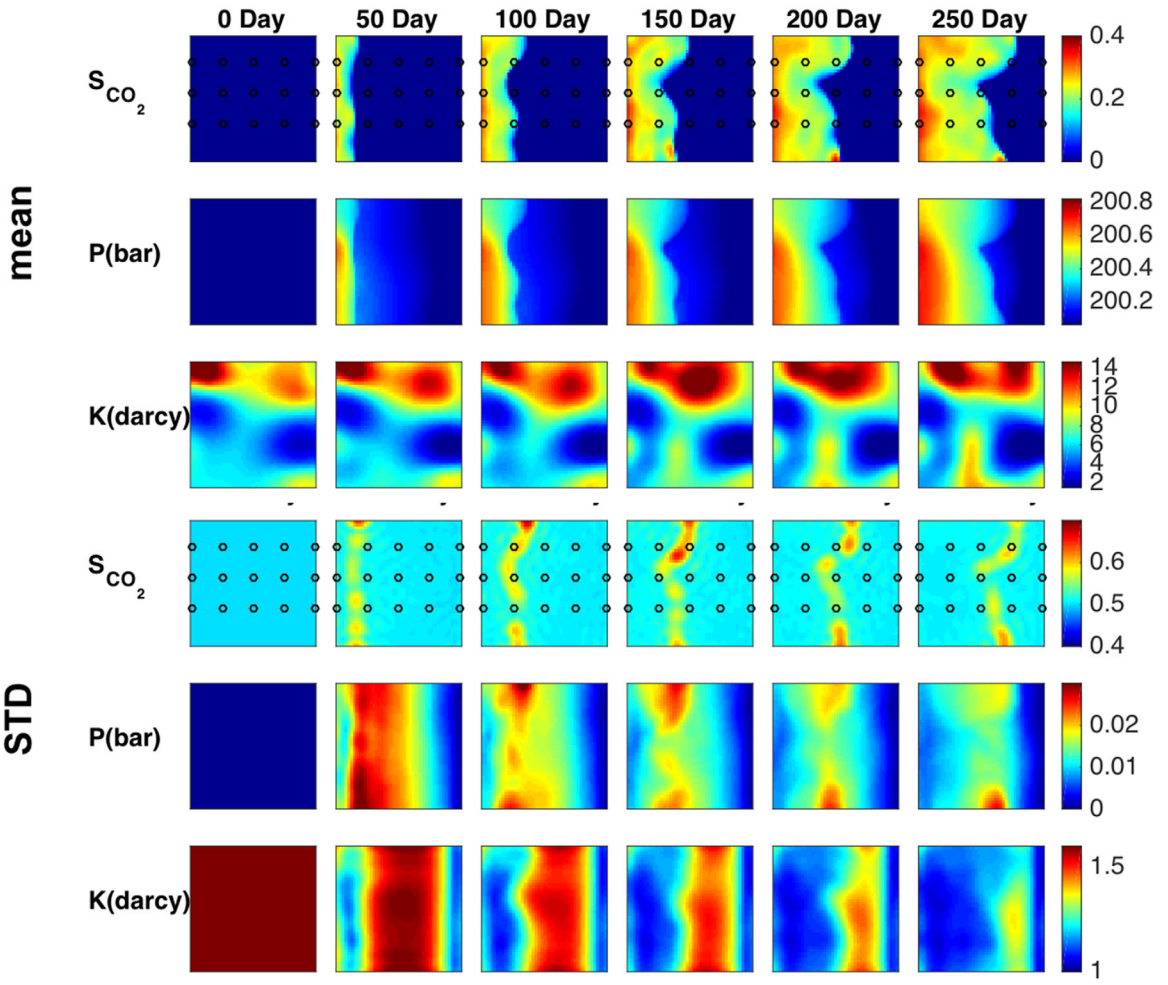

Figure 6. Posterior mean and standard deviations of unknown states and parameters for case B given by CSKF by assimilating observations every 50th day. Comparison to Figure 5 (same experiment with worse initial guess) indicates that overshoots can be mitigated with a better initial guess of unknowns.

Overall, our results indicate that the final estimate of sCSKF is not affected by overshooting, even in the least favorable case where the initial guess of model parameters was far from the truth. One step ahead smoothing improved the initial conditions and model parameters; with this information, the model prediction provided an accurate distribution of the CO_2 saturation. By design, states and parameters are consistent by running the improved physical model as the final step of the estimation. Note that both CSKF and sCSKF used 100 basis functions to represent the variability of the saturation field and both filters matched the observations equally well, hence the overcorrection is not due to insufficient rank or misfit in the observations.

Next, to investigate whether the overshooting observed for CSKF was a result of the bad initial guess of permeability, we conducted the same experiments for case B, for which the initial guess of the permeability field is closer to the true heterogeneity. An initial guess of the unknown that is closer to the truth, makes the linear correction less prone to errors, since assumption of local linearity holds. As shown in Figure 6 for case B, CSKF gives reasonable estimates of the reservoir states, and shows fewer overcorrections in the saturation estimates compared to case A. Also, the uncertainty associated with a good initial guess is smaller in magnitude than the case with a bad initial guess (Figure 5). These results show that the uncertainty in the hyperbolic saturation variable is strongly related to the quality of the model parameters.

The two benchmarks used to compare the filters represent a scenario that is rather difficult for Kalman filtering. Even though the initial conditions and the true permeability field are Gaussian, the true probability density of the state variables as outputs from the nonlinear dynamic equation is highly non-Gaussian. When the model parameters are highly uncertain (case A), the predicted probability density of saturation is very

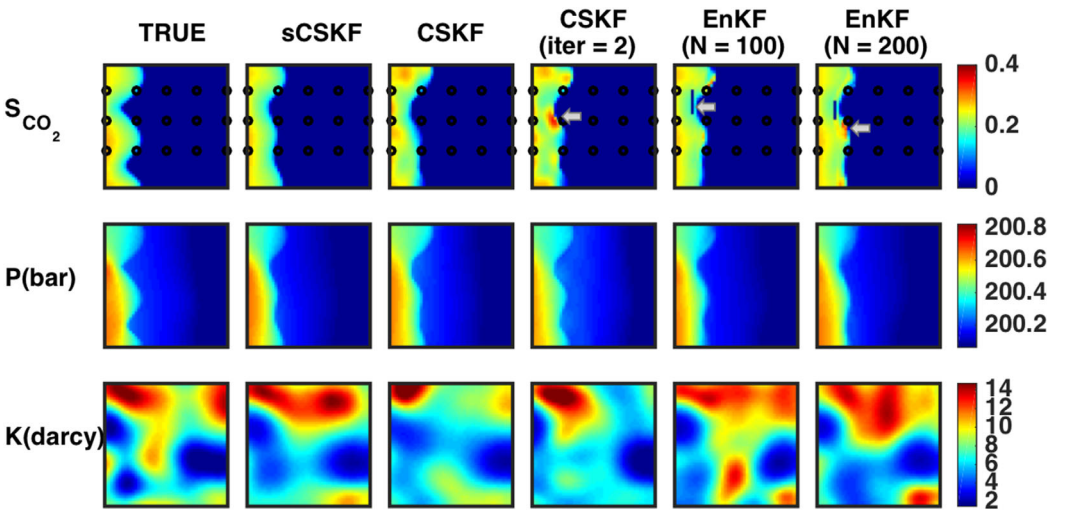

Figure 7. Comparison of posterior mean given by sCSKF, CSKF, iterative CSKF with two iteration and EnKF after assimilating data for 2 time steps. Overshoots in saturations are marked by arrows.

broad and has multiple modes, posing difficulties for the traditional Kalman Filter that linearizes the problem by approximating the density with a unimodal Gaussian. In contrast, the proposed sCSKF is not affected as much by the bad initial guess as the latter is first corrected using data, and then used to constrain the predictive probabilistic density of the saturation. This suggests that one-step ahead smoothing is an effective means of alleviating overcorrections due to nonlinearities and providing physically consistent estimates for systems exhibiting sharp fronts and hyperbolic behavior.

Finally, we perform a comparison of sCSKF with commonly used filters that do not employ one step ahead smoothing. The objective of the comparison is to investigate whether overshooting can be alleviated with methods other than one step ahead smoothing at the same computational cost. Two classes of filters that have been proposed in the literature as being able to better handle nonlinear data assimilation are iterative methods and ensemble-based methods. We compare the sCSKF with an EnKF variant with improved sampling and with an iterative CSKF with two iterations. The computational cost of the compared methods is the same, and we compare the results for overall accuracy and overshooting.

A comparison of sCSKF with CSKF, iterative CSKF and EnKF is shown in Figure 7 for case A where a bad initial guess of model parameters causes strong nonlinearity. Here, we only compare the state estimates after two steps, at day 100. This is because the EnKF became unstable after day 100; unphysical corrections in

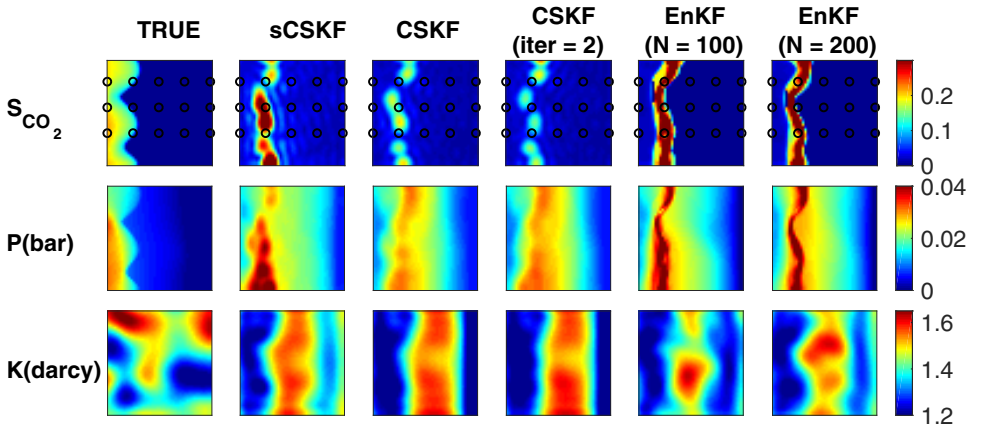

Figure 8. Comparison of posterior standard deviation given by sCSKF, CSKF, iterative CSKF with two iteration and EnKF after assimilating data for 2 time steps.

Table 2. Percentage of True Values of the Reservoir States Captured by the Predicted 95% Confidence Interval for Five Different Methods

Method	Pressure	Saturation	Permeability
Percentage of True Values Captured by 95% Confidence Interval			
Case A: bad initial guess, strong nonlinearity			
sCSKF	97.0%	96.2%	91.9%
CSKF	85.1%	63.7%	81.8%
CSKF (iter = 2)	76.6%	49.0%	76.7%
EnKF (N = 100)	97.5%	95.2%	90.0%
EnKF (N = 200)	97.4%	95.1%	90.5%
Case B: good initial guess, weak nonlinearity			
CSKF (case B)	91.2%	92.0%	85.6%

saturation prevented the model from converging in the next step. It can be observed that all methods give similar estimates of pressure. However, for saturations all filters except sCSKF yielded unphysical corrections (see arrows in Figure 7). It is acknowledged that two iterations are not enough for the iterative CSKF to converge, however more iterations would have resulted in higher computational costs. sCSKF is compared with the EnKF with improved sampling, where $N+1$ ensemble members are generated from the low-dimensional subspace spanned by the first N eigenvectors [see *Li et al*., 2015 for more details]. Interestingly, the EnKF with 100 ensemble members (a typical ensemble size suggested for EnKF) gives nonphysical estimates of saturation and permeability, and increasing the ensemble size from 100 to 200 did not result in significant improvements in the final estimates. This suggests that the overshoots in saturation are not a result of insufficient ensemble size, indicating EnKF may not be able to handle hyperbolic nonlinear dynamics effectively. Moreover, because of the overshoots in saturation, the forward CO_2 simulation failed to converge and the EnKF stopped after two assimilation steps.

Figure 8 shows the uncertainty estimated by sCSKF and the other four filtering approaches. Table 2 shows the percentage of true values that falls into the 95% confidence interval predicted by each method. A low percentage means that the method underestimates the uncertainty, which is commonly observed for low-rank filters like EnKF [*Sætrom and Omre*, 2011]. sCSKF gives robust 95% confidence interval estimates as it includes 97.0% of the saturation, 96.2% of the pressure, and 91.9% of the permeability true values. In contrast, the CSKF for the poor initial guess case (case A) underestimates the uncertainty. With a better initial guess (case B) the CSKF gives more reliable predictions of the confidence interval. The iterative CSKF with two iterations does not improve the quality of the error bounds, as the filter presumably requires more iterations to converge. The error bounds given by EnKF are similar to sCSKF, where the 95% confidence interval captures more than 90% of the true state values. However, EnKF cannot give results for more than two steps as the filter fails to converge due to overshooting in saturations.

To summarize, the CO_2 benchmark used in this work showed that at the same computational costs, the proposed algorithm sCSKF that uses one-step ahead smoothing is more accurate and more robust than other nonlinear filters. States estimated by sCSKF are not affected by overshoots and unphysical corrections contrary to other approaches, even when the initial guess of the parameters is far away from the true values. The improved performance of the sCSKF comes at the cost of one additional forward model run for every run needed for CSKF at each data assimilation cycle, a cost that is comparable to the cost of an EnKF with 200 ensemble members or an iterative approach with two iterations. Based on our 1 and 2-D benchmarks, we expect that the advantages of sCSKF demonstrated in this study also hold for other hyperbolic-type state estimation problems with sharp fronts where most of the uncertainty is caused by unknown parameters.

5. Conclusions

In this work, we presented the smoothing-based compressed state Kalman filter (sCSKF), a computationally efficient Kalman filter for joint state and parameter estimation, which uses one-step-backward smoothing to eliminate unphysical corrections in the states and parameters. By design, the proposed approach produces physically consistent state and parameter estimates. The method is derived following a statistically rigorous Bayesian sequence and is specifically applicable for systems where the model uncertainty is predominantly due to unknown model parameters, a common situation in hydrogeologic applications.

Our analysis indicates that for state variables subject to hyperbolic conservation laws that exhibit sharp fronts, overshoots and nonphysical corrections commonly produced by the traditional filtering sequence are manifestations of linearization errors. Such errors are caused by approximating the true multimodal distribution of the state using a unimodal Gaussian distribution. Hyperbolically varying state models stretch

the limits of the Kalman Filter sequence which assumes local linearity and Gaussianity. For hyperbolic nonlinear models, even if the initial states are Gaussian, the output states after the model forecast have a broad and multimodal density. As a result, the linear Gaussian assumption leads to overshoots in the final estimation, and the additional uncertainty introduced by the model parameters exaggerates the overshoots resulting in nonphysical corrections. The introduction of one-step-ahead smoothing is beneficial because it assimilates data to improve the model parameters and the initial conditions of the state before obtaining a model forecast. In doing that, a better model prediction is produced and physical inconsistencies in the final estimates are eliminated by passing the states through the physical model.

The proposed smoothing-based filter is appropriate for large-scale problems as it uses covariance compression to control the computational cost. With the compression, the improved performance of sCSKF is made possible at a cost equal to twice the effective rank of the unknowns' variability, which is a cost comparable to ensemble-based approaches, and smaller than the cost of iterative approaches. Unlike iterative and ensemble approaches that aim to provide improved solutions to general nonlinear problems, the proposed method focuses specifically on solving the problem of physical inconsistency and overshooting in hyperbolic-type variables characterized by sharp changes in systems where model parameters are unknown, a typical scenario in hydrogeologic applications. Building on previous findings, one step ahead smoothing was indeed found to be beneficial for data assimilation problems with these characteristics, but not as a general rule. This knowledge can help guide future work in Bayesian data assimilation in the geosciences, where sharp fronts are commonly encountered.

Appendix A

We would like to acknowledge the following sources referenced in Supporting Information S1: *Lee and Kitanidis* [2014] and *Nerger et al.* [2005].

Acknowledgments
We would like to thank the two anonymous reviewers and Olaf Cirpka for their valuable feedback and for helping to improve the quality of the manuscript. Funding for this work was provided by "US Department of Energy, National Energy Technology Laboratory" (DOE, NETL) under the awardDE-FE0009260: "An Advanced Joint Inversion System for CO_2 Storage Modeling with Large Data Sets for Characterization and Real-Time Monitoring," and also by the "National Science Foundation"—Division of Mathematical Sciences—under the award 1228275. The data for the this paper can be reproduced using TOUGH2 (http://esd1.lbl.gov/research/projects/tough/) and are available upon request to the corresponding author Judith Li (yuel@alumni.stanford.edu).

References

Anderson, B., and J. Moore (1979), *Optimal Filtering*, Prentice Hall, Englewood Cliffs, N. J.
Bar-Shalom, Y., and X.-R. Li (1993), *Estimation and Tracking: Principles, Techniques, and Software*, Artech House, Norwood, Mass.
Burgers, G., P. van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, *126*, 1719–1724, doi:10.1175/1520-0493(1998)126⟨1719:ASITEK⟩2.0. CO;2.
Desbouvries, F., Y. Petetin, and B. Ait-El-Fquih (2011), Direct, prediction- and smoothing-based Kalman and particle filter algorithms, *Signal Process.*, *91*(8), 2064–2077, doi:10.1016/j.sigpro.2011.03.013.
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. (2004), Sampling strategies and square root analysis schemes for the ENKF, *Ocean Dyn.*, *54*(6), 539–560.
Gharamti, M., B. Ait-El-Fquih, and I. Hoteit (2015), An iterative ensemble Kalman filter with one-step-ahead smoothing for state-parameters estimation of contaminant transport models, *J. Hydrol.*, *527*, 442–457, doi:10.1016/j.jhydrol.2015.05.004.
Gu, Y., and D. S. Oliver (2007), An iterative ensemble Kalman filter for multiphase fluid flow data assimilation, *SPE J.*, *12*(4), 438–446, doi:10.2118/108438-pa.
Hendricks Franssen, H. J., and W. Kinzelbach (2008), Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem, *Water Resour. Res.*, *44*, W09408, doi:10.1029/2007WR006505.
Ho, Y., and R. Lee (1964), A Bayesian approach to problems in stochastic estimation and control, *IEEE Trans. Autom. Control*, *9*(4), 333–339, doi:10.1109/TAC.1964.1105763.
Kitagawa, G. (1991), A nonlinear smoothing method for time series analysis, *Stat. Sin.*, *1*(2), 371–388.
Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, *31*, 2411–2419, doi:10.1029/95WR01945.
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.
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, doi:10.1002/2015WR017203.
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.
Man, J., W. Li, L. Zeng, and L. Wu (2016), Data assimilation for unsaturated flow models with restart adaptive probabilistic collocation based Kalman filter, *Adv. Water Resour.*, *92*, 258–270, doi:10.1016/j.advwatres.2016.03.016.
Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. R. Houser (2005), Dual state–parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, *28*(2), 135–147, doi:10.1016/j.advwatres.2004.09.002.
Naevdal, G., L. Johnsen, S. Aanonsen, and E. Vefring (2003), Reservoir monitoring and continuous model updating using ensemble Kalman filter, SPE-84372-PA, *SPE J.*, *10*(1), 66–74, doi:10.2118/84372-PA.
Nowak, W. (2009), Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator, *Water Resour. Res.*, *45*, W04431, doi:10.1029/2008WR007328.
Pruess, K. (1991), TOUGH2: A general-purpose numerical simulator for multiphase fluid and heat flow, *NASA STI/Recon Tech. Rep.*, Lawrence Berkeley Laboratory Report LBNL-29400, Berkeley, Calif.
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:10.2172/877331.

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.

Shapiro, A. (1996), Estimation of effective porosity in fractured crystalline rock by controlled tracer tests, in Joint US geological survey, U.S. Nuclear Regulatory Commission Workshop on Research Related to Low-level Radioactive Waste Disposal, U.S. Geol. Surv. Water Resour. Invest. Rep., 94–4015, pp. 185–190.

Snodgrass, M. F., and P. K. Kitanidis (1997), A geostatistical approach to contaminant source identification, *Water Resour. Res.*, *33*, 537–546.

Wan, E. A., and A. T. Nelson (2001), Dual extended Kalman filter methods, in *Kalman Filtering and Neural Networks*, edited by S. Haykin, pp. 123–173, John Wiley, New York, doi:10.1002/0471221546.ch5.

Wen, X.-H., and W. H. Chen (2006), Real-time reservoir model updating using ensemble Kalman filter with confirming option, *SPE J.*, *11*(4), 431–442, doi:10.2118/92991-pa.

Wen, X.-H., and W. H. Chen (2007), Some practical issues on real-time reservoir model updating using ensemble Kalman filter, *SPE J.*, *12*(2), 156–166, doi:10.2118/111571-pa.

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, *Groundwater*, *48*(4), 494–514, doi:10.1111/j.1745-6584.2009. 00657.x.