Use of Phase Streamlines for Covariance Localization in Ensemble Kalman Filter for Three-Phase History Matching

Shingo Watanabe, SPE, Akhil Datta-Gupta, SPE, Texas A&M University

Summary

The ensemble Kalman filter (EnKF) has gained increased popularity for history matching and continuous reservoir-model updating. It is a sequential Monte Carlo approach that works with an ensemble of reservoir models. Specifically, the method uses cross covariance between measurements and model parameters estimated from the ensemble. For practical field applications, the ensemble size needs to be kept small for computational efficiency. However, this leads to poor approximations of the cross covariance and can cause loss of geologic realism from unrealistic model updates outside the region of the data influence and/or loss of variance leading to ensemble collapse. A common approach to remedy the situation is to limit the influence of the data through covariance localization.

In this paper, we show that for three-phase-flow conditions, the region of covariance localization strongly depends on the underlying flow dynamics as well as on the particular data type that is being assimilated, for example, water cut, oil cut, or gas/oil ratio (GOR). This makes the traditional distance-based localizations suboptimal and, often, ineffective. Instead, we propose the use of water- and gas-phase streamlines as a means for covariance localization for water-cut and GOR-data assimilation. The phase streamlines can be computed on the basis of individual-phase velocities which are readily available after flow simulation. Unlike the total velocity streamlines, phase streamlines can be discontinuous. We show that the discontinuities in water-phase and gas-phase streamlines naturally define the region of influence for water-cut and GOR data and provide a flow-relevant covariance localization during EnKF updating.

We first demonstrate the validity of the proposed localization approach using a waterflood example in a quarter-five-spot pattern. Specifically, we compare the phase streamline trajectories with cross-covariance maps computed using an ensemble size of 2,000 for both water-cut and GOR data. The results show a close correspondence between the time evolution of phase streamlines and the cross-covariance maps of water-cut and GOR data. A benchmark uncertainty quantification (the PUNQ-S3) (Carter 2007) model application shows that our proposed localization outperforms the distance-based localization method. The updated models show improved forecasts while preserving geologic realism.

Introduction

There has been a great deal of progress in developing automatic-history-matching methods for reservoir-characterization problems during the last 20 years. A relatively recent development that combines the uncertainty in the reservoir description and the reservoir performance predictions is the EnKF (Evensen 1994). Although a vast amount of literature is overwhelming, a review of the method in the scope of petroleum engineering was recently published by Aanonsen et al. (2009).

In the EnKF framework, an ensemble of model realizations is progressively updated as observation data become available using an assimilation sequence comprising a forecast step that propagates the ensemble forward in time and an update step that modifies the reservoir variables in order to match the current observations.

Although the EnKF has been extensively applied to field-scale reservoir-characterization and history-matching studies (Naevdal et al. 2005; Skjervheim et al. 2007; Seiler et al. 2009; Chen and Oliver 2010a, b), there are several outstanding difficulties associated with the use of the EnKF. The cause of the difficulties is rooted in the severe nonlinearity of the problem and nonuniqueness of the ill-posed inverse problem in general. Especially in the context of a history-matching problem, Gaussian statistics are the underlying assumption for updating models. These assumptions can be violated in many situations for reservoir characterization such as multimodal multifacies or channelized permeability distribution. Some reparameterization techniques were proposed to address this issue (Jafarpour and McLaughlin 2009; Sarma and Chen 2009).

Several authors have addressed the nonlinearity issue and have developed variants of EnKF, such as iterative formulations (Gu and Oliver 2007; Li and Reynolds 2007) and hybrid formulations (Watanabe et al. 2009). In this paper, however, we use the conventional EnKF formulation to investigate the fundamental issue of inadequate estimation of cross covariance because of small ensemble size. The results from this study also apply to other forms of EnKF.

The EnKF relies on an accurate representation of the ensemble-derived statistical measures, such as the cross covariance between the reservoir responses (e.g., flow rates, bottomhole pressures, GORs, and water cuts) and the reservoir variables (e.g., porosity and permeability), to update geological models. However, sampling errors in the ensemble-based estimates can significantly degrade the quality of model updating, especially with modest ensemble sizes. As a consequence, the ensemble model responses can collapse toward a single response, leading to “ensemble collapse” and/or the final model responses deviating from the true model trajectory in a phenomenon known as “filter divergence.” One approach to mitigate these effects is through covariance localization (Houtekamer and Mitchell 1998).

The commonly applied “distance-dependent localization” originated in atmospheric-science literature. Houtekamer and Mitchell (1998) first used a cutoff distance such that only the parameters within the specified distance of the observation were updated. In most practical applications, the distance-dependent localization uses a fifth-order compactly supported correlation function of Gaspari and Cohn (1999) to eliminate spurious correlations far away from the observation points.

In reservoir-characterization applications, Arroyo-Negrete et al. (2008) and Devegowda et al. (2010) introduced a streamline-based localization approach that uses streamline trajectories to identify the region of influence associated with the observed data. Such flow-relevant localization has been shown to naturally follow the prior-geological-model heterogeneity and the underlying displacement phenomena. By targeting and limiting the model updates, the streamline-based approach could also preserve the prior-geological-model characteristics and mitigate the parameter over- and undershooting problems.

Anderson (2007) proposed a statistical localization approach on the basis of a hierarchical EnKF. The goal was to minimize the

Copyright © 2012 Society of Petroleum Engineers
sampling error in the Kalman gain using groups of ensembles to estimate Kalman-gain error. A minimization of the estimation error results in localization multipliers. One disadvantage of the approach is the computational cost for practical applications because of the need for multiple ensembles.

Previous studies on cross-covariance estimation with localization have been mostly limited to two-phase-flow conditions. Chen and Oliver (2010a, b) analyzed cross-correlation profiles for different data types and assimilation times in an attempt to define appropriate distance-based localization functions. They concluded that distance-based localization can eliminate spurious correlations with the knowledge of data sensitivity, the prior covariance for model variables, and the past history of data assimilation. However, no clear criterion was proposed to define an appropriate distance measure for localization under dynamic conditions. Emerick and Reynolds (2011) combined well drainage areas and prior covariance information to define localization for a two-phase synthetic example. Their approach is related to the streamline-based localization approach; however, they alter the well drainage regions, potentially making the localization scheme inconsistent with the underlying flow dynamics.

In this paper, we focus on three-phase data assimilation using EnKF. We show that for three-phase-flow conditions, the region of covariance localization strongly depends on the underlying flow dynamics and on the particular data type that is being assimilated in terms of water cut orGOR. This makes the traditional distance-based localizations suboptimal and, often, ineffective. Unlike previous streamline-based localization that uses total velocity streamlines, we propose the use of water- and gas-phase streamlines as a means for covariance localization for water-cut andGOR-data assimilation. The phase streamlines can be computed on the basis of individual-phase velocities, which are readily available after flow simulation. Unlike the total velocity streamlines, phase streamlines can be discontinuous. We show that the discontinuities in water-phase and gas-phase streamlines naturally define the region of influence for water-cut andGOR data and provide a flow-relevant covariance localization during EnKF updating.

The outline of this paper is as follows. First, we give mathematical background to introduce the EnKF formulation and brief descriptions of various localization methods. Second, we describe and validate our proposed localization approach by using an illustrative three-phase synthetic example. Finally, we apply EnKF history matching for a waterflood example and the benchmark PUNQ-S3 model (Barker et al. 2001). We compare different localization techniques and examine the performance of each localization approach to demonstrate the utility and advantage of our proposed approach.

**Background and Methodology**

This section briefly discusses the EnKF formulation and introduces various localization methods. We first review the classical EnKF formulation and introduce the relevant terminologies. We then discuss localization approaches, mainly categorized into three groups: distance-dependent localization, streamline-based localization, and statistically derived localization, or hierarchical EnKF.

**EnKF Formulation.** The EnKF, first introduced by Evensen (1994, 2003), is a sequential Monte Carlo technique for data assimilation. In the EnKF approach, an ensemble of model states is recursively conditioned to dynamic data as they become available. The details of the derivation of the EnKF can be found in Evensen (2006). In the following, we will focus on some of the key features of the EnKF equations.

In the EnKF formulation, each ensemble member or realization is represented by a state vector \( \mathbf{x}_k \) at time \( k \) and containing the following: a vector of static variables \( \mathbf{m}^e_k \) (e.g., permeability, porosity) of length \( N_p \), a vector of dynamic variables \( \mathbf{m}^d_k \) (e.g., pressure, phase saturations) of length \( N_d \), and a vector of model predictions \( \mathbf{d}_k \) (e.g., bottomhole pressure, water cut, andGOR at the wells) of length \( M \).

The superscript \( p \) denotes the prediction state. The model predictions at time \( k \) are related to the state vector through the use of a measurement matrix \( \mathbf{H} \), as follows:

\[
\mathbf{d}_k = \mathbf{H} \mathbf{y}^p_k
\]

Thus the mapping matrix \( \mathbf{H} \) is given by Eq. 3, as follows:

\[
\mathbf{H} = \begin{bmatrix} \mathbf{0}_{N_p} & \mathbf{0}_{N_d} & \mathbf{I}_M \end{bmatrix}
\]

where \( \mathbf{0}_{N_p} \) and \( \mathbf{0}_{N_d} \) are zero matrices of size \( M \times N_p \) and \( M \times N_d \), respectively, and \( \mathbf{I} \) is the identity matrix of size \( M \times M \). The EnKF works with an ensemble of state vectors denoted as

\[
\Psi_k = \{ \mathbf{y}^p_{k,1}, \mathbf{y}^p_{k,2}, \ldots, \mathbf{y}^p_{k,N_e} \}
\]

where \( N_e \) is the ensemble size. Each state vector represents an individual member of an ensemble of possible states that are consistent with the initial measurements from cores, well logs, seismic surveys, and geologic interpretation studies.

**EnKF Forecast and Update.** The EnKF comprises two main steps, a forecast step and an update step. The forecast step can be written as

\[
\begin{bmatrix} \mathbf{m}^e_k \newline \mathbf{d}_k \end{bmatrix} = g(\mathbf{m}^e_{k-1}, \mathbf{m}^p_{k-1})
\]

where the forward model operator \( g(\cdot) \) represents a numerical solution of the porous-media fluid-flow equations moving forward from time \( k-1 \) to time \( k \), when new observations become available. At this time, the update step modifies the reservoir state vector for each ensemble member, \( j = 1, 2, \ldots, N_e \), using the well-known Kalman update equation, as follows (Evensen 2003):

\[
\mathbf{y}^p_{k,j} = \mathbf{y}^p_j + \mathbf{K}_k (\mathbf{d}_{obs,k,j} - \mathbf{H} \mathbf{y}^p_j)
\]

The superscript \( u \) denotes the updated model. The matrix \( \mathbf{K}_k \) is known as the Kalman gain matrix and relates the data misfit to the changes required in the reservoir state vector. In Eq. 6, \( \mathbf{d}_{obs,k,j} \) represents a vector of perturbed observations as defined by the following equation:

\[
\mathbf{d}_{obs,k,j} = \mathbf{d}_{obs,k} + \mathbf{e}_j
\]

where \( \mathbf{e}_j \) represents the noise in the observation for the ensemble member \( j \). The noise associated with the measurements, \( \mathbf{e} \), is assumed to be Gaussian with a zero mean and covariance, \( \mathbf{C}_e \).

The Kalman gain matrix \( \mathbf{K}_k \) is expressed as follows:

\[
\mathbf{K}_k = \mathbf{C}^p_{\psi,k} \mathbf{H}^T (\mathbf{H} \mathbf{C}^p_{\psi,k} \mathbf{H}^T + \mathbf{C}_d)^{-1}
\]

where \( \mathbf{C}^p_{\psi,k} \) represents an estimate of the state vector covariance matrix at time \( k \) and does not need to be computed from the ensemble explicitly. Instead, we compute the products of the cross-covariance matrix \( \mathbf{C}^p_{\psi,k} \mathbf{H}^T \) and the prediction-error covariance matrix \( \mathbf{H} \mathbf{C}^p_{\psi,k} \mathbf{H}^T \). From Eq. 8 in particular, it is obvious that EnKF updating is largely governed by the cross-covariance term \( \mathbf{C}^p_{\psi,k} \mathbf{H}^T \), estimated from the ensemble by the following expression:

\[
\mathbf{C}^p_{\psi,k} \mathbf{H}^T = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\mathbf{y}^p_j - \bar{\mathbf{y}}^p_k)(\mathbf{y}^p_j - \bar{\mathbf{y}}^p_k)^T
\]
where

\[ \mathbf{y}_k^e = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{y}_{kj}^e \]  

and

\[ \mathbf{H}_k^e = \frac{1}{N_e} \sum_{j=1}^{N_e} \mathbf{H}_{kj}^e. \]

Cross-Covariance Localization: Current Approaches. The main aim of covariance localization schemes is to eliminate spurious terms in the cross-covariance matrix arising from sampling errors caused by finite and small ensemble sizes and to increase the effective number of ensemble members (Hamill et al. 2001). In the absence of localization, the covariance matrix is rank-deficient, which leads to spurious correlation and loss of variance (ensemble collapse). Mathematically, the EnKF update equation with covariance localization can be expressed on the basis of Eq. 6 as

\[ \mathbf{y}_{kj}^e = \mathbf{y}_{kj}^p + (\mathbf{p} \ast C_{0,k}^e \mathbf{H}^T) (\mathbf{H} C_{0,k}^e \mathbf{H}^T + C_D)^{-1} (\mathbf{d}_{obs,k,j} - \mathbf{H}_{kj}^e), \]

where the localization function \( \mathbf{p} \) operates on the cross-covariance matrix. The operator \( \ast \) is an element-by-element multiplication, also called the Schur product. The various localization functions differ in the way we calculate this multiplier \( \mathbf{p} \). In this paper, we investigate three categories of localization methods: distance-dependent localization, streamline-based localization, and statistically derived hierarchical localization methods. It is worthwhile to mention that in the absence of covariance localization, the solution from the EnKF updating is restricted to a linear combination from the initial ensemble members (Evensen 2003). However, with covariance localization as in Eq. 12, a much larger basis can be accessed, and the solution is no longer limited to a linear combination of the initial members (Evensen 2003). Brief descriptions of each localization method are next.

Distance-Dependent Localization. Distance-based covariance localization schemes (Houtekamer and Mitchell 2001; Hamill et al. 2001) rely on the assumption that the correlation between model grid cells and observation data is a function of the distance between the grid cell and the observation location. Various distance-based correlation functions have been discussed by Gaspari and Cohn (1999). In this paper, we have used the following form of the localization function:

\[ \Omega(a,b) = \begin{cases} 
\frac{1}{6} b^5 + \frac{1}{2} b^4 + \frac{5}{8} b^3 + \frac{5}{3} b^2 + 1, & 0 \leq b \leq a; \\
\frac{1}{12} b^5 - \frac{1}{2} b^4 + \frac{5}{8} b^3 + \frac{5}{3} b^2, & a < b < 2a; \\
-5 \left( \frac{b}{a} \right) + 4 - \frac{2}{3} \frac{b}{a}, & a > 2a; \\
0, & b > 2a.
\end{cases} \]

Hierarchical Ensemble-Filter Localization. Anderson (2007) proposed a hierarchical ensemble filter that estimates sampling errors in the Kalman gain matrix from a group of \( N_e \) ensembles and computes a regression confidence (weighting) factor \( a_{\text{min}} \) between zero and unity to minimize sampling error at each assimilation timestep \( n \). The factor \( a_{\text{min}} \) is defined to minimize the expected root-mean-square (RMS) difference in Kalman gain among the group as

\[ \min O(a_n) = \min \left\{ \sum_{k=1}^{N_k} \sum_{l=1}^{N_k} \left( \frac{a_{\text{min}} K_k - K_l}{a_{\text{min}} K_k} \right)^2 \right\}, \]

where \( K_k \) is an element of the Kalman gain matrix for the \( k \)th ensemble at time \( n \). This minimization results in the following:

\[ a_{\text{min},n} = \max \left\{ \left\{ \left[ \sum_{k=1}^{N_k} K_k \right] / \sum_{l=1}^{N_k} K_l \right\} \right\} - 1 \]

\[ / (N_k - 1), 0 \].

The localization multiplier matrix \( \mathbf{p} \) is constructed element by element using \( a_{\text{min},n} \). The modified Kalman update for each group of the ensemble is given as

\[ \mathbf{y}_{kj}^e(i) = \mathbf{y}_{kj}^p(i) + \mathbf{p} \cdot \mathbf{d}_{obs,k,j} - \mathbf{H}_{kj}^e(i), \]

where \( i = 1, \ldots, N_e \) and the localization multiplier is acting not on the cross covariance but on the Kalman gain matrix directly.

**Approach**

The main objectives of our proposed localization scheme are summarized as follows:

- Eliminating spurious cross-correlation calculations arising from small-ensemble statistics
- Localizing the observation data to regions that are based on underlying flow dynamics
- Preserving geological realism through targeted parameter updating
- Maintaining model variability and preventing ensemble collapse and filter divergence

The major steps in our proposed approach are outlined as follows:

- **Ensemble Prediction Step.** Given an ensemble of static reservoir-simulation model realizations generated from prior geologic information and/or geostatistical analysis, we conduct
reservoir simulation on each ensemble member up to the next available observation time. The outcome of this step is the prediction of the observed data for each ensemble member.

- **Phase-Streamline Tracing.** On the basis of the forward simulation results for each ensemble member, we extract individual-phase fluxes for each gridblock. Using the phase fluxes at the grid-block faces, we trace phase-streamline trajectories for water, oil, and gas phases. For streamline tracing, we used the algorithm proposed by Jimenez et al. (2007) for its ease of implementation and applicability to corner-point grids. The streamline trajectories from each ensemble member are stored and used to identify the flow-relevant regions contributing to the observed data at the current time for localization purposes.

- **Localizing Cross-Covariance Matrix.** Using the ensemble model predictions, model parameters, and the flow-relevant regions identified by the phase-streamline trajectories, we compute the cross-covariance matrix only for the localized gridblocks.

- **Ensemble-Correction Step.** Finally, we update the ensemble members using the localized cross-covariance matrix and the Kalman update equation. We repeat all steps for the next available observation data.

**An Illustrative Example.** This section illustrates the importance of covariance localization for EnKF using a simple 2D homogeneous reservoir model with a mesh size of $21 \times 21$ having permeability of 8.12 md and porosity of 0.1, as shown in Fig. 1. To generate the initial ensemble of realizations, the permeability at each gridblock is perturbed with an uncorrelated Gaussian noise having a variance of 2 md. We generate a total of 2,000 realizations, some of which are shown in Fig. 1. This set of 2,000 realizations will constitute our reference case, with which we will compare various other ensemble sizes.

As an initial condition, the reservoir is saturated with oil at the bubblepoint pressure of 3,000 psi. A quarter-five-spot well configuration with an injector and a producer is used for this example. The wells are controlled by reservoir rate constraints, with a production rate of 50 RB/D and an injection rate of 45 RB/D. Our goal is to create three-phase-flow conditions by causing the reservoir pressure to decline because of the excess production, resulting in the liberation of the solution gas. The simulation timestep is fixed at 150 days for a total of 1,500 days. The reference-model responses are shown in Fig. 2.

In Fig. 2, we have a water breakthrough at approximately 250 days and water cut increases as production proceeds. A comparison of the bottomhole pressure and GOR response indicates that in the early period of production, gas production is mainly caused by dissolved gas from oil production and results in a flat value of the response. Once the free gas evolves in the reservoir, the gas

![Fig. 1—Homogeneous reference model and ensemble of realizations of permeability fields.](image1)

![Fig. 2—Reference-model responses of water cut at producer (top left), bottomhole pressure at producer (top right), bottomhole pressure at injector (bottom left), and GOR at producer (bottom right).](image2)
production is from both the dissolved-gas and the free-gas volume that results in increasing GOR.

**Spurious Correlations.** The fundamental problem of the EnKF that we address here is the accuracy of the cross-correlation calculations with respect to the ensemble size. Specifically, we examine the quality of the sample cross-covariance calculations between the gridblock permeabilities and three different observation data types (e.g., water cut, bottomhole pressure, and GOR) by changing the ensemble size.

The cross-covariance map is calculated at every assimilation step of the EnKF. From these, we select three times for comparison purposes: early time (300 days), intermediate time (750 days), and late time (1,350 days). The cross covariance is computed by the standard statistical means, using the ensemble members and the model predictions (see Eq. 9). We varied the ensemble size to 20, 200, and 2,000 members, and the results are shown in Figs. 3 and 4.

In Fig. 3, we can see that the bottomhole pressure and water-cut covariance shows a severe degradation for the ensemble size of 20 members. This is obvious when we compare it with the results from 2,000 members (the reference case). The 20-member results show significant spurious correlations and fail to capture much of the dominant characteristics of the reference case. If we apply this cross covariance in the Kalman update, the updating will result in parameter over- and undershooting (Arroyo-Negrete et al. 2008). For the 200-member case, we see some correspondence in the spatial features with the 2,000-member reference case. We apply this cross covariance in the Kalman update, the updating will result in parameter over- and undershooting (Arroyo-Negrete et al. 2008).

The cross-covariance map is calculated at every assimilation step of the EnKF. From these, we select three times for comparison purposes: early time (300 days), intermediate time (750 days), and late time (1,350 days). The cross covariance is computed by the standard statistical means, using the ensemble members and the model predictions (see Eq. 9). We varied the ensemble size to 20, 200, and 2,000 members, and the results are shown in Figs. 3 and 4.

The GOR-correlation profile is more complex and contains transitions over time. At early times, there is a negative correlation band moving radially outward. The inside and outside of the band have positive correlation values. To explain this correlation profile, we examine the corresponding pressure and saturation profiles shown in Fig. 6. Also, recall that the GOR is calculated by

\[
\text{GOR} = \frac{Q_g}{Q_o} = \frac{Q_{g,\text{free}} + R_s Q_o}{Q_o} = \frac{Q_{g,\text{free}}}{Q_o} + R_s = \frac{k_{rg} B_o}{k_{ro} B_g} + R_s.
\]

Inside the positive-correlation region near the production well, free mobile gas exists because the pressure has fallen below the bubblepoint pressure of 3,000 psi. Thus, the increasing permeability in this region results in increasing free-gas production and

---

**Fig. 3—Cross-covariance map comparisons with various ensemble sizes (20, 200, and 2,000 members from left). Permeability and bottomhole pressure of P1 at 300 days (top row), permeability and bottomhole pressure of I1 at 300 days (middle row), and permeability and water cut of P1 at 300 days (bottom row).**
Fig. 4—Cross-covariance map comparisons with various ensemble sizes (20, 200, and 2,000 members from left). Permeability and GOR of P1 at 300 days (top row), permeability and GOR of P1 at 750 days (middle row), and permeability and GOR of P1 at 1,350 days (bottom row).

Fig. 5—Cross-covariance map comparisons with ensemble size of 2,000 members at different times (300, 750, and 1,350 days, from top). From left, permeability and bottomhole pressure of P1, permeability and bottomhole pressure of I1, permeability and water cut of P1, and permeability and GOR of P1.
increased GOR; hence, the correlation is positive. In Fig. 6, we can see that the negative-correlation band approximately corresponds to a gas-saturation contour of 0.03. This value is the critical gas saturation in this example. Thus, the negative-correlation band is located at the boundary of the mobile and immobile gas. Increasing the permeability in this region induces faster oil flow, and oil production will increase relative to gas production. Therefore, the correlation is negative. Outside the negative-correlation band toward the injection well, we have only oil and water phases mobile, and increasing permeability results in preferential water displacement followed by reduced oil production relative to gas production. This produces the positive correlation outside the band. As production proceeds, the location of the negative band moves toward the injector, because the boundary between the mobile and immobile gas propagates toward the injector. By the time the mobile-gas saturation appears everywhere (within 1,200 days),

Fig. 6—Cross covariance of permeability and GOR map, as well as pressure, gas-saturation, and water-saturation contour maps at different times (300, 600, 900, and 1,200 days, from left).

Fig. 7—Cross covariance of permeability and GOR map and gas-phase streamlines at different times (300, 600, 900, and 1,200 days, from left).
days), the negative-correlation band no longer exists and the GOR-correlation profile shows close correspondence with that of water cut, because the flow regime becomes free-gas convective transport along preferential paths between wells.

Finally, if we compare the GOR-correlation map with the gas-phase streamlines shown in Fig. 7, we can clearly see the correspondence between the area covered by gas-phase streamline trajectories and the progression of the location of the negative-correlation band over time. This is because the gas streamlines are traced naturally from the producer to the boundary between the mobile- and immobile-gas saturations. For comparison purposes, we have shown in Fig. 8 the total streamlines as well as the gas streamlines at different times. The total streamlines cover the entire region between the injector and producer at all times, and we do not see the correspondence with the cross-covariance profile as seen with the gas streamlines (Fig. 8). This is the primary motivation in using phase streamlines for localization as opposed to total streamlines, as was done in our previous studies (Arroyo-Negrete et al. 2008).

Applications

In this section, we first present a comprehensive comparison of various localization methods using a nine-spot waterflooding example with three-phase-flow conditions. This example highlights the benefits of using phase-streamline localization over other localization methods. A variety of quality checks of the EnKF performance is presented to compare the localization methods. We also demonstrate the power and practical feasibility of the phase-streamline localization approach using a field-scale three-phase history-matching application.

Nine-Spot Synthetic Model. A heterogeneous reference synthetic model was generated by sequential Gaussian simulation (Deutsch and Journel 1992) with high-permeability continuity and a prior anisotropy direction, as shown in Fig. 9. The reservoir-model domain is discretized into \(51 \times 51\) gridblocks \((1,530 \times 1,530\) ft) with eight producers and one injector in a nine-spot well pattern, as shown in Fig. 9. Similar to the previously discussed example, three-phase-flow conditions are set up by starting with undersaturated conditions and letting the reservoir pressure fall below the bubblepoint pressure. The total simulation time is 4,000 days, which is split into the assimilation period of 2,000 days and the prediction period for the rest. Observation data are water cut, bottomhole pressure, and GOR. The data are assimilated at every 50 days for 40 time intervals. For EnKF, we used 50 ensemble members. Some examples of the initial members are shown in Fig. 9. The initial ensemble model predictions are shown in Fig. 10.
as gray lines and are compared with the reference-model response denoted as a red line in the figures.

Comparison of EnKF With and Without Localization. We conduct the EnKF with and without localizations for this synthetic case, and compare their performances.

First, we examine the distance-based localization, the most commonly used method in the literature. Because we generally do not have prior knowledge of the correlation distance, we examine two choices for $a$ in Eq. 13 for the localization function: (Case 1) water cut $= 1,000$ ft, bottomhole pressure $= 100$ ft, GOR $= 1,000$ ft; (Case 2) water cut $= 300$ ft, bottomhole pressure $= 100$ ft, GOR $= 300$ ft. Our rationale for these values is based on the previous observation that bottomhole-pressure-data influence is mainly near the well locations, and water-cut and GOR data cross correlation tends to cover the region between the well-pair configuration.

For hierarchical EnKF, we also experiment with combinations of the number of ensemble group and the subensemble member size for two cases with a total ensemble size of 200 members: (Case 1) 10 groups of 20 members; (Case 2) four groups of 50 members. For the phase-streamline-based localization, we define a localization region for each observation data type for each ensemble member. Unlike the total-streamline-based approach (Arroyo-Negrete et al. 2008), the localization region now varies for each data type (e.g., water cut, GOR) on the basis of the intersection of the phase streamlines with the grid cells at the time of interest. A common localization region is defined by the intersection of the localization regions of all individual members, and we require that a grid cell be part of the localization region for a minimum of 10% of the total ensemble members for it to be included in the common localization region.

In the results that follow, we denote the conventional EnKF as Plain EnKF, while EnKF with localizations are denoted as follows: distance-dependent localization as EnKF-DT1 for Case 1 and EnKF-DT2 for Case 2; hierarchical EnKF as EnKF-HC1 for Case 1 and EnKF-HC2 for Case 2; and phase-streamline-based localization as EnKF-PST.

The water-cut, bottomhole-pressure, and GOR history-matching comparisons for the Plain EnKF and EnKF-PST are shown in Fig. 11. The Plain EnKF shows evidence of ensemble collapse. Also, the GOR predictions for P5 show a systematic bias. The EnKF-PST results show reduced spread from the initial models and no systematic bias for the prediction period. The updated permeability fields are shown in Fig. 12. The Plain EnKF clearly results in ensemble collapse, where all ensemble models became almost identical. The EnKF-PST updated ensemble models retain the variability of the permeability distribution and capture the high-permeability streak in the left upper corner.

For various localization methods, the history-match results are shown in Fig. 13, and the updated permeability fields are shown in Fig. 14. For history-matching comparisons, we consider both matching of the data and the prediction quality for computing the misfit from the true-model response in terms of RMS for the entire simulation period of 4,000 days. Fig. 15 compares the ensemble mean log-permeability histograms for various localizations. The RMS error summary is listed in Table 1 and shown in Fig. 16. Overall, the RMS error results show EnKF-DT1 as the best, followed by EnKF-PST and EnKF-HC1. When we use a shorter correlation distance as in Case 2 (EnKF-DT2), the RMS becomes worse than EnKF-PST. This clearly indicates that the choice of correlation length is critical for distance-based localization performance. All localization schemes seem to prevent ensemble collapse in this example. Eigen spectrum of the updated permeability covariance matrix is plotted for each case in Fig. 17. The results show that all localization methods maintain the model variability throughout the updating period. In terms of the updated permeability distributions, EnKF-DT1 and EnKF-HC1 fail to reproduce the continuity of the high-permeability streak in Fig. 14. Similar results are also obtained for Case 2. In Fig. 18a, we have compared the localization regions for the three different
schemes for computing the covariance of permeability and water cut at 1,200 days. Similarly, in Fig. 18b, we have shown the localization region for computing the covariance of permeability and GOR at 500 days. Clearly, the distance-based localization does not distinguish between the water cut and GOR and uses the same localization region for both. The phase-streamline-based localization accounts for the underlying physics of flow and seems to capture the dynamic change in the localization region with respect to

Fig. 11—History-matching comparisons of Plain EnKF and EnKF-PST. True-model response (red line) and initial- and updated-ensemble model responses (gray lines) for water cut of P5, bottomhole pressure of P5, and GOR of P5.

Fig. 12—Updated permeability comparisons of Plain EnKF and EnKF-PST. True-model permeability field (top left) and three realizations of initial permeability fields (top row), Plain EnKF updated permeability fields (middle row), and EnKF-PST updated permeability fields (bottom row).
time and space naturally. Finally, because of its statistical nature, the localization region for EnKF-HC1 appears scattered and somewhat noisy, although there is some underlying resemblance with the phase-streamline-based localization region.

Thus, localization is necessary to prevent ensemble collapse in EnKF updating, particularly when the ensemble size is relatively small. The proposed phase-streamline localization can capture the interactions between the flow dynamics and the permeability heterogeneity better than other localization schemes. Also, the method does not require the choice of optimal correlation length as in distance-based localization, because streamlines naturally define the flow-relevant regions.

PUNQ-Model Application. We now demonstrate the applicability of the proposed phase-streamline localization for the benchmark PUNQ-S3 model, which is designed after a real field (Barker et al. 2001). The model contains $19 \times 28 \times 5$ gridblocks, of which 1,761 are active. The top-structure map of the field shows

![Image](image-url)
that the field is bounded to the east and south by a fault, and is supported by a fairly strong aquifer in the north and west (see Fig. 19a). A small gas cap exists in the center of the dome-shaped structure. The field initially contains six production wells drilled around the gas/oil contact. Because of the strong aquifer, no injection wells are present. A geostatistical approach has been used to generate the reservoir-permeability distribution, consistent with the geological model in each of the five layers. The production

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Plain EnKF</th>
<th>EnKF-PST</th>
<th>EnKF-DT1</th>
<th>EnKF-DT2</th>
<th>EnKF-HC1</th>
<th>EnKF-HC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCT</td>
<td>2.77E+03</td>
<td>2.61E+01</td>
<td>3.01E+02</td>
<td>8.31E+01</td>
<td>8.39E+02</td>
<td>3.58E+02</td>
<td>2.46E+02</td>
</tr>
<tr>
<td>BHP</td>
<td>3.34E+07</td>
<td>5.64E+06</td>
<td>1.12E+07</td>
<td>4.25E+06</td>
<td>1.27E+07</td>
<td>8.34E+06</td>
<td>6.00E+06</td>
</tr>
<tr>
<td>GOR</td>
<td>1.29E+03</td>
<td>4.93E+02</td>
<td>8.32E+02</td>
<td>6.16E+02</td>
<td>1.27E+03</td>
<td>1.16E+03</td>
<td>1.05E+03</td>
</tr>
</tbody>
</table>

Fig. 16—RMS of history-matching and forecasting comparisons of water cut (left), bottomhole pressure (right), and GOR (bottom). From left, initial models, Plain EnKF, EnKF-PST, EnKF-DT1 (Case 1), EnKF-DT2 (Case 2), EnKF-HC1 (Case 1), and EnKF-HC2 (Case 2).
The schedule contains a first year of extended well testing, followed by a 3-year shut-in period before field production starts. The well testing consists of four 3-month production periods, each having its own production rate. During field production, 2 weeks of each year are used for a shut-in test to collect shut-in pressure data at each well.

A reference “true” model was used to generate the production history. For the production data, Gaussian noise was added to mimic measurement errors. The total simulation period was 16.5 years. The bottomhole-pressure, water-cut, and GOR data were generated for each of the wells. Of the total time, we take a history-matching period of 8 years (0 to 2,936 days) and a forecast period of 8.5 years (2,937 to 6,025 days).

The simulation model is shown in Fig. 19b and has a corner-point grid. We used 40 ensemble members in the EnKF application. Fig. 20 shows the phase streamlines for this case. From the gas-streamline trajectories, we can clearly see the communication between the gas cap and the wells with gas breakthrough. The aquifer support is depicted by the water streamlines. We conducted Plain EnKF, EnKF-PST, and EnKF-DT for this case. Observation data for the assimilation are water cut, bottomhole pressure, and GOR at the production wells with 20 assimilation times. The

![Fig. 17—Eigen spectrum of updated parameter-covariance matrix: Plain EnKF, EnKF-PST, EnKF-DT1 (Case 1), EnKF-DT2 (Case 2), EnKF-HC1 (Case 1), and EnKF-HC2 (Case 2), from top legend.](image)

![Fig. 18—Localization multiplier for EnKF-DT1 (Case 1), EnKF-HC1 (Case 1), and EnKF-PST. (a) Localization multiplier for cross covariance of permeability and water cut of P1, P2, and P4 at 1,200 days. (b) Localization multiplier for cross covariance of permeability and GOR of P1, P2, and P5 at 500 days.](image)

![Fig. 19—PUNQ-S3 model: (a) top-structure map and (b) simulation model.](image)
model parameter updated is the gridblock permeability. For the distance-based localization, we assigned the same correlation length, $a = 500$ m, for all data types. History-matching results are compared in Fig. 21, and the RMS errors for all three data types as well as oil-production rate (OPR) are summarized in Table 2 and Fig. 22. Updated ensemble average permeabilities for EnKF-PST and EnKF-DT are shown in Fig. 23, and compared with the reference model. Plain EnKF updated permeability fields suffer from over- and undershooting problems and are not shown here. Overall, the results seem to be consistent with the nine-spot synthetic example discussed previously. As for history-matching and forecasting quality, Plain EnKF performs poorly for water-cut matching, shows evidence of ensemble collapse, and results in unsatisfactory forecast. The EnKF-DT has slightly better RMS errors than EnKF-PST, but is almost comparable for this case. However, a comparison of the updated permeability fields in Fig. 23 shows that the EnKF-PST outperforms EnKF-DT in terms of capturing the continuity of the permeability barriers and the channels. In particular, the EnKF-PST is able to reproduce the high-permeability streaks in Layer 3 and Layer 5, contributing to aquifer support better than EnKF-DT. Fig. 24 compares the corresponding layer-permeability histograms. The results show that the
EnKF-PST results in a broader spread in permeability compared with the EnKF-DT. Additionally, in order to analyze the sensitivity of the choice of correlation length for distance-based localization, we rerun EnKF with a shorter correlation length, $a = 200$ m, for all data types. We call this case EnKF-DT1. History-matching results in terms of RSM become worse, as shown in Table 2. The corresponding updated permeability fields are shown in Fig. 23. These

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Plain EnKF</th>
<th>EnKF-PST</th>
<th>EnKF-DT</th>
<th>EnKF-DT1</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCT</td>
<td>4.02E+02</td>
<td>3.40E+02</td>
<td>2.90E+02</td>
<td>2.46E+02</td>
<td>3.19E+02</td>
</tr>
<tr>
<td>BHP</td>
<td>3.98E+07</td>
<td>2.10E+06</td>
<td>2.10E+06</td>
<td>1.84E+06</td>
<td>2.61E+06</td>
</tr>
<tr>
<td>GOR</td>
<td>3.61E+07</td>
<td>3.32E+06</td>
<td>7.47E+06</td>
<td>4.75E+06</td>
<td>1.42E+07</td>
</tr>
<tr>
<td>OPR</td>
<td>7.27E+07</td>
<td>1.89E+06</td>
<td>3.16E+06</td>
<td>3.56E+06</td>
<td>6.34E+06</td>
</tr>
</tbody>
</table>

**Table 2—RMS of History-Matching and Forecasting Comparisons**

Fig. 22—RMS of history-matching and forecasting comparisons of water cut (top left), bottomhole pressure (top right), GOR (bottom left), and OPR (bottom right). (From left: initial models, Plain EnKF, EnKF-PST, and EnKF-DT.)

Fig. 23—Updated-ensemble mean permeability comparisons of EnKF-PST and EnKF-DT for selected layers (from left: reference model, permeability field, EnKF-PST updated permeability fields, EnKF-DT with correlation length $a = 500$ m updated permeability fields, and EnKF-DT1 with correlation length $a = 200$ m updated permeability fields).
results indicate that the selection of correlation length is critical for distance-based localization and for dynamically changing production conditions; this is not a trivial task.

Conclusions
In this paper, we have performed a series of numerical experiments to demonstrate that for three-phase-flow conditions, the region of covariance localization strongly depends on the underlying flow dynamics as well as on the particular data type that is being assimilated, for example, water cut or GOR. Thus, we advocate that the covariance localization should be physically motivated, taking into account underlying physics of flow. For our example cases, the distance-based localization worked well for incorporating bottomhole-pressure data. However, for water-cut and GOR data, the phase-streamline-based localization seems to be a better choice because it does not rely on the selection of correlation length for localization. Also, the region of covariance localization changes with changing flow conditions such as infill drilling, and there is no natural way to account for this in distance-based localization. The hierarchical localization can be computationally demanding and tends to introduce statistical variations in the localization region.

The following is a summary of the major features of this paper:
• We have proposed a novel methodology for covariance localization based on the reservoir dynamics for three-phase-flow conditions. Specifically, we propose the use of water- and gas-phase streamlines as a means for covariance localization for water-cut- and GOR-data assimilation. The phase streamlines can be computed on the basis of individual-phase velocities, which are readily available after flow simulation.
• Total velocity streamlines have been used in the past for covariance localization in two-phase flow. Unlike the total streamlines, phase streamlines can be discontinuous. We show that the discontinuities in water-phase and gas-phase streamlines naturally define the region of influence for water-cut and GOR data and provide a flow-relevant covariance localization during EnKF updating.
• Our numerical experiments show that the statistical sample cross correlation for water cut and GOR encompasses the underlying physics of the reservoir flow, and their spatial patterns can be explained by the phase-streamline trajectories. Also, phase-streamline-based localization can naturally capture the dynamically changing localization regions because of changing field conditions such as infill drilling and pattern conversions.
• Our proposed approach improved EnKF performance in terms of (1) adequately matching the data while preserving the model variability without ensemble collapse and (2) quantifying the uncertainty in the forecasting period. In the updated permeability distributions, the phase-streamline-based localization seems to preserve geologic continuities affecting the flow dynamics better than other localization methods.

• The power and utility of our proposed method have been demonstrated through a synthetic model and the benchmark PUNQ-S3 model.

Nomenclature
\[ C_{\Psi,k} = \text{joint model parameter data state vector covariance matrix at time } k \]
\[ C_{\Psi,H} = \text{model data cross-covariance matrix at time } k \]
\[ C_{W} = \text{posterior model parameter covariance matrix} \]
\[ C_{D} = \text{data-observation covariance matrix} \]
\[ d_{in,k,k} = \text{coarse-scale permeability-constraint data at time } k \]
\[ d_{obs,k} = \text{observation data at time } k \]
\[ H = \text{measurement matrix} \]
\[ H_{C_{\Psi,H}} = \text{calculated data-error covariance matrix at time } k \]
\[ K_{k} = \text{Kalman gain matrix at time } k \]
\[ m_{l} = \text{static variable vector at time } k \]
\[ m_{u} = \text{dynamic variable vector at time } k \]
\[ N_{p} = \text{number of members in the ensemble} \]
\[ y_{k} = \text{prior joint model parameter data state vector at time } k \]
\[ y_{o} = \text{mean of the prior joint model parameter data state vector} \]
\[ y_{k}^{o} = \text{posterior joint model parameter data state vector at time } k \]
\[ y_{k}^{p} = \text{posterior joint model parameter data state vector at time } k \text{ from the first assimilation step} \]
\[ \Psi_{k} = \text{ensemble of joint model data state vectors at time } k \]
\[ \epsilon_{i} = \text{white-random-noise data observation} \]

Acknowledgments
The authors acknowledge the sponsors of Model Calibration and Efficient Reservoir Imaging and the Texas A&M University Brazos High Performance Computing (HPC) cluster that contributed to the research reported here.

References
Carter, J. 2007. PUNQ-S3 Test Case. Online dataset comparison study, Department of Earth Science and Engineering, Imperial College


Shingo Watanabe is a PhD candidate in petroleum engineering at Texas A&M University in College Station, Texas, USA. His current research interests include data assimilation (EnKF), numerical reservoir simulation, and reservoir characterization. Watanabe holds a BS degree in mineral resources and environmental engineering from Waseda University in Tokyo and an MS degree in petroleum engineering from Texas A&M University.

Akhil Datta-Gupta is Regents Professor and holder of the L.F. Peterson ’36 Endowed Chair in Petroleum Engineering at Texas A&M University in College Station, Texas, USA. Previously, he worked for BP Exploration and Research and for the Lawrence Berkeley National Laboratory. Datta-Gupta holds a PhD degree from the University of Texas at Austin. He is the recipient of the 2009 SPE John Franklin Carl Award and the 2003 SPE Lester C. Uren Award. He also received the SPE Cedric K. Ferguson Certificate twice (2000 and 2006) and the AIME Rossiter W. Raymond Memorial Award (1992). Datta-Gupta is an SPE Distinguished Member and a member of the US National Academy of Engineering.