A KALMAN FILTER POWERED BY $\mathcal{H}$-MATRICES FOR QUASI-CONTINUOUS DATA ASSIMILATION PROBLEMS

Y. J. LI* , SIVARAM AMBIKASARAN† , ERIC F. DARVE‡†, AND PETER K.KITANIDIS¶

Abstract. Continuously tracking the movement of a fluid or a plume in the subsurface is a challenge that is often encountered in applications, such as tracking a plume of injected CO_2 or of a hazardous substance. Advances in monitoring techniques empower collecting measurements with a high frequency while the plume moves, which has the potential advantages of providing continuous high-resolution images of fluid flow with the aid of data assimilation. However, the applicability of this approach is limited by the high computational cost associated with having to analyse large data sets within the time constraints imposed by real-time monitoring. Existing data assimilation methods have computational requirements that increase super-linearly with the size of the problem m. In this paper, we present HiKF, a new Kalman filter (KF) variant powered by hierarchical matrix approach, which dramatically reduces the computational and storage cost of the standard KF from $\mathcal{O}(m^2)$ to $\mathcal{O}(m)$, while reproducing its results accurately. HiKF takes advantage of the so-called random walk dynamical model, which is tailored to a class of data assimilation problems in which measurements are collected quasi-continuously. The proposed method has been applied to a realistic CO_2 injection model and compared with the ensemble Kalman filter (EnKF). Numerical results show that HiKF can provide estimates that are more accurate than EnKF, and also demonstrate the usefulness of modelling the system dynamics as a random walk in this context.

Key words. Kalman filter, Hierarchical matrix, Random walk forecast model

AMS subject classifications. 15A15, 15A09, 15A23

1. Introduction. Monitoring the progress of injected fluid fronts in the subsurface is essential to many field operations, such as storing CO_2 underground for greenhouse gas mitigation [8, 16], tracking infiltration [15], injecting steam for enhanced oil recovery [32], and targeting hazardous substances [29]. Time-lapse geophysical monitoring provides a cost-effective and non-invasive approach to image fluid flow in the volumetric region that cannot be sampled by wells. Traditional time-lapse geophysical monitoring uses a large temporal sampling interval. The vast improvements in monitoring technology allow sampling subsurface quasi-continuously [16]. Sampling the subsurface with high frequency means that we can better separate signal (e.g., CO_2 leakage) from noise, and identify important events as early as possible [6]. Therefore, an efficient data assimilation method is desired to efficiently exploit the temporal redundancy resulted from fast data acquisition, which also poses computational challenges associated with analysing large data sets as they arrive to obtain high-resolution fluid flow images in real time.

The Kalman filter (KF) is a powerful statistical tool for processing data as they arrive to continuously improve our knowledge about a dynamical system. The original Kalman filter [30] and its variants have been widely applied in the field of numerical weather forecasting, medical imaging, etc., as well as for geophysical monitoring [34, 38]. Recognizing that multiple solutions are consistent with noisy data, the data conditioning process of KF applies Bayesian inference, which combines prior

*DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING, STANFORD UNIVERSITY, STANFORD, CALIFORNIA, USA.

†INSTITUTE FOR COMPUTATIONAL AND MATHEMATICAL ENGINEERING, JENHSUN HUANG ENGINEERING CENTER, STANFORD UNIVERSITY, STANFORD, CALIFORNIA, USA.

‡DEPARTMENT OF MECHANICAL ENGINEERING, STANFORD UNIVERSITY, STANFORD, CALIFORNIA, USA.

knowledge and measurements to produce an optimal probability density function (PDF) prediction. Unlike the Tikhonov approach that gives a single best solution, the solutions given by KF are a range of possible estimates with uncertainty quantification. However, KF is computationally prohibitive for high dimensional problems because it has to store and operate on a large error covariance matrix. The computational constraints limit the applications of KF to coarse-scale models, which is incapable of capturing the heterogeneity of the subsurface, and loses information on the variations at fine scale. This prevents assimilating additional data in the future to refine the model.

Various KF algorithms have been developed to address the computational challenges. One type of method reduces the effective dimension of model state vector through applying model reduction [21]. Another type of method retains the state space, but reduces the "working space" by projecting the error covariance on a basis with reduced dimension [17, 42, 36]. One such method, the ensemble Kalman filter (EnKF) [17, 12], implements a Monte Carlo-based, low-rank approximation of the full Kalman filter. EnKF requires a large ensemble size to obtain a solution space with small statistical error, especially when a large number of measurements are assimilated [18]. An alternative randomized low rank approximation of KF is given in [36], which expresses the Kalman filter equations in terms of low rank perturbations of its steady state limits. These low rank methods are characterized by the trade-off between the computational cost and the accuracy of approximation, both of which are a function of the effective rank.

For a class of covariance matrices, their matrix-vector products can be computed very accurately using fast linear algebra methods [35, 23, 11], yet with a much lower cost ($\mathcal{O}(m \log^{\alpha} m)$, $\alpha \geq 0$ is a small integer depending on the method) than conventional methods, which require $\mathcal{O}(m^2)$ run time. FFT-based algorithms have been used to efficiently compute matrix-vector products from large-scale kriging problems [35, 20], however, restricting the unknowns to be discretized on regular equispaced grids. Recently, hierarchical matrix [24, 25, 10] that does not require regular and Cartesian grids has been introduced to solve large scale linear geostatistical inverse problems at $\mathcal{O}(m)$ [3, 4]. Our idea extends the work of [3] to accelerate KF. Software packages related to the fast methods used in this article, including the Black Box Fast Multipole Method and Fast Linear Inversion PACKage, can be operated as a black box and are available at https://github.com/sivaramambikasaran/BBFMM2D and https://github.com/sivaramambikasaran/FLIPACK respectively.

In this article, we present a computationally efficient Kalman filter powered by $\mathcal{H}$-matrices (HiKF) that scales linearly with the number of unknowns. Our specific KF employs a random-walk forecast model, which is widely adopted in the medical-imaging literatures to describe system dynamics [43], and more recently, in geophysical monitoring [34, 38] with rapid data acquisition. HiKF relies on two novel ideas. First, it takes advantage of the hierarchical nature of matrices involved to accelerate the computation of dense matrix vector products. Second, the Kalman filtering equations are rewritten in a computationally efficient manner, which enables the use of these fast methods. These two new ideas reduce the computational cost of KF from $\mathcal{O}(m^2)$ to $\mathcal{O}(m)$, while accurately reproducing the linear minimum mean square error (LMMSE) solutions given by the exact KF. For instance, the proposed approach reduces the time of processing 10^5 unknowns on a single core CPU from over 4 hours to within a few minutes, making it feasible for real-time data assimilation.

The paper is organized as follows. The first part introduces the state-space rep-

resentations of the physical system and solutions given by Kalman filtering. Then the derivation of HiKF is presented. In the second part, we demonstrate HiKF on a synthetic crosswell seismic monitoring problem, in which the new algorithm is compared to the standard KF and EnKF in terms of accuracy and computational cost.

2. Theoretical Development.

2.1. Linear state space model with random walk forecast model. Most linear dynamical systems can be examined in the light of estimating a hidden Markov random process. The system state $\mathbf{x}_t$ of interest is a Markov random process governed by the forecast model $f(\mathbf{x}_t|\mathbf{x}_{t-1})$, where the unknown state x_t is a vector of size m at time step t. Markov process indicates that the value of the current state $\mathbf{x}_t$ only depends on previous state $\mathbf{x}_{t-1}$. The measurement $\mathbf{z}_t$ is a vector of size n, and is related to the unknown state through a measurement operator $h(\mathbf{z}_t|\mathbf{x}_t)$. The data assimilation problem is to recover the unobserved quantity $\mathbf{x_t}$ from a sequence of observations $\mathbf{z_1}, \mathbf{z_2}, ..., \mathbf{z_t}$ that are collected at discrete time steps.

We assume that the system dynamics is governed by a linear state space model, which is given as

$$\mathbf{x}_t = F_t\mathbf{x}_{t-1} + \mathbf{w}_t \tag{2.1}$$

$$\mathbf{z}_t = H_t\mathbf{x}_t + \mathbf{v}_t \tag{2.2}$$

State evolution equation (2.1) represents our knowledge about the temporal behavior of system state $\mathbf{x}_t \in \mathbb{R}^{m\times 1}$, where $F_t \in \mathbb{R}^{m\times m}$ is the state transition matrix. The input noise $\mathbf{w}_t \in \mathbb{R}^{m\times 1}$ is the model or input error. The measurement equation (2.2) relates the observations, $\mathbf{z}_t \in \mathbb{R}^{n\times 1}$ to the state vector $\mathbf{x}_t \in \mathbb{R}^{m\times 1}$, through the linear measurement matrix, $H_t \in \mathbb{R}^{n\times m}$. The vector $\mathbf{v}_t$ represents the measurement noise.

In many applications, the state evolution is described by a random walk model [43, 31], in which the transition matrix F is taken to be an identity matrix

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{w}_t \tag{2.3}$$

The random-walk model is useful for practical applications in which data are acquired in rapid succession. Therefore, it is reasonable to assume that changes between subsequent states are small, and the incremental change can be approximated using a white-noise random process $\mathbf{w}_t$. We show later that the solution to the state-space equations can be computed very efficiently when the random-walk forecast model (2.3) is adopted with our fast algorithm.

The noise process $\mathbf{v}_t$ and $\mathbf{w}_t$ are specified as white Gaussian processes with zero mean and known covariances given by

$$\mathbb{E}\left[\mathbf{v}_t\mathbf{v}_t^T\right] = R_t \tag{2.4}$$

$$\mathbb{E}\left[\mathbf{w}_t\mathbf{w}_t^T\right] = Q_t \tag{2.5}$$

The initial condition of the system state usually has no data as constraints. Therefore it is assumed to be random Gaussian variables with known mean μ_0 and known covariance P_0. That is,

$$\mathbb{E}\left[\mathbf{x}_0\right] = \mu_0 \tag{2.6}$$

$$\mathbb{E}\left[(\mathbf{x}_0 - \mu_0)(\mathbf{x}_0 - \mu_0)^T\right] = P_0 \tag{2.7}$$

Given $\mathbf{x}_0$, $\mathbf{v}_t$ and $\mathbf{w}_t$ are independent Gaussian process, the system state $\mathbf{x}_t$ and observation $\mathbf{z}_t$ are also jointly Gaussian.

2.2. Kalman filter. The Kalman filter [30] can be used to compute the linear minimum mean square error (LMMSE) state estimates $\hat{\mathbf{x}}_t$ compatible with the probabilistic model for state evolution (2.1) and measurement process (2.2). The Kalman filter consists of two steps: (i) Predict and (ii) Update. The prediction step predicts the state of the system at the next time step using only information from the current time step. The estimate obtained from the prediction step is termed as the *a priori* state estimate. The update step combines this prediction with the measurements obtained at the next time step, to refine the estimate of the state of the system at the next time step. The estimate obtained from the update step is termed as the *a posteriori* state estimate. The *a posteriori* state estimate can be thought of as a weighted average of the *a priori* state estimate and the estimate of the system obtained from the measurements.

Algorithm 1 Conventional KF algorithm for a random walk forecast model ($m \gg n$)

Prediction:

Operation		Cost
***a priori* state estimate**	$\hat{x}_{t+1\|t} = \hat{x}_{t\|t}$	-
***a priori* covariance estimate**	$\hat{P}_{t+1\|t} = \hat{P}_{t\|t} + Q$	$\mathcal{O}(m^2)$

Correction:

Operation		Cost
Kalman gain	$K_{t+1} = \hat{P}_{t+1\|t} H^T \left(H \hat{P}_{t+1\|t} H^T + R\right)^{-1}$	$\mathcal{O}(nm^2)$
***a posteriori* state estimate**	$\hat{x}_{t+1\|t+1} = \hat{x}_{t+1\|t} + K_{t+1}\left(y_{t+1} - H\hat{x}_{t+1\|t}\right)$	$\mathcal{O}(nm)$
***a posteriori* covariance estimate**	$\hat{P}_{t+1\|t+1} = \hat{P}_{t+1\|t} - K_{t+1} H \hat{P}_{t+1\|t}$	$\mathcal{O}(nm^2)$

The conventional Kalman filtering algorithm with a random walk forecast model is shown in Algorithm 1, which provides solutions to equations from (2.2) to (2.7). In all cases, we assume R, Q and H are stationary matrices, and hence remove the subscript t. Let $\hat{\mathbf{x}}_{t_1|t_2}$ and $\hat{P}_{t_1|t_2}$ denote the estimate of the state, $\hat{\mathbf{x}}_{t_1}$, and the covariance, P_{t_1}, of the system at time step t_1, given the measurements up to the time step t_2. As most data assimilation problems in geosciences are under-determined, i.e., the number of measurements n is much smaller than the number of unknowns m, the computational cost is measured by leading-order terms of m. The major computational loads of the random walk KF come from computing Kalman gain and propagating covariance matrix, which all involves $\mathcal{O}(m^2)$ operations on a matrix of size $m \times m$. Therefore, in its standard form the Kalman filter is computationally prohibitive for high-dimensional modeling domains.

2.3. Ensemble Kalman filter. The ensemble Kalman filter, originally proposed by [17], adopts a Monte Carlo approach to approximate the LMMSE estimates given by the KF when the number of unknowns m is large enough that it is computationally intractable for KF to operate on a error covariance of size $m \times m$. Begin with a small number of independent samples (ensemble) of size N from the initial probability distribution $N(\bar{\mathbf{x}}_0, P_0)$, EnKF propagates and corrects each ensemble member $\tilde{\mathbf{x}}^j$ using the same procedures as KF. The sample mean $\widetilde{\mathbf{x}}$ and sample covariance $\widetilde{\Sigma}$ are used to approximate the mean $\hat{\mathbf{x}}$ and covariance Σ given by the Kalman filter in section 2.2, which are defined as

$$\hat{\mathbf{x}} \approx \tilde{\mathbf{x}} = \frac{1}{N}\sum_{j=1}^{N} \tilde{\mathbf{x}}^j \tag{2.8}$$

$$\Sigma \approx \tilde{\Sigma} = \frac{1}{N-1}\sum_{j=1}^{N} (\tilde{\mathbf{x}}^j - \tilde{\mathbf{x}})(\tilde{\mathbf{x}}^j - \tilde{\mathbf{x}})^T \tag{2.9}$$

The sample Kalman gain is computed by

$$\widetilde{K}_t = \widetilde{\Sigma}_{t|t-1} H^T (H \widetilde{\Sigma}_{t|t-1} H^T + R)^{-1} \tag{2.10}$$

The sample covariance $\widetilde{\Sigma}$ is never constructed explicitly during implementation, which allows EnKF and other ensemble square root filters [41] to circumvent the computational bottleneck of Kalman filter and apply to high-dimensional systems. In this article we will compare our methods against the perturbation-based EnKF described in [27], the solution of which will converge to the LMMSE estimates given a large enough ensemble size N [12]. In practice, *ad hoc* techniques, e.g., localization [26, 28], covariance inflation [5] are used to stabilize the filter performance. However, their effectiveness in one type of problems may not be applied in others. Comparison with other versions of EnKF, e.g., ensemble adjustment Kalman filter [5], ensemble transform Kalman filter [9] are left for future investigations.

2.4. Hierarchical Matrices. The $\mathcal{H}$-matrix approach provides a data-sparse representation of dense matrices arising in many engineering applications, e.g., the boundary element method (BEM) for discretizing integral equations [7], dense covariance matrices [3], etc. The data-sparse representation relies on the fact that these matrices can be recursively sub-divided based on a tree structure, and most matrix sub-blocks at different levels in the tree can be well-approximated using a low-rank block. There are many different hierarchical matrices depending on the tree structure and algorithms to obtain the low-rank blocks. For these hierarchical matrices, dense linear algebra operations, like matrix-vector products, matrix factorizations, and solving linear equations [2, 33, 14, 13], can be performed *in almost linear complexity.* In this article, we will restrict our attention to a class of hierarchical matrices known as $\mathcal{H}^2$ matrices, which efficiently represent covariance matrices arising out of analytic kernels [3]. $\mathcal{H}^2$-matrices are closely related to the fast multiple method [23, 19], which reduces the complexity of storage and matrix-vector products to linear complexity.

One of the ingredients of the hierarchical matrix approach is to have fast low-rank factorization algorithms. There are quite a few low-rank factorization techniques available. For instance, if the entries of the matrix arise from a Green's function or some

analytic function (in some suitable region), the low-rank approximation can be computed using the multipole expansion, as in the traditional fast multipole method [23], or using other analytic techniques such as Taylor series expansions [10] or Chebyshev interpolation [19]. However, if the entries in the matrix are only known algebraically, then fast algebraic techniques, including adaptive cross approximation (ACA) [39] and pseudo-skeletal approximations [22], can be used to obtain low-rank factorizations in almost linear complexity. In the current article, we relied on Chebyshev interpolation [19, 2] to construct fast low-rank factorizations.

2.5. Hierarchical Kalman filter. In this subsection, we introduce a computationally efficient Kalman filter algorithm powered by $\mathcal{H}$-matrices (HiKF). The computational tractability is achieved by taking advantage of a couple of key observations from Algorithm 1.

The first observation is that the initial covariance matrix can be well represented as a $\mathcal{H}^2$-matrix. This enables us to compute the matrix product QH^T in $\mathcal{O}(nm)$ as opposed to standard $\mathcal{O}(nm^2)$, thereby reducing the computational cost of the first step in Algorithm 1. We refer the readers to [3] for more details on how to accelerate computing QH^T using the $\mathcal{H}^2$-matrix approach. It is also worth noting that [3] also exploits the sparsity of the measurement operator H in reducing the run time. The second key observation that enables us to reduce the computational cost of the subsequent steps in the Kalman filtering algorithm is the fact that it is enough to store and compute the cross-covariance matrices, i.e., $\hat{P}_{t+1|t}H^T$ and $\hat{P}_{t+1|t+1}H^T$ rather than storing and computing with the entire covariance matrices $\hat{P}_{t+1|t}$ and $\hat{P}_{t+1|t+1}$. The proposed algorithm is given in Algorithm 2. Let $\hat{P}'$ denote $\hat{P}H^T$ and Q' denote QH^T.

Algorithm 2 HiKF algorithm for a random walk forecast model ($m \gg n$)

Precomputation:

Operation		Cost
Cross-covariance matrix (the $\mathcal{H}^2$-matrix approach)	$Q' = QH^T$	$\mathcal{O}(nm)$

Prediction:

Operation		Cost
***a priori* state**	$\hat{s}_{t+1\|t} = \hat{s}_{t\|t}$	-
***a priori* cross-covariance**	$\hat{P}'_{t+1\|t} = \hat{P}'_{t\|t} + Q'$	$\mathcal{O}(nm)$

Correction:

Operation		Cost
Kalman gain	$K_{t+1} = \hat{P}'_{t+1\|t}\left(H\hat{P}'_{t+1\|t} + R\right)^{-1}$	$\mathcal{O}(n^2m)$
***a posteriori* state**	$\hat{s}_{t+1\|t+1} = \hat{s}_{t+1\|t} + K_{t+1}\left(y_{t+1} - H\hat{s}_{t+1\|t}\right)$	$\mathcal{O}(nm)$
***a posteriori* cross-covariance**	$\hat{P}'_{t+1\|t+1} = \hat{P}'_{t+1\|t} - K_{t+1}H\hat{P}'_{t+1\|t}$	$\mathcal{O}(n^2m)$
***a posteriori* variance**	$\delta^2_{t+1\|t+1} = \delta^2_{t+1\|t} - \sum_{j=1}^{n}(K_{t+1})_{ij}(\hat{P}'_{t+1\|t})_{ij}$	$\mathcal{O}(nm)$

Therefore, the overall cost of the Kalman filtering algorithm for a random walk forecast model is $\mathcal{O}(n^2m)$ as opposed to $\mathcal{O}(nm^2)$. Recall that for most problems $n \ll m$. Therefore, the solutions can be obtained in linear run time.

3. Numerical Benchmark. In this section, we illustrate the performance of our proposed HiKF by comparing it against the conventional KF and EnKF. The

data assimilation methods are implemented to continuously track CO_2 plume in the subsurface from seismic travel times, resulting in a dynamical system described using a linear state space model with random walk forecast model.

3.1. Synthetic TOUGH2 CO_2 Monitoring Experiment. A detailed reservoir flow model has been developed for the Frio-II brine pilot CO_2 injection experiment [16] using a state-of-the-art model reservoir simulator TOUGH2 [37]. The pilot injected 380 tons of CO_2 into a 17m-thick deep brine aquifer at the depth of 1657m. The brine aquifer is assumed to have a dip of 18 degree and a layered permeability and porosity distribution, based on well logs and core measurements. The flow simulation predicted the spatial distributions of CO_2 saturation and pressure for 5 days. The seismic wave velocity model is built from the flow model using a patchy petrophysical relationship [45]. The baseline velocities and time-lapse velocity reductions of compressional seismic waves are shown in Figure 3.1. CO_2 can be monitored seismically by mapping the time-varying CO_2-induced velocity reductions from measurements of travel-time delays [44]. We conduct synthetic cross well seismic surveys every 3 hours, with 6 sources deployed at the injection well and 48 receivers deployed at the observation well 30 meters apart. The acquisition geometry follows [1] and remains fixed during the monitoring experiment.

The travel-times are simulated by integrating the seismic slowness (reciprocal of velocities) along the ray path on which seismic waves propagate. In reality, ray path will change in response to CO_2 plume migration. For simplicity, we assume that the ray path follows a straight line connecting the source-receiver pair, which is the high-frequency limiting case of the wave equation. By assuming the ray path l has not been significantly altered due to CO_2 injection, the travel-time $\mathbf{y}_t$ can be expressed as a linear function of the unknown slowness $\mathbf{s}_t$, which varies with location $\mathbf{r}$. That is,

$$\mathbf{y}_t = \int_l \mathbf{s}(\mathbf{r}) dl = H\mathbf{s}_t \tag{3.1}$$

The time-invariant linear observation operator H embeds the information about the area sampled by the ray path, with each element H_{ij} represents the length of the i^{th} ray path within the j^{th} grid. CO_2-induced low velocity zone is imaged from travel-time delays $\mathbf{\Delta y_t}$ relative to the baseline travel-time, which can be obtained by subtracting the baseline travel-time from the measurement equation (3.1). That is,

$$\Delta\mathbf{y}_t = H\Delta\mathbf{s}_t \tag{3.2}$$

The variable of interest $\Delta\mathbf{s}_t$ is the perturbation of the background slowness at time step t. The differential tomography approach applies spatial and temporal regularizations directly on the slowness perturbations instead of slowness itself. The observation $\mathbf{z}_t$ in equation (2.2) is simulated as travel-time delay $\Delta\mathbf{y}_t$ contaminated with white noise $\mathbf{v}_t$, resulting in a 65dB signal-to-noise ratio (SNR) defined as

$$SNR = 10\log_{10}\frac{\|\Delta\mathbf{y}\|_2^2}{\|\sigma\|_2^2} \tag{3.3}$$

where $\Delta\mathbf{y}$ is the measurement signal, and $\sigma^{\mathbf{2}}$ is the measurement variance. The measurement noise $\mathbf{v}$ are realizations from $N(0, \sigma^2 I)$.

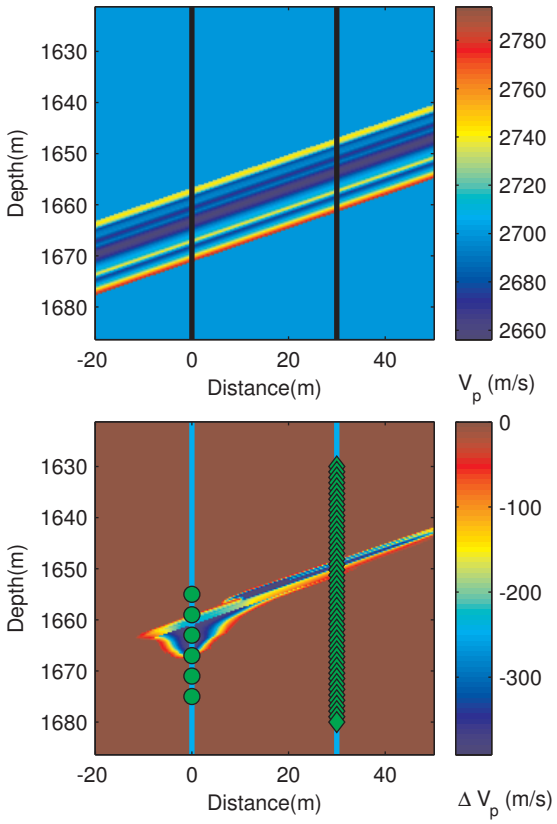

FIG. 3.1. *The base velocity model (top) before CO_2 injection and the simulated CO_2-induced velocity reduction (bottom) 5 days after injection at FRIO II site. Data provided courtesy of Jonathan B. Ajo-Franklin, Thomas M. Daley and Christine Doughty from Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States.*

3.2. State-space model parameter selection. The state-space representations of the dynamic travel-time tomography problem is given in equation (2.2) to (2.7). The major error sources in the mathematical model come from using the random-walk model (2.3) to approximate the true complex system dynamics, as well as using noisy data sets. If the data assimilation method is based on Monte Carlo, e.g. EnKF, then sampling errors become another significant error source. The parameterization of the noise structures incorporates our prior knowledge of the dynamical process, which is given by

$$R = \sigma^2 I \tag{3.4}$$

$$Q = \theta \exp(-\frac{h^p}{l^p}) \tag{3.5}$$

The observation noise $\mathbf{v}_t \sim N(0, R)$ are assumed to be uncorrelated, where $I \in \mathbb{R}^{n\times n}$ is an identity matrix, and σ^2 is the observation variance. The noise process

$\mathbf{w}_t \sim N(0, Q)$ accounts for the lag errors due to using a random-walk forecast model. The model error covariance Q is parameterized as an exponential kernel function, which possesses a hierarchical structure described in section 2.4. Figure 3.2 shows how the correlation decays as the spatial distance h increases for various power p and length scale l. For example, small p indicates sharp changes and a long-range correlation. The lag errors can be estimated approximately by selecting reasonable values for p and l. The regularization parameter θ and σ controls the relative influence of the lag errors and the observational errors. If the slowness changes are significant, then the lag errors can be assumed to prevail, therefore more weight should be put on measurements. Although Q is a large dense matrix, it can be efficiently represented in a data-sparse manner as a hierarchical matrix, more specifically as a $\mathcal{H}^2$ matrix [3, 40]. Therefore, the product QH^T can be computed at $\mathcal{O}(mn)$, where $m \gg n$.

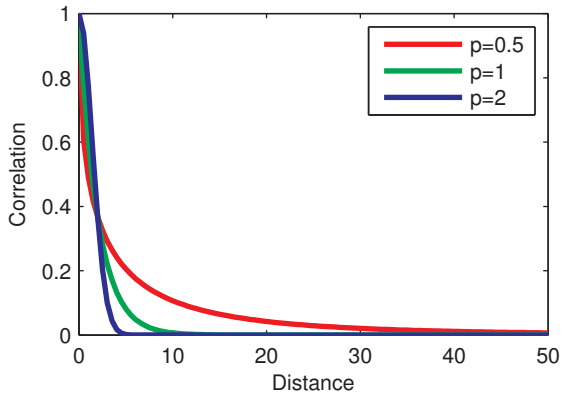

FIG. 3.2. *Covariance kernel specified in equation* (3.5) *plotted against separation h with different p values. Other parameters are given by* $l = 2, \theta = 1$.

As we assume no CO_2 is present before the injection, the initial guess for the slowness perturbation $\mathbf{x}_0$ is a zero vector, with zero initial error covariance P_0. The effects of the choice of P_0 will die out relatively fast as more data has been assimilated [43].

3.3. Dynamic inversion results. The slowness changes induced by CO_2 injection have been estimated from simulated travel-times acquired at 41 time frames, using KF, HiKF and EnKF described in previous sections. All prediction filters adopt the same random-walk state space equations from (2.2) to (2.7) to describe the dynamical system, with same noise structures and regularization parameters. Therefore the difference in the estimates results solely from the choice of filter. Methods are compared in terms of accuracy, uncertainty quantification, and computational cost. CO_2-induced slowness changes are estimated at three different resolutions, with 59×55, 117×109, 234×217 pixels respectively. The accuracy of the inversion results is measured in terms of relative estimation error $\mathbf{e}_t$ defined as

$$\mathbf{e}_t = \frac{\|\mathbf{x}_{\text{est}} - \mathbf{x}_{\text{ref}}\|_2}{\|\mathbf{x}_{\text{ref}}\|_2} \tag{3.6}$$

The reference value $\mathbf{x}_{\text{ref}}$ for measuring estimation errors is the true solution $\mathbf{x}_{\text{true}}$ or the LMMSE solution $\mathbf{x}_{\text{KF}}$ given by KF depending on the context. The true image

as well as the reconstructed image of synthetic CO_2 plume are shown in Figure 3.3 at low resolution (59×55). The HiKF solution is also shown in Figure 3.4 at high resolution (234×217). The estimation errors relative to the true image as well as to the LMMSE estimate are plotted as a function of time in Figure 3.5. As shown from the figures, HiKF reproduces the LMMSE estimates given by KF very accurately while EnKF produces noisy estimates even with an ensemble size of 600. The predictions become increasingly accurate over time as more data have been assimilated. Notice that after the CO_2 breakthrough (2 days after injection), the volume of CO_2 between the wells ceases to change. The cross well seismic survey becomes less informative, therefore yielding less improvements in estimation errors.

Figure 3.6 shows the estimated variance plotted as a function of location at Day 5 after CO_2 injection. The area with more ray coverage has less uncertainty comparing to the remainder of the picture. As shown in the right figure, variance estimated by EnKF has a large variability due to sampling errors. It is expected that with larger sample size, the variance will approach to the value given by KF, however, at the expense of greatly increasing computational burdens.

The computational cost of performing a single measurement update are summarized in Table 3.1 for each data assimilation method and plotted against the number of unknowns in Figure 3.7 for comparison. All computations are performed using C++ on a PC with 2.40 GHz single-core CPU. Compared to conventional KF, HiKF reduces the time of processing 10^5 unknowns from 4 hours to around 2 minutes, making it a promising tool for fast online data assimilation. As shown from Figure 3.7, the computational time and storage cost of KF grows quadratically with number of unknowns, which involves propagating and storing a full error covariance of size $m \times m$. In comparison to KF, the overall cost of HiKF is $\mathcal{O}(m)$ because it operates on a cross covariance of size $m \times n$. The cost of HiKF comprises offline and online parts. The offline expense measures the cost of forming QH^T from a dense covariance matrix Q, which needs to be computed only once in each monitoring event before any data is processed. The cost is normally $\mathcal{O}(m^2)$, however, can be reduced to $\mathcal{O}(m)$ by efficiently representing Q as a $\mathcal{H}^2$ matrix [3]. The online portion only consists forming matrix-vector product associated with cross covariance, thus both the update and storage cost is $\mathcal{O}(m)$. The computational and storage cost of EnKF also scale linearly with the number of unknowns, as it propagates errors using an ensemble consists of N realizations of state vectors of size $m \times 1$ instead of a large covariance matrix. However, EnKF is computationally more expensive than HiKF in our case, as a very large ensemble size is required to produce accurate solutions.

TABLE 3.1
Comparison of time taken on a 2.40 GHz single-core CPU for our proposed computationally efficient Kalman filter (HiKF) with other Kalman filters

Grid size	**Time taken (minutes)**			**Storage Cost (MB)**		
	KF	EnKF	HiKF	KF	EnKF	HiKF
59×55	1.18	1.05	0.14	84.2	15.0	7.5
117×109	19.8	4.10	0.57	1331.2	57.9	29.4
234×217	263.7	16.3	2.25	21125.1	230.5	117
Complexity	$\mathcal{O}(m^2)$	$\mathcal{O}(m)$	$\mathcal{O}(m)$	$\mathcal{O}(m^2)$	$\mathcal{O}(m)$	$\mathcal{O}(m)$

We further investigate about why EnKF does not perform well in our example by plotting the eigenvalue spectrum of posterior covariance at last time step in Figure 3.8.

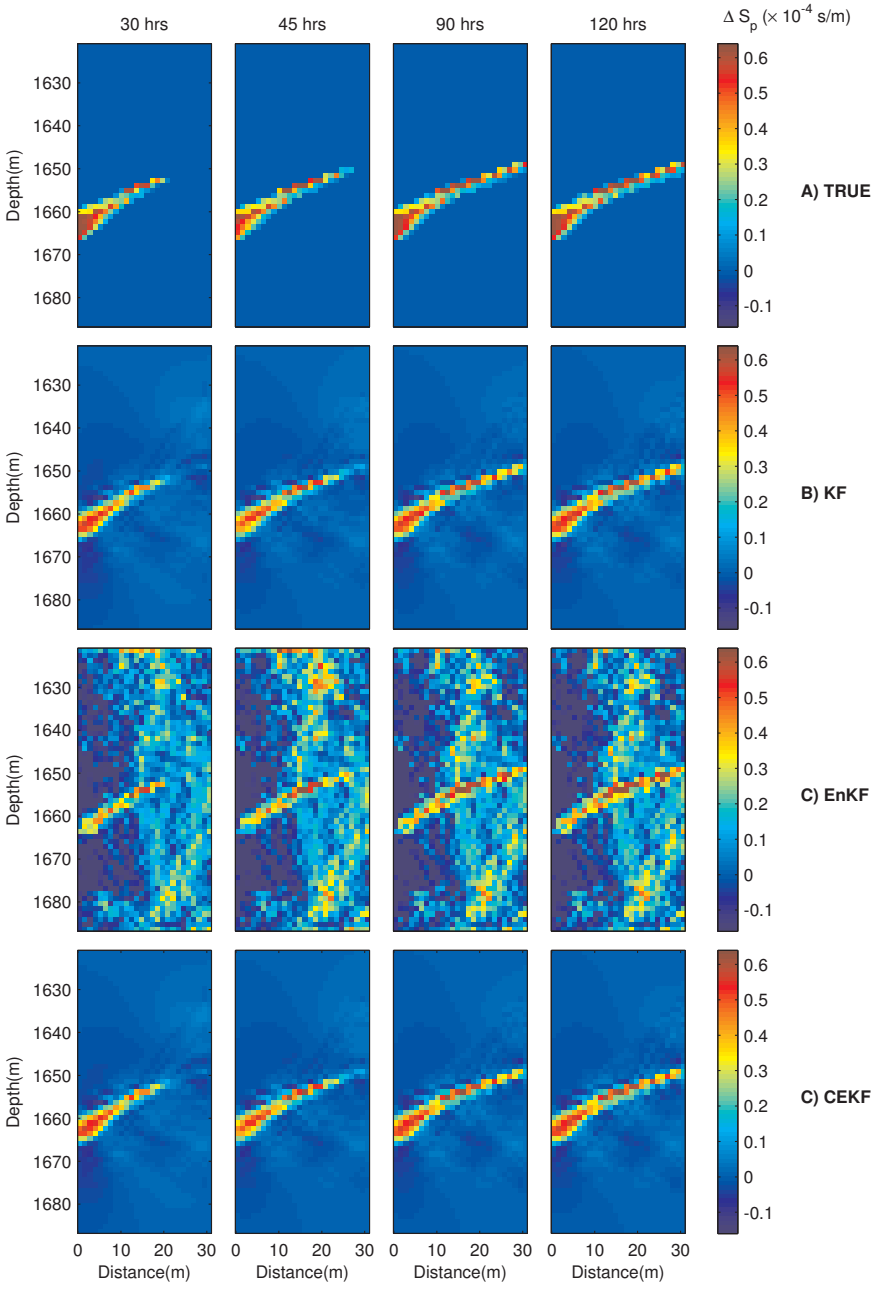

FIG. 3.3. *True and estimated CO_2-induced changes in slowness (reciprocal of velocity) between two wells at low resolution. EnKF results are obtained with 600 samples averaged over 50 trials. Each state vector* $\mathbf{x}$ *represents a* 59×55 *image, each observation vector* $\mathbf{z}$ *contains 288 simulated tomographic travel time measurements with 65-dB SNR at each time instance. SNR is defined in* (3.3)*. Data is acquired at 41 time instances, lasting for 5 days.*

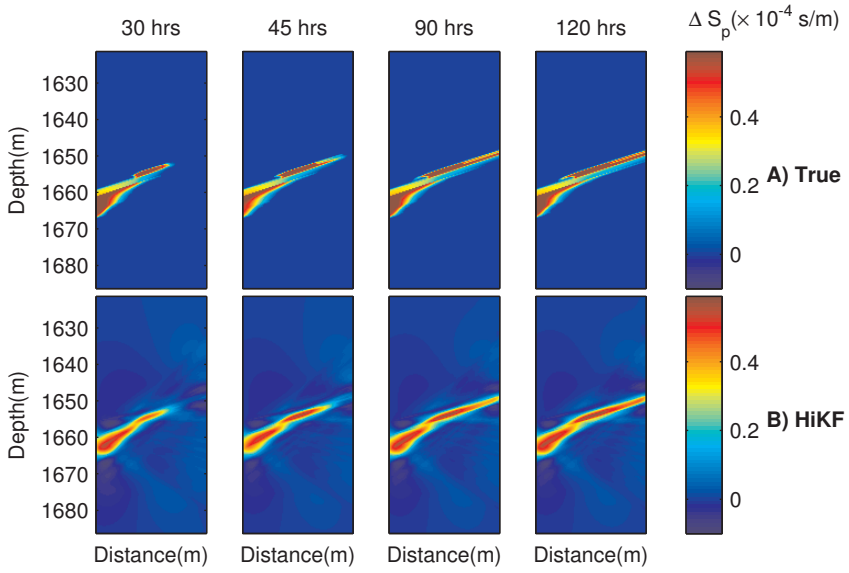

FIG. 3.4. *True slowness change and results given by HiKF at high resolution. Each state vector* $\mathbf{x}$ *represents a* 234×217 *image.*

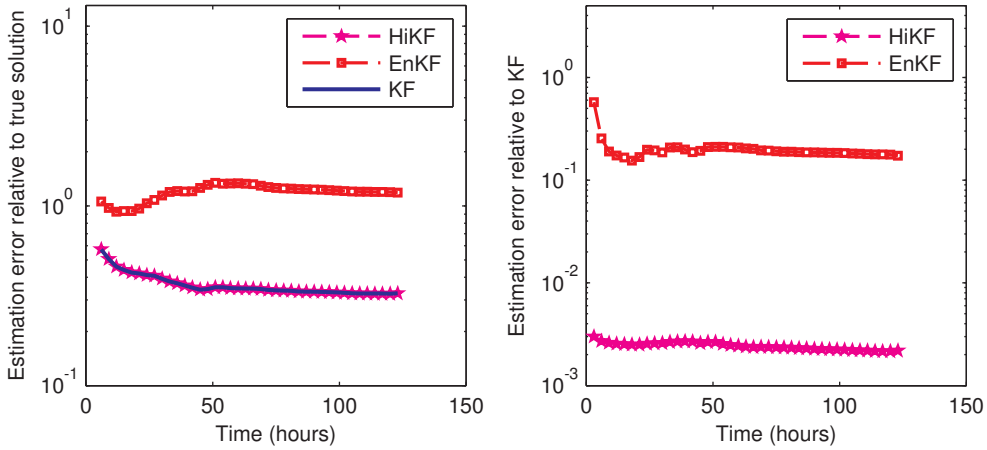

FIG. 3.5. *Estimation error relative to the true solution (left) and the LMMSE solution by KF (right). Definition is given in equation* (3.6).

In comparison with the eigenvalue spectrum given by KF, which is flat and has a long tail, EnKF yields a steep eigenvalue spectrum with excess variance associated with the leading eigenvectors and insufficient variance with the rest. The same observation is reported in [26]. By increasing the ensemble size from 400 to 600, the variance in the tails of the spectrum increases yet the magnitude is still much lower than KF. Figure 3.8 also shows the effective rank of the posterior covariance for each assimilation time step. The effective rank is defined as the least number of principle eigenvalues needed to explain 95% of the total variance. The plot suggests that EnKF is suffered from rank deficiency associated with using insufficient ensemble

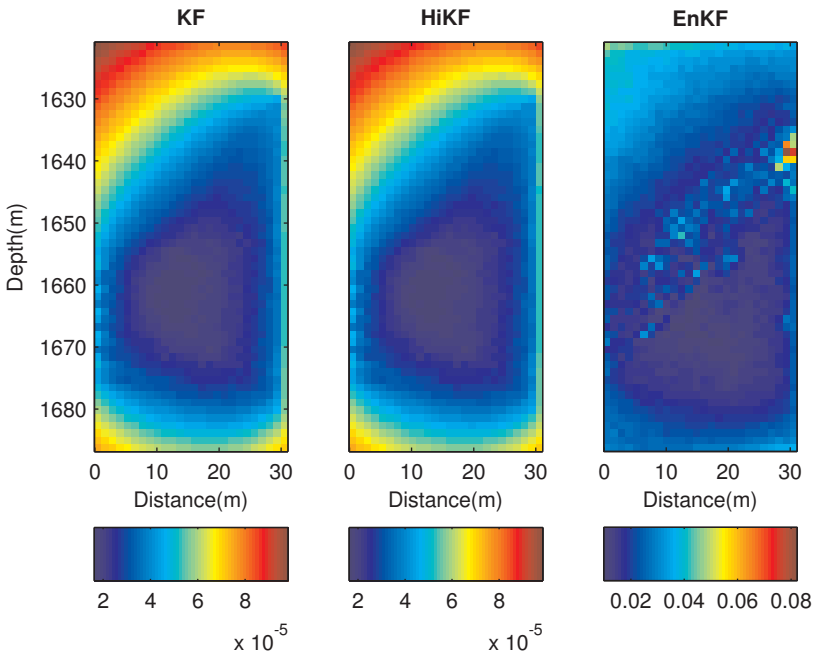

FIG. 3.6. *Gridblock variance estimated by KF, HiKF and EnKF at last assimilation time step corresponding to Figure 3.3.*

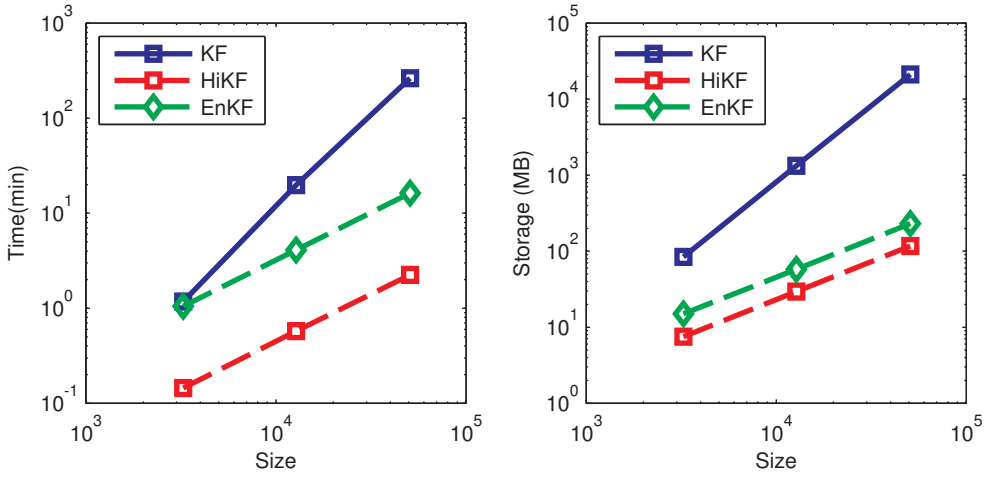

FIG. 3.7. *Computational time (minutes) and storage cost (MB) of each data assimilation method plotted as a function of problem size.*

size. Overall, the results show that the information embedded in the covariance of KF cannot be adequately represented by the proposed ensemble size. Therefore, for our particular case EnKF cannot give satisfactory solutions with an ensemble size as large as 600.

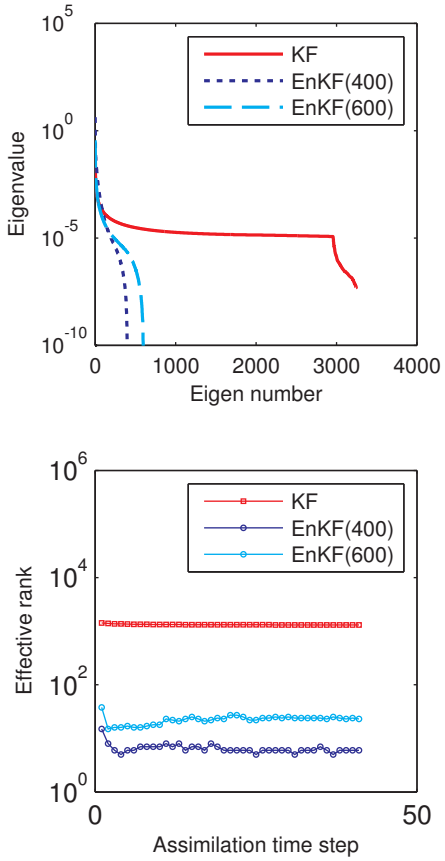

FIG. 3.8. *Eigenvalue spectrums of the posterior covariance at last time step (top). Covariance given by EnKF are computed from 400- and 600-member ensembles respectively. Effective rank of the posterior covariance is plotted for each assimilation step (bottom).*

4. Conclusions. We have presented a novel fast and accurate Kalman filter variant, HiKF, which is a powerful data assimilation tool for real-time data assimilation. The algorithm has $\mathcal{O}(m)$ complexity, and is therefore scalable for large-scale problems. It relies on $\mathcal{H}$-matrix algebra. The low-rank structure of sub-blocks in the model error covariance matrix Q is used to reduce the computational cost and storage requirements.

A comprehensive comparison of HiKF, KF and EnKF can be found in Table 4.1. Whenever the condition for using a random walk model is met, i.e., fast data acquisition, HiKF can provide a fast and accurate alternative to existing algorithms. It reduces the time and storage cost taken by KF dramatically, and accurately reproduces the LMMSE estimates given by KF. To achieve the same accuracy, EnKF requires a large ensemble size at the expense of greatly increasing the computational cost.

Acknowledgments. This work was supported by "US Department of Energy, National Energy Technology Laboratory" (DOE, NETL) under the award number –DE-FE0009260: "An Advanced Joint Inversion System for CO_2 Storage Modeling

TABLE 4.1
Comparison of our proposed computationally efficient Kalman filter with other Kalman filters

Criteria	**KF**	**HiKF**	**EnKF**
Uncertainty propagation	Covariance	Cross covariance	Ensembles
Estimation error	LMMSE	LMMSE	Large error for small ensemble size
Update cost	$\mathcal{O}(m^2)$	$\mathcal{O}(m)$	$\mathcal{O}(m)$
Storage cost	$\mathcal{O}(m^2)$	$\mathcal{O}(m)$	$\mathcal{O}(m)$
Reliable risk analysis	Yes	Variance and cross covariance	Require a large ensemble size
Generality	Linear forecast model	Random walk model	Any forecast model

with Large Data Sets for Characterization and Real-Time Monitoring", and also by the "National Science Foundation" —Division of Mathematical Sciences —under the award number: 1228275. The authors would also like to thank Dr. Jonathan B. Ajo-Franklin, Thomas M. Daley, and Christine Doughty from the Lawrence Berkeley Lab for sharing TOUGH2 and rock physics simulation data.

REFERENCES

[1] Jonathan B Ajo-Franklin. Optimal experiment design for time-lapse traveltime tomography. *Geophysics*, 74(4):Q27–Q40, 2009.
[2] S Ambikasaran and E Darve. An $\mathcal{O}(N \log N)$ fast direct solver for partial hierarchically semi-separable matrices. *Journal of Scientific Computing*, pages 1–25, 2013.
[3] S. Ambikasaran, J.Y. Li, E. Darve, and P. K. Kitanidis. Large-scale stochastic linear inversion using hierarchical matrices. *Computational Geosciences*, 2013.
[4] S. Ambikasaran, A. K. Saibaba, E.F. Darve, and P. K. Kitanidis. *Fast algorithms for Bayesian Inversion, The IMA Volumes in Mathematics and its Applications, vol.* 156. Springer-Verlag, New York, 2013.
[5] Jeffrey L Anderson. An ensemble adjustment Kalman filter for data assimilation. *Monthly weather review*, 129(12):2884–2903, 2001.
[6] Adeyemi Arogunmati and Jerry M Harris. An approach for quasi-continuous time-lapse seismic monitoring with sparse data. In *79th Annual Meeting and International Exposition*, 2009.
[7] Mario Bebendorf. Approximation of boundary element matrices. *Numerische Mathematik*, 86(4):565–589, 2000.
[8] S. Benson. Monitoring carbon dioxide sequestration in deep geological formations for inventory verification and carbon credits. In *SPE Annual Technical Conference and Exhibition*, 2006.
[9] Craig H Bishop, Brian J Etherton, and Sharanya J Majumdar. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly weather review*, 129(3):420–436, 2001.
[10] S. Börm, L. Grasedyck, and W. Hackbusch. Hierarchical matrices. *Lecture notes 21*, 2003.
[11] Steffen Börm, Lars Grasedyck, and Wolfgang Hackbusch. Introduction to hierarchical matrices with applications. *Engineering Analysis with Boundary Elements*, 27(5):405–422, 2003.
[12] G. Burgers, P.J. van Leeuwen, and G. Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126:1719–1724, 1998.
[13] Shiv Chandrasekaran, Patrick Dewilde, Ming Gu, W Lyons, and T Pals. A fast solver for HSS representations via sparse matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):67–81, 2006.
[14] Shiv Chandrasekaran, Ming Gu, and T Pals. A fast ULV decomposition solver for hierarchically semiseparable representations. *SIAM Journal on Matrix Analysis and Applications*, 28(3):603–622, 2006.
[15] W. Daily, A. Ramirez, D. LaBrecque, and J. Nitao. Electrical resistivity tomography of vadose water movement. *Water Resources Research*, 28(5):1429–1442, 1992.
[16] T.M. Daley, R.D. Solbau, J.B. Ajo-Franklin, and S.M. Benson. Continuous active-source seismic monitoring of injection in a brine aquifer. *Geophysics*, 72(5):A57–A61, 2007.
[17] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99:10143–10162, 1994.
[18] Abul Fahimuddin, Sigurd Ivar Aanonsen, and Trond Mannseth. Effect of large number of

measurements on the performance of Enkf model updating. In *Proceedings of the 11th European Conference on the Mathematics of Oil Recovery, Bergen, Norway*, pages 1–14, July 2008.

[19] W. Fong and E. Darve. The black-box fast multipole method. *Journal of Computational Physics*, 228(23):8712–8725, 2009.

[20] J Fritz, I Neuweiler, and W Nowak. Application of FFT-based algorithms for large-scale universal kriging problems. *Mathematical Geosciences*, 41(5):509–533, April 2009.

[21] I. Fukumori and P. Malanotte-Rizzoli. An approximate Kalman filter for ocean data assimilation; an example with an idealized gulf stream model. *Journal of Geophysical Research*, 1994.

[22] Sergei A Goreinov, Eugene E Tyrtyshnikov, and Nickolai L Zamarashkin. A theory of pseudoskeleton approximations. *Linear Algebra and its Applications*, 261(1):1–21, 1997.

[23] Leslie Greengard and Vladimir Rokhlin. A fast algorithm for particle simulations. *Journal of computational physics*, 73(2):325–348, 1987.

[24] W. Hackbusch. A sparse matrix arithmetic based on $\mathcal{H}$-matrices. part I: Introduction to $\mathcal{H}$-matrices. *Computing*, 62(2):89–108, 1999.

[25] W. Hackbusch and B.N. Khoromskij. A sparse $\mathcal{H}$-matrix arithmetic. part II: Application to multi-dimensional problems. *Computing*, 64(1):21–47, 2000.

[26] T.M. Hamill, J.S. Whitaker, and C. Snyder. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11):2776–2790, 2001.

[27] P.L. Houtekamer and H.L. Mitchell. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3):796–811, 1998.

[28] P.L. Houtekamer and H.L. Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129:123–137, 2001.

[29] S.S. Hubbard, J. Chen, J. Peterson, E.L. Majer, K.H. Williams, D.J. Swift, B. Mailloux, and Y. Rubin. Hydrogeological characterization of the south oyster bacterial transport site using geophysical data. *Water Resources Research*, 37(10):2431–2456, 2001.

[30] RE Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[31] J.H. Kim, M.J. Yi, S.G. Park, and J.G. Kim. 4-D inversion of DC resistivity monitoring data acquired over a dynamically changing earth model. *Journal of Applied Geophysics*, 68(4):522–532, 2009.

[32] S.K. Lazaratos and B.P. Marion. Crosswell seismic imaging of reservoir changes caused by CO_2 injection. *The Leading Edge*, 16(9):1300–1308, 1997.

[33] Per-Gunnar Martinsson and Vladimir Rokhlin. A fast direct solver for boundary integral equations in two dimensions. *Journal of Computational Physics*, 205(1):1–23, 2005.

[34] V. Nenna, A. Pidlisecky, and R. Knight. Application of an extended Kalman filter approach to inversion of time-lapse electrical resistivity imaging data for monitoring recharge. *Water Resources Research*, 47(10), 2011.

[35] W. Nowak, S. Tenkleve, and O.A. Cirpka. Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities. *Mathematical Geology*, 35(1):53–66, 2003.

[36] Eftychios A Pnevmatikakis, Kamiar Rahnama Rad, Jonathan Huggins, and Liam Paninski. Fast Kalman filtering and forward-backward smoothing via a low-rank perturbative approach. *Journal of Computational and Graphical Statistics*, January 2013.

[37] K. Pruess. TOUGH2: A general-purpose numerical simulator for multiphase fluid and heat flow. *NASA STI/Recon Technical Report N*, 92(14316), 1991.

[38] Y. Quan and J.M. Harris. Stochastic seismic inversion using both waveform and traveltime data and its application to time-lapse monitoring. In *2008 SEG Annual Meeting*, 2008.

[39] Sergej Rjasanow. Adaptive cross approximation of dense matrices. In *Int. Association Boundary Element Methods Conf., IABEM*, pages 28–30, 2002.

[40] A.K. Saibaba, S. Ambikasaran, J.Y. Li, P.K. Kitanidis, and E.F Darve. Application of hierarchical matrices to linear inverse problems in geostatistics. *OGST Revue d'IFP Energies Nouvelles*, 67(5):857–875, 2012.

[41] Michael K Tippett, Jeffrey L Anderson, Craig H Bishop, Thomas M Hamill, and Jeffrey S Whitaker. Ensemble square root filters*. *Monthly Weather Review*, 131(7):1485–1490, 2003.

[42] Dinh Tuan Pham, Jacques Verron, and Marie Christine Roubaud. A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16(3-4):323–340, October 1998.

[43] M. Vauhkonen, P.A. Karjalainen, and J.P. Kaipio. A Kalman filter approach to track fast

impedance changes in electrical impedance tomography. *Biomedical Engineering, IEEE Transactions on*, 45(4):486–493, 1998.

[44] Zhijing Wang, Michael E Cates, and Robert T Langan. Seismic monitoring of a CO_2 flood in a carbonate reservoir: A rock physics study. *Geophysics*, 63(5):1604–1617, 1998.

[45] JE White. Computed seismic speeds and attenuation in rocks with partial gas saturation. *Geophysics*, 40(2):224–232, September 2001.