RESEARCH ARTICLE

10.1002/2013WR014830

Key Points:

- Chemical data are important for constraining flow models
- Dual estimation scheme outperforms the standard joint estimation scheme

Correspondence to:

I. Hoteit,
ibrahim.hoteit@kaust.edu.sa

Citation:

Gharamti, M. E., A. Kadoura, J. Valstar, S. Sun, and I. Hoteit (2014), Constraining a compositional flow model with flow-chemical data using an ensemble-based Kalman filter, *Water Resour. Res.*, 50, 2444–2467, doi:10.1002/2013WR014830.

Received 1 OCT 2013

Accepted 4 MAR 2014

Accepted article online 10 MAR 2014

Published online 19 MAR 2014

Constraining a compositional flow model with flow-chemical data using an ensemble-based Kalman filter

M. E. Gharamti¹, A. Kadoura¹, J. Valstar², S. Sun^{1,3}, and I. Hoteit^{1,3}
¹Department of Earth Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, ²Subsurface and Groundwater Systems, Utrecht, Netherlands, ³Department of Applied Mathematics and Computational Sciences, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Abstract Isothermal compositional flow models require coupling transient compressible flows and advective transport systems of various chemical species in subsurface porous media. Building such numerical models is quite challenging and may be subject to many sources of uncertainties because of possible incomplete representation of some geological parameters that characterize the system's processes. Advanced data assimilation methods, such as the ensemble Kalman filter (EnKF), can be used to calibrate these models by incorporating available data. In this work, we consider the problem of estimating reservoir permeability using information about phase pressure as well as the chemical properties of fluid components. We carry out state-parameter estimation experiments using joint and dual updating schemes in the context of the EnKF with a two-dimensional single-phase compositional flow model (CFM). Quantitative and statistical analyses are performed to evaluate and compare the performance of the assimilation schemes. Our results indicate that including chemical composition data significantly enhances the accuracy of the permeability estimates. In addition, composition data provide more information to estimate system states and parameters than do standard pressure data. The dual state-parameter estimation scheme provides about 10% more accurate permeability estimates on average than the joint scheme when implemented with the same ensemble members, at the cost of twice more forward model integrations. At similar computational cost, the dual approach becomes only beneficial after using large enough ensembles.

1. Introduction

1.1. Background on Inverse Methodology in Subsurface Reservoirs

Subsurface models, such as groundwater flow, multiphase compositional flow, and contaminant transport models, are widely used to describe the physics and dynamics of subsurface hydrological systems. Such numerical models are often based on many physical assumptions and approximations, such as the use of the "simplified" Darcy's law to represent reservoir models [Bailey and Baú, 2010]. Aside from the physical constraints, the parameters that define subsurface processes (e.g., permeability and porosity) are poorly understood, often involving large uncertainties. The simple and deterministic structure of these hydrological models reduces their accuracy and affects their capability to simulate responses of the system they are designed to represent. Assessing the quality of the output of these models then becomes a challenge [Hendricks-Franssen and Kinzelbach, 2008; Gharamti et al., 2013].

Inverse modeling is often used as a way to reduce the uncertainties associated with subsurface reservoir models. The idea behind the inverse approach is to find a set of calibrated, time-invariant, model parameters using optimization-based techniques such as the pilot point method (PPM), the self-calibrated method (SCM), orthogonal matching pursuit, and the representer method (RM) [e.g., Oliver et al., 1997; Gupta et al., 1998; Thiemann et al., 2001; Valstar, 2001; Alcolea et al., 2006; Altaf et al., 2013; Elsheikh et al., 2013a, 2013b]. These inverse methods, which are generally adjoint-based and/or Monte Carlo (MC)-type approaches, seek for some model parameters that minimize an objective function of the long-term model-data misfits based on historical batches of pressure data, phase saturations, contaminant concentrations, etc. In the RM, for instance, one needs to establish a set of nonlinear Euler-Lagrange equations that constitute the gradients of the objective functions and then to solve them using a set of unknown basis functions called representer. As stated by Valstar et al. [2004], the representer is the cross-covariance terms between the measured data points and the unknown state variables. The RM has a nice feature of acting as a smoother, carrying information both forward and backward in time, although this requires building a tangent linear model and

its adjoint. The PPM, on the other hand, is a geostatistical calibration technique that aims at estimating the parameters values using a set of arbitrary pilot points that best fit the aquifer's response as measured by available observations [Alcolea *et al.*, 2006]. Recently, Altaf *et al.* [2013] and Elsheikh *et al.* [2013a] proposed ensemble-based and cost-efficient approaches to compute approximate gradients. Despite their wide use, these inverse methods have some limitations and may suffer from slow convergence rates. One primary hurdle is the extensive computational burden and data storage as in the RM case where the nonlinear nature of Euler-Lagrange equations generally requires an intensive iterative procedure that is equivalent to two model runs for each additional measurement [Valstar *et al.*, 2004]. Additionally, inverse methodologies cannot easily take into account different sources of uncertainty, other than those associated with the inverted parameters. One could also cite the lack of an efficient way to update the estimated parameters with new incoming observations. Additionally, the convergence of such techniques is not always guaranteed.

Another inverse approach is based on statistical conditioning and, unlike the optimization methods, it relies on the correlation between the system parameters and the dynamic fluxes to adjust the parameter values. This stochastic inverse modeling, also known as data assimilation, has the advantage of accounting for model errors that are not only present in the uncertain parameters but also in the model structure and inputs, such as external forcings [Hendricks-Franssen and Kinzelbach, 2008]. Data assimilation can be described as a Bayesian estimation problem in which prior information about a dynamical system is combined with available observations to produce an updated posterior estimate of the parameters and state variables [Hoteit *et al.*, 2012]. Various data assimilation schemes have been developed to tackle estimation problems in meteorology, oceanography, and hydrology. These schemes vary in structure and complexity depending on the nonlinearity, complexity, and dimension of the system. An assimilation scheme that recently drew attention is the particle filter (PF), which is capable of handling any type of statistical distribution, Gaussian or not, making it well suited for strongly nonlinear systems [Montzka *et al.*, 2011; Chang *et al.*, 2012; Moradkhani *et al.*, 2012]. The PF requires, however, a large set of particles to accurately sample the distribution of the state and parameters, making this scheme computationally intensive for large-scale hydrological applications. A more popular assimilation technique for such applications is the ensemble Kalman filter (EnKF). The EnKF, as described by Burgers *et al.* [1998], operates sequentially following two main steps, the forecast step and the analysis step. In the forecast step, an ensemble of state vectors is integrated forward in time using a nonlinear prediction model. Once data become available, an update, based on the Bayes rule in a Gaussian framework, i.e., the Kalman filter update, is made [Evensen, 2003].

Despite its Kalman-based update step, the EnKF and its (square-root) variants have been successfully used in many hydrology and reservoir applications [e.g., Naevadal *et al.*, 2003; Moradkhani *et al.*, 2005; Chen and Zhang, 2006; Gu and Oliver, 2006; Wen and Chen, 2007; Hendricks-Franssen and Kinzelbach, 2008; Sun *et al.*, 2009; Bailey and Baú, 2010; Gharamti *et al.*, 2013; Xu *et al.*, 2013; Gharamti and Hoteit, 2014]. These previous studies agreed that the EnKF is an efficient estimation algorithm because of its central processing unit (CPU) efficiency, its robustness against nonlinearities, and its ease of implementation and flexibility. The EnKF further readily provides a general framework for state-parameter estimation. Some studies compared the performance of the EnKF to optimization-based techniques, e.g., Hendricks-Franssen and Kinzelbach [2009], and concluded that the EnKF can be as accurate as the SCM while imposing a significantly smaller computational burden.

Several studies recently considered the EnKF in the context of non-Gaussian distributions to further enhance its performances. For instance, Zhou *et al.* [2011] and Schöniger *et al.* [2012] suggested an additional transformation step to the state variables and parameters to ensure that the distribution was Gaussian before applying the filter update. Other studies proposed to constrain the EnKF solution by the physical laws of the system [Wang *et al.*, 2009], to use an efficient iterative formulation of the EnKF [Wen and Chen, 2007; Hendricks-Franssen and Kinzelbach, 2008], or to base the filter update on Gaussian-mixture distributions [Hoteit *et al.*, 2008, 2012].

1.2. Motivation and Objectives

Contamination of the subsurface environment by nonaqueous phase liquids (NAPL), such as gasoline and other petroleum products and industrial wastes is a serious hazard to groundwater resources. Such contamination generally leads to a remediation process that requires attention in hydrology and reservoir

engineering. Beyond remediation, storing CO₂ in water-flooded oil reservoirs while increasing oil recovery is a topic of interest. In petroleum engineering, secondary oil recovery is performed by increasing a reservoir's pressure by injecting water or natural gas at the bottom of the active well [Moortgat *et al.*, 2011]. A critical step in understanding the impact of subsurface release of NAPL and remediation and oil recovery is to model fluid migration, transport, and the fate of the essential constituents. Several numerical models designed to capture these processes have been developed with various levels of sophistication. Reeves *et al.* [1994] proposed an iterative compositional formulation to model nonequilibrium phase partitioning and to provide a flexible framework for organic mixtures with different numbers of components. They found that iterative compositional models yield more accurate solutions than those derived from standard compositional formulations. Panday *et al.* [1995] presented a comprehensive nonisothermal compositional model for rigorous treatment of strongly coupled and highly nonlinear flow and transport applications. Analysis of their model formulation suggests that if interface mass transfer is neglected, then the governing equations are simplified to the standard, but substantially faster, multiphase flow scenario. Higher-order compositional modeling of three-phase flow in three-dimensional (3-D) fractured porous media was presented by Moortgat and Firoozabadi [2013]. To achieve stable solutions, the authors properly adopted the mixed-hybrid finite element method to approximate the convective Darcy fluxes and pressure equation.

Despite the broad applications in hydrocarbon reservoir engineering and hydrology, data assimilation has not yet been applied to compositional flow models. History matching studies have so far relied on pressure, phase saturation and seismic data. Yet the use of chemical composition data of the fluid species to improve the prediction of compositional models has not yet been investigated. There are good reasons to believe, however, that integrating such chemical information into an EnKF assimilation system on top of other data sets would improve the calibration of reservoir parameters. First, fluid components are usually characterized by certain critical pressure and temperature values that make them behave differently depending on the conditions of the reservoir. Second, the molecular weight of the fluid components can be highly variable and this generally affects the migration of the fluid because of gravitational forces. Moreover, the chemical information that can be retrieved from the fluid composition can be essential to calibrate the thermodynamic state of the reservoir. A better response representation may be obtained from the models if they include fluid composition data. Assessing the relevance of the chemical composition data to reservoir history matching is one of the main goals of this study.

The EnKF can be used in compositional flow models to estimate the dynamic state variables and the most relevant hydrological parameters following a joint or a dual state-parameter estimation framework. In the joint approach, the parameters and the state variables are concurrently estimated in a single augmented vector [Naevadal *et al.*, 2003; Chen and Zhang, 2006; Hendricks-Franssen and Kinzelbach, 2009; Li *et al.*, 2012]. The dual approach employs two parallel filters, one operating on the parameters and the other on the state [Nelson, 1999; Moradkhani *et al.*, 2005; Lü *et al.*, 2011]. It remains unclear from the assimilation literature which of the dual or the joint approach is more suitable for history matching. In a recent study, Gharamti *et al.* [2013] introduced the dual approach for the estimation of the states (not state parameters) of one-way coupled models and found that a dual filtering scheme provides a more efficient framework than joint approach for estimating the states of a coupled flow and contaminant transport hydrological system.

In this work, we explore the feasibility of employing the EnKF to condition reservoir states and parameters using a two-dimensional (2-D) isothermal compositional flow model for oil recovery. Assimilation experiments are carried out using joint and dual estimation schemes. Estimates of pressure, fluid composites, and subsurface permeability from both schemes are compared on the basis of method efficiency and robustness. The importance of assimilating two sets of system response data is demonstrated. Specifically, pressure measurements and chemical data of the fluid components are used to condition the permeability field in a single-phase compositional flow system. Different observational scenarios are tested to infer the significance of the assimilated data sets. The effect of the location of the production site on the accuracy of the permeability and state estimates is also studied. The remainder of this paper is organized as follows. Section 2 presents the compositional flow model. Section 3 briefly describes the EnKF algorithm and the joint and dual state-parameter estimation strategies. In section 4, synthetic assimilation experiments with various observational scenarios and configurations are presented. Numerical results are presented and discussed in section 5. Conclusions are offered in section 6.

2. The Compositional Flow Model

Isothermal compositional flow models require coupling a transient compressible flow with advective transport systems of various chemical species in subsurface porous media. This dynamic coupling involves solving for Darcy velocities, pressure fields, and molar fractions of available components in the fluid mixture. In this work, we consider a single-phase flow and we ignore the effects of chemical reactions between the components of the fluid phase, radioactive decay, biodegradation, and growth due to bacterial activities that might increase or decrease the quantity of components.

2.1. Governing Equations

Unlike in solute transport equations, molar fractions are used in the description of mass conservation rather than mass concentration. Consider a gas mixture consisting of N_c chemical species and define the molar fraction of each individual component, ξ_i as:

$$\xi_i = n_i / n, \quad i = 1, \dots, N_c, \quad (1)$$

where n_i is the number of moles of the i th species and n is the total number of moles of the mixture. Note that the sum over all molar fractions in the mixture should be equal to one. The mass balance equation for each component can be written as [Chen, 2007]:

$$\frac{\partial(\phi c \xi_i)}{\partial t} = -\nabla \cdot (c \xi_i u + G_i) + F_i, \quad i = 1, \dots, N_c, \quad (2)$$

where ϕ is the field porosity, c is the molar density (mol/m^3) of the fluid, t is time (s), u is the Darcy velocity (m/s), and G_i is the Fickian diffusive mass flux of the i th component. F_i is the source term for each component ($\text{mol}/\text{m}^3 \text{ s}$) which can be used to represent the well flow rate per unit volume. Darcy's law in the presence of gravity for a single fluid phase is given by:

$$u = -\frac{k}{\mu} (\nabla P - \rho g). \quad (3)$$

We denote pressure by P , permeability by k , and viscosity by μ . ρ is the fluid mass density and g is the gravitational acceleration. Starting with the mass balance equation and ignoring the Fickian diffusion term for clarity, one can multiply equation (2) by the partial molar volume, $\bar{v}_i$, and sum over all components to obtain:

$$\sum_{i=1}^{N_c} \bar{v}_i \frac{\partial(\phi c \xi_i)}{\partial t} + \sum_{i=1}^{N_c} \bar{v}_i \nabla \cdot (c \xi_i u) = \sum_{i=1}^{N_c} \bar{v}_i F_i. \quad (4)$$

Formation deformations due to change in the fluid volume (V_f) can be expressed as follows, where V_b stands for bulk volume and C_r stands for the compressibility of the rock:

$$C_r = \frac{1}{\phi} \left(\frac{\partial \phi}{\partial P} \right)_T, \quad (5)$$

$$dV_f = V_b d\phi = V_b C_r \phi dP. \quad (6)$$

These deformation effects, on the other hand, are balanced by changes within the fluid phase itself under isothermal conditions:

$$\begin{aligned} dV_f &= \left(\frac{\partial V_f}{\partial P} \right)_{T, \bar{n}} dP + \sum_{i=1}^{N_c} \left(\frac{\partial V_f}{\partial n_i} \right)_{T, P, \bar{n}_i} dn_i, \\ &= -V_f C_f dP + \sum_{i=1}^{N_c} \bar{v}_i dn_i, \end{aligned} \quad (7)$$

where C_f is the fluid compressibility. By equating the physical changes of equations (6) and (7), it can be shown that, with further rearrangement,

$$\sum_{i=1}^{N_c} \bar{v}_i \frac{\partial(\phi c_{\xi_i})}{\partial t} = \phi C_t \frac{\partial P}{\partial t}, \quad (8)$$

where C_t is the total compressibility. However, in most cases, the rock is assumed to be incompressible, which means that the fluid compressibility can replace C_t and hence the mass balance equation reduces to:

$$\phi C_f \frac{\partial P}{\partial t} + \sum_{i=1}^{N_c} \bar{v}_i \nabla \cdot (c_{\xi_i} u) = \sum_{i=1}^{N_c} \bar{v}_i F_i. \quad (9)$$

The velocity term in (9) can be replaced by Darcy's law in equation (3), which leads to the generalized pressure equation:

$$\phi C_f \frac{\partial P}{\partial t} + \sum_{i=1}^{N_c} \bar{v}_i \nabla \cdot \left[-\frac{k}{\mu} (\nabla P - \rho g) c_{\xi_i} \right] = \sum_{i=1}^{N_c} \bar{v}_i F_i. \quad (10)$$

Once the pressure field is obtained, the velocity field is updated and the new molar fractions and molar densities are evaluated for the next time step using separate mass balance equations for each of the components.

2.2. Fluid Properties Evaluation

Different equations of state (EOS) are employed in subsurface simulators, but all are designed to yield consistent compositions, densities, and molar volumes. In this work, we utilize the well-known Peng Robinson EOS (PR-EOS). Given the molar fractions of each of the components, with temperature, T , and pressure, P , at each cell, the compressibility factor, Z , is computed using the cubic form of the PR-EOS:

$$Z^3 - (1 - B)Z^2 + (A - 3B^2 - 2B)Z - (AB - B^2 - B^3) = 0, \quad (11)$$

with

$$A = \frac{aP}{R^2 T^2}, \quad B = \frac{bP}{RT}, \quad (12)$$

where R is the universal gas constant and a and b are material dependent parameters that can be obtained based on the critical properties of the substance of interest. The parameters a and b are first obtained for pure components as in equation (13):

$$a_i = 0.45724 \frac{R^2 T_{c,i}^2}{P_{c,i}} \alpha(T_{r,i}, \omega_i), \quad b_i = 0.07780 \frac{RT_{c,i}}{P_{c,i}}. \quad (13)$$

Here $T_{r,i}$ is the reduced temperature of the i th component, ω_i is the acentric factor for component i , and δ_{ij} is a binary interaction parameter between components i and j . $T_{c,i}$ and $P_{c,i}$ are the critical temperature and pressure for component i , respectively. a_i and b_i are empirical factors for component i . Using the mixing rules in equation (14), the parameters a and b are then evaluated for the entire mixture:

$$a = \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} \xi_i \xi_j (1 - \delta_{ij}) \sqrt{a_i a_j}, \quad b = \sum_{i=1}^{N_c} \xi_i b_i. \quad (14)$$

The interaction parameter, δ_{ij} account for molecular interactions between two unlike molecules. Ideally, δ_{ij} is zero when i and j represent the same component, small when i and j represent components that do not differ by much and large when the components are substantially different [Firoozabadi, 1999].

The fluid compressibility, C_f and the partial molar volume, $\bar{v}_i$, can be evaluated based on their relations with respect to the obtained Z from PR-EOS. The following relations are derived from the definition of each property:

$$\begin{aligned} C_f &= -\frac{1}{V_f} \left(\frac{\partial V_f}{\partial P} \right)_{T, \bar{n}} = -\frac{1}{v} \left(\frac{\partial v}{\partial P} \right)_{T, \bar{n}} \\ &= -\frac{1}{Z} \left[\left(\frac{\partial Z}{\partial A} \right)_B \left(\frac{\partial A}{\partial P} \right)_{T, \bar{n}} + \left(\frac{\partial Z}{\partial B} \right)_A \left(\frac{\partial B}{\partial P} \right)_{T, \bar{n}} \right] + \frac{1}{P}, \end{aligned} \quad (15)$$

and

$$\begin{aligned} \bar{v}_i &= \left(\frac{\partial V_f}{\partial n_i} \right)_{T, P, \bar{n}_i} = \left(\frac{\partial (nv)}{\partial n_i} \right)_{T, P, \bar{n}_i} = v \left(\frac{\partial n}{\partial n_i} \right)_{T, P, \bar{n}_i} + n \left(\frac{\partial v}{\partial n_i} \right)_{T, P, \bar{n}_i} \\ &= v + \frac{nRT}{P} \left[\left(\frac{\partial Z}{\partial A} \right)_B \left(\frac{\partial A}{\partial n_i} \right)_{T, P, \bar{n}_i} + \left(\frac{\partial Z}{\partial B} \right)_A \left(\frac{\partial B}{\partial n_i} \right)_{T, P, \bar{n}_i} \right], \end{aligned} \quad (16)$$

where v is the molar volume computed as:

$$v = \frac{ZRT}{P} = \frac{1}{c}. \quad (17)$$

Regarding the viscosity of the available fluid phase, different analytical formulas have been considered in the literature, mostly depending on the chemical characteristics of the phase as being light, medium or heavy. In this study, we adopt the most widely used correlations of the modified Lohrenz-Bray-Clark (LBC) model [Al-Syabi et al., 2001].

3. Ensemble-Based Data Assimilation

Consider the following discrete nonlinear dynamical system:

$$x_k = \mathcal{M}_k(x_{k-1}) + \eta_k, \quad (18)$$

where $x_k \in \mathbf{R}^{N_x}$ denotes a state vector of N_x model variables at time t_k , $\mathcal{M}_k : \mathbf{R}^{N_x} \rightarrow \mathbf{R}^{N_x}$ is the dynamical operator that integrates the model state from t_{k-1} to t_k , $\eta_k \in \mathbf{R}^{N_x}$ is a model error (system noise). It is often assumed that η_k follows a Gaussian distribution, $\mathcal{N}(0, Q_k)$. The observations of the state variables are assumed to be available through the following observational system:

$$y_k = \mathcal{H}_k(x_k) + \varepsilon_k, \quad (19)$$

where $y_k \in \mathbf{R}^{N_y}$ is a vector of N_y observations at time t_k , $\mathcal{H}_k : \mathbf{R}^{N_x} \rightarrow \mathbf{R}^{N_y}$ is the observation operator including grid interpolations, eventually nonlinear. The observation errors $\varepsilon_k \in \mathbf{R}^{N_y}$ consist of instrumental and representativeness errors having a Gaussian distribution with zero mean and covariance R_k . Model and observational errors are assumed to be independent.

In a Bayesian filtering setting, the objective is to determine the probability density function (pdf) of the state x_k given available observations y_k :

$$\begin{aligned} p(x_{0:N} | y_{0:N}) &\equiv p(x_0, x_1, \dots, x_N | y_0, y_1, \dots, y_N), \\ &= p(x_0 | y_0) \cdot p(x_1 | x_0, y_{0:1}) \cdots p(x_N | x_{N-1}, y_{0:N}), \end{aligned} \quad (20)$$

where N is the final simulation time. Once the pdf is determined, one may compute the maximum a posteriori (MAP) estimate of the system which maximizes the density (20). Sequentially, the observations $y_{0:k}$ are fed back to the model at each time they become available and a new estimate is produced. The current estimate is then used to compute the future prediction. This procedure is widely known as sequential data assimilation. For a linear system with Gaussian errors, this algorithm reduces to the well-known Kalman filter (KF) [e.g., Kalman, 1960; Gharamti et al., 2012].

3.1. The Ensemble Kalman Filter (EnKF)

The ensemble Kalman filter (EnKF) is a sequential Monte Carlo method that was introduced as an alternative to the KF for large dimensional and nonlinear data assimilation problems [Evensen, 2003]. The EnKF provides a simple conceptual formulation and ease of implementation, only requiring forward model integrations [Burgers *et al.*, 1998]. The EnKF relies on samples (or ensemble of realizations) to represent the state distributions in the recursive Bayesian algorithm. At every forecast step, ensemble members are integrated forward in time with the dynamical model (18). The state and the associated covariance matrix are then estimated as:

$$\hat{x}_k^f = \frac{1}{N_e} \sum_{j=1}^{N_e} x_k^{f,j}, \quad (21)$$

$$\begin{aligned} \hat{P}_k^f &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (x_k^{f,j} - \hat{x}_k^f) (x_k^{f,j} - \hat{x}_k^f)^T, \\ &= X_k^f (X_k^f)^T, \end{aligned} \quad (22)$$

where N_e is the number of ensemble members and the i th column of the covariance anomaly X_k^f is $(N_e - 1)^{-1/2} (x_k^{f,i} - \hat{x}_k^f)$. In practice, only the covariance anomalies need to be computed instead of the full covariance, P_k^f , which enables the implementation of the filter with large-scale applications. In the analysis step, every ensemble member is updated using the linear KF update formula [Burgers *et al.*, 1998; Evensen, 2003]:

$$x_k^{a,j} = x_k^{f,j} + X_k^f (H_k X_k^f)^T \left[(H_k X_k^f) (H_k X_k^f)^T + R_k \right]^{-1} (y_k^j - H_k x_k^{f,j}). \quad (23)$$

The updated first and second moments of the posterior distribution are then evaluated as the sample mean and covariance of the analyzed ensemble, $x_k^{a,j}$, as in equations (21) and (22). One important feature of the EnKF is the treatment of the observations as random variables with the mean equal to the actual observation and a covariance R_k , i.e., in equation (23) the observation ensemble is as follows:

$$y_k^j = y_k + \epsilon_k^j, \quad (24)$$

with $\epsilon_k^j \sim \mathcal{N}(0, R_k)$. Other ensemble-based Kalman filters that do not require perturbing the observations are referred to as ensemble square-root filters [Tippett *et al.*, 2003].

3.2. State-Parameter Estimation With the EnKF

In contaminant hydrology and reservoir engineering, permeabilities and porosities are the parameters that require accurate characterization for reliable simulations of water pressures, oil-water phase saturations, and different contaminant concentrations. These parameters, which represent 2-D or 3-D fields, are fed to the dynamic model as follows:

$$x_k = \mathcal{M}_k(x_{k-1}, \theta) + \eta_k, \quad (25)$$

where $\theta \in \mathbf{R}^{N_\theta}$ represents these large field parameters in a vectorized form. Since the parameters are time-invariant quantities, θ does not have any time subscript. In the next sections, we include the time evolution subscripts with the parameters (considering identity as a dynamical operator) to be consistent with the sequential formulation of the EnKF. If the MAP estimate is selected, the objective of the sequential Bayesian filtering approach is then to maximize the joint density of the parameters and state variables at a given time conditioned on the available state observations; i.e.,

$$p(x_k, \theta_k | y_{0:k}). \quad (26)$$

This maximization problem can be tackled with a recursive assimilation scheme using either a joint or a dual technique.

3.2.1. Joint State-Parameter Estimation

The joint state-parameter approach is straightforward and is based on defining a new joint state-space representation as:

$$z_k = \begin{bmatrix} x_k^T & \theta_k^T \end{bmatrix}^T. \quad (27)$$

Maximizing the density, $p(z_k | y_{0:k})$, would thus be equivalent to maximizing the joint conditional density in (26). The prediction step in the EnKF becomes:

$$z_k^{f,i} = \tilde{\mathcal{M}}_k(z_{k-1}^{a,i}) = \begin{bmatrix} \mathcal{M}_k(x_{k-1}^{a,i}, \theta_{k-1}^{a,i}) \\ \theta_{k-1}^{a,i} \end{bmatrix}, \quad (28)$$

where $\tilde{\mathcal{M}}_k$ is the augmented nonlinear dynamic operator for the joint system. The sample background error covariance matrix is obtained using similar computations as in (22) but for the augmented ensemble vectors, $z_k^{f,i}$. This covariance would then consist of the state and parameter covariances in addition to cross-covariance terms as follows:

$$\hat{P}_k^f = \begin{bmatrix} \hat{P}_{x_k x_k}^f & \hat{P}_{x_k \theta_k}^f \\ \hat{P}_{\theta_k x_k}^f & \hat{P}_{\theta_k \theta_k}^f \end{bmatrix} = \begin{bmatrix} X_{x_k}^f (X_{x_k}^f)^T & X_{x_k}^f (X_{\theta_k}^f)^T \\ X_{\theta_k}^f (X_{x_k}^f)^T & X_{\theta_k}^f (X_{\theta_k}^f)^T \end{bmatrix}. \quad (29)$$

The analyzed state is then computed using the current observation ensemble, y_k^i :

$$z_k^{a,i} = z_k^{f,i} + \hat{P}_k^f \tilde{H}_k^T (\tilde{H}_k \hat{P}_k^f \tilde{H}_k^T + R_k)^{-1} [y_k^i - \tilde{H}_k z_k^{f,i}], \quad (30)$$

$$\equiv \begin{cases} x_k^{f,i} + \hat{P}_{x_k x_k}^f H_k^T (H_k \hat{P}_{x_k x_k}^f H_k^T + R_k)^{-1} [y_k^i - H_k x_k^{f,i}] \\ \theta_k^{f,i} + \hat{P}_{\theta_k x_k}^f H_k^T (H_k \hat{P}_{x_k x_k}^f H_k^T + R_k)^{-1} [y_k^i - H_k x_k^{f,i}] \end{cases}, \quad (31)$$

where $\tilde{H}_k = [H_k \quad O]$ is the augmented observational operator, with the O zeros matrix. The parameters ensemble, θ_k^i , is updated with state observations through the state-parameters cross correlation represented by $\hat{P}_{\theta_k x_k}^f$.

The joint approach has been widely used in many applications, but some studies have also reported some limitations that may degrade its performance. *Moradkhani et al.* [2005] argued that the large degrees of freedom of the joint formulation might lead to unstable and intractable solutions because of the difficulty of fitting a relatively large number of observations with limited size ensembles. Other studies, such as *Wen and Chen* [2006], found that jointly updating the states and the parameters may introduce important inconsistencies, especially in strongly heterogeneous formations and strongly nonlinear models.

3.2.2. Dual State-Parameter Estimation

The dual state-parameter estimation was first introduced in the context of the extended Kalman filter (EKF) by *Nelson* [1999] in neural networks designed for estimating the signal (as the state) and the weight (as parameters). Later, the approach was examined in the framework of the EnKF and applied to a conceptual rainfall-runoff model [*Moradkhani et al.*, 2005], and by many others afterward. In the dual approach, the joint density in (26) is separated into two terms for which the derivation of the MAP estimate is performed separately on x_k and θ_k , i.e.,

$$p(x_k, \theta_k | y_{0:k}) = p(x_k | \theta_k, y_{0:k}) \cdot p(\theta_k | y_{0:k}), \quad (32)$$

such that,

$$\theta_k^{MAP} = \arg \max_{\theta_k} p(\theta_k | y_{0:k}), \quad x_k^{MAP} = \arg \max_{x_k} p(x_k | \theta_k, y_{0:k}). \quad (33)$$

The motivation behind this separation is explained by the fact that the marginal density term, $p(\theta_k | y_{0:k})$, is the relevant quantity to estimate, rather than the full joint $p(x_k, \theta_k | y_{0:k})$. Based on this formulation, two interactive filters are employed; one for the parameters and one for the state, as summarized below.

1. Parameter Filter:

1. Propagate the parameter ensemble from t_{k-1} to t_k to compute the transition density, $p(\theta_k | \theta_{k-1}, y_{0:k-1})$:

$$\theta_k^{f,j} = \theta_{k-1}^{a,j}. \quad (34)$$

One could also consider a random walk model for the parameters using a kernel smoothing perturbing procedure [Moradkhani et al., 2005].

2. Update the transition density using the incoming observation at t_k to compute $p(\theta_k | y_{0:k})$. To do so, first integrate the states ensemble with the forecast parameters of (34):

$$x_k^{f,j} = \tilde{\mathcal{M}}_k \left(x_{k-1}^{a,j}, \theta_k^{f,j} \right), \quad (35)$$

then, update the parameters using:

$$\theta_k^{a,j} = \theta_k^{f,j} + \hat{P}_{\theta_k x_k}^f H_k^T \left(H_k \hat{P}_{x_k x_k}^f H_k^T + R_k \right)^{-1} \left[y_k^j - H_k x_k^{f,j} \right]. \quad (36)$$

2. State Filter:

1. Integrate the state ensemble forward in time to t_k using the updated parameters:

$$x_k^{f,j} = \tilde{\mathcal{M}}_k \left(x_{k-1}^{a,j}, \theta_k^{a,j} \right). \quad (37)$$

The resulting density would be $p(x_k | x_{k-1}, \theta_k, y_{0:k-1})$.

2. Use available observations to update the forecast state ensemble and compute the desired density, $p(x_k | \theta_k, y_{0:k})$:

$$x_k^{a,j} = x_k^{f,j} + \hat{P}_{x_k x_k}^f H_k^T \left(H_k \hat{P}_{x_k x_k}^f H_k^T + R_k \right)^{-1} \left[y_k^j - H_k x_k^{f,j} \right]. \quad (38)$$

Following this procedure, a consistent estimation of both state variables and parameters is expected. Given the dependency of the state variables on the parameters, the dual algorithm assimilates measurements of the state variables to first improve the parameters estimate, and then uses these in the forecast model before updating the forecast ensemble with the data.

3.2.3. Computational Complexity and Implementation

Following the joint formulation, a full EnKF assimilation cycle would require N_e runs of the forecast model and another N_e Kalman updates for each ensemble member. The computational complexity for updating the ensemble members is generally small as compared to integrating the forecast model. The dual approach, on the other hand, is computationally more intensive, requiring $2N_e$ model runs for forecasting and $2N_e$ Kalman updates (one for the N_e parameter ensemble and one for the N_e state ensemble), which is practically double the computational effort of the joint scheme. In terms of practical implementation, however, the dual approach offers a more flexible and simpler framework as it separates the estimation of the parameters from the state variables. This could be beneficial in the debugging process. Furthermore, the dual scheme operates with smaller ensemble arrays, reducing computational requirements in high-dimensional systems. In Table 1, the approximate computational complexities and memory storage for the EnKF (state estimation), joint-EnKF (state-parameter estimation) and dual-EnKF (state-parameter estimation) are summarized. The tabulated complexities for each method are valid under the assumption that $N_y \ll N_x$, which is generally the case in subsurface models.

Table 1. Approximate Computational Complexities of the Standard EnKF, the Joint-EnKF, and the Dual-EnKF Algorithms^a

Method	Forecast	Analysis	Storage
EnKF (State only)	$NN_e C_m$	$NN_e C_h + NN_e^2 N_x$	$2NN_e S_x$
Joint EnKF	$NN_e C_m + NN_e C_{m'}$	$NN_e C_h + NN_e^2 (N_x + N_\theta)$	$2NN_e S_x + 2NN_e S_\theta$
Dual EnKF	$2NN_e C_m + NN_e C_{m'}$	$2NN_e C_h + NN_e^2 (N_x + N_\theta)$	$2NN_e S_x + 2NN_e S_\theta$

^aNotations are as follows: N_x : number of state variables, N_θ : number of parameter variables, N_y : number of observations, N : number of assimilation cycles, N_e : ensemble size, C_m : state model cost ($=N_x^2$ is the linear KF), $C_{m'}$: parameter model cost (usually free), C_h : observation operator cost ($=N_x N_y$ in the linear KF), $\tilde{C}_h$: modified observation operator cost ($=N_y (N_x + N_\theta)$ in the linear KF), S_x : storage volume for one state vector, S_θ : storage volume for one parameter vector.

4. Experimental Design

4.1. Model Physics and Configuration

We consider a 2-D single-phase compositional flow model as a synthetic modeling case for oil recovery in a shallow marine depositional environment. The migration, transport, and fate of natural gas are modeled in a rectangular reservoir of clayey sandstone. The hydrocarbon constituents of the modeled gas

are methane, CH_4 , and propane, C_3H_8 . Methane and propane are popular fuel sources and they are both by-products of processing natural gas and refining crude oil. In the oil industry, methane is typically cheaper in terms of the cost per amount of energy released, whereas propane is an energy-rich gas capable of releasing high amounts of heat. Physically, methane is lighter than air and thus tends to rise if released, whereas propane is heavier than air and tends to sink to the bottom of the reservoir when it escapes.

The longitudinal and transverse lengths of the reservoir domain are both equal to 100 m. Gravitational effects are included in the north-south orientation of the domain. To recover oil, injection and production wells are placed in the domain according to the sketch in Figure 1. The reservoir is assumed to be enclosed so that all boundaries are impermeable. The spatial domain is discretized into 30×30 grid cells of 3.3 m $\times$ 3.3 m. The upwind cell-centered finite difference (CCFD) scheme is used to discretize the pressure and molar fraction equations [Chen, 2007] in space. The system of equations is then solved using the standard Implicit Pressure Explicit Component transport (IMPEC) scheme [Moortgat et al., 2011].

The total flow rate in the injection well is 63.1 m³/yr and the prescribed pressure in the production well is set equal to the atmospheric conditions, i.e., 101,325 Pa. The gas migration takes place at a temperature of 400 K. The universal gas constant is 8.3145 Pa m³/(mol K) and the gravitational acceleration is 9.807 m/s². The critical temperature and pressure for methane is 190.58 K and 4.604×10^6 Pa, respectively. Propane has a critical temperature of 369.82 K and a critical pressure of 4.25×10^6 Pa. Methane has a molecular

weight of 16 g/mol, whereas that of propane is 44 g/mol. The porosity of the geologic domain is considered to be uniform and equal to 0.20 (i.e., 20% of the total rock volume). The permeability field (Figure 2) is generated with sequential Gaussian simulation using the GCOSIM3D code [Gómez-Hernández and Journel, 1993] with a mean of $-25 \log(\text{m}^2)$, a variance of 1.24 $\log(\text{m}^4)$, and a Gaussian anisotropical variogram (anisotropy ratio = 0.4) with a range equal to 50 m in the x and y directions. The permeability field has two main regions, the south-west area with large permeability values allowing the quick transport of species and the north-west part with small permeability values.

Figure 1. Sketch of the compositional flow simulation with boundary conditions. Injection well is located at the south-west corner in Experiments 1 and 2. The production well is placed in the north-east corner and in the south-east corner in Experiments 1 and 2, respectively. Gravitational forces are south-oriented (negative Y axis).

4.2. Assimilation Experiments and Scenarios

Joint and dual state-parameter estimation experiments are performed with the EnKF following a twin-experiment framework.

Figure 2. The reference log-permeability field obtained with a Gaussian variogram using "GCOSIM3D" with mean $-25 \log(\text{m}^2)$ and variance $1.24 \log(\text{m}^4)$.

First, a reference model run is carried out and considered as the "truth." Then, we assume that some quantities (e.g., initial conditions and model parameters) are not perfectly known in the forecast model and we perform the filter forecast step with the imperfect model. The goal is thus, to estimate the system's unknowns using this imperfect forecast model and data extracted from the reference run. By doing so, all quantities are known by design, which allows for full evaluation of the assimilation system. The dynamic states (x) to be estimated are the pressure P , the molar fraction of methane, ξ_m , and the molar fraction of propane, ξ_p . The parameters (θ) are the permeability field values (in a logarithmic sense) at every grid cell, i.e.:

$$x = \begin{bmatrix} P^T & \xi_m^T & \xi_p^T \end{bmatrix}^T, \quad (39)$$

$$\theta = \log(k). \quad (40)$$

The augmented state vector to be estimated is then:

$$z = \begin{bmatrix} x^T & \theta^T \end{bmatrix}^T = \begin{bmatrix} P^T & \xi_m^T & \xi_p^T & \log(k^T) \end{bmatrix}^T. \quad (41)$$

4.2.1. Reference Run

In the reference run, we consider that the domain is full of propane and the goal is to recover this propane by continuous injection of methane at the south-west corner of the domain. Propane recovery is accomplished via two experimental procedures:

1. *Experiment 1:* The production well is placed at the north-east corner.
2. *Experiment 2:* The production well is placed at the south-east corner.

Obviously, under the effect of gravity, different recovery rates of propane are expected in these two experiments. Because methane is very light, it will tend to escape to the top of the reservoir after injection. Therefore, over the same simulation period, more propane can be recovered from Experiment 1 than from Experiment 2. The reference run in both experiments is conducted for 3 years with a time step of 12 h (a total of 2190 time steps). Figure 3 plots the final distribution of pressure and molar fraction of propane from the reference run for Experiment 1. As expected, after some time, propane gets trapped at the bottom of the reservoir and this is reflected in the stratified form of the pressure field. Small amounts of propane also get trapped in low permeability parts of the domain.

4.2.2. Observations and Forecast Model

Once the reference run is completed, state measurements are collected from a number of hydraulic wells. The three state variables, P , ξ_m , and ξ_p , are observed at 25 wells distributed regularly in the domain as shown in Figure 4. Chemical composition data from exploration wells can be obtained through special core analysis using resistivity logs and electrical parameterization techniques. Four simulation scenarios are considered for which different types of measurement data are assimilated:

1. *Scenario-1 (S1):* The geostatistical simulation [Gómez-Hernández and Journel, 1993] is used to condition the three measured permeability values as shown by the circles in Figure 4.

Figure 3. (a) The reference pressure field and (b) the chemical distribution of propane in the reservoir domain after 3 years.

2. *Scenario-2 (S2)*: On top of the permeability measurements of S1, dynamic pressure data are used in the EnKF assimilation every day.

3. *Scenario-3 (S3)*: On top of the permeability measurements of S1, dynamic chemical composition data of methane and propane are used in the EnKF assimilation every four months (a total of nine assimilation cycles over three years).

4. *Scenario-4 (S4)*: All observations from scenarios 1, 2, and 3 are assimilated.

These observational scenarios are selected to assess the significance and contribution of every data set.

All observations are perturbed with Gaussian noise of zero mean and a standard deviation equal to 10% of the temporal mean, assuming the observational errors are known. For general situations, *Leisenring and Moradkhani* [2012] recently introduced a method to characterize the observational covariance using a variable variance multiplier (VVM) technique. This procedure is based on selecting perturbation factors for the

observations that adjust the spread of the state and parameter ensembles based on the prediction error of previous time steps.

To perform the assimilation in a more realistic setting, we impose some perturbations on the reference model and run the assimilation system with the resulting imperfect forecast model. First, we initialize the model with the mean of the reference run (here, the mean value corresponds to the average of all system outputs (methane and propane) from the reference run over the entire 3 years) so that initially both gas components are present in the domain. Second, the flow rate of methane at the injection site is decreased by 5%, i.e., $u^* = 59.945 \text{ m}^2/\text{yr}$. Third, the porosity is increased by 10% ($\phi^* = 0.22$) to slow the transport of the

Figure 4. Observation network. Squares indicate wells where pressure and chemical composition data are sampled. Circles correspond to locations where permeability is measured. Diamonds indicate control wells.

Table 2. AAE (Mean of All Time) and AES (at Final Time, $N = 3$ Years) for Pressure and Methane Using Different Scenarios From Experiment 1^a

Scheme/Errors	Scenario 1	Scenario 2	Scenario 3	Scenario 4
J-EnKF: $\overline{AAE}(P)$	109.3284	27.5211	48.4271	25.4108
J-EnKF: $\overline{AAE}(\xi_m)$	0.2043	0.1098	0.0891	0.0501
J-EnKF: $AES(P)_N$	57.8916	14.1395	9.5028	7.6321
J-EnKF: $AES(\xi_m)_N$	0.1492	0.0654	0.0376	0.0205
D-EnKF: $\overline{AAE}(P)$	109.1621	23.7190	44.4672	23.1545
D-EnKF: $\overline{AAE}(\xi_m)$	0.2041	0.1069	0.0857	0.0496
D-EnKF: $AES(P)_N$	58.2286	15.3777	10.0976	9.0024
D-EnKF: $AES(\xi_m)_N$	0.1484	0.0804	0.0388	0.0214

^aJ-EnKF stands for joint EnKF strategy and D-EnKF stands for dual EnKF strategy. The mean AAEs of the J-EnKF in Scenario 1 are slightly different than those of the D-EnKF due to model convergence and since the model is run twice in the dual approach, we see a small difference of order $10^{-1} - 10^{-2}$.

components across the reservoir domain. Finally, 5% perturbation is imposed on the temperature of the fluid, setting it at $T^* = 420$ K. In the imperfect model, the updated molar fractions for methane and propane are checked to make sure they are nonnegative and smaller than one. Cell values that fall out of this constraint are set either to zero (if negative) or one (if larger than one) to obtain a physically meaningful solution [Li *et al.*, 2012; Gharamti *et al.*, 2013].

4.2.3. Initial Ensembles and Evaluation

Prior to assimilation, a sensitivity analysis is performed to select an appropriate

ensemble size. In our experimental setup, we found that an ensemble of $N_e = 500$ members could provide accurate estimation results at a reasonable computational cost. The initial state ensemble is obtained by randomly selecting 500 snapshots from a long run of the imperfect forecast model. For the parameters, 500 $\log(k)$ realizations are sampled with the geostatistical algorithm GCOSIM3D using the same parameters of the reference permeability field but conditioned on three measurements as discussed in the previous section. The filter estimates are evaluated based on their average absolute errors (AAE) and average ensemble spread (AES):

$$AAE = N_x^{-1} N_e^{-1} \sum_{j=1}^{N_e} \sum_{i=1}^{N_x} |x_i^R - x_{j,i}^E|, \quad (42)$$

$$AES = \sqrt{N_x^{-1} \sum_{i=1}^{N_x} \sigma_{x_i}^2}, \quad (43)$$

where x_i^R is the reference value of the variable at cell i , $x_{j,i}^E$ is the ensemble value of the variable, and $\sigma_{x_i}^2$ is the ensemble variance at location i . AAE (the AAE as shown in equation (42) might be influenced by the ensemble variance because the residuals are summed over all ensemble members) measures the estimate-truth misfit and AES measures the ensemble spread, or the confidence in the estimated values. Spatially, the estimated fields are assessed based on their correlation factors with the reference solution. Linear regression is also applied to statistically evaluate the permeability estimates as:

$$\frac{1}{N_e} \sum_{j=1}^{N_e} \log(k_j^E) = \alpha + \beta \log(k^R), \quad (44)$$

where k_j^E is the j th permeability ensemble realization, k^R is the reference permeability, and α and β are regression coefficients.

Table 3. AAE (Mean of All Time) and AES (at Final Time, $N = 3$ Years) for $\log(k)$ Using Different Scenarios From Experiment 1^a

Scheme/Errors	Scenario 1	Scenario 2	Scenario 3	Scenario 4
J-EnKF: $\overline{AAE}$	1.0694	0.9672	0.8885	0.6693
J-EnKF: AES_N	0.9304	0.8358	0.4762	0.4438
D-EnKF: $\overline{AAE}$	1.0694	0.8401	0.7841	0.6280
D-EnKF: AES_N	0.9304	0.7970	0.4724	0.4345

^aJ-EnKF stands for joint EnKF strategy and D-EnKF stands for dual EnKF strategy.

5. Results and Discussion

5.1. Significance of Pressure and Chemical Composition Data

In the first experiment, the compositional flow and reservoir permeability outputs are adjusted using different data sets. Observations are not assimilated in the first scenario to quantify the significance of assimilating data into the

Figure 5. Evolution of log-permeability AAE values in time using the dual EnKF estimation strategy in Experiment 1.

model. Assimilation is performed over three years using the joint and dual EnKF schemes.

Estimation errors and ensemble spread values for pressure and methane state variables are listed in Table 2. The reported AAE values correspond to the mean of all computed errors for all 2190 cycles. Clearly, with more assimilated data, the accuracy and the confidence in the state estimates increases. For instance, using the joint EnKF scheme, the mean AAE values associated with the recovered gas pressure are reduced between scenarios S1 and S4 by about 75%. Moreover, the confidence in the predicted pressure after 3 years is increased by 75%, 83%, and 87% in scenarios S2, S3, and S4, respectively. This suggests that the imperfect forecast model alone cannot produce accurate

Figure 6. Cross plots of permeability from Experiment 1 (all scenarios) using the joint EnKF algorithm. Red lines indicate the regression fitting lines and dashed black lines indicate the best fit.

Table 4. Parameters of Regression Lines and Correlation Coefficients (at Final Time, $N = 3$ Years) for $\log(k)$ Cross Plots of Experiment 1^a

Scheme/Parameters	Scenario 1	Scenario 2	Scenario 3	Scenario 4
J-EnKF: λ	0.5876	0.6462	0.7535	0.8571
J-EnKF: α	-21.0857	-14.9494	-8.3414	-1.7203
J-EnKF: β	0.1446	0.4027	0.6623	0.9312
D-EnKF: λ	0.5876	0.8673	0.7793	0.8802
D-EnKF: α	-21.0857	-3.1064	-6.2862	0.6831
D-EnKF: β	0.1446	0.8877	0.7412	1.0285

^aJ-EnKF stands for joint EnKF strategy and D-EnKF stands for dual EnKF strategy.

the pressure data are most useful to improving pressure predictions and reducing its uncertainty, although the temporal density of the data sets in scenarios 2 and 3 is different. Likewise, direct chemical composition data have larger impact on updating methane and propane molar fractions than the pressure data. When all data are assimilated in S4, the composition fields are best characterized.

The permeability estimation results from all scenarios are presented in Table 3. The performances of both the joint and dual schemes are similar in producing accurate log-permeability fields. Again, the best results are obtained when all data types are assimilated. Conditioning the reservoir permeability with geostatistical simulation data (S1) uncovers the main patterns of the permeability field but is clearly not enough to obtain an accurate permeability. Incorporating the full data set of S4 reduces the ensemble spread and reduces the estimation error by 40%. This is noticeable from the AAE time series curves plotted in Figure 5. As time passes, pressure and chemical composition data provide significant information, guiding the permeability estimates toward the true solution. Pressure data are very important in the first couple of months (~ 3 months) when the pressure field is rapidly changing. Once the pressure is stratified under the effect of gravity as shown in Figure 3a, the improvements to the permeability field are small and the AAE curve almost stagnates at $0.8 \log(\text{m}^2)$. The continuous transport of the gas species, on the other hand, leads to larger improvements in the estimated permeability field attaining an AAE of $0.68 \log(\text{m}^2)$ after 3 years.

We also performed statistical regression analysis to evaluate the resulting permeability estimates. To do so, we established a linear relation between the reference log-permeability field and the ensemble mean at the end of the assimilation window. We interpreted the joint EnKF results for all scenarios using 2-D cross plots (Figure 6) and regression parameters (Table 4). In the first scenario, the scatter plot indicates a poor correlation ($\lambda = 0.59$) between the estimated and

the reference field. The orientation of the fitting line is far from the best fit with incorrect parameters at $\alpha = -21.09$ and $\beta = 0.14$. Ideally, a good fit should have α close to 0 and β close to 1. When assimilating pressure data, the orientation of the fitting line improves and the correlation increases to 0.65. Integrating chemical data in the third scenario also reduces the misfit and displays a better distribution of the scattered permeability points than S1. In the last scenario, a very good correlation is observed ($\lambda = 0.86$) and the regression parameters are very close to the ideal ones where $\alpha = -1.72$ and $\beta = 0.93$. Similar results are obtained with the dual EnKF as shown in Table 4.

The importance of pressure and chemical composition data is also tested when estimating propane recovery at the production well located in the north-east

Figure 7. Estimated propane recovery as compared to the reference recovery of Experiment 1 (all scenarios) using the dual EnKF estimation strategy.

Figure 8. Recovered ensemble mean fields for propane after 3 years using joint and dual estimation schemes. The results are shown for all scenarios of Experiment 1.

corner of the reservoir. In the reference run, the final propane recovery is 48%, which is almost half the initial amount in the reservoir (Figure 7). We compute and plot the estimated propane recovery rates in all scenarios using the dual EnKF. As seen, with no state observations, the estimated recovery is much higher

Figure 9. Analysis (updated) pressure time series of the reference field (red) and ensemble realizations (gray) for S4 as compared to those of S1. Black thick lines correspond to the mean of the ensemble.

(~60%) than the true recovery because of the perturbed initial conditions. Once pressure and molar fraction data are assimilated, the estimated rates quickly move toward the true recovery, producing highly accurate estimates. The plot shows that propane recovery rate is approximated accurately when only pressure data are assimilated. Pressure seems to have the largest effect on the estimation of propane recovery rates.

5.2. Joint Versus Dual Estimation

We compare the joint and dual schemes quantitatively, in Experiment 1 based on the algorithm's efficiency and the accuracy of the estimates. Starting with the molar fractions of methane gas, and using the AAE results in Table 1, we notice that the dual scheme performs slightly better than the joint scheme in all assimilation scenarios. On average, the dual algorithm provides around 2%, 8%, and 1% additional improvement over the joint algorithm in scenarios 2, 3, and 4, respectively. The resulting ensemble spread from both schemes after 2190 cycles is almost the same. To visualize this, we plot the final propane ensemble means from all scenarios in Figure 8. The 2-D configuration of the molar fractions of propane from both joint and dual schemes look similar and show high correlation with the reference solution, especially in S4. The ensemble means of the dual EnKF algorithm in S3 and S4 better delineate the main patterns of variability and, at the same time, unlike scenarios S1 and S2, also exhibit some small-scale patterns. Concerning pressure part, the slight advantage of the dual scheme over the joint is as follows: 4% in S2, 3% in S3, and 2% in S4. The reference pressure time series and the ensemble realizations at the control well from the first and the fourth scenarios are plotted in Figure 9. Clearly, when all data are assimilated, both algorithms produce accurate pressure reproduction ($\sim 1017 \times 10^5$ Pa) with a remarkable uncertainty reduction, in contrast with the estimate of S1. To this end, one could claim that for state estimation, the accuracy of the dual and the joint EnKF algorithms is, to certain extent, close with a little advantage to the dual approach.

Figure 10. Recovered log-permeability ensemble mean fields after 3 years using joint and dual EnKF schemes of Experiment 1 (all scenarios).

Concerning the permeability estimates, the dual EnKF strategy outperforms the joint EnKF strategy in all scenarios as shown in Table 2. The results suggest that, on average, the dual scheme yields permeability maps that are about 10% more accurate than the maps from the joint estimate. This percentage is important for identifying the overall spatial patterns of the rock permeability. To elaborate, we plot in Figure 10 the final log-permeability ensemble means in all scenarios using both estimation strategies. In the second scenario, for instance, the low permeability region in the north-west part of the domain is captured by the dual algorithm, but not by the joint algorithm. In S4, the south-east corner of the domain is better represented by the dual algorithm, yet the joint scheme displays slight overshooting in this region. From a statistical point of view, the final estimated log-permeability map using the dual scheme is also better correlated with the reference field as the correlation coefficients and regression parameters from Table 4 suggest. Figure 11 plots the ensemble variance maps of the 500 log-permeability realizations from all scenarios at the end of the 3 years. A typical bull's eye look is observed with zero variance at the sampled three locations and increasing variance away from these locations. The uncertainty of the estimated permeability with the

Figure 11. Variance maps (at $N = 3$ years) of the estimated log-permeability for Experiment 1 using joint and dual estimation schemes.

joint scheme is slightly larger, especially in S2 as can be seen in the upper left area of the domain. The variance resulting from both schemes in S3 and S4 is small and quite comparable.

One may conclude from these results that when implemented with the same number of ensemble members: (1) the dual EnKF algorithm provides around 3.5% more accurate solutions than the joint scheme on the state estimates; (2) the dual EnKF algorithm outperforms the joint EnKF and further reduces the mean estimation error of the parameters by about 10%. We should, however, note that the dual algorithm is computationally more intensive when compared to the joint. On a 6-Core Intel Xeon 2.66 GHz machine, the overall CPU time required to run the assimilation with the dual EnKF, using 500 ensemble members, was approximately 12 h, whereas the joint EnKF required only 6 h to produce the state and parameters estimates. This reported CPU effort matches the computational complexity of the two schemes reported in Table 1. In the forecast and the analysis steps of the stochastic EnKF, the dual algorithm roughly requires

Figure 12. Mean AAE resulting from the joint and dual EnKF filters for methane and permeability as function of computational complexities.

doubling the CPU effort ($2N_e C_m$ from Table 1) of the joint algorithm. This could be a disadvantage of the dual algorithm but this should not be a major limiting factor with the continuous advances in computing resources.

To study this further, we compared the estimates resulting from both assimilation strategies when implemented with similar computational requirements as shown in Figure 12. At first glance, we can easily see that the difference between the joint and dual state estimates is not as important compared to the parameter estimates. For both state and parameters, the joint scheme provides more accurate estimates with smaller ensembles. Beyond certain complexity rates ($450N_e C_m$ for the parameters and $350N_e C_m$ for the state), the dual scheme outperforms the joint scheme. For instance, when the computational requirement is $100N_e C_m$ (roughly, using 100 members in the joint scheme and 50 members in the dual scheme), the joint and the dual EnKF algorithms yield mean AAE($\log(k)$)s of 0.99 and 1.12, respectively. However, for a computational complexity of $1000N_e C_m$ and beyond, the dual EnKF produces better estimates than the joint EnKF with mean AAE($\log(k)$)s equal to 0.89 and 0.78, respectively. These results suggest that large enough ensembles are needed in the EnKF before one may notice the benefits of using the dual approach for state-parameter estimation.

5.3. Effects of the Production Site on the Estimation

Under the same conditions of Experiment 1, we conduct assimilation runs in Experiment 2 with different observational scenarios using both dual and joint estimation schemes. The goal is to assess whether changing the production site would affect the forecast of the compositional model. The new location of the production well, being at the south-east corner, favors higher recovery rates because of the south-oriented gravitational force. Consequently, after some time, the pressure field becomes stratified as in Experiment 1 but with smaller pressure values at each cell of the domain. In Table 5, the mean AAE values and the final AES for pressure and methane gas are shown. For scenario S4, the estimation results

Table 5. AAE (Mean of All Time) and AES (at Final Time, $N = 3$ Years) for Pressure and Methane Using Different Scenarios From Experiment 2^a

Scheme/Errors	Scenario 1	Scenario 2	Scenario 3	Scenario 4
J-EnKF: $\overline{AAE(P)}$	109.1621	29.2621	40.3294	24.3592
J-EnKF: $\overline{AAE(\xi_m)}$	0.2739	0.1140	0.0784	0.0764
J-EnKF: $AES(P)_N$	47.2836	17.2157	29.7205	8.5139
J-EnKF: $AES(\xi_m)_N$	0.0873	0.0737	0.0234	0.0251
D-EnKF: $\overline{AAE(P)}$	109.1621	28.2803	35.6312	24.2473
D-EnKF: $\overline{AAE(\xi_m)}$	0.2041	0.1110	0.0767	0.0685
D-EnKF: $AES(P)_N$	58.2286	13.1155	10.3238	6.4583
D-EnKF: $AES(\xi_m)_N$	0.1484	0.0758	0.0233	0.0267

^aJ-EnKF stands for joint EnKF strategy and D-EnKF stands for dual EnKF strategy.

For scenario S4, the estimation results

Figure 13. Different recovery rates and log-permeability estimates (Dual-S2) after 3 years of Experiments 1 and 2.

are similar to those obtained in Experiment 1 (Table 2); nonetheless, the state estimates of S2 and S3 are quite different. Pressure and molar fraction estimates of S2 become worse, while the estimates of S3 are now improved. Log-permeability estimates are also affected in the same way where the mean AAE for the

dual scheme in S2 increased by 14% in Experiment 2 and decreased by 3% in S3. With the change of producer well location, the pressure data seems to be more informative in Experiment 1 than in Experiment 2. To illustrate, we plot in Figure 13 the reference recovery rates from Experiments 1 and 2 in addition to the estimated permeability fields (using dual EnKF-S2) and their statistical analysis. At the end of the third year, the final recovery of propane is 48% in Experiment 1 and 53% in Experiment 2. Assimilating pressure data leads, unlike Experiment 1, to a poorly correlated ($\lambda_2=0.74 < 0.87=\lambda_1$) log-permeability map with the reference field. This is clearly observed from the computed regression parameters in which $\alpha=-15.20$ and $\beta=0.39$. The major spatial patterns of the permeability are better mapped in the first experiment. Although similar observational errors are imposed on the pressure data, more accurate assimilation results are obtained when the production site is located in the north-east corner, i.e., in opposite gravity direction. This feature suggests a clear influence of the production well's location on the accuracy of the states and parameters estimates.

6. Summary and Conclusions

This study investigated the use of the ensemble Kalman filter (EnKF) for state-parameter estimation of single-phase compositional flow models. The EnKF is now recognized as an efficient and reliable framework for various reservoir history matching and groundwater contaminant problems. Previous assimilation studies have successfully assimilated different types of data, such as pressure, seismic impedance, electromagnetic (EM), concentrations, and saturations for mapping subsurface permeability and porosity. In this work, chemical composition data of phase species is exploited for estimating phase pressure, molar fractions of species, and permeability of the reservoir. We use two well-known estimation strategies for state-parameter estimation in the context of the EnKF, namely the joint and the dual strategies. The joint algorithm concurrently estimates the state variables and parameters whereas, the dual algorithm separates the estimation of the dynamic variables from the parameters by employing two parallel filters. We performed quantitative statistical analysis to compare both estimation approaches on the basis of algorithm's performance and efficiency. We choose a 2-D comprehensive reservoir flooding case in which methane gas is inserted at the bottom of the reservoir to produce the energy-rich propane components. The migration and transport of gas species is modeled with gravity in a twin-experiments setup for three years.

Using four observational scenarios, we found that log-permeability estimates are best characterized when both pressure and composition data are assimilated. Besides, the use of composition data provides a more accurate permeability representation than those resulting from assimilation of pressure data only. The largest updates to the pressure distribution results from the direct assimilation of pressure observations. Conversely, direct and indirect observations have an important impact on the predicted molar fractions of methane and propane. In our experiments, the joint EnKF algorithm provided state estimates that are closely similar to those resulting from the dual algorithm. The dual scheme provided up to 10% more accurate parameter estimates than the joint scheme in all tested scenarios. This demonstrates the relevance of the dual technique, which first updates the parameters before integrating the dynamical system. Using the same ensemble size, however, the dual scheme is more expensive to run than the joint algorithm, requiring double the computational effort. We have found that under the same computational cost, the dual estimation strategy outperforms the standard joint strategy when a large enough ensemble is used.

This study demonstrated the importance of using of compositional data for reservoir characterization, significantly enhancing the estimates of the reservoir state and parameters. In terms of future research, this work can be considered as a basis to investigate compositional two-phase, or even multiphase, flow systems. It is possible to model density-driven convection of CO_2 in aquifers as two-phase miscible fluids by simulating CO_2 as a gas and water as a liquid. As such, pressure, chemical composition, and also phase saturation data could be used to condition the unknown state variables and parameters. Potential further research includes evaluating more advanced nonlinear filtering schemes and applications to real-world modeling scenarios.

Acknowledgment

This publication is based on work supported in part by funds from the KAUST OCRF Collaborative Fellow program.

References

- Alcolea, A., J. Carrera, and A. Medina (2006), Pilot points method incorporating prior information for solving the groundwater flow inverse problem, *Adv. Water Resour.*, 29, 1678–1689, doi:10.1016/j.advwatres.2005.12.009.
- Al-Syabi, Z., A. Danesh, B. Tohidi, A. C. Todd, and D. H. Tehrani (2001), A residual viscosity correlation for predicting the viscosity of petroleum reservoir fluids over wide ranges of pressure and temperature, *Chem. Eng. Sci.*, 56, 6997–7006, doi:10.1016/S0009-2509(01)00334-7.

- Altaf, M. U., M. E. Gharamti, A. W. Heemink, and I. Hoteit (2013), A reduced adjoint approach to variational data assimilation, *Comput. Methods Appl. Mech. Eng.*, **254**, 1–13, doi:10.1016/j.cma.2012.10.003.
- Bailey, R., and D. Baú (2010), Ensemble smoother assimilation of hydraulic head and return flow data to estimate hydraulic conductivity distribution, *Water Resour. Res.*, **2**, W12543, doi:10.1029/2010WR009147.
- Burgers, G., P. J. vanLeeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, **126**, 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.
- Chang, S., T. Chowhan, and S. Latif (2012), State and parameter estimation with an sir particle filter in a three-dimensional groundwater pollutant transport model, *J. Environ. Eng.*, **138**(11), 1114–1121, doi:10.1061/(ASCE)EE.1943-7870.0000584.
- Chen, Z. (2007), *Reservoir Simulation: Mathematical Techniques in Oil Recovery*, 220 pp., Soc. Ind. Appl. Math., Philadelphia, Penn., ISBN-10: 0898716403.
- Chen, Y., and D. Zhang (2006), Data assimilation for transient flow in geologic formations via Ensemble Kalman Filter, *Adv. Water Resour.*, **29**, 1107–1122, doi:10.1016/j.advwatres.2005.09.007.
- Elsheikh, A. H., M. F. Wheeler, and I. Hoteit (2013a), Clustered iterative stochastic ensemble method for multi-modal calibration of subsurface flow models, *J. Hydrol.*, **491**, 40–55, doi:10.1016/j.jhydrol.2013.03.037.
- Elsheikh, A. H., M. F. Wheeler, and I. Hoteit (2013b), Sparse calibration of subsurface flow models using nonlinear orthogonal matching pursuit and an iterative stochastic ensemble method, *Adv. Water Resour.*, **56**, 14–26, doi:10.1016/j.advwatres.2013.02.002.
- Evensen, G. (2003), The Ensemble Kalman Filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, **53**, 343–367, doi:10.1007/s10236-003-0036-9.
- Firoozabadi, A. (1999), *Thermodynamics of Hydrocarbon Reservoirs*, 335 pp., McGraw-Hill, New York.
- Gharamti, M. E., and I. Hoteit (2014), Complex step-based low-rank extended Kalman filtering for state-parameter estimation in subsurface transport models, *J. Hydrol.*, **509**, 588–600, doi:10.1016/j.jhydrol.2013.12.004.
- Gharamti, M. E., I. Hoteit, and S. Sun (2012), Low rank Kalman filtering for efficient state estimation of subsurface advective contaminant transport models, *J. Environ. Eng.*, **138**(4), 446–457, doi:10.1061/(ASCE)EE.1943-7870.0000484.
- Gharamti, M. E., I. Hoteit, and J. Valstar (2013), Dual states estimation of a subsurface flow-transport coupled model using ensemble Kalman filtering, *Adv. Water Resour.*, **60**, 75–88, doi:10.1016/j.advwatres.2013.07.011.
- Gómez-Hernández, J. J., and A. G. Journel (1993), Joint sequential simulation of multi-Gaussian fields, *Geostat. Troia*, **1**, 85–94, doi:10.1007/978-94-011-1739-5-8.
- Gu, Y., and D. Oliver (2006), The ensemble Kalman filter for continuous updating of reservoir simulation models, *J. Energy Resour. Technol.*, **128**, 79, doi:10.1115/1.2134735.
- Gupta, V. K., S. Sorooshian, and P. O. Yapo (1998), Towards improved calibration of hydrological models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, **34**, 751–763, doi:10.1029/97WR03495.
- Hendricks Franssen, H. J., and W. Kinzelbach (2008), Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem, *Water Resour. Res.*, **44**, W09408, doi:10.1029/2007WR006505.
- Hendricks Franssen, H. J., and W. Kinzelbach (2009), Ensemble Kalman filtering versus sequential self-calibration for inverse modeling of dynamic groundwater flow systems, *J. Hydrol.*, **365**(3–4), 261–274, doi:10.1016/j.jhydrol.2008.11.033.
- Hoteit, I., D. T. Pham, G. Triantafyllou, and G. Korres (2008), A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography, *Mon. Weather Rev.*, **136**, 317–334, doi:10.1175/2007MWR1927.1.
- Hoteit, I., X. Luo, and D. T. Pham (2012), Particle Kalman filtering: A nonlinear Bayesian framework for ensemble Kalman filters, *Mon. Weather Rev.*, **140**, 528–542, doi:10.1175/2011MWR3640.1.
- Kalman, R. (1960), New Approach to linear filtering and prediction problems, *J. Basic Eng.*, **82**, 35–45, doi:10.1115/1.3662552.
- Leisenring, M., and H. Moradkhani (2012), Analyzing the uncertainty of suspended sediment load prediction using sequential data assimilation, *J. Hydrol.*, **468–469**, 268–282, doi:10.1016/j.jhydrol.2012.08.049.
- Li, L., H. Zhou, J. J. Gómez-Hernández, and H. J. Hendricks Franssen (2012), Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble Kalman filter, *J. Hydrol.*, **428–429**, 152–169, doi:10.1016/j.jhydrol.2012.01.037.
- Lü, H., Z. Yu, Y. Zhu, S. Drake, Z. Hao, and E. A. Sudicky (2011), Dual state-parameter estimation of root zone soil moisture by optimal parameter estimation and extended Kalman filter data assimilation, *Adv. Water Resour.*, **34**, 395–406, doi:10.1016/j.advwatres.2010.12.005.
- Moortgat, J., and A. Firoozabadi (2013), Higher-order compositional modeling of three-phase flow in 3D fractured porous media based on cross-flow equilibrium, *J. Comput. Phys.*, **250**, 425–445, doi:10.1016/j.jcp.2013.05.009.
- Moortgat, J., S. Sun, and A. Firoozabadi (2011), Compositional modeling of three-phase flow with gravity using higher-order finite element methods, *Water Resour. Res.*, **47**, W05511, doi:10.1029/2010WR009801.
- Montzka, C., H. Moradkhani, M. Weihermuller, M. Canty, H. J. Hendricks Franssen, and H. Vereecken (2011), Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter, *J. Hydrol.*, **399**(3–4), 410–421, doi:10.1016/j.jhydrol.2011.01.020.
- Moradkhani, H., S. Sorooshian, H. Gupta, and R. P. Houser (2005), Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, **28**, 135–147, doi:10.1016/j.advwatres.2004.09.002.
- Moradkhani, H., C. M. DeChant, and S. Sorooshian (2012), Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method, *Water Resour. Res.*, **48**, W12520, doi:10.1029/2012WR012144.
- Nelson, A. T. (1999), *Nonlinear estimation and modeling of noisy time series by dual Kalman filtering methods*, PhD dissertation, 267 pp., Oreg. Grad. Inst., Portland, Ore. <http://drl.ohsu.edu/cdm/ref/collection/etd/id/211>.
- Naevdal, G., L. V. Johnsen, S. I. Aanonsen, and E. H. Vefring (2003), Reservoir monitoring and continuous model updating using ensemble Kalman filter, *SPEJ Soc. Petrol. Eng. J.*, **10**, 66–74, doi:10.2118/84372-PA.
- Oliver, D., L. Cunha, and A. Reynolds (1997), Markov Chain Monte Carlo methods for conditioning a permeability field to pressure data, *J. Math. Geol.*, **29**(1), 61–91, doi:10.1007/BF02769620.
- Panday, S., P. A. Forsyth, R. W. Falta, Y. S. Wu, and P. S. Huyakorn (1995), Considerations for robust compositional simulations of subsurface nonaqueous phase liquid contamination and remediation, *Water Resour. Res.*, **31**, 1273–1289, doi:10.1029/95WR03288.
- Reeves, H. W., and L. M. Abriola (1994), An iterative compositional model for subsurface multiphase flow, *J. Contam. Hydrol.*, **15**, 249–276.
- Schöniger, A., W. Nowak, and H. J. Hendricks Franssen (2012), Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography, *Water Resour. Res.*, **48**, W04502, doi:10.1029/2011WR010462.
- Sun, X., A. P. Morris, and S. Mohanty (2009), Comparison of deterministic ensemble Kalman filters for assimilating hydrogeological data, *Adv. Water Resour.*, **32**, 280–292, doi:10.1016/j.advwatres.2008.11.006.

- Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*(10), 2521–2535, doi:10.1029/2000WR900405.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker (2003), Ensemble square root filters, *Mon. Weather Rev.*, *131*, 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.
- Valstar, J. (2001), Inverse Modeling of groundwater flow and transport, PhD dissertation, 250 pp., Delft Univ. Technol., Netherlands, <http://repository.tudelft.nl/view/ir/uuid:6ce5fc45-ea9e-428a-8743-2e06890fdeb1/>.
- Valstar, J., D. B. McLaughlin, C. B. M. te Stroet, F. C. van Geer (2004), A representer-based inverse method for groundwater flow and transport applications, *Water Resour. Res.*, *40*, W05116, doi:10.1029/2003WR002922.
- Wang, D., Y. Chen, and X. Cai (2009), State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter, *Water Resour. Res.*, *45*, W11416, doi:10.1029/2008WR007401.
- Wen, X. H., and W. H. Chen (2006), Real-time reservoir updating using ensemble Kalman filter: The confirming approach, *SPEJ Soc. Pet. Eng. J.*, *11*, 431–442, doi:10.2118/92991-PA.
- Wen, X. H., and W. H. Chen (2007), Some practical issues on real-time reservoir updating using ensemble Kalman filter, *SPEJ Soc. Pet. Eng. J.*, *12*, 156–166, doi:10.1118/111571-PA.
- Xu, T., J. J. Gómez-Hernández, L. Li, and H. Zhou (2013), Parallelized ensemble Kalman filter for hydraulic conductivity characterization, *Comput. Geosci.*, *52*, 42–49, doi:10.1016/j.cageo.2012.10.007.
- Zhou, H., J. J. Gómez-Hernández, H. J. Hendricks 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, doi:10.1016/j.advwatres.2011.04.014.