Comparison of ensemble filtering algorithms and null-space Monte Carlo for parameter estimation and uncertainty quantification using CO_2 sequestration data

Reza Tavakoli,[1] Hongkyu Yoon,[2] Mojdeh Delshad,[1,3] Ahmed H. ElSheikh,[1] Mary F. Wheeler,[1] and Bill W. Arnold[4]

[1] Geological storage of CO_2 requires multiphase flow models coupled with key hydrogeologic features to accurately predict the long-term consequences. The prediction uncertainty during geological CO_2 storage requires a computationally efficient and practically useful framework. This paper presents a comparative study between ensemble-based filtering algorithms (En-As) and calibration-constrained null-space Monte Carlo (NSMC) methods. For the En-As, we use the ensemble Kalman filter (EnKF), ensemble smoother (ES), ES with multiple data assimilation (ES-MDA), and EnKF and ES with the pilot point method. For the NSMC calibrated models with various parameterization, schemes are tested and single and multiple NSMC (M-NSMC) methods are used. A synthetic case with two layers was developed to mimic an actual CO_2 injection pilot test where one injection and two observation wells are located within a short distance. Observed data include bottom hole pressure at injection well and gas saturation (S_g) at two observation wells in the upper layer. Model parameters include horizontal permeability and porosity. Comparison of results shows that both methodologies yield good history match and reasonable prediction results in a computationally efficient way. In particular, the ES-MDA and M-NSMC resulted in smaller objective function values and lower prediction uncertainties of S_g profiles compared to other variants tested in this work. The En-As with the pilot point method have higher variability of permeability compared to those without one, but the En-As show smoother permeability fields compared to the NSMC methods. This is because stochastic randomness at a grid scale was included to generate NSMC fields. Both ensemble-based and NSMC algorithms are unable to correct the structural orientation of the prior ensemble members using only the sparse dynamic data from wells, while they obtain reasonable history match, suggesting that structural uncertainty should be incorporated into prior information. Overall, the ES-MDA has an advantage in terms of computational efficiency, but at the expense of additional computation M-NSMC shows applicability for highly nonlinear problems such as multiphase flow problems.

Citation: Tavakoli, R., H. Yoon, M. Delshad, A. H. ElSheikh, M. F. Wheeler, and B. W. Arnold (2013), Comparison of ensemble filtering algorithms and null-space Monte Carlo for parameter estimation and uncertainty quantification using CO_2 sequestration data, *Water Resour. Res.*, *49*, 8108–8127, doi:10.1002/2013WR013959.

[1]Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas, USA.

[2]Geoscience Research and Applications, Sandia National Laboratories, Albuquerque, New Mexico, USA.

[3]Center for Petroleum and Geosystems Engineering, University of Texas at Austin, Austin, Texas, USA.

[4]Advanced Nuclear Energy, Sandia National Laboratories, Albuquerque, New Mexico, USA.

Corresponding author: R. Tavakoli, Institute for Computational Engineering and Sciences, University of Texas at Austin, 201 East 24th St., Austin, TX 78712, USA. (tavakoli@ices.utexas.edu)

1. Introduction

[2] Parameter estimation and model prediction for multiphase systems can be very difficult, mainly due to both the computational cost and the nonlinearities of coupled processes encountered in multiphase flow solutions. In particular, emerging subsurface problems like geological CO_2 storage are often represented by simplified conceptual models with a small number of parameters corresponding to a limited set of observations and prior data. Modeling and simulation of multiphase flow and transport processes also requires proper characterization of uncertainty in model parameters and predictions for the engineering assessment of project risk and the proper design of operational conditions. In reservoir engineering literature, the process of calibrating model parameters with nonlinear dynamic data is

known as "history matching" [*Oliver et al.*, 2008]. The recent developments in reservoir history matching are presented in detail in a review paper by *Oliver and Chen* [2010]. A single calibrated model cannot address reservoir characterization and risk management because the history matching is an ill-posed inverse problem with several non-unique solutions that match the measurements but produce different predictions.

[3] In Bayesian statistics, the calibration process is equivalent to drawing samples from the conditional probability density function (PDF) of uncertain model parameters conditional to noisy measurements. It is well known that the stochastic approaches, such as Markov chain Monte Carlo (MCMC) [*Hastings*, 1970; *Gilks et al.*, 1996], bootstrap filtering [*Smith and Gefland*, 1992], or particle filters [*Doucet et al.*, 2001; *Doucet and Johansen*, 2009], will sample the posterior PDF correctly in the limit as they explore all the possible model outcomes with many (theoretically infinite) forward runs. Over the past decades, Monte Carlo (MC) methods for predictive uncertainty analysis have been developed to improve the efficiency of sampling. For example, representative techniques are the generalized likelihood uncertainty estimation (GLUE) method [*Beven and Binley*, 1992] and the calibration-constrained MC and MCMC methods [*Harvey and Gorelick*, 1995; *Kitanidis*, 1996; *Woodbury and Ulrych*, 2000; *Vrugt et al.*, 2008; *Elsheikh et al.*, 2012]. For large-scale nonlinear problems, however, these algorithms usually require a large number of simulations to sample the posterior PDF correctly. Therefore, the application of these extensive algorithms can be computationally very expensive for realistic history matching of CO_2 sequestration projects. Alternatively, approximate sampling methods can be used to reduce computational cost [*Tarantola*, 1987; *Kitanidis*, 1995; *Oliver et al.*, 1997]. In practice, uncertainty quantification techniques can be categorized into two general groups: (1) methods that sample multiple models from the conditional PDF [*Bonet-Cunha et al.*, 1998; *Evensen*, 1994; *Liu and Oliver*, 2003; *Tavakoli and Reynolds*, 2011] and (2) methods that generate a single calibrated "best" model which is used to generate multiple conditional samples with the covariance matrix and/or the null-space of the sensitivity matrix [*Shah et al.*, 1978; *Lépine et al.*, 1999; *Tonkin and Doherty*, 2009]. In this work, we implement and compare several algorithms from each of these two categories to provide a computationally efficient and practically useful framework.

[4] Data assimilation is the process of continuously integrating observed data into dynamic models to determine the best estimate of poorly known parameters and/or states. The ensemble Kalman filter (EnKF) algorithm was introduced by *Evensen* [1994, 2005] for data assimilation in oceanographic and atmospheric models. Since its introduction, it has been used extensively in many applications including oceanography [*Bertino et al.*, 2003; *Keppenne*, 2000], weather forecasting [*Houtekamer et al.*, 2005], hydrology [*Reichle et al.*, 2002; *Chen et al.*, 2006], and petroleum reservoir history matching [*Aanonsen et al.*, 2009; *Oliver and Chen*, 2010]. The EnKF has been widely used mainly due to the ease of its implementation (no sensitivity computation), its sequential updating scheme, which provides multiple simultaneous conditional models, and its

computational efficiency. Moreover, both model parameters (e.g., porosity and permeability) and primary variables (e.g., pressure and saturation) are updated in the form of an augmented state vector. Alternatively, the ensemble smoother (ES) updates only the model parameters and assimilates all the measurements at once using a single analysis step [*van Leeuwen and Evensen*, 1996]. By avoiding the frequent modification of simulator primary variables files (i.e., restart files), the ES performs much faster than the EnKF method. However, it has been shown that the EnKF outperforms the ES method with regard to accuracy [*van Leeuwen and Evensen*, 1996; *Evensen and van Leeuwen*, 2000] because the sequential updates in the EnKF modify the ensemble members toward two-point statistics (Gaussian) models. Therefore, the later update steps do not suffer too much from the Gaussian assumption. Recently, *Emerick and Reynolds* [2012] proposed a methodology to improve the performance of the ES by assimilating the same data multiple times (ES-MDA) with an inflated measurement error covariance matrix. The ES-MDA could be considered an iterative form of the ES method. They concluded that the ES-MDA resulted in better data matches than the EnKF with a comparable computational cost. *Nowak* [2009] used a similar idea of measurement error inflation in a geostatistically based Levenberg-Marquardt algorithm to stabilize the inverse problem while avoiding information loss. It is important to note that the (ensemble) Kalman filtering algorithms are only optimal if the prior state vector is linearly related to the predicted data and if the joint distribution of the prior state vector is multivariate Gaussian [*Zafari and Reynolds*, 2007]. These assumptions are certainly not true for multiphase flow and they can result in degraded performance of ensemble-based algorithms. Several authors have tried to develop techniques to improve the performance of these algorithms for highly non-Gaussian and nonlinear problems [*Reynolds et al.*, 2006; *Agbalaka and Oliver*, 2008; *Moreno and Aanonsen*, 2007; *Jafarpour and McLaughlin*, 2009; *Sarma and Chen*, 2009; *Saad and Ghanem*, 2009; *He et al.*, 2011; *Zhou et al.*, 2011; *Dovera and Rossa*, 2011]. In a review paper, *Aanonsen et al.* [2009] present a summary of the EnKF applications for history matching in reservoir engineering. The sampling performance of different ensemble-based algorithms with a simple, yet highly nonlinear reservoir model has been investigated in *Emerick and Reynolds* [2013] and the results are compared with the results of a very long chain of a MCMC method.

[5] Parameterization of heterogeneous spatial fields using the pilot point method has been used to calibrate groundwater flow models [*Doherty*, 2003; *Alcolea et al.*, 2006; *Doherty et al.*, 2010a]. Recently, the pilot point parameterization was utilized in the EnKF for parameter estimation and history matching [*Heidari et al.*, 2013]. In this approach, pilot points are used to determine model parameters and parameter values over the entire model domain. Model parameters are typically interpolated from the pilot point values based on a measured or assumed covariance [*de Marsily et al.*, 1984, 1999]. In model calibration with the pilot point method, it is now conceived that, among parameters at pilot points, limited number of linear combinations of parameters accounts for the majority of the sensitivity to the calibration data. Hence, the model

parameterization may be decomposed into a calibration solution space (i.e., sensitive to the calibration data) and a calibration null-space [*Moore and Doherty*, 2005]. The null-space is comprised of all combinations of parameters that are not sensitive to the calibration data during model calibration, and conceptually it represents all differences between the parameter field simplified in a model and the real world complexity. Hence, the null-space often pertains to fine-scale information such as local heterogeneities resulting from conceptual, temporal, and spatial simplifications. A single set of parameters is extracted from a gradient-based calibration run. Then a Monte Carlo sampling is applied to the null-spaces of the extracted solution to form an ensemble of solutions conforming to the calibration data. A recently developed null-space Monte Carlo (NSMC) method combines the calibration solution space parameters with the ensemble of null-space parameters, creating sets of calibration-constrained parameters for input to follow-on flow and/or transport predictions [*Tonkin and Doherty*, 2009]. *Zimmerman et al.* [1998] and *Hendricks-Franssen et al.* [2009] present unique comparisons of different inverse approaches for identification and estimation of hydrogeologic properties of groundwater flow.

[6] The NSMC provides a large number of calibration-constrained parameter sets that can be used for prediction uncertainty analysis in a computationally efficient manner. Since the NSMC is constrained by a calibrated parameter field, the performance of calibration-constrained NSMC fields is strongly dependent upon the starting point for the Jacobian (i.e., sensitivity matrix) and the proportion of parameters that lie in the null-space [*Tonkin and Doherty*, 2009; *Doherty et al.*, 2010b]. *Yoon et al.* [2013] demonstrated that the NSMC fields based on a single calibrated model were significantly constrained by the calibrated solution space, and the resulting distribution of advective travel times was biased toward the travel time from the single calibrated field. To overcome this, a multiple calibration-constrained NSMC approach (M-NSMC) was evaluated by comparing M-NSMC with multiple starting points (MSP) methods using a highly parameterized ($>$1300 parameters) model of the Culebra dolomite previously developed for the Waste Isolation Pilot Plant (WIPP) project in New Mexico. The M-NSMC and MSP methods resulted in similar results in terms of the quality of the data fit and predictive uncertainty. The performance of the NSMC was also compared to Markov chain Monte Carlo (MCMC) for a highly nonlinear groundwater model using a simplified surrogate model [*Keating et al.*, 2010], and both methods provided similar results. Due to the computational intensity of the MCMC method, only the NSMC method possesses the computational efficiency necessary to explore the predictive uncertainty of the highly parameterized groundwater model.

[7] Emerging subsurface problems, such as multiphase flow and reactive transport, require a computationally efficient and practical solution for predictive uncertainty analysis. Recently, several new methods, such as an iterative ES algorithm based on the Levenberg-Marquardt (LM) method [*Chen and Oliver*, 2013] and polynomial chaos expansion (PCE) [*Zhang and Sahinidis*, 2013; *Oladyshkin et al.*, 2013] have been developed for application to geological CO_2 storage systems. *Chen and Oliver* [2013] applied

the LM-ensemble randomized maximum likelihood (LM-EnRML) for 1-D and 3-D multiphase flow problems and demonstrated that LM-EnRML provides an efficient and accurate method of uncertainty quantification for highly nonlinear problems. *Zhang and Sahinidis* [2013] built surrogate models using PCE and applied models for quantifying the uncertainties of injection CO_2 and determining the optimal CO_2 injection rate in a 2-D aquifer system. *Espinet and Shoemaker* [2013] also compared different optimization algorithms, including the derivative-based methods, the heuristics-based dynamically dimensioned search, the Radial Basis Function (RBF), surrogate response surfaces, and stochastic RBF for parameter estimation of multiphase flow models under different CO_2 sequestration scenarios. They showed that, for highly heterogeneous cases, surrogate response surfaces algorithms perform best. Although these new methods are promising to provide a computationally efficient mean for complex subsurface problems, it is not straightforward to apply these methods in a model-independent and computationally efficient way. Recently, *Tavakoli et al.* [2012] demonstrated that a parallel ensemble-based framework for reservoir history matching and uncertainty characterization can provide a model-independent and computationally efficient method for a two-phase reservoir model. However, a number of questions remain regarding the consistency of different algorithms, such as ensemble-based algorithms and gradient-based NSMC methods, which are both computationally efficient and model independent.

[8] The objective of this work is to examine the consistency of parameter estimates and model predictions between ensemble-based filtering algorithms and the NSMC methods using a CO_2 sequestration simulation example. In particular, variants of these approaches, such as the EnKF, ES, ES-MDA, and single and multiple calibration-constrained NSMCs are evaluated. The simulation model is inspired by a CO_2 sequestration pilot test held in Cranfield, Mississippi [*Nicot et al.*, 2009; *Hosseini et al.*, 2012; *Delshad et al.*, 2013], where CO_2 is injected into the water leg of the reservoir below the water-oil contact. In particular, the ensemble of the permeability fields, model fit to the observed data, and prediction of CO_2 breakthrough at the observation wells are compared to perform prediction uncertainty analysis and evaluate computational efficiency.

[9] This paper is structured as follows: In section 2, we introduce the forward model and present the descriptions of each method that we used in a separate subsection. In section 3, we present the computational results of our study. We start this section by description of the case study, followed by presenting results of different algorithms. After that, we present an overall discussion of the results in section 4. In section 5, we summarize and conclude the paper.

2. Methodology

[10] The application of ensemble-based data assimilation algorithms for parameter estimation and uncertainty quantification has been steadily increasing over the past decade. In these algorithms, an ensemble of realizations is used to represent the uncertainty by an approximation of error covariance for the state vector. The ensemble of model

states is integrated forward in time to predict the uncertainty. Whenever observation data are available, each realization is updated to assimilate the information from new measurements. On the other hand, gradient-based algorithms have been widely investigated and used in model calibration and predictive uncertainty estimation in different fields, such as hydrology, oceanography, and petroleum reservoir modeling. In this section, we first introduce the forward compositional model used to simulate the CO_2 sequestration process. Then, we provide brief descriptions and formulations of the algorithms considered in this paper.

2.1. Compositional Flow Model

[11] The governing equations of multiphase flow (e.g., water and CO_2) in porous media are given by a set of partial differential equations that represent conservation of mass, momentum, and energy together with equations of state that describe the fluid property as a function of pressure, saturation, and composition. Let i and α represent component and phase indices, respectively. The material balance equations for n_c components in the multiphase flow are given by the system of partial differential equations

$$\frac{\partial\left(\phi\sum_\alpha x_i^\alpha \rho_\alpha S_\alpha\right)}{\partial t} + \sum_\alpha \nabla\cdot\left(x_i^\alpha \rho_\alpha u_\alpha\right) = q_i, \quad \text{for } i=1,2,\dots,n_c,$$

(1)

where ϕ is porosity, x_i^α is the mole fraction of component i in phase α, ρ_α is the molar density of phase α, S_α is phase α saturation, u_α is the velocity of phase α, and q_i is the source/sink term of component i. The velocity of phase α is given by Darcy's law as

$$u_\alpha = \frac{k_{r\alpha}}{\mu_\alpha}\cdot(\nabla P_\alpha - \gamma_\alpha \nabla D),$$

(2)

where $k_{r\alpha}$ denotes relative permeability of phase α, μ_α is the viscosity of phase α, $\gamma_\alpha = \rho_\alpha g$, where ρ_α is the fluid density and g is the gravity, and the phase pressures, P_α, may be eliminated by using the capillary pressure definition, $P_\alpha = P - P_{c\alpha}$, where $P_{c\alpha}$ is the capillary pressure and P is the reference pressure. We assume capillary pressure and relative permeability are functions of saturation. The pure component properties are expressed in terms of pressure and temperature. For isothermal processes, the primary variables in equation (1) are the gas saturation (S_g), pressure, and mole fraction. It is assumed both that the gaseous phase and the aqueous phase are in thermodynamic equilibrium and that the mole fractions are determined using thermodynamic equations of state(s) (EOS) [e.g., *Peng and Robinson*, 1976; *Soave*, 1972] for phase equilibrium, which in turn yield phase properties from appropriate mixing rules. It remains to decouple the system and solve for reference pressure, P, phase saturation, and phase compositions, x_i^α. In the general compositional reservoir simulator (GEM) of CMG [*CMG*, 2013], the flow equations are discretized using the adaptive-implicit method because it encompasses both the explicit-transmissibility method and the fully implicit method as particular cases. Also, the finite difference method is used for the space discretization. A detailed description of these techniques is beyond the scope of this paper, but it can be found elsewhere [*CMG*, 2013; *Collins et al.*, 1992].

2.2. Ensemble-Based Filtering Algorithms

2.2.1. Ensemble Kalman Filter

[12] The EnKF formulation is presented by defining the N_y-dimensional state vector, y^n, at the nth data assimilation time step

$$y^n = \begin{bmatrix} m^n \\ p^n \end{bmatrix},$$

(3)

where m^n is the N_m-dimensional vector of reservoir model parameters (such as permeability and porosity) and p^n is the N_p-dimensional vector of dynamical model state variables (such as grid block pressures and saturations of the reservoir simulator). Thus, $N_y = N_m + N_p$. The EnKF analysis equation for each ensemble member j can be written as

$$y_j^{n,a} = y_j^{n,f} + C_{YD}^{n,f}\left(C_{DD}^{n,f} + C_D^n\right)^{-1}\left(d_{uc,j}^n - d_j^{n,f}\right) \quad \text{for } j=1,2,\dots,N_e.$$

(4)

[13] In this equation, $C_{YD}^{n,f}$ denotes the $N_y \times N_{d^n}$ cross-covariance matrix between the forecast state vector $y_j^{n,f}$ and predicted data $d_j^{n,f}$. The forecast is equivalent to propagating the forward model (reservoir simulator) in time for each of the ensemble members. The $C_{DD}^{n,f}$ denotes the $N_{d^n} \times N_{d^n}$ auto-covariance matrix of predicted data; C_D^n denotes the $N_{d^n} \times N_{d^n}$ observed data covariance matrix (measurement error covariance matrix); $d_{uc,j}^n$ is a sample of the normal distribution $N\left(d_{\text{obs}}^n, C_D^n\right)$; d_{obs}^n represents the N_{d^n}-dimensional vector of observed data at the nth data assimilation time step; $d_j^{n,f}$ denotes the corresponding N_{d^n}-dimensional vector of predicted data for the jth ensemble member; N_{d^n} is the number of data assimilated at the nth data assimilation time step; N_e is the number of ensemble members. The superscript f denotes forecast; the superscript a denotes analysis, and the subscript uc denotes an unconditional perturbation of observed data. The above EnKF update equation is similar to the traditional Kalman filter update equation [*Kalman*, 1960], with the difference being the way the covariance matrices are propagated. The predicted data, $d_j^{n,f}$, can be obtained by running the reservoir simulator forward, which is represented by the nonlinear equation $d_j^{n,f} = g(m_j^n)$ for $j=1,2,\dots,N_e$. The cross-covariance matrix, $C_{YD}^{n,f}$, and the autocovariance matrix, $C_{DD}^{n,f}$, can be expressed by

$$C_{YD}^{n,f} = \frac{1}{N_e-1}\sum_{j=1}^{N_e}(y_j^{n,f} - \bar{y}^{n,f})(d_j^{n,f} - \bar{d}^{n,f})^T,$$

(5)

and

$$C_{DD}^{n,f} = \frac{1}{N_e-1}\sum_{j=1}^{N_e}(d_j^{n,f} - \bar{d}^{n,f})(d_j^{n,f} - \bar{d}^{n,f})^T,$$

(6)

where $\bar{y}^{n,f}$ and $\bar{d}^{n,f}$ denote the mean of the N_e predicted state vectors and the mean of the N_e predicted data at assimilation time step n, respectively. The superscript T

denotes the transpose of a matrix. Note in the implementation of the EnKF analysis equation (4) that the $N_y \times N_{d^n}$ matrix $C_{YD}^{n,f}$ as defined by (5) is never explicitly computed or stored.

2.2.2. Ensemble Smoother

[14] The smoother algorithms update both the current and the past state vectors of a dynamic model which is different than the filtering approach that uses sequential updates in time. The ES method was introduced in oceanography by *van Leeuwen and Evensen* [1996] and recently has been used to approach the history matching problems by *Skjervheim et al.* [2011]. In Bayesian statistics, the EnKF and ES methods use similar update formulations. In the former, a recursive update is performed at specific measurement times based on current ensemble statistics, while in the latter, a global update is calculated over the entire space-time domain based on prior ensemble statistics. The assumptions necessary for the ES method are less strict, because the measurement errors in the observed data need not be statistically independent in time. It is known that when the prior PDFs are Gaussian and the relationship between the model parameters and the observation data is linear, the ES and EnKF methods produce the same result [*Evensen and van Leeuwen*, 2000; *Evensen*, 2004]. The implementation of the ES is much simpler, though, because one does not need to update simulator primary variables (e.g., pressure and saturation) and only the model parameters are updated. Therefore, the restart files are completely eliminated. A comparison of the computational efficiency of the EnKF and ES is presented by *Tavakoli et al.* [2012] for a synthetic history matching problem and *Crestani et al.* [2013] compared the EnKF and ES capabilities to retrieve the hydraulic conductivity spatial distribution in a groundwater flow and transport modeling framework.

[15] The ES method only updates model parameters m, instead of the full state vector y^n required for the EnKF. Therefore, the ES implementation is a parameter-estimation problem that is consistent with traditional history matching problems. Thus, the updates for primary variables of simulator are obtained by running reservoir simulation from time zero using the updated model parameters m. The ES update equation for each ensemble member is given by

$$m_j^a = m_j^f + C_{MD}^f \left(C_{DD}^f + C_D \right)^{-1} \left(D_{uc,j} - D_j^f \right) \quad \text{for } j = 1, 2, \ldots, N_e, \quad (7)$$

where $D_{uc,j}$ is the N_d-dimensional vector of perturbed observed data that contains all dynamic data for ensemble member j, i.e.,

$$D_{uc,j} = \left[d_{uc,j}^{1,T} d_{uc,j}^{2,T} \ldots d_{uc,j}^{N,T} \right]^T, \quad (8)$$

where N denotes the total number of assimilation time step. We let N_d denote the total number of measurements assimilated. The vector D_j represents the N_d-dimensional predicted data corresponding to $D_{uc,j}$. Since all the data are being assimilated simultaneously, we suppress the time superscript n in equation (7). The $N_m \times N_d$ cross-covariance matrix between the vector of model parameters m_j and the vector of all predicted data, D_j, is defined as

$$C_{MD}^f = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (m_j^f - \overline{m}^f)(D_j^f - \overline{D}^f)^T, \quad (9)$$

where $\overline{m}^f$ and $\overline{D}^f$ denote the mean of the prior model parameters and the predicted data, respectively. In the ES update (equation (7)), C_D denotes the observed data covariance matrix corresponding to vector $D_{obs} = \left[d_{obs,j}^{1,T} d_{obs,j}^{2,T} \ldots d_{obs,j}^{N,T} \right]^T$, containing all measurements. Therefore, there is no need to assume that individual data are uncorrelated in time.

[16] Similar to the EnKF, the ES algorithm consists of two main steps, the forecast step and the analysis step. In the forecast step, each of the ensemble members is propagated in time from $t = 0$ to $t = t^N$, the time of last data assimilation. In the analysis step, the vector of predicted data, D_j, corresponding to all measured data, D_{obs}, is used in equations (7) to update the model parameters for each of the ensemble members.

2.2.3. Ensemble Smoother With Multiple Data Assimilation

[17] Although the implementation of the ES algorithm is much faster than the EnKF method, the results are not as satisfactory as those obtained with the EnKF. The performance of the ES and EnKF has been compared in an oceanography model by *van Leeuwen and Evensen* [1996] and for a Lorenz model by *Evensen and van Leeuwen* [2000]. In both cases, the EnKF performed better than the ES. *van Leeuwen and Evensen* [1996] concluded that the relatively poor performance of the ensemble smoother compared to the EnKF is due to the poor prior estimate of uncertainty. In the filter, the sequential updates "pull" the ensemble member toward the data at every assimilation time step. Therefore, the posterior ensemble has a much lower variance. It is well known that the EnKF is similar to applying a single Gauss-Newton iteration at each data assimilation time step with a full step size and average sensitivity matrix estimated from the ensemble [*Reynolds et al.*, 2006; *Gu and Oliver*, 2007]. Motivated by this idea, *Emerick and Reynolds* [2012] proposed a methodology to improve the performance of the ES with assimilating the same data multiple times (ES-MDA) with an inflated measurement error covariance matrix. In their procedure, *Emerick and Reynolds* [2012] showed the equivalence between single and multiple data assimilation for linear models where the number of data assimilation must be selected a priori in order to determine the inflation coefficients. The update equation in the ES-MDA is similar to the update equation in the ES (i.e., equation (7)), however, C_D is replaced by $\alpha_i C_D$. The inflation coefficients, α_i, must be selected such that

$$\sum_{i=1}^{N_a} \frac{1}{\alpha_i} = 1, \quad (10)$$

where N_a denotes the number of data assimilation cycles or iterations.

2.2.4. Ensemble-Based Algorithms With Pilot Points

[18] The pilot point method can be utilized to improve the efficiency of inverse methods for subsurface flow models [*de Marsily et al.*, 1984; *RamaRao et al.*, 1995; *Gómez-*

Hernández et al., 1997; *de Marsily et al.*, 2000; *Oliver and Chen*, 2010]. Parameter values over the numerical domain can be interpolated from the parameters estimated at the pilot points using interpolation methods such as the ordinary kriging method with a variogram model.

[19] In ensemble-based data assimilation algorithms, the prior ensemble members are generated using a geostatistical model that describes the spatial variability of the model parameters. The prior geostatistical model, inferred from geological data, is characterized by mean, variance, and variogram model. It is easy to show that the updated models obtained with the linear Kalman filtering equation (equations (4) or (7)) are linear combinations of the initial models. Therefore, the correct characterization of prior models is crucial for improving the performance of the parameter-estimation algorithms such as the EnKF and ES. One efficient approach in preserving the prior geostatistical properties among updated realizations is to combine the EnKF and ES algorithms with the pilot point technique. In this approach, the pilot points are considered to be the parameters of calibration. The Kalman filtering update equation is used to update values of petrophysical properties at these locations. Then kriging or conditional simulation methods (e.g., sequential Gaussian simulation) is used to generate updated realizations conditioned to measurement data and property values at pilot point locations. Therefore, the updated realizations honor the measurements at their locations, preserve the variogram properties, and also match the observation data. Moreover, in comparison with the standard implementation of ensemble-based algorithms, more observation data can be assimilated with the pilot point parameterization without encountering the ensemble collapse problem [*Houtekamer and Mitchell*, 1998; *Emerick and Reynolds*, 2010].

2.3. Model Calibration and Null-Space Monte Carlo Analysis

2.3.1. Calibration

[20] Model calibration and predictive uncertainty analysis are performed based on gradient-based algorithms. The overall objective function to be minimized consists of the measurement objective function (i.e., data mismatch) and the Tikhonov regularization objective function (i.e., model mismatch) [*Doherty*, 2003; *Medina and Carrera*, 2003]. The overall objective function can be expressed as

$$O(m) = W_{\text{obs}} (d - d_{\text{obs}})^T C_D^{-1} (d - d_{\text{obs}}) \\ + W_{\text{reg}} (m - m_{\text{prior}})^T C_M^{-1} (m - m_{\text{prior}}), \tag{11}$$

where W_{obs} and W_{reg} are weighting factors for the observation and regularization mismatch, respectively; d is the simulated data corresponding to d_{obs}; m_{prior} denotes the prior model parameters for regularization; m is the vector of current parameter values; and C_M is the prior covariance matrix. For a Gaussian prior and Gaussian measurement errors with $W_{\text{obs}} = 1/2$ and $W_{\text{reg}} = 1/2$, it can be shown that the maximum a posteriori (MAP) estimate is obtained by minimizing the objective function given by equation (11) [*Tarantola*, 1987; *Oliver et al.*, 2008]. For traditional weighted least squares minimization, a posteriori statistical analysis can be used to obtain the optimal values for W_{obs}

and W_{reg} [*Medina and Carrera*, 2003; *Alcolea et al.*, 2006] and a constrained minimization can be used to obtain W_{reg} similar to a Lagrange multiplier with W_{obs} equal to unity [*Doherty*, 2003]. Tikhonov regularization incorporates the soft information (e.g., preferred values or prior estimates) into the calibration process and prevents an overfit to the calibration data in a nonphysical sense [*Doherty and Hunt*, 2010].

[21] In this work, the pilot point method described in the previous section is used to estimate model parameters at pilot points using the gradient-based search. For model calibration, the parameter set can be decomposed into the calibration solution space and null-space based on the sensitivity (or Jacobian) matrix [*Tonkin and Doherty*, 2005]. Pilot points are a convenient and an effective way to describe the difference between solution and null-spaces of the parameter set at pilot points since the solution space represents bulk properties of a system constrained by the observed data, but the fine details of the system that cannot be resolved by the observed data belong to the null-space. For a computationally efficient and numerically stable truncated singular value decomposition (TSVD) application, the "SVD-assist" technique [*Tonkin and Doherty*, 2005; *Doherty and Hunt*, 2010] is used to estimate linear combinations of parameters spanning the calibration solution space. The SVD-assist is different from the TSVD in that iteration of parameter estimation for nonlinear problems is processed using the Jacobian matrix calculated only with respect to linear combinations of parameters spanning the calibration solution space, while the pure TSVD requires updating the Jacobian matrix with respect to all base parameters. Hence, the number of parameters to be calibrated (i.e., computational expense) in the SVD-assist technique is dramatically reduced. The potential limitation of the SVD-assist is that it is best suited for more linear problems where the sensitivities calculated at initial parameter values represent those at optimal values well, but this limitation can be overcome by updating the calibration solution space iteratively during calibration at the expense of additional computation [*Welter et al.*, 2012]. Although *Doherty and Hunt* [2010] suggest the additional use of Tikhonov regularization with TSVD/SVD-Assist, it was not implemented in this work with a gradient-based method as in *Espinet and Shoemaker* [2013]. Instead, the fixed pilot points are employed to honor observed data; hence, the objective function is the data mismatch part only in equation (11) (i.e., $W_{\text{obs}} = 1$ and $W_{\text{reg}} = 0$).

2.3.2. Null-Space Monte Carlo

[22] For uncertainty quantification, the NSMC method uses the calibrated model to generate calibration-constrained NSMC parameter sets by incorporating the impact of stochastic nature of heterogeneity resulting from conceptual, temporal, and spatial simplifications [*Tonkin and Doherty*, 2009; *Doherty et al.*, 2010b]. Model predictions with the NSMC fields constrained by a single calibrated model can be influenced by the calibration solution space, which is likely to be a local optimum obtained using a gradient-based algorithm (e.g., Marquardt-Levenberg algorithm). To overcome this, we also evaluate a multiple calibration-constrained NSMC (M-NSMC) approach [*Yoon et al.*, 2013]. The detailed description and theoretical

Table 1. List of Investigated Algorithms and Different Types of Results Compared[a]

Algorithm	Symbol	$ln(k)$ and S_g at Obs-1 and Obs-2	STD of $ln(k)$	Mean and STD of S_g	S_g PDF at Obs-1 and Obs-2 (5100 days)	$ln(k)$ With Incorrect Anisotropy Direction
Ensemble Kalman filter	EnKF	X	X	X	X	X
Ensemble smoother	ES	X				
Ensemble smoother with multiple data assimilation	ES-MDA (ES4)	X			X	X
Ensemble Kalman filter with pilot point	EnKF-PP	X	X	X		
ES4 with pilot point	ES4-PP	X				
Null-space Monte Carlo[b]	NSMC	X	X	X	X	X
Multiple calibration-constrained NSMC	M-NSMC	X	X	X	X	X

[a]For uncertainty quantification, 200 ensemble members are evaluated except for M-NSMC with 150 ensemble members.
[b]NSMC methods require model calibration. For model calibration, the number of pilot points (30, 56, 108), inclusion of fixed pilot points at the observation and injection wells (three fixed pilot points per each horizontal layer), and the porosity distribution (true values, constant value, correlation between $ln(k)$ and porosity) were evaluated. For single constrained NSMC, three calibrated models (PP30-fixed, PP30-nofixed, and PP30-fixed-coregion) were used. For all other cases, PP30-fixed was used as a calibrated model to generate calibration-constrained NSMC and M-NSMC sets.

background of the NSMC method can be found in *Tonkin and Doherty* [2009] and *Keating et al.* [2010].

[23] In the NSMC method, stochastic realizations generated from the prior covariance matrix (C_M) are not constrained by measurements of system state (i.e., calibrated model). To generate calibration-constrained NSMC fields, differences between stochastic parameter realizations and the calibrated parameter values are projected onto the calibration null-space. In particular, the fine-scale heterogeneity that is not estimable with the calibration data set is incorporated into the prediction uncertainty analysis through stochastic random fields. The parameter values (i.e., $ln(k)$) at pilot points are estimated from the stochastic fields generated using the sequential Gaussian simulation (RANDPAR) [*Doherty et al.*, 2010b], so that the $ln(k)$ field interpolated over the numerical domain from estimated pilot point parameters provides the best fit to each stochastic random field in the least squares sense. The parameters obtained through this step are projected onto the calibration null-space (i.e., zero projection onto the calibration solution space). The differences between stochastic fields and fields interpolated from pilot points are added to the calibrated parameters to retain the small-scale heterogeneity reflected in the stochastic random field. The perturbation of the null-space components does not significantly change the overall calibration, and the solution space components are retained. Due to nonlinearity and the existence of nonzero truncated singular values, however, the NSCM fields depart from the calibrated state slightly. This departure requires further adjustment of each NSMC parameter set with one or two additional recalibration iterations [*Tonkin and Doherty*, 2009; *Yoon et al.*, 2013]. A detailed description of the NSMC procedure is given by *Doherty et al.* [2010b] and a description of M-NMSC is given by *Yoon et al.* [2013].

[24] Overall, eight different methods described in this section are compared. A brief summary and symbols are listed in Table 1. The ensemble Kalman filter (EnKF), ensemble smoother (ES), ES with multiple data assimilation (ES-MDA), and three ensemble-based algorithms with pilot points (PP) are included and single and multiple calibration-constrained NSMC methods are compared. For model calibrations, six different parameterizations with different number of pilot points, inclusion of fixed pilot

points, and different permeability and porosity relationships are evaluated. Details of test cases will be described in the section 3.

3. Computational Results

3.1. Problem Descriptions

[25] In this example, we considered a numerical model with a uniform grid of $96 \times 72 \times 2$ with gridblock sizes of 7.62 m $\times$ 7.62 m $\times$ 6.09 m (25 ft $\times$ 25 ft $\times$ 20 ft) as a basis for our comparison. This model is inspired by an actual pilot field test to study different aspects of CO_2 sequestration problems: Cranfield [*Nicot et al.*, 2009; *Hosseini et al.*, 2012; *Delshad et al.*, 2013]. There are one CO_2 injection well and two observation wells 60.96 m (200 ft) and 106.68 m (350 ft) from the injector. An open boundary condition was specified by placing six constant pressure wells at the sides of the numerical model. The reference (true solution) horizontal permeability for each layer is shown in Figure 1. The true horizontal log-permeability and porosity fields are generated using an anisotropic spherical covariance function with different geostatistical parameters for each layer of the reservoir. The main geostatistical parameters used to generate the truth are summarized in Table 2. In this table, $ln(k)_{mean}$ and ϕ_{mean} denote the prior mean of horizontal log-permeability and porosity, respectively. The standard deviations of log-permeability and porosity are denoted by $\sigma_{ln(k)}$ and σ_ϕ. Here, $\rho_{\phi,ln(k)}$ denotes the correlation coefficient between porosity and log-permeability; α is the angle measured counterclockwise from the x axis to the principal direction of the covariance function; l_1 is the correlation range in the principal direction and l_2 is the correlation range in the orthogonal direction. Note that the prior means of horizontal log-permeability and porosity fields are uniform in each layer. We assume that the ratio of vertical to horizontal permeability is 0.3.

[26] The carbon dioxide (CO_2) is injected at a constant rate of 283.17 m^3/d (10,000 Standard Cubic Feet (SCF)/d). Injection starts at the beginning of the simulation and continues for 5200 days. The top views of the simulated gas saturation distributions at 3000 and 5200 days are given in Figure 2. The observation data were generated by running
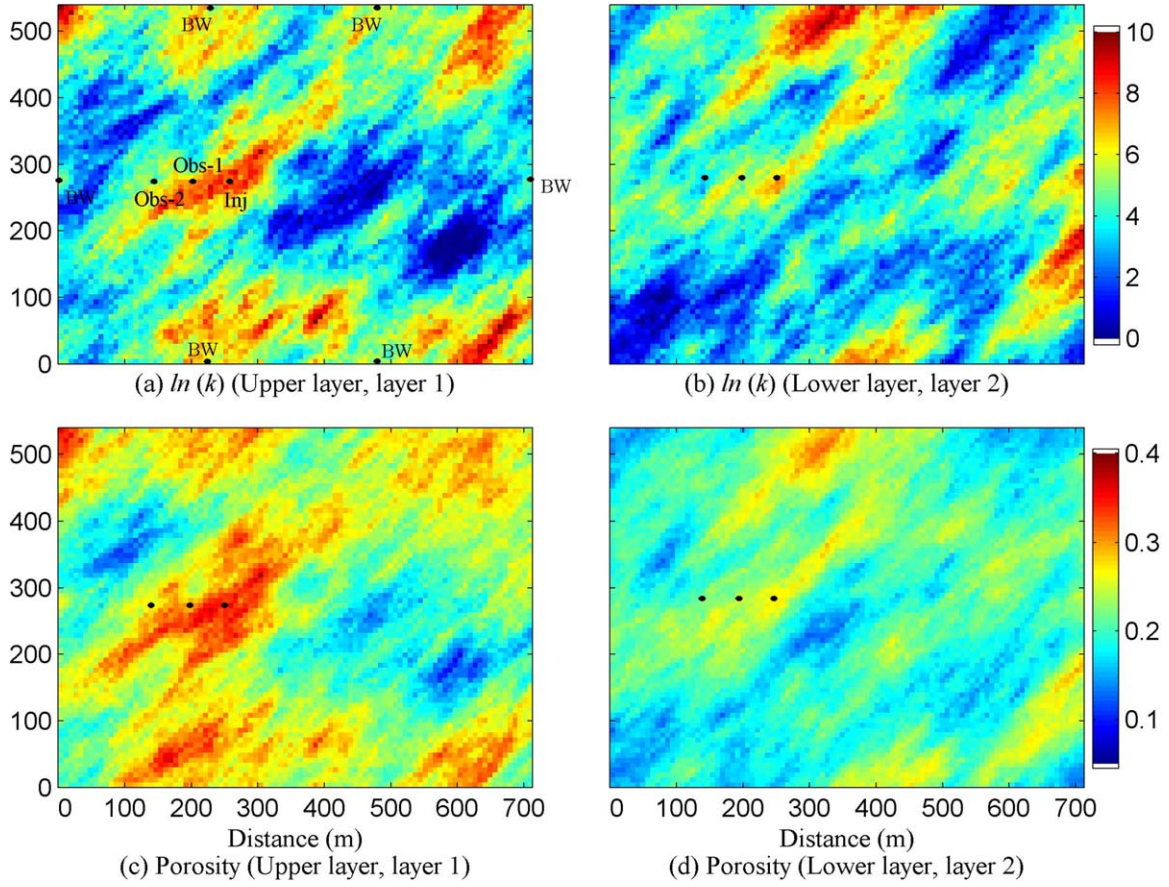
Figure 1. Reference horizontal (top row) $ln(k)$ and (bottom row) porosity with location of wells.

the GEM-CMG [*CMG*, 2013] simulator with a reference permeability field and perturbing the simulator output at the wells with Gaussian random noise accounting for measurement and modeling errors. The data includes injection bottomhole pressure (BHP), pressure, and CO_2 gas saturation (S_g) at observation wells. The standard deviations of the Gaussian noise are equal to 5% of the true (reference) data for all different types of measurements. The data are recorded in variable time intervals, ranging from 50 to 3000 days. For all eight algorithms in Table 1, $ln(k)$ field and S_g profiles at two observations (Obs-1 and Obs-2) are compared and an objective function (equation (11)) is computed after assimilation of all measurements up to 3000 days. Other measures such as statistics of $ln(k)$ field and CO_2 gas saturation (S_g) distribution and probability density function (PDF) of S_g at Obs-1 and Obs-2 at 5100 days are selectively shown to highlight model comparison as summarized in Table 1. In particular, prediction uncertainty analysis is performed to evaluate the predictability of each method on S_g profiles at Obs-1 and Obs-2 which are not used as the observed data for history matching (i.e., after 3000 days).

3.2. EnKF and ES Results

[27] We have used a parallel ensemble-based framework for this synthetic example to perform parameter estimation for the log-permeability and porosity fields and prediction uncertainty analysis for S_g profiles at Obs-1 and Obs-2. The detailed description of this framework can be found in

Tavakoli et al. [2012]. The assimilation step is followed by a prediction phase until 5200 days. We have used 200 realizations generated with the same geostatistical parameters as the reference case to integrate data with the standard EnKF and ES algorithms. The ES algorithm assimilates all observation data at the same time in one step. To improve the results of the ES algorithm, we have performed multiple data assimilation steps. In this work, we used the ES-MDA procedure as proposed by *Emerick and Reynolds* [2012] with $N_a = 4$ and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 4.0$.

[28] The final estimates of the ensemble mean of log-permeability fields for two layers of the aquifer obtained with the EnKF are shown in Figures 3b and 3f. The prior ensemble mean of $ln(k)$ for each layer is uniform. In comparison to the true fields (Figures 3a and 3e), it is clear that the EnKF has captured the main features of the permeability near injection and observation wells. These results also indicate that the first layer shows more similarity to the truth than the second layer. This is because the majority of the injected CO_2 is in the top layer where the observation data are less sensitive to permeabilities in bottom layers.

Table 2. Geostatistical Parameters Used to Generate Reference Model and Prior Realizations

Parameters	$[ln(k)]_{mean}$	ϕ_{mean}	$\sigma_{ln(k)}$	σ_ϕ	$\rho_{\phi,ln(k)}$	α	l_1 (ft)	l_2 (ft)
Layer 1	4.70	0.25	2.10	0.05	0.82	45.0	1200	400
Layer 2	4.20	0.20	1.75	0.035	0.83	45.0	1200	400

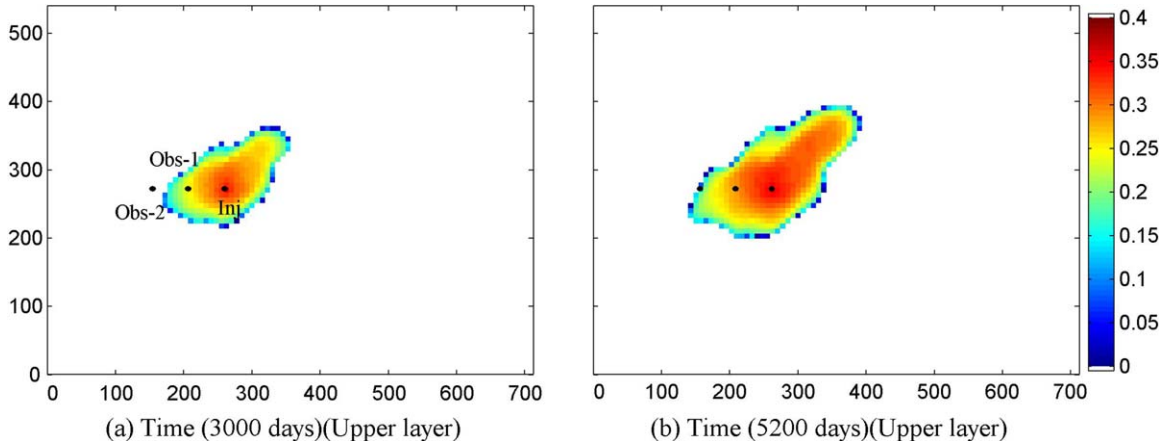(a) Time (3000 days)(Upper layer) (b) Time (5200 days)(Upper layer)

Figure 2. Top view of CO_2 saturation distribution for reference model at two different times.

The comparison of log-permeability fields obtained with the EnKF and ES algorithms with multiple data assimilation (iterations) is shown in Figure 3 for the first iteration, ES1 (third column) and the fourth iteration, ES4 (fourth columns). It is clear that the final mean of $ln(k)$ obtained with the ES resembles the truth but it is much smoother than the one obtained by the EnKF. The ES results improve as we assimilate data multiple times. As can be seen from this figure, the results of log-permeability fields after four iterations of the ES are qualitatively better than the EnKF results. The EnKF and ES results of porosity fields (not shown) are also qualitatively similar to those of $ln(k)$ fields.

[29] The first (left) column of Figure 4 presents the history matching and prediction performance of gas saturation (S_g) of Obs-1 and Obs-2 obtained by the EnKF using final updated models and running simulations from time zero. As shown in this figure, the ensemble spread has been reduced after data assimilation, and the mean prediction from the final models matches the reference (truth) results. The gas breakthrough from posterior ensemble members (blue curves) also matches the truth for Obs-1, with some degree of variability among the realizations. In contrast, the

predicted gas saturation at well Obs-2 (Figure 4e) varies considerably among the updated ensemble members, reflecting the relative lack of information conveyed to the model by observational data at this location during the historical period. The BHP data match and prediction (not shown) are also very good with the EnKF. The history match and forecast of gas saturation (S_g) at the observation wells for the ES-MDA algorithm are shown in Figure 4 (second column to fourth column). These results show large uncertainties (variance) in gas saturation from the prior ensemble members, and the prior ensemble mean is far from the truth. While the prior uncertainty has reduced with the first ES run (second column of Figure 4), the updated ensemble mean (green curve) of gas saturation for Obs-1 deviates from the reference results (red curve) in the prediction phase. As expected from the results of the permeability fields, the calibrated results are improved significantly by the second and fourth ES runs, and we obtain comparable results to those of the EnKF. These results suggest that simply performing the ES multiple times will improve the results significantly with comparable computational cost. It is worth mentioning that the final ensemble
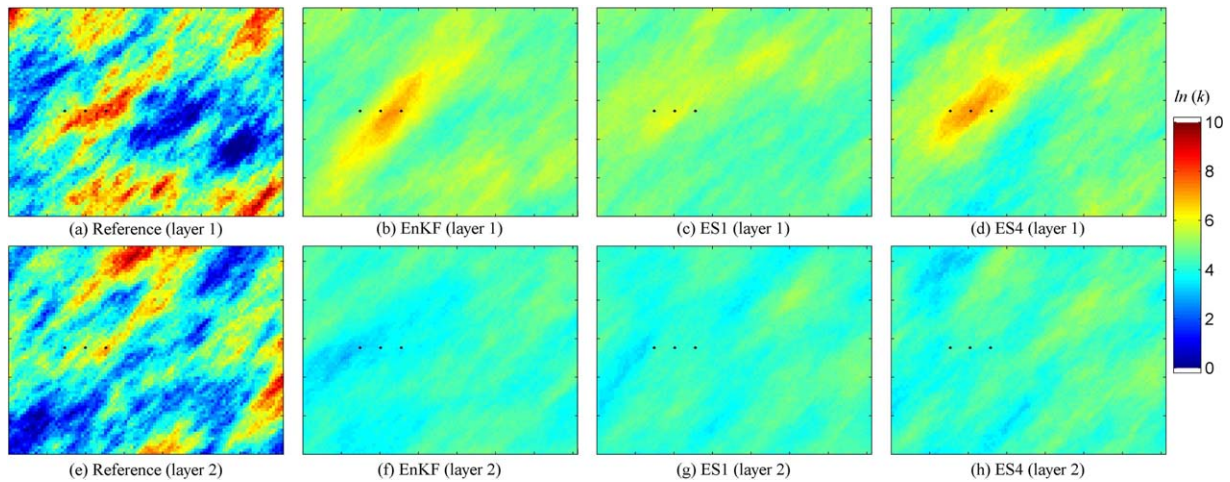

(a) Reference (layer 1) (b) EnKF (layer 1) (c) ES1 (layer 1) (d) ES4 (layer 1)

(e) Reference (layer 2) (f) EnKF (layer 2) (g) ES1 (layer 2) (h) ES4 (layer 2)

Figure 3. Final updated ensemble mean of horizontal $ln(k)$ fields (second column) with EnKF and (third column and fourth column) with ES algorithm after first and fourth assimilation. (first row) Layer 1 and (second row) layer 2.
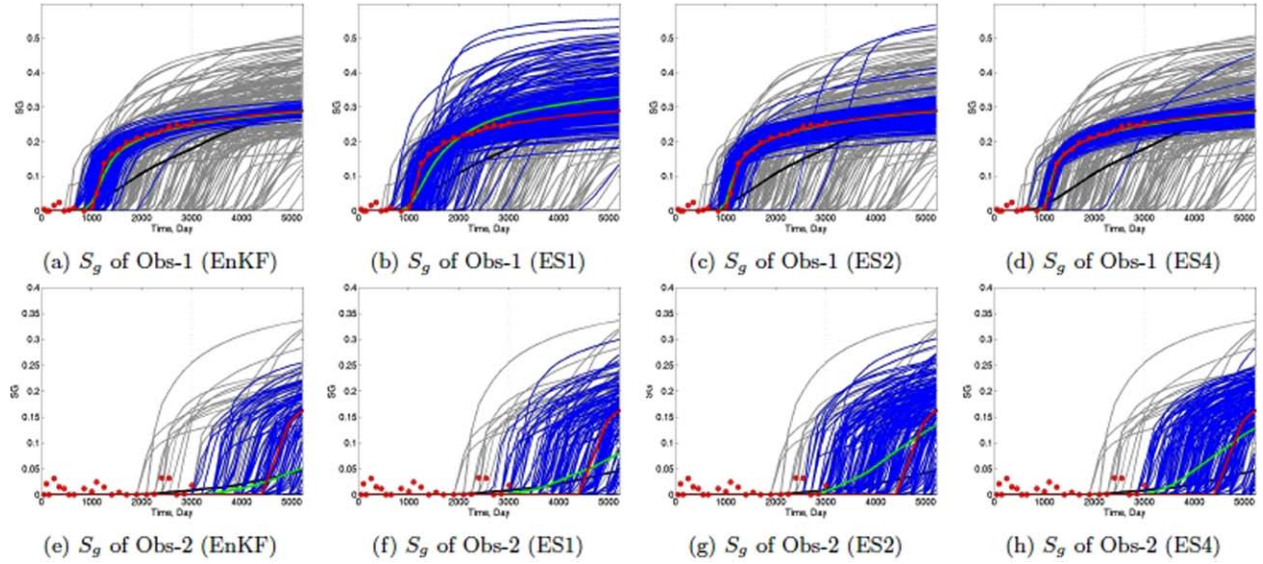
Figure 4. History matching and production forecasts from the final updated ensembles using EnKF and ES method with $N_e=200$. Open red circles are observed data, red curve is from the true model, green curve is the updated ensemble mean, blue curves represent updated ensemble predictions, gray curves represent prior ensemble predictions, and black curve is the prior ensemble mean. The vertical line marks the end of historical data (3000 days).

means of gas saturation for Obs-1, obtained with the EnKF and ES with assimilating data four times, match the CO_2 breakthrough time. However, the ensemble mean of prediction of gas saturation of the second observation well deviated from the true prediction. This is mainly because the CO_2 saturation has not reached this location for the reference model during the whole assimilation period. Therefore, it is not possible to have an update for permeability/porosity in the vicinity of the second observation well.

3.3. NSMC Results

3.3.1. Model Calibration

[30] As in ensemble-based algorithms, parameter estimation for the log-permeability and porosity fields and prediction uncertainty analysis for S_g profiles at Obs-1 and Obs-2 were performed using the SVD-assist technique implemented in the parameter-estimation package PEST [*Doherty*, 2012; *Doherty and Hunt*, 2010; *Hunt et al.*, 2010]. As described earlier, model calibration is required to generate calibration-constrained NSMC fields. For model calibrations, the impact of the number of pilot points (PP), inclusion of fixed pilot points at the observation and injection wells, and the porosity distribution were evaluated. Due to the small number of observation and injection wells (i.e., a total three wells), the parameters at many locations will lie in the null-space, which is typical at CO_2 sequestration sites where installing the observation well is prohibitively expensive. For porosity distribution, three different approaches are tested: the constant value of porosity per each layer (constant), the original distribution of the true field, and the porosity distribution based on the linear model of coregionalization (coregion) [e.g., *Goovaerts*, 1997]. For coregionalization, the estimated $ln(k)$ field was used to generate a random field for porosity with the same covariance structure as the permeability field with a speci-

fied $ln(k)-\phi$ correlation coefficient (Table 2). In addition, a different anisotropy direction ($\alpha = 135$) is used to test the impact of structural error.

[31] The S_g profiles at Obs-1 and Obs-2 in the upper layer for several calibration cases are compared with the observed data in Figure 5. For all cases in Figure 5, the mean value of $ln(k)$ for each layer was assigned as the initial value at each pilot point. The solution space
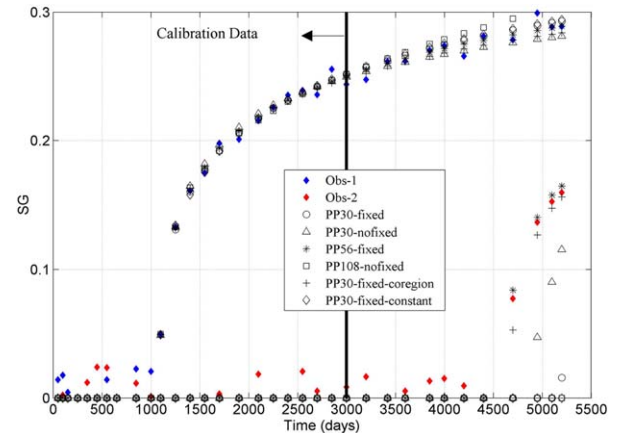

Figure 5. Comparison of CO_2 saturations (S_g) at two observation wells for different calibrated models. Calibration data sets include gas saturation up to 3000 days. In the legend, PP** represents the number of pilot points per each layer, "fixed" includes three fixed pilot points per each layer at one injection and two observation wells, "nofixed" does not, "coregion" generates the porosity distribution using coregionalization, "constant" used a mean porosity value of each layer, and for other cases the porosity distribution of the reference (true) case is used.
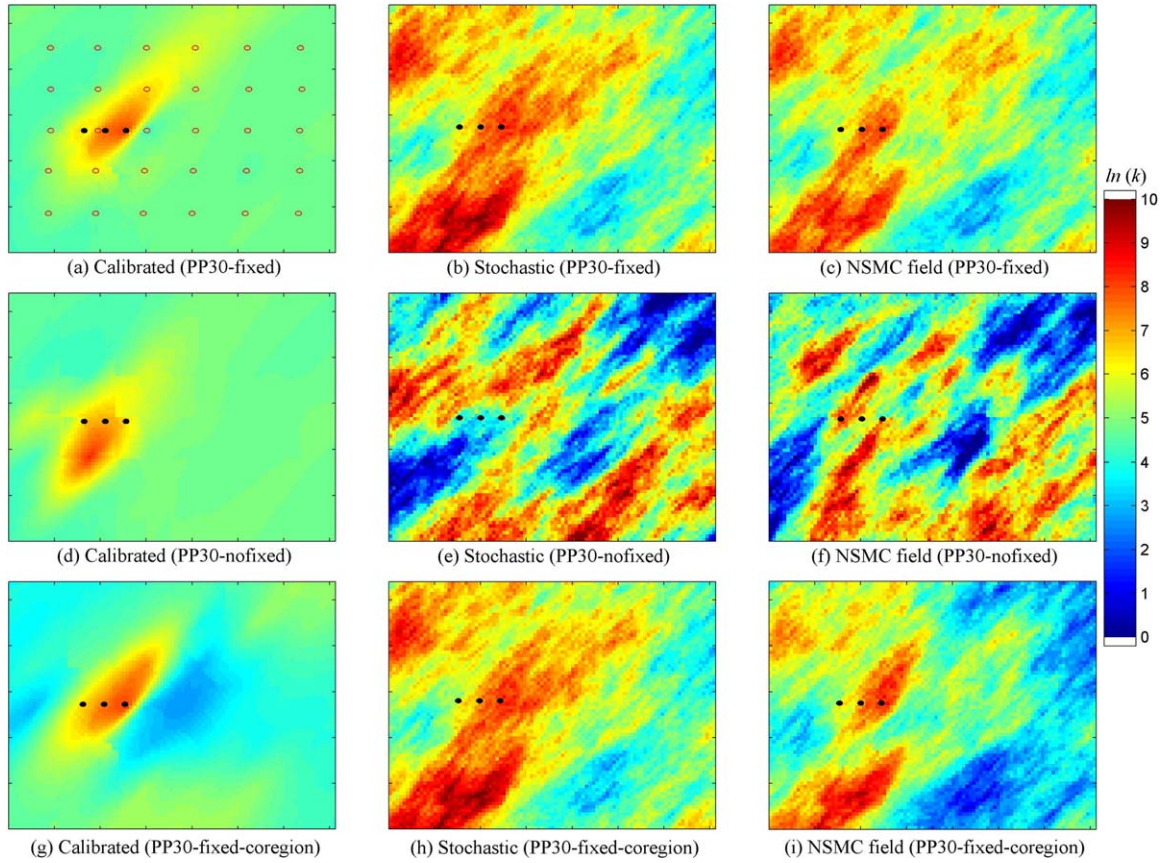
Figure 6. Horizontal $ln(k)$ fields of first layer for (left) calibrated models, (middle) a stochastic random field, and (right) calibration-constrained NSMC field. The red circles in (a) show the location of pilot points where parameter values are estimated.

dimensionality was also selected to be 7 for 30 pilot points (PP30) cases and 9 for the PP56 and PP108 cases, based on the quality of the model fit and the ratio of the singular value to the highest one (10^{-4}).

[32] Figure 5 shows that, regardless of different parameterizations, all calibration results match the S_g profile at Obs-1 very well, even after 3000 days. However, model predictions at Obs-2 are quite different for the investigated six cases. The fact that the case with the largest number of pilot points (PP108) does not predict the S_g profile at Obs-2 reveals the importance of a proper number and location of pilot points [see *Doherty and Hunt*, 2010, for pilot point placement guidelines]. In this particular case, the additional pilot points resulted in point-to-point distances shorter than the characteristic length of $ln(k)$, and a majority of pilot points are redundantly located around Inj-1 and Obs-1, resulting in misrepresentation of the petrophysical properties between Obs-1 and Obs-2. Overall, inclusion of fixed pilot points at the wells improves the prediction of S_g at Obs-2, but it is also affected by other parameterizations. For the porosity distribution, three different approaches have very different predictions, and the true porosity distribution (PP30-fixed) does not have a better prediction than the correlated case (PP30-fixed-coregion).

[33] Calibrated $ln(k)$ fields in layer 1 for three cases are shown in Figure 6. The location of pilot points is shown in Figure 6a. As in the EnKF and ES results, all calibrated fields capture the directional trend of high-permeability values near the well locations, but they have much smoother spatial variability, implying that the calibrated data set does not provide detailed information of the true heterogeneity. It is noted that the PP30-fixed-coregion case (Figure 6g) particularly has the permeability contrast similar to the true case when compared to the PP30-fixed with the true porosity distribution (Figure 6a). As in *Yoon and McKenna* [2012], this comparison shows that the proper relationship of $ln(k)-\phi$ would be more appropriate than adjusting only $ln(k)$ with the fixed porosity distribution.

3.3.2. Prediction Uncertainty Analysis

[34] Three calibrated models with 30 pilot points (Figure 6) were used to generate 200 single calibration-constrained NSMC fields for each case. To evaluate the sampling efficiency of M-NSMC fields, 50 NSMC fields generated from three calibrated models with the PP30-fixed were combined to construct 150 M-NSMC fields. Following a practical strategy to find multiple calibrated fields [*Yoon et al.*, 2013], a single forward model run with 200 prior ensemble fields was performed to compute CO_2 arrival times at Obs-1, and three prior fields spanning the range of the arrival times including fastest, 50th percentile, and slowest cases were selected for the M-NSMC analysis. For model calibration, the initial $ln(k)$ values at pilot points were obtained from the least squares approximation of stochastic random fields as described in *Doherty* [2012].
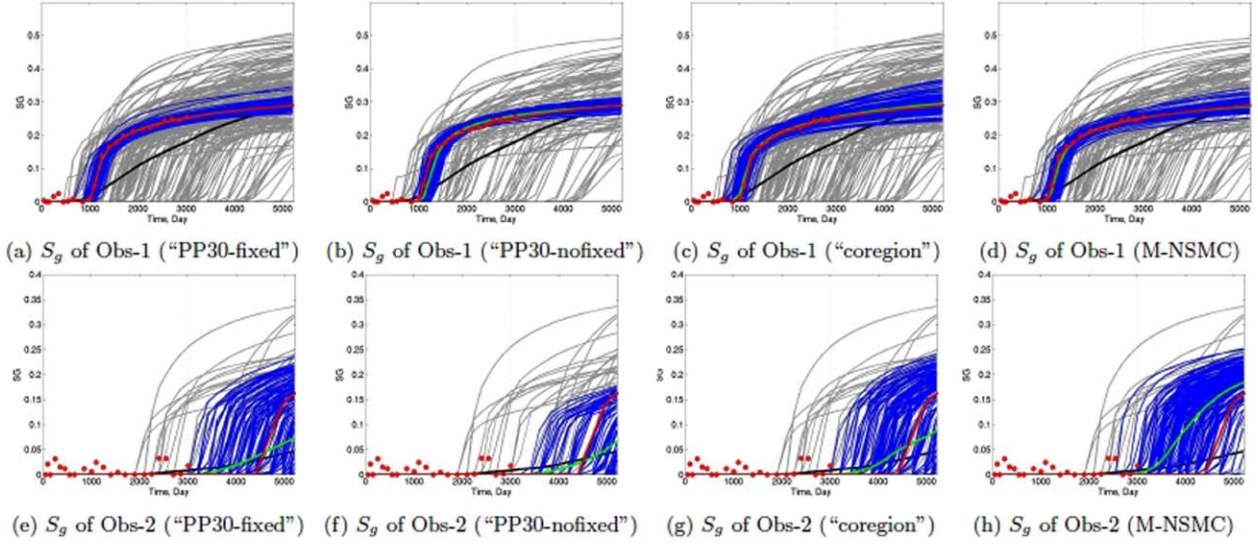
Figure 7. History matching and prediction profiles from the single calibration-constrained NSMC and M-NSMC fields. The meaning of the curves in this figure is similar to Figure 4.

[35] A calibrated field, a stochastic field (one of prior ensemble fields), and a null-space projected field with difference field (i.e., NSMC field) for three single calibration-constrained NSMC cases are shown in Figure 6. Three calibrated models include PP30-fixed, PP30-nofixed, and PP30-fixed-coregion in Figure 5. The stochastic field in Figure 6e is different from the two other fields because it has no fixed pilot point. Compared to the stochastic fields in Figure 6 (middle), all calibrated models (left), and NSMC fields (right) have a permeability contrast around the injection well similar to the true field in Figure 1a. This high contrast in all three calibrated models is partly due to the pilot point location close to the injection well (Figure 6a), which is more clearly reflected in the two cases with the fixed pilot points at the injection well. As discussed, the permeability patterns around the wells are relatively well informed by the calibration data set, highlighting the importance of observation well(s) location and the parameterization for calibration [*Doherty et al.*, 2010b].

[36] S_g profiles at the Obs-1 and Obs-2 for three NSMC cases and one M-NSMC case are shown in Figure 7. All three NSMC fields (blue curves) reduce the prediction uncertainty dramatically compared to the unconstrained prior fields at Obs-1 (Figures 7a–7c). As in posterior ensemble members from the EnKF and ES results (Figure 4), however, predicted S_g profiles at Obs-2 vary considerably. In particular, for the PP30-nofixed case the spread of S_g profiles at Obs-2 is much smaller, and the number of NSMC fields with no breakthrough at 5200 days is higher compared to the two other cases with fixed pilot points. This indicates that hard data (e.g., permeability and porosity data from well logs) together with geological structures (e.g., anisotropy and stratigraphic data) can improve prediction capability significantly. Although Thikonov regularization is not included in this synthetic case, this shows the need to include the regularization.

[37] For the M-NSMC, three calibrated models used to construct M-NSMC fields have very similar $O(m)$ values (equation (11)) and mean arrival times at Obs-1 but different mean arrival times (3420, 4000, and 5100 days) at Obs-2. The predicted S_g profile at Obs-1 has a small spread over the entire simulation period, indicating that M-NSMC fields are constrained by their calibrated fields. However, M-NSMC fields (blue curves) show an earlier mean arrival time and slightly wider distributions of the S_g profile at Obs-2 (Figure 7h) compared to single calibration-constrained NSMC results (Figures 7e–7g). This demonstrates both that M-NSMC fields are able to span the range of observed data that are not used in the model calibration process (e.g., S_g profile at Obs-2 after 3000 days) and that M-NSMC fields have a better sampling efficiency for spanning a range of null-space. The characteristics of the mean $ln(k)$ distribution are similar for both cases in Figure 8 (first and second columns) and the variability of $ln(k)$ distribution of M-NSMC is higher (shown in the next section), possibly causing the wider distributions of the S_g profiles at Obs-2.

3.4. Results of EnKF and ES With Pilot Points

[38] For the EnKF and ES algorithms with pilot points, the same pilot point configurations as in the NSMC framework (i.e., PP30-fixed, 33 pilot points per each layer) were employed (Figure 6a). Therefore, the model parameters vector, m, has a dimension of 132 when both $ln(k)$ and porosity fields are updated compared to a dimension of 27,648 with traditional grid-based implementations. Figure 8 (third column and fourth column) shows the final ensemble mean of $ln(k)$ obtained with the EnKF and ES-MDA algorithms. The pilot points located around the injection and observation wells show similarity to the truth field, whereas the pilot points far from the observed data location do not show any similarity to the reference permeability fields. It is noted that only the values at pilot points are conditioned to the observation data and the rest of model domain are interpolated using kriging without randomizing as in calibration process with the gradient-based method. This may cause very similar mean of $ln(k)$ in the EnKF-PP (Figure 8c) and calibrated model (Figure 6a). The data
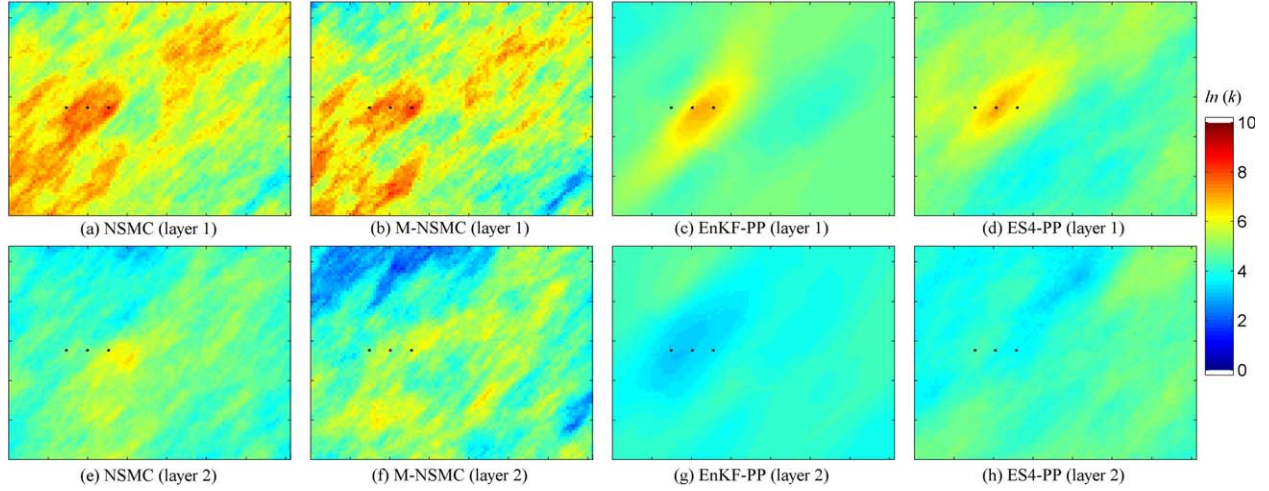
Figure 8. Final updated mean of horizontal $ln(k)$ fields with single calibration-constrained NSMC fields, 150 M-NSMC fields, and EnKF and ES-MDA (ES4) using pilot points. (first row) layer 1 and (second row) layer 2.

matches and future predictions obtained with the EnKF and ES-MDA combined with pilot point (not shown) are also similar to the results of the standard EnKF and ES-MDA presented in Figure 4. These results indicate that considering pilot point parameters in the Kalman filter update give model parameters with higher variability among ensemble members. Moreover, the history matching results of the EnKF-PP and ES-PP (ES4) are similar to the results obtained with the NSMC method (Figure 7). The difference outside the well locations between M-NSMC and ES4-PP (Figures 8b and 8d) is mainly due to the incorporation of the fine-scale heterogeneity in the NSMC fields described earlier.

3.5. Standard Deviation Results

[39] The $ln(k)$ and gas saturation (S_g) distributions from the final ensemble fields are compared to evaluate the spatial patterns of $ln(k)$ fields and their impact on CO_2 plume migration. The SD of $ln(k)$ fields after data assimilation is presented in Figure 9 for selected ensemble-based methods and NSMC methods. Since the SD distributions from the EnKF and ES-MDA are similar, only the EnKF results are shown. Comparison of the EnKF and NSMC methods shows that two different methodologies have similar distributions of SD; the reduction in uncertainty of model parameters (small values of SD) has occurred in the regions close to the injection and observation wells because of the high sensitivity of the observed data to the calibration parameters. The SD values in regions far from the observation locations are close to the ones used to generate prior ensemble members (Table 2). Specifically, the EnKF-PP has a lower range of SD distribution (Figures 9c and 9d) compared to other cases, mainly due to the effect of kriging interpolation. As discussed, the stochastic difference field was incorporated into the null-space projected field for the NSMC methods, resulting in higher SD distributions far from wells compared to the EnKF-PP.

[40] Figure 10 shows the ensemble mean and STD of S_g of the upper layer of reservoir obtained with different
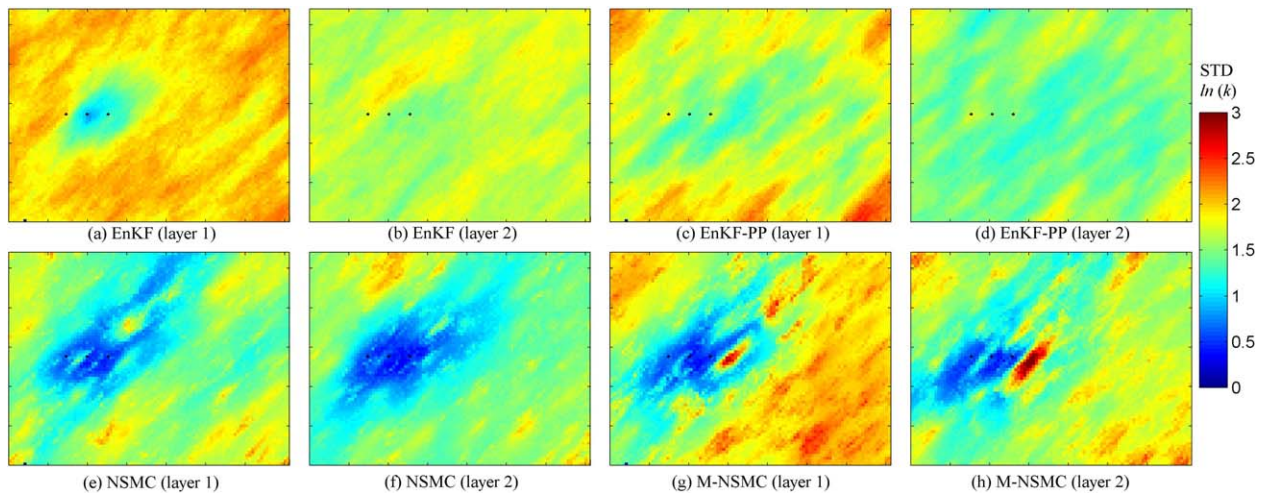

Figure 9. Comparison of the standard deviation (STD) of final horizontal $ln(k)$ field after data assimilation using different algorithms. For cases with pilot points, fixed pilot points are used.
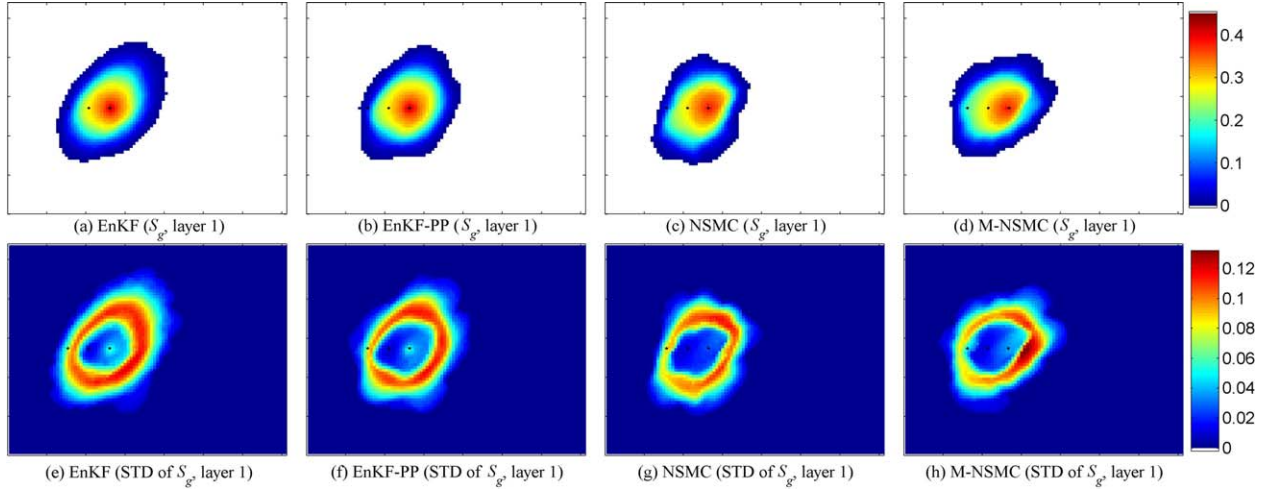
Figure 10. Comparison of the ensemble mean and STD of S_g of layer 1 after data assimilation at time 5200 days. For cases with pilot points, fixed pilot points are used.

algorithms at time 5200 days. All algorithms capture the plume shape around the wells as shown in the true case (Figure 2b), but do not reflect the plume advance in the upper right direction. This comparison indicates the degree of information that can be obtained from the observed data at the observation wells, which is the main reason for relatively high SD distribution far from the wells.

4. Comparison of Ensemble Filtering Algorithms and NSMC Methods

4.1. Metric Comparison

[41] To quantify the results of our different approaches, a normalized objective function using equation (11) is computed after assimilation of all measurements up to 3000 days. The normalized objective function is defined as

$$O_N(m) = \frac{O(m)}{N_d}, \qquad (12)$$

where N_d denotes the total number of measured data ($N_d = 161$). Figure 11 presents the box plot of the normalized objective function, O_N, obtained for each case after data assimilation. These box plots present the results of 200 final ensemble members of key algorithms considered in this study. According to the results in Figure 11, the ES gives high values of the objective function as compared to the EnKF. This difference is a result of the poor data matches (Figures 4b and 4f). Multiple data assimilation with the ES (ES-MDA) improved the quality of data matches, resulting in lower $O_N(m)$ values after four iterations (ES4) than the EnKF. This is similarly observed in the EnKF-PP and ES-PP results. Both the NSMC results show relatively low values and fewer outliers, probably due to good history matching results (Figure 7) of the calibrated models. Overall, ES4 has the lowest mean value of the objective function and both the ES-MDA and M-NSMC approaches have similarly lower objective functions, corresponding to excellent matches of S_g at Obs-1 and Obs-2 until 3000 days (Figures 4d, 4h, 7d, and 7h).

[42] Four additional statistical metrics are evaluated to assess the performance of the algorithms using the final updated models [*Hendricks-Franssen et al*., 2009]. For the state variables (i.e., $ln(k)$, porosity, and S_g), we compute root mean square error (RMSE), ensemble spread (S), mean absolute error (MAE), and mean ensemble standard deviation (MESD) defined as follows:

$$\text{RMSE}(y) = \frac{1}{N_m} \sum_{i=1}^{N_m} \sqrt{\frac{1}{N_e} \sum_{j=1}^{N_e} (y_{i,j} - y_{\text{true},i})^2}, \qquad (13)$$

$$\text{S}(y) = \frac{1}{N_m} \sum_{i=1}^{N_m} \sqrt{\frac{1}{N_e} \sum_{j=1}^{N_e} (y_{i,j} - \bar{y}_i)^2}, \qquad (14)$$
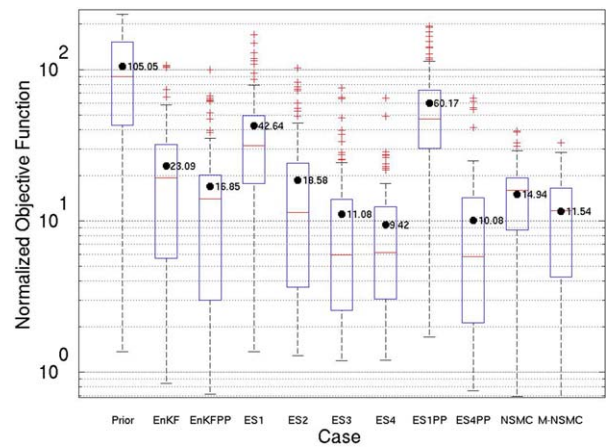

Figure 11. Comparison of the normalized objective function based on 200 final ensemble members for different cases. The line in the box is the median, the bounds of the box are 25% and 75% quantiles, the whiskers are the extremes, the black dot is mean, and the red pluses are outliers. The numbers next to the boxes correspond to the value of the mean.

Table 3. Summary of Final Metric (RMSE, S, MAE, and MESD) Values for $ln(k)$ and S_g of Layer 1 of Observation Wells (Obs-1 and Obs-2) Obtained With Different Algorithms[a]

Algorithm	Log-Permeability				S_g of Obs-1			S_g of Obs-2		
	RMSE	S	MAE	MESD	RMSE	S	MAE	RMSE	S	MAE
Prior	2.457	1.904	1.461	1.913	1.612	1.49	0.352	0.034	0.031	0.015
EnKF	2.38	1.79	1.368	1.802	0.27	0.182	0.174	0.02	0.015	0.012
ES	2.424	1.84	1.37	1.85	0.579	0.47	0.237	0.024	0.021	0.012
ES4	2.347	1.743	1.378	1.748	0.309	0.216	0.180	0.025	0.022	0.011
EnKF-PP	2.091	1.438	1.340	1.441	0.282	0.172	0.179	0.019	0.014	0.012
ES-PP	2.266	1.630	1.378	1.635	0.617	0.488	0.252	0.030	0.026	0.013
ES4-PP	2.243	1.552	1.426	1.556	0.785	0.685	0.233	0.024	0.021	0.01
NSMC	2.334	1.401	1.665	1.408	0.292	0.112	0.266	0.030	0.021	0.0303
M-NSMC	2.509	1.609	1.725	1.615	0.348	0.201	0.279	0.029	0.021	0.019

[a]The prior unconditional metric values are also listed. RMSE: root mean square error, S: spread, MAE: mean absolute error, MESD: mean ensemble standard deviation.

$$\mathrm{MAE}(y) = \frac{1}{N_m} \sum_{i=1}^{N_m} |\bar{y}_i - y_{\mathrm{true},i}|, \qquad (15)$$

$$\mathrm{MESD}(y) = \frac{1}{N_m} \sum_{i=1}^{N_m} \sigma_{y_i}, \qquad (16)$$

where y is state variable, either model parameters m (e.g., $ln(k)$ or porosity) or simulator primary variables (e.g., S_g), $\bar{y}$ is the mean of N_e state vectors, y_{true} is the true state vector, and σ_{y_i} is the standard deviation of state variable y at grid block i. The RMSE and ensemble spread values will be similar if the mean of the model parameters is close to the true values. The smallest value of MESD among different algorithms represents the largest amount of reduction of prior uncertainty. However, this reduction does not necessarily mean a good characterization. The predictive capabilities of the considered algorithms are also measured based on computing RMSE, S, and MAE of S_g at observation wells as follows:

$$\mathrm{RMSE}(S_g) = \frac{1}{N_T} \sum_{i=1}^{N_T} \sqrt{\frac{1}{N_e} \sum_{j=1}^{N_e} \left(S_{g_{i,j}} - S_{g,\mathrm{true}_i} \right)^2}, \qquad (17)$$

$$\mathrm{S}(S_g) = \frac{1}{N_T} \sum_{i=1}^{N_T} \sqrt{\frac{1}{N_e} \sum_{j=1}^{N_e} \left(S_{g_{i,j}} - \bar{S}_{g,i} \right)^2}, \qquad (18)$$

$$\mathrm{MAE}(S_g) = \frac{1}{N_T} \sum_{i=1}^{N_T} |\bar{S}_{g,i} - S_{g,\mathrm{true}_i}|, \qquad (19)$$

where N_T is the total number of steps that observation data are available. In our computation, we consider all data from time zero until the end of prediction phase (5200 days). Table 3 summarizes the final values of different metrics for $ln(k)$ and S_g at Obs-1 and Obs-2 (upper layer). The prior unconditional metric values are also listed. According to the results of Table 3, the differences between the algorithms are generally small. For $ln(k)$ all metrics except MAE values of the NSMC cases are lower than those of the prior, indicating that conditioning on observation data improves the characterization of the $ln(k)$ field. The higher MAE values of the NSMC methods are expected because the regularization term is not included in the objective function and the fine-scale heterogeneity through the null-space projection was incorporated. It should be noted that

in this synthetic case a high fraction of MAE values are originated outside the influence of the well systems. In a real world application it can be dramatically reduced with zonation with prior information.

[43] For all algorithms, the metric values of S_g are substantially reduced, compared to the prior values. Therefore, conditioning on observation data is more beneficial to the characterization of the dynamic variables (such as output at wells) than to the characterization of the static parameters (such as permeability). This is also demonstrated in *Chen and Oliver* [2013] who compared the iterative ES algorithms with and without the model mismatch and showed the reduction of the objective function from the data mismatch. It should be noted that the RMSE and S values of S_g at both observation wells approaches together after data assimilation implying that the forecast estimates are less biased.

4.2. Posterior Distribution of Gas Saturation

[44] Figure 12 shows the probability density function (PDF) of S_g (t = 5100 days) at Obs-1 and Obs-2 to evaluate the predictability of four different algorithms (EnKF, ES-MDA (ES4), NSMC, and M-NSMC). All results are obtained by simulating the final ensemble members from time zero. For all four methods, the highest posterior probability of S_g at Obs-1 matches the truth value (red line), indicating that history data up to 3000 days at Obs-1 were informative to accurately predict the behavior at 5100 days. At Obs-2 the most probable value of S_g deviates slightly from the truth since CO_2 plume did not reach at Obs-2 during history matching period (3000 days), showing lower predictability of final realizations. As indicated in Table 3, other ensemble-based methods are qualitatively similar. Overall, this comparison shows that two computationally efficient methods obtain reasonably good data matches and reliable estimations of uncertainty in predicting future CO_2 plume migration.

4.3. Computational Cost

[45] Table 4 summarizes the computational costs of each algorithm in terms of the number of equivalent simulation runs and the central processing unit (CPU) time to obtain 200 conditional ensemble members. As shown in *Tavakoli et al.* [2012], for the ensemble-based algorithms, the analysis step(s) takes only about 1% of total cumulative time,
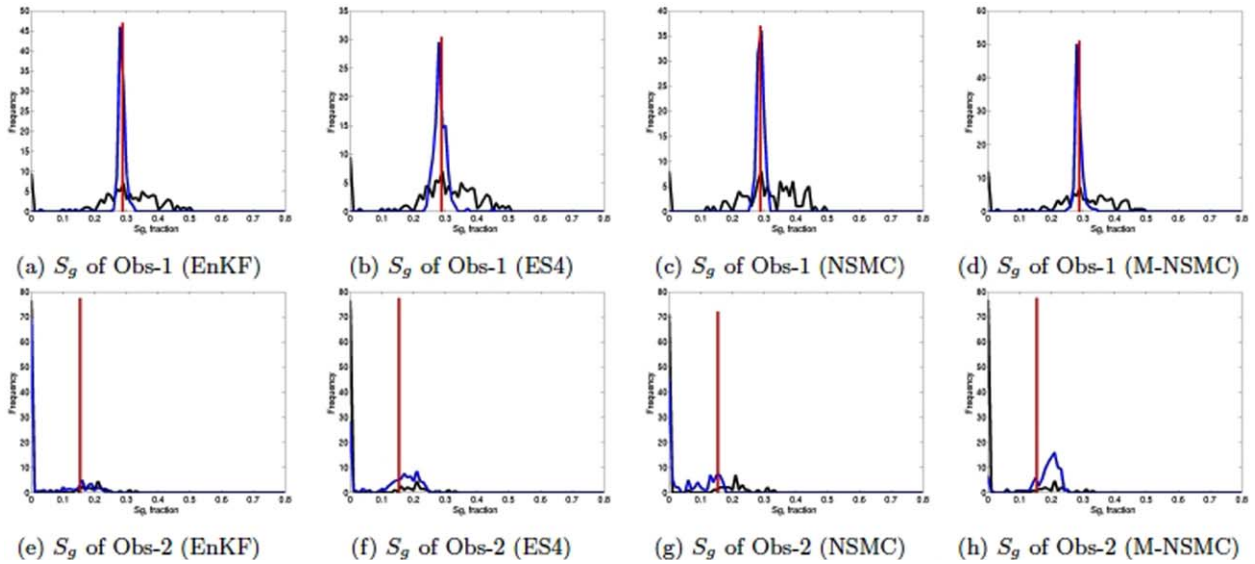
Figure 12. Posterior probability density function (pdf) of gas saturation of (top row) Obs-1 and (bottom row) Obs-2 at time 5100 days. Black curve is prior distribution, blue is posterior distribution, and the red line represents the location of truth. Note that the scale of vertical axes is different for each figure.

and forward simulation runs require 99% of total computational time. For the gradient-based calibration employed in the NSMC analysis, the computational time of matrix calculations was small due to a small number of overall parameters (60) and superparameters (7) in this study. Therefore, we present the computational cost of each algorithm in terms of the number of equivalent simulation runs. One equivalent simulation run is equal to running a reservoir simulator from time zero to the end of an historical period, i.e., 3000 days for the presented example in this paper. The approximate simulation run time for the considered problem in this study is about 80 s using eight processors in a parallel version of the GEM-CMG simulator [*CMG*, 2013]. Note that for an actual CO_2 sequestration test the reservoir simulation requires a very large amount of CPU time, and usually it is not affordable to run a large number of forward simulations (i.e., more than a thousand), which is typical in robust algorithms such as MCMC or particle filter. The measured CPU time of each algorithm is presented relative to the measured CPU time of the ES method, which is the fastest algorithm.

[46] Both the EnKF and ES require 200 simulation runs, even though the EnKF simulation runs are associated with many restarts. The consequence of these restarts is an increase in CPU time for the EnKF application that is 8.4 times larger than the ES. The ES-MDA with four data assimilations (ES4) requires 800 simulation runs and four times the computational cost of the ES method. The use of pilot points in the EnKF and ES-MDA slightly increases the total computational costs such that the CPU time is 9.3 and 4.4 times larger than the ES for the EnKF-PP and ES4-PP, respectively. As described earlier, the first step in the NSMC framework is to obtain a calibrated model that requires the iterative minimization of an objective function. Overall, the calibration process constructs the sensitivity matrix (X) twice (i.e., at the beginning and at the end of cal-

ibration), requiring 122 forward runs. The parameter solution dimensionality was 7, and the minimum objective function values ($O(m)$) were generally achieved within 6–10 iterations, including an additional three consecutive iterations for calibration convergence, which requires 57–133 equivalent simulation runs. For calibration-constrained NSMC fields, one additional iteration was performed for recalibration. For example, the single calibration-constrained 200 NSMC fields require 634 forward runs for recalibration, resulting in a total of 833 runs (Table 4). For M-NSMC, the convergence of three calibrated models required slightly more iterations due to different starting points constructed from the prior random fields. In addition, it requires 100 forward runs to select the initial starting fields. Hence, the overall computational cost of M-NSMC was higher than NSMC, as expected. It should be noted that M-NSMC may require fewer realizations to represent heterogeneity than the NSMC because it contains more heterogeneity from different staring fields. For this example, the NSMC has a computational cost similar to ES4, and M-NSMC has a cost similar to the EnKF.

Table 4. Summary of Computational Cost in Terms of Number of Simulation Runs to Generate 200 Conditional Realizations for Different Algorithms[a]

Algorithm	Equivalent Number of Simulation Runs	Relative CPU Time
EnKF	200	8.4
ES	200	1.0
ES-MDA (ES4)	800	4.0
EnKF-PP	200	9.3
ES4-PP	800	4.4
NSMC	833 (511)	4.2 (2.6)
M-NSMC	1507	7.5

[a]The number in parentheses represents the computational cost for the case with 100 NSMC fields.

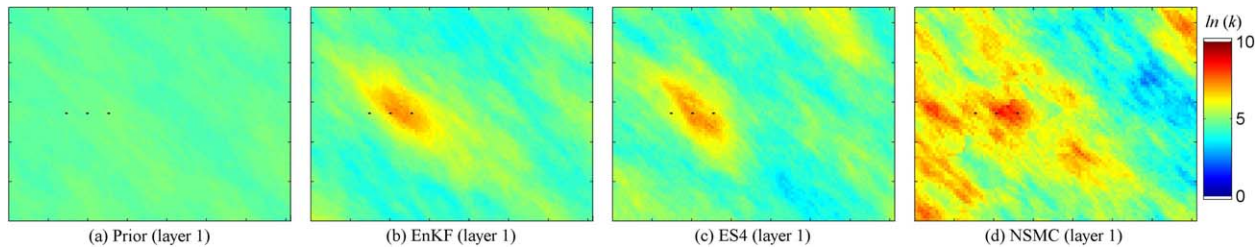(a) Prior (layer 1) (b) EnKF (layer 1) (c) ES4 (layer 1) (d) NSMC (layer 1)

Figure 13. Prior and final mean of horizontal $ln(k)$ fields of first layer obtained with different algorithms using inconsistent prior realizations.

4.4. Inconsistent Prior Realizations

[47] The results presented in the previous sections were obtained using the initial model parameters, i.e., $ln(k)$ fields, with the same geostatistical parameters as the reference model (Table 2). Here, we present the results of an extreme case, where a different anisotropy direction ($\alpha = 135°$) is used to test the capabilities of different algorithms in correcting the structural error. Figure 13 shows the prior (uniform) and final ensemble mean of $ln(k)$ fields ($N_e = 200$) of the first layer of a reservoir obtained with different algorithms. Although the EnKF and ES-MDA (ES4) algorithms have captured high $ln(k)$ patterns of reference model near the wells, they failed to correct the inaccurate structural orientation of the prior ensemble members. Similarly, the NSMC was unable to correct the biased structures. On the other hand, the history matching and prediction results obtained in this case (not shown) are very good and they are comparable to those obtained in the previous cases (Figures 7 and 4). This can be explained by the ill-posed nature of the history matching problems, which implies the existence of many nonunique solutions. In particular, the ensemble mean of the normalized objective function (equation (12)) is 109.3 for the prior realizations and has been reduced to 22.99, 8.78, and 22.4 using the EnKF, ES-MDA (ES4), and NSMC algorithms, respectively. This example shows that the sparse dynamic wells information (e.g., BHP, injection rate, S_g) is not enough on its own to recover the main geostatistical parameters of a reservoir model, regardless of the implemented algorithms (either ensemble based or gradient based). This is also observed in *Jafarpour and Tarrahi* [2011] who extensively investigated the impact of variogram uncertainty on the performance of the EnKF. They suggested that it is preferable to use prior members with a wide range of variability in property description. Other soft data such as 4-D seismic and well logs will also alleviate the impact of the structural uncertainty. For the NSMC methods, the structural uncertainty such as anisotropic direction and ratio can be easily parameterized for model calibration.

5. Summary and Conclusions

[48] We compared two recently developed and computationally efficient approaches including ensemble-based algorithms and calibration-constrained null-space Monte Carlo (NSMC) methods coupled with gradient-based calibration. For ensemble-based algorithms, we implemented ensemble Kalman filter (EnKF), ensemble smoother (ES), ES with multiple data assimilation (ES-MDA), and the EnKF and ES-MDA with the pilot point parameterizations. For the NSMC analysis, single and multiple NSMC (M-NSMC) methods are compared and the SVD-Assist technique is used to reduce the dimension of the calibration solution space during both calibration and recalibration phases. Eight different algorithms from two methodologies were applied for inverse modeling of a 3-D synthetic case with two vertical layers developed to mimic an actual CO_2 inject pilot test. The performance of variants of two algorithms has been evaluated based on a comparison of data matches, evaluation of uncertainty in predictions, prior and final values of several metrics, such as objective function and root mean square error (RMSE), and computational cost. Comparison of results leads to the following conclusions:

[49] 1. Both methodologies yield good history match and reasonable prediction results in a computationally efficient way. In particular, the results with an incorrect prior assumption (i.e., different anisotropy direction) also produce relatively low objective function values for both algorithms and match the observed S_g profile, but do not match the feature of the log-permeability field. This highlights the importance of incorporating the structural uncertainty into prior members with a range of variability or parameterizations for calibration model.

[50] 2. The ES-MDA with four repeated data assimilations (ES4) resulted in a smaller objective function value and better history matches with lower computational cost than the EnKF. The implementation of pilot points with ensemble-based algorithms slightly improved the results in terms of data matches and induced more heterogeneities in model parameters. There is a need to improve computational efficiency of the ES-MDA with large data mismatch cases [*Chen and Oliver*, 2013], which is typical in field CO_2 sequestration sites.

[51] 3. The M-NSMC method gave comparable results to one with the ES-MDA. Although the M-NSMC method is computationally more expensive than single-constrained NSMC, the M-NSMC was able to capture heterogeneity and resulted in slightly better predictions than the NSMC. This work demonstrates that the M-NSMC approach with multiple (e.g., three) starting points can be applied for multiphase flow problems in a computational efficient way. Enhanced parameterization such as regularized pilot points method or Thikonov regularization will improve the applicability of the NSMC for very complex nonlinear problems.

[52] 4. Overall, the ES-MDA has an advantage in terms of computational efficiency, but at the expense of additional computation M-NSMC shows applicability for

highly nonlinear problems such as multiphase flow problems [*Keating et al.*, 2010; *Yoon et al.*, 2013].

[53] To our knowledge, this work is the first comparison of ensemble-based algorithms and NSMC methods on CO_2 sequestration simulations. Both approaches are model-independent and can be applied for complex and CPU-intensive problems where required computational demands are high for forward simulations. In general, the ensemble-based algorithms obtained smoother $ln(k)$ fields than the NSMC methods. The computational efficiency of the NSMC is mainly due to the use of the pilot point and TSVD parameterizations in the context of SVD-Assist. For large-scale problems as in real world applications, the ensemble-based algorithms can be more efficient than the NSMC approaches. However, incorporation of soft data as in the Thikonov regularization and parameterization of structural uncertainty will enhance the NSMC performance without the dramatic increase of computational expense. Unlike the synthetic case considered in this work, hydrogeological parameters are poorly known at most of actual CO_2 sequestration fields, and knowledge of geological structures is also limited. Hence, it will be critical to quantify the uncertainty of this information and evaluate the impact of the uncertainty (e.g., through the covariance of the parameters) on prediction uncertainty analysis in order to support the decision-making process. In the future work, the framework developed in this study will be applied to a CO_2 injection test at Cranfield, Mississippi as a real world test case.

References

Aanonsen, S. I., G. Nævdal, D. S. Oliver, A. C. Reynolds, and B. Vallès (2009), The ensemble Kalman filter in reservoir engineering—A Review, *SPE J.*, *14*(3), 393–412, doi:10.2118/117274-PA.

Agbalaka, C., and D. S. Oliver (2008), Application of the EnKF and localization to automatic history matching of facies distribution and production data, *Math. Geosci.*, *40*(4), 353–374.

Alcolea, A., J. Carrera, and M. Medina (2006), Inversion of heterogeneous parabolic type equations using the pilot points method, *Int. J. Numer. Methods Fluids*, *51*, 963–980, doi:10.1002/fld.1213.

Bertino, L., G. Evensen, and H. Wackernagel (2003), Sequential data assimilation techniques in oceanography, *Int. Stat. Rev.*, *71*(2), 223–241.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*(3), 279–298, doi:10.1002/hyp.3360060305.

Bonet-Cunha, L., D. S. Oliver, R. A. Redner, and A. C. Reynolds (1998), Hybrid Markov chain Monte Carlo method for generating permeability fields conditioned to multiwell pressure data and prior information, *SPE J.*, *11*(3), 261–271.

Chen, S., G. Li, and A. C. Reynolds (2006), An in situ test for the estimation of relative permeabilities: An approximate analytical solution for the wellbore pressure, in TUPREP Res. Rep. 23, The Univ. of Tulsa, Tulsa, Okla.

Chen, Y., and D. S. Oliver (2013), Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification, *Comput. Geosci.*, *17*, 689–703, doi:10.1007/s10596-013-9351-5.

CMG (2013), *Computer Modeling Group*, Calgary, Alberta, Canada. [Available at http://www.cmgl.ca/.]

Collins, D. A., L. X. Nghiem, Y. K. Li, and J. E. Grabonstotter (1992), An efficient approach to adaptive-implicit compositional simulation with an equation of state (SPE 15133), *SPE Reservoir Eng.*, *7*(2), 259–264.

Crestani, E., M. Camporese, D. Bau, and P. Salandin (2013), Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation, *Hydrol. Earth Syst. Sci.*, *17*, 1517–1531, doi:10.5194/hess-17-1517-2013.

de Marsily, G., G. Lavedan, M. Boucher, and G. Fasanino (1984), Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model, in *Geostatistics for Natural Resources Characterization*, Part 2, pp. 831–849, D. Reidel, London.

de Marsily, G., J. P. Delhomme, F. Delay, and A. Buoro (1999), 40 years of inverse problems in hydrogeology, *Earth Planet. Sci.*, *329*(2), 73–87.

de Marsily, G., J. P. Delhomme, A. Coudrain-Ribstein, and A. M. Lavenue (2000), Four decades of inverse problems in hydrogeology, in *Theory, Modeling and Field Investigation in Hydrogeology: A Special Volume in Honor of Shlomo P. Neuman's 60th Birthday*, edited by D. Zhang and C. L. Winters, *Geol. Soc. of Am. Spec. Pap. 348*, pp. 1–17.

Delshad, M., X. Kong, R. Tavakoli, S. A. Hosseini, and M. F. Wheeler (2013), Modeling and simulation of carbon sequestration at Cranfield incorporating new physical models, *Int. J. Greenhouse Gas Control.*, *18*, 463–473.

Doherty, J. (2003), Ground water model calibration using pilot points and regularization, *Ground Water*, *41*(2), 170–177.

Doherty, J. (2012), *PEST: Model Independent Parameter Estimation*, Watermark Numer. Comput., Corinda, Queensland, Australia. [Available at http://www.pesthomepage.org.]

Doherty, J., and R. J. Hunt (2010), Approaches to highly parameterized inversion: A guide to using PEST for groundwater-model calibration, Tech. Rep., U.S. Geol. Surv. Sci. Invest. Rep. 2010–5169, U.S. Geol. Surv., Middleton, Wis.

Doherty, J., M. N. Fienen, and R. J. Hunt (2010a), Approaches to highly parameterized inversion: Pilot-point theory, guidelines, and research directions, Tech. Rep., U.S. Geol. Surv. Sci. Invest. Rep. 2010–5168, U.S. Geol. Surv., Middleton, Wis.

Doherty, J., R. J. Hunt, and M. J. Tonkin (2010b), Approaches to highly parameterized inversion: A guide to using PEST for model-parameter and predictive-uncertainty analysis, Tech. Rep., U.S. Geol. Surv. Sci. Invest. Rep. 2010–5211, U.S. Geol. Surv., Middleton, Wis.

Doucet, A., and A. M. Johansen (2009), A tutorial on particle filtering and smoothing: Fifteen years later, in *Oxford Handbook of Nonlinear Filtering*, vol. 12, edited by D. Crisan and B. Rozovsky, pp. 656–704, Oxford Univ. Press.

Doucet, A., N. D. Freitas, and N. Gordon (2001), *Sequential Monte Carlo Methods in Practice,* Stat. for Eng. and Inf. Sci., Springer, New York.

Dovera, L., and E. D. Rossa (2011), Multimodal ensemble Kalman filtering using Gaussian mixture models, *Comput. Geosci.*, *15*(2), 307–323.

Elsheikh, A. H., M. D. Jackson, and T. C. Laforce (2012), Bayesian reservoir history matching considering model and parameter uncertainties, *Math. Geosci.*, *44*(5), 515–543, doi:10.1007/s11004-012-9397-2.

Emerick, A., and A. Reynolds (2010), Combining sensitivities and prior information for covariance localization in the ensemble Kalman filter for petroleum reservoir applications, *Comput. Geosci.*, *15*(2), 251–269.

Emerick, A., and A. Reynolds (2012), Ensemble smoother with multiple data assimilation, *Comput. Geosci.*, *55*, 3–15, doi:10.1016/j.cageo.2012.03.011.

Emerick, A., and A. Reynolds (2013), Investigation of the sampling performance of ensemble-based methods with a simple reservoir model, *Comput. Geosci.*, *17*, 1–26.

Espinet, A. J., and C. Shoemaker (2013), Comparison of optimization algorithms for parameter estimation of multi-phase flow models with application to geological carbon sequestration, *Adv. Water Resour.*, *54*, 133–148.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162.

Evensen, G. (2004), Sampling strategies and square root analysis schemes for the EnKF, *Ocean Dynamics*, *54*, 539–560, doi:10.1007/s10236-004-0099-2.

Evensen, G., and B. Laurent (2005), The combined parameter and state estimation problem, *Computational Geosciences*, 1–39.

Evensen, G., and P. J. van Leeuwen (2000), An ensemble Kalman smoother for nonlinear dynamics, *Mon. Weather Rev.*, *128*(6), 1852–1867.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (Eds.) (1996), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, New York.

Gómez-Hernández, J. J., A. Sahuquillo, and J. E. Capilla (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 1. Theory, *J. Hydrol.*, *203*, 162–174.

Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, Oxford Univ. Press, New York.

Gu, Y., and D. S. Oliver (2007), An iterative ensemble Kalman filter for multiphase fluid flow data assimilation, *SPE J.*, *12*(4), 438–446.

Harvey, C. F., and S. M. Gorelick (1995), Mapping hydraulic conductivity: Sequential conditioning with measurements of solute arrival time, hydraulic head, and local conductivity, *Water Resour. Res.*, *31*(7), 1615–1626.

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*(1), 97–109.

He, J., P. Sarma, and L. Durlofsky (2011), Use of reduced-order models for improved data assimilation within an EnKF context (SPE-141967), paper presented at SPE Reservoir Simulation Symposium, Woodlands, Tex.

Heidari, L., V. Gervais, M. L. Ravalec, and H. Wackernagel (2013), History matching of petroleum reservoir models by the ensemble Kalman filter and parameterization methods, *Comput. Geosci.*, *55*, 84–95, doi:10.1016/j.cageo.2012.06.006.

Hendricks-Franssen, H., A. Alcolea, M. Riva, M. Bakr, N. van der Wiel, F. Stauffer, and A. Guadagnini (2009), A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments, *Adv. Water Resour.*, *32*(6), 851–872, doi:10.1016/j.advwatres.2009.02.011.

Hosseini, S. A., H. Lashgari, J. W. Choi, J. P. Nicot, J. Lu, and S. D. Hovorka (2012), Static and dynamic reservoir modeling for geological CO2 sequestration at Cranfield, Mississippi, U.S.A., *Int. J. Greenhouse Gas Control*, *18*, 449–462, doi:10.1016/j.ijggc.2012.11.009.

Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filtering technique, *Mon. Weather Rev.*, *126*, 796–811.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen (2005), Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations, *Mon. Weather Rev.*, *133*(3), 604–620.

Hunt, R. J., J. Luchette, W. A. Shreuder, J. Rumbaugh, J. Doherty, M. J. Tonkin, and D. Rumbaugh (2010), Using the cloud to replenish parched groundwater modeling efforts, *Ground Water*, *48*(3), 360–365, doi:10.1111/j.1745–6584.2010.00699.x.

Jafarpour, B., and D. B. McLaughlin (2009), Estimating channelized reservoir permeabilities with the ensemble Kalman filter: The importance of ensemble design, *SPE J.*, *14*(2), 374–388.

Jafarpour, B., and M. Tarrahi (2011), Assessing the performance of the ensemble Kalman filter for subsurface flow data integration under variogram uncertainty, *Water Resour. Res.*, *47*, W05537, doi:10.1029/2010WR009090.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, *82*, 35–45.

Keating, E. H., J. Doherty, and Q. Kang (2010), Optimization and uncertainty assessment of strongly non-linear groundwater models with high parameter dimensionality, *Water Resour. Res.*, *46*, W10517, doi:10.1029/2009WR008584.

Keppenne, C. L. (2000), Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter, *Mon. Weather Rev.*, *128*(6), 1971–1981.

Kitanidis, P. (1996), On the geostatistical approach to the inverse problem, *Adv. Water Resour.*, *19*(6), 333–342, doi:10.1016/0309–1708(96)00005-x.

Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, *31*, 2411–2419.

Lépine, O. J., R. C. Bissell, S. I. Aanonsen, I. C. Pallister, and J. W. Barker (1999), Uncertainty analysis in predictive reservoir simulation using gradient information, *SPE J.*, *4*(3), 251–259.

Liu, N., and D. S. Oliver (2003), Evaluation of Monte Carlo methods for assessing uncertainty, *SPE J.*, *8*(2), 188–195.

Medina, A., and J. Carrera (2003), Geostatistical inversion of coupled problems: Dealing with computational burden and different types of data, *J. Hydrol.*, *281*(4), 251–264.

Moore, C., and J. Doherty (2005), Role of the calibration process in reducing model predictive error, *Water Resour. Res.*, *41*, 1–14, doi:10.1029/2004WR003501.

Moreno, D., and S. I. Aanonsen (2007), Stochastic facies modelling using the level set method, in *Extended Abstracts Book of Petroleum Geostatistics 2007*, vol. A18, Cascais, Portugal, EAGE Publications BV, Utrecht, Netherlands.

Nicot, J. P., J. Choi, T. Meckel, C. Y. Chang, S. D. Hovorka, and S. Solano (2009), Results of numerical investigations at SECARB Cranfield, MS fielt test site, in *Eighth Annual Conference on Carbon Capture and Sequestration*, 11 p., Dep. of Energy/Natl. Energy Technol. Lab.

Nowak, W. (2009), Best unbiased ensemble linearization and the quasi-linear Kalman ensemble generator, *Water Resour. Res.*, *45*, W04431, doi:10.1029/2008WR007328.

Oladyshkin, S., C. Holger, and W. Nowak (2013), Bayesian updating via bootstrap filtering combined with data-driven polynomial chaos expansions: Methodology and application to history matching for carbon dioxide storage in geological formations, *Comput. Geosci.*, *17*(4), 671–687, doi:10.1007/s10596-013-9350-6.

Oliver, D. S., and Y. Chen (2010), Recent progress on reservoir history matching: A review, *Comput. Geosci.*, *15*(1), 185–221.

Oliver, D. S., L. B. Cunha, and A. C. Reynolds (1997), Markov chain Monte Carlo methods for conditioning a permeability field to pressure data, *Math. Geol.*, *29*(1), 61–91.

Oliver, D. S., A. C. Reynolds, and N. Liu (2008), *Inverse Theory for Petroleum Reservoir Characterization and History Matching*, Cambridge Univ. Press, Cambridge, U. K.

Peng, D. Y., and D. B. Robinson (1976), A new two-constant equation of state, *Ind. Eng. Chem. Fundam.*, *15*(1), 59–64.

RamaRao, B. S., A. M. LaVenue, G. de Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, 1. Theory and computational experiments, *Water Resour. Res.*, *31*, 475–493.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, *130*(1), 103–114.

Reynolds, A. C., M. Zafari, and G. Li (2006), Iterative forms of the ensemble Kalman filter, paper presented at 10th European Conference on the Mathematics of Oil Recovery, Amsterdam, .

Saad, G., and R. Ghanem (2009), Characterization of reservoir simulation models using a polynomial chaos-based ensemble Kalman filter, *Water Resour. Res.*, *45*, W04417, doi:10.1029/2008WR007148.

Sarma, P., and W. H. Chen (2009), Generalization of the ensemble Kalman filter using kernels for non-Gaussian random fields, paper presented at the SPE Reservoir Simulation Symposium, SPE 119177, Woodlands, Tex., 2–4 Feb.

Shah, P. C., G. R. Gavalas, and J. H. Seinfeld (1978), Error analysis in history matching: The optimum level of parameterization, *Soc. Pet. Eng. J.*, *18*(6), 219–228.

Skjervheim, J.-A., G. Evensen, J. Hove, and J. G. Vabo (2011), An ensemble smoother for assisted history matching, paper presented at the SPE Reservoir Simulation Symposium, SPE 141929, Woodlands, Tex., 21–23 Feb.

Smith, A., and A. Gefland (1992), Bayesian statistics without tears: A sampling-resampling perspective, *Am. Stat.*, *46*(2), 84–88.

Soave, G. (1972), Equilibrium constants from a modified Redlich-Kwong equation of state, *Chem. Eng. Sci.*, *27*(6), 1197–1203.

Tarantola, A. (1987), *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, Amsterdam.

Tavakoli, R., and A. A. Reynolds (2011), Monte Carlo simulation of permeability fields and reservoir performance predictions with SVD parameterization in RML compared with EnKF, *Comput. Geosci.*, *15*(1), 99–116.

Tavakoli, R., G. Pencheva, M. F. Wheeler, and B. Ganis (2012), A parallel ensemble-based framework for reservoir history matching and uncertainty characterization, *Comput. Geosci.*, *17*(1), 83–97.

Tonkin, M., and J. Doherty (2009), Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, *Water Resour. Res.*, *45*, W00B10, doi:10.1029/2007WR006678.

Tonkin, M. J., and J. Doherty (2005), A hybrid regularized inversion methodology for highly parameterized environmental models, *Water Resour. Res.*, *41*, W10412, doi:10.1029/2005WR003995.

van Leeuwen, P. J., and G. Evensen (1996), Data assimilation and inverse methods in terms of a probabilistic formulation, *Mon. Weather Rev.*, *124*, 2898–2913.

Vrugt, J. A., C. J. F. T. Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing

hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, *44*, W00B09, doi:10.1029/2007WR006720.

Welter, D. E., J. Doherty, R. Hunt, C. Muffels, M. Tonkin, and W. Schreder (2012), Approaches in Highly Parameterized Inversion: PEST++, A Parameter Estimation Code Optimized for Large Environmental Models, book 7, section C5, technical report, U.S. Geol. Surv. Tech. and Methods.

Woodbury, A., and T. Ulrych (2000), A full-Bayesian approach to the groundwater inverse problem for steady state flow, *Water Resour. Res.*, *36*, 2081–2093, doi:10.1029/2000WR900086.

Yoon, H., and S. A. McKenna (2012), Highly parameterized inverse estimation of hydraulic conductivity and porosity in a three-dimensional, heterogeneous transport experiment, *Water Resour. Res.*, *48*, W10536, doi:10.1029/2012WR012149.

Yoon, H., D. B. Hart, and S. A. McKenna (2013), Parameter estimation and predictive uncertainty in stochastic inverse modeling of groundwater flow: Comparing null-space Monte Carlo and multiple starting point methods, *Water Resour. Res.*, *49*, 536–553, doi:10.1002/WRCR.20064.

Zafari, M., and A. C. Reynolds (2007), Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter, *SPE J.*, *12*(3), 382–391.

Zhang, Y., and N. Sahinidis (2013), Uncertainty quantification in CO2 sequestration using surrogate models from polynomial chaos expansion, *Ind. Eng. Chem. Res.*, *52*(9), 3121–3132.

Zhou, H., J. Gomez-Hernandez, H.-J. H. Franssen, and L. Li (2011), An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering, *Adv. Water Resour.*, *34*, 844–864.

Zimmerman, D. A., et al. (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resour. Res.*, *34*, 1373–1413.