AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2013WR013842

Key Points:

- Bayesian model to integrate multiscale data, including surface seismic data
- Reactive facies approach to parameterize heterogeneous subsurface properties
- Estimate flow and transport properties along a 300 m transect at high resolution

Correspondence to:
H. M. Wainwright,
hmwainwright@lbl.gov

Citation:
Wainwright, H. M., J. Chen, D. S. Sassen, and S. S. Hubbard (2014), Bayesian hierarchical approach and geophysical data sets for estimation of reactive facies over plume scales, *Water Resour. Res.*, *50*, 4564–4584, doi:10.1002/2013WR013842.

Received 18 MAR 2013
Accepted 6 MAY 2014
Accepted article online 12 MAY 2014
Published online 4 JUN 2014

Bayesian hierarchical approach and geophysical data sets for estimation of reactive facies over plume scales

Haruko M. Wainwright[1], Jinsong Chen[1], Douglas S. Sassen[1,2], and Susan S. Hubbard[1]

[1]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA, [2]Now at ION Geophysical, Houston, Texas, USA

Abstract A stochastic model is developed to integrate multiscale geophysical and point data sets for characterizing coupled subsurface physiochemical properties over plume-relevant scales, which is desired for parameterizing reactive transport models. We utilize the concept of reactive facies, which is based on the hypothesis that subsurface units can be identified that have distinct reactive-transport-property distributions. To estimate and spatially distribute reactive facies and their associated properties over plume-relevant scales, we need to (1) document the physiochemical controls on plume behavior and the correspondence between geochemical, hydrogeological, and geophysical measurements; and (2) integrate multisource, multiscale data sets in a consistent manner. To tackle these cross-scale challenges, we develop a hierarchical Bayesian model to jointly invert various wellbore and geophysical data sets that have different resolutions and spatial coverage. We use Markov-chain Monte-Carlo sampling methods to draw many samples from the joint posterior distribution and subsequently estimate the marginal posterior distribution of reactive-facies field and their associated reactive transport properties. Synthetic studies demonstrate that our method can successfully integrate different types of data sets. We tested the framework using the data sets collected at the uranium-contaminated Savannah River Site F-Area, including wellbore lithology, cone penetrometer testing, and crosshole and surface seismic data. Results show that the method can estimate the spatial distribution of reactive facies and their associated reactive-transport properties along a 300 m plume centerline traverse with high resolution (1.2 m by 0.305 m).

1. Introduction

Understanding and predicting subsurface contaminant-plume evolution and natural attenuation require characterization of heterogeneous flow and reactive transport properties in high resolution over plume-relevant scales. Although numerical modeling of flow, geochemical, and microbial processes have greatly advanced in the last decade [e.g., *Hammond et al.*, 2011; *Yabusaki et al.*, 2011], the difficulty in characterizing flow and reactive transport properties in high resolution and over large scales has hindered the ability to accurately predict plume evolution and to design optimal remediation strategies [e.g., *Scheibe and Chien*, 2003; *Scheibe et al.*, 2006; *Li et al.*, 2010].

Geophysical methods hold great potential for improving the subsurface characterization of both physical and geochemical properties [e.g., *Chen et al.*, 2004; *Scheibe et al.*, 2006]. Although there has been a significant advance in hydrogeophysical characterization [e.g., *Hubbard and Rubin*, 2005; *Vereecken et al.*, 2006], many studies have been limited to field experiments having a domain smaller than ~100 m. Recent studies have recognized the need to advance methods that can be used to characterize subsurface properties over large spatial regions, as are needed to assist with the management of water resources, agrosystems, ecosystems, and remediation of large subsurface plumes [e.g., *Hubbard and Linde*, 2011].

Recently, *Sassen et al.* [2012] documented the utility of a reactive facies concept—a concept based on the hypothesis that subsurface units exist with distinct distributions of coupled physiochemical properties influencing reactive transport, such as effective surface area, mineralogy, and hydraulic conductivity. Because geophysical techniques can also be used to identify such subsurface units having different physical and geochemical properties that often covary, this concept allows us to take advantage of both geophysical and lithological data sets for estimating spatially distributed reactive transport parameters. The reactive facies concept builds upon previous hydrofacies and lithofacies approaches [e.g., *Fogg et al.*, 1998; *Klingbeil et al.*, 1999; *Weissmann et al.*, 2002; *Heinz et al.*, 2003; *Yabusaki et al.*, 2011] to include consideration of

© 2014. American Geophysical Union. All Rights Reserved.

geochemical properties that play a significant role in reactive transport. An extensive review of zonation-based approaches for characterizing the subsurface was provided by *Sassen et al*. [2012]. *Sassen et al*. [2012] developed a Bayesian method for spatially distributing reactive facies and their associated physiochemical properties over local scales (~10 m) and tested it using wellbore and crosshole geophysical data sets at the Savannah River Site (SRS) F-Area, where this study is carried out.

Through data mining and analysis, *Sassen et al*. [2012] identified the presence of two reactive facies at the SRS F-Area site, which were coincident with two depositional facies: a Barrier Beach and a Lagoonal facies [*Jean et al*., 2002, 2004]. The Barrier Beach reactive facies had lower fines, higher hydraulic conductivity, and lower Al: Fe ratio (i.e., proxy for the mineral ratio between kaolinite and goethite) compared to the Lagoonal reactive facies. Laboratory analysis confirmed that the physiochemical differences exhibited by the two depositional facies were important for uranium flow and transport in the groundwater at the F-Area [*Dong et al*., 2012], and thus could be considered reactive facies. *Sassen et al*. [2012] also documented the geophysical identifiability of the facies: the Barrier Beach reactive facies had higher seismic and radar crosshole velocities and lower radar attenuation than the Lagoonal reactive facies. With this information, and using a Bayesian approach, they were able to estimate the spatial distribution of reactive facies over the local scale (or the extent of the crosshole data set, ~10 m) and their associated flow and transport properties. The estimated properties were used to parameterize a local-scale reactive transport model. Simulations were performed to document uranium transport in the presence of this coupled physiochemical heterogeneity, and to explore how it is compared to domains where the physical or chemical heterogeneity (or both) were uniform over the domain. The simulation results showed the value of the reactive facies for improving predictions of plume mobility at the local scale.

This study builds on *Sassen et al*. [2012] with a focus on estimating the spatial distribution of reactive facies over large scales at high resolution, using both hydrogeological and geophysical data sets. Extension to large scales is needed to improve reactive transport simulations at scales where large contaminant plumes, ecosystems, and watersheds are managed. In this study, we refer to two spatial scales: *local (or small) scale* (~10 s of m) and *large scale* (~100 s of m). We define *fine (or high) resolution* and *coarse (or low) resolution* to represent the resolution of the data sets based on their measurement support scale or inversion discretization [*Gotway and Young*, 2002; *Hubbard et al*., 1999]. Fine resolution refers to data sets with resolutions of ~10 cm to 1 m (e.g., core and crosshole geophysical data), while coarse resolution refers to data sets with resolutions larger than 1 m (e.g., surface seismic data).

Fine-resolution characterization over large regions is typically hindered by the simultaneous lack of spatial coverage and resolution of the available data sets. Datasets that provide fine resolution are typically representative of only a small spatial region. Datasets that provide good spatial coverage (such as surface seismic data) usually provide coarse-resolution information, where each pixel in a coarse grid field represents effective or averaged properties. The challenge is to develop effective methods for combining the multiscale data sets (e.g., wellbore, crosshole data, and surface geophysical data) in a consistent manner. Previous studies have proposed several methods for integrating multiscale data sets [e.g., *Wikle et al*., 2001; *Gotway and Young*, 2002; *Ines and Mohanty*, 2009], some of which have also considered geophysical data [*Chen and Hoversten*, 2012]. *Wikle et al*. [2001] integrated multiscale meteorological data sets to estimate the surface wind field using a Bayesian hierarchical model. *Ines and Mohanty* [2009] integrated remote sensing data and in situ measurements for estimating soil hydraulic parameters using a genetic algorithm approach. *Chen and Hoversten* [2012] developed a Bayesian model to jointly invert seismic AVA and CSEM data for estimating reservoir parameters based on data-driven correlations. *Gotway and Young* [2002] reviewed multiple methods for combining spatial data with different support volume (i.e., different resolution), such as block kriging, tree models [e.g., *Willsky*, 2002; *Ferreira and Lee*, 2007], and hierarchical Bayesian models.

Among those, we have adopted a Bayesian hierarchical model to estimate reactive facies and associated transport properties over plume-relevant scales, since it is flexible and extensible to include multiple complex processes, and also to take into account different scales of data sets explicitly. The model consists of three statistical submodels: a data model, a process model, and a prior model, which are expressed following the notation of *Wikle et al*. [2001]:

1. The data model provides a probabilistic linkage between the multiple geophysical data sets and the spatially distributed geophysical attributes through linear or nonlinear forward models.

 © 2014. American Geophysical Union. All Rights Reserved.

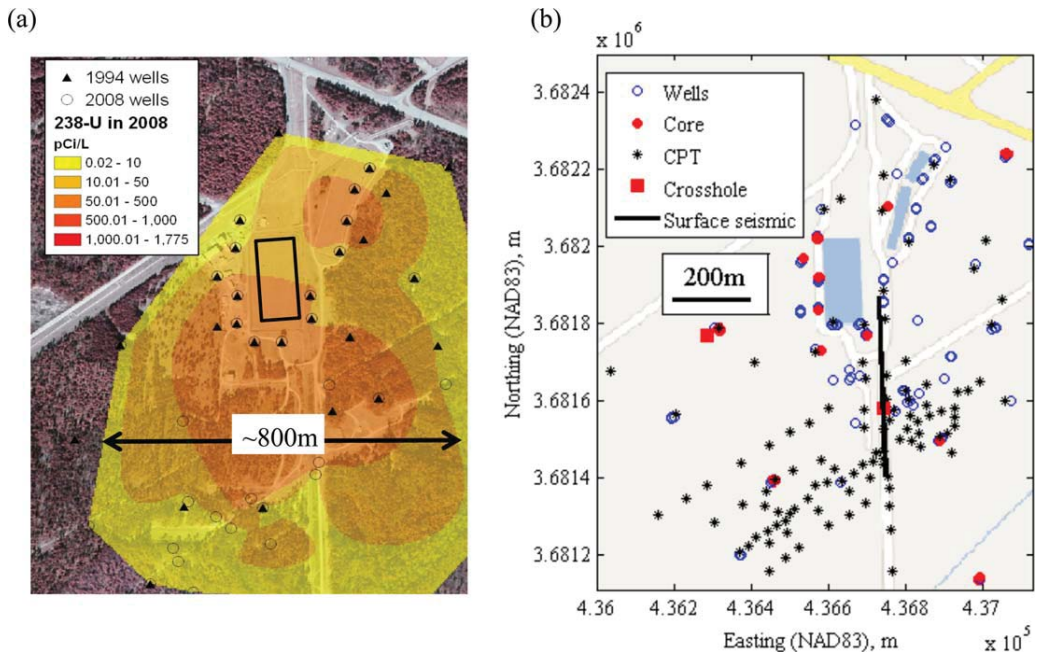

Figure 1. (a) SRS F-Area site map and interpolated U(VI)−238 concentration map over the area in 2008 and (b) data locations data surrounding the F-Area Seepage Basins (shown in blue rectangles).

2. The process model describes the spatial distribution of reactive facies and geophysical attributes as spatial random processes (or fields) described by geostatistical and petrophysical parameters.

3. The prior model is constructed for geostatistical and model parameters based on the information from geologically similar sites.

Once each submodel is properly defined, Markov-chain Monte-Carlo (MCMC) sampling methods can be used to compute the marginal posterior distribution of the reactive facies field and associated parameters for stochastic reactive transport simulations.

We apply our framework to data sets collected at the Savannah River Site (SRS), a former nuclear weapons production site in South Carolina, U.S.A. The SRS F-Area basin was used to dispose of plutonium production waste, which has led to nitric acid and uranium U(VI) plumes in the groundwater that extend several hundred meters downstream from the basins (see Figure 1a). Since sorption of U(VI) is a function of pH, pH-based remediation strategies are now under way at the site to immobilize the contaminant. Although this approach is effective, it is also expensive to maintain for long durations. As such, monitored natural attenuation is the desired closure strategy for the site. This strategy is based on a conceptual model that rainwater will eventually neutralize the lingering mineral surface acidity, causing an increase in pH, which will lead to sorption of U(VI) and thus natural immobilization at the trailing end of the plume. Understanding the long-term H^+ and U(VI) sorption behavior at the site in the presence of natural heterogeneity is critical for assessing the long-term viability of any active or natural remediation strategy [*Denham and Vangelas*, 2008; *Wan et al*., 2012]. *Dong et al*. [2012] performed laboratory studies on site sediments and documented the controls of kaolinite and goethite on pH-dependent U(VI) sorption at the F-Area. Their study suggested that characterization of kaolinite, goethite, and fine content is critical for understanding sorption, and thus long-term plume behavior. Their findings motivate our field-scale estimation challenge: to estimate these controlling parameters on reactive transport, or reactive facies, which are described by unique distributions of transport parameters—in high resolution yet over large scales. The key reactive facies transport parameters that we estimate include fine content, hydraulic conductivity, and Al:Fe ratio, which serves as a proxy for the relative percent of kaolinite to goethite.

Because the SRS F-Area site has been under active remediation for the last 20 years, extensive historic hydrogeological and geochemical data sets are available for this estimation study. Our project has

 © 2014. American Geophysical Union. All Rights Reserved.

augmented this historical database with recently acquired geochemical [*Dong et al*., 2012; *Wan et al*., 2012], crosshole geophysical [*Sassen et al*., 2012], and surface geophysical data. Our subsurface characterization objective is to systematically combine these new geochemical and geophysical data sets with the abundant historical data sets, including wellbore lithology data and cone penetrometer testing (CPT) data.

In section 2, we describe a general multiscale estimation methodology that can be applied to many different sites. We describe the SRS F-Area data sets and the site-specific geophysical forward models in sections 3 and 4, respectively. In section 5, we test our estimation method using a synthetic study, followed by application of the method to the SRS F-Area field data sets.

2. Methodology

In this section, we first introduce the overall model structure and then describe each component of the model in the later subsections. Since we use a Bayesian hierarchical model, we first formulate the probabilistic forward models in sections 2.2 and 2.3, and then explain the MCMC sampling methods for inversion in section 2.4. In this section, we describe a general model, which is applicable to many other sites and different types of geophysical data sets. The site-specific geophysical forward models and petrophysical relationships at the Savannah River Site are discussed in section 4.

2.1. Model Structure

As described earlier, a typical Bayesian hierarchical model consists of three main statistical submodels: (1) a data model: $p(data|field, \phi)$, (2) a process model: $p(field|\theta)$, and (3) a prior model: $p(\phi,\theta)$, in which the random vectors ϕ and θ represent model parameters that will be specified later [*Wikle et al*., 2001]. In our case, the process model describes the spatial distribution of reactive facies and geophysical attributes (referred to hereafter as the reactive-facies field and geophysical-attribute field, respectively) conditioned on the process model parameters θ, which includes both geostatistical parameters and petrophysical correlation parameters. The data model links the actual data to the reactive-facies field and geophysical-attribute field for given model parameters ϕ. The overall model, decoupled into a series of conditional models, is flexible enough to include complex physical processes or observations. Once all the conditional models are specified, we can estimate the marginal posterior distribution for the reactive facies field $p(field|data)$, using MCMC sampling methods.

Figure 2 shows a graphical model representation of the overall model. Without losing generality, we consider two different levels of scales in terms of data resolution and spatial coverage. We define three random vectors, each of which is a random field: (1) a reactive-facies field $\{f_i, i = 1,\ldots, n\}$, (2) a coarse-resolution field of geophysical attributes $\{u_j, j = 1, \ldots, m\}$ (e.g., surface-seismic P-wave velocity, seismic slowness or acoustic impedance), and (3) a fine-resolution field of geophysical attributes $\{v_i, i = 1, \ldots, n\}$ (e.g., crosshole-seismic P-wave velocity, seismic slowness, or acoustic impedance). Each pixel of the reactive facies field $\{f_i\}$ is an indicator variable representing the presence or absence of a particular facies. Data used in the estimation include crosshole geophysical data $\boldsymbol{z}_t$ (e.g., first arrival time in crosshole tomography) depending on the fine-resolution geophysical-attribute field $\{v_i\}$, surface geophysical data $\boldsymbol{z}_s$ (e.g., seismic full waveform responses) as a function of the coarse-resolution geophysical-attribute field $\{u_j\}$, and facies data $\boldsymbol{z}_c$ as point measurements of the facies along boreholes.

As shown in Figure 2a, each pixel of the coarse-resolution field is connected to several pixels of the fine-resolution fields, which represents multiple resolutions of geophysical attributes. Following *Willsky* [2002] and *Ferreira and Lee* [2007], we call the coarse-resolution pixel in this connection the *parent pixel*, and the fine-resolution pixel the *child pixel*. We define a vector of indices $\boldsymbol{\alpha}(j)$ to represent the child pixels of Pixel j, and define a scalar index $\gamma(i)$ to represent the parent pixel of Pixel i.

Figure 2b is a graphical model describing a series of conditional relationships. Each connection represents a state of conditional dependency—Markovian in the sense that the two nodes are conditionally independent if they are not connected directly, given their intermediate nodes. For example, $\{f_i\}$ and $\{u_j\}$ are conditionally independent given $\{v_i\}$. The process model parameters θ include (1) geostatistical parameters θ_f, which describe the spatial correlation structure of the reactive-facies field $\{f_i\}$, (2) petrophysical parameters θ_p, which describe the relationship between the reactive-facies field $\{f_i\}$ and fine-resolution geophysical-attribute field $\{v_i\}$, and (3) upscaling parameters θ_u, which are part of an upscaling function from $\{v_i\}$ to $\{u_j\}$.

© 2014. American Geophysical Union. All Rights Reserved.

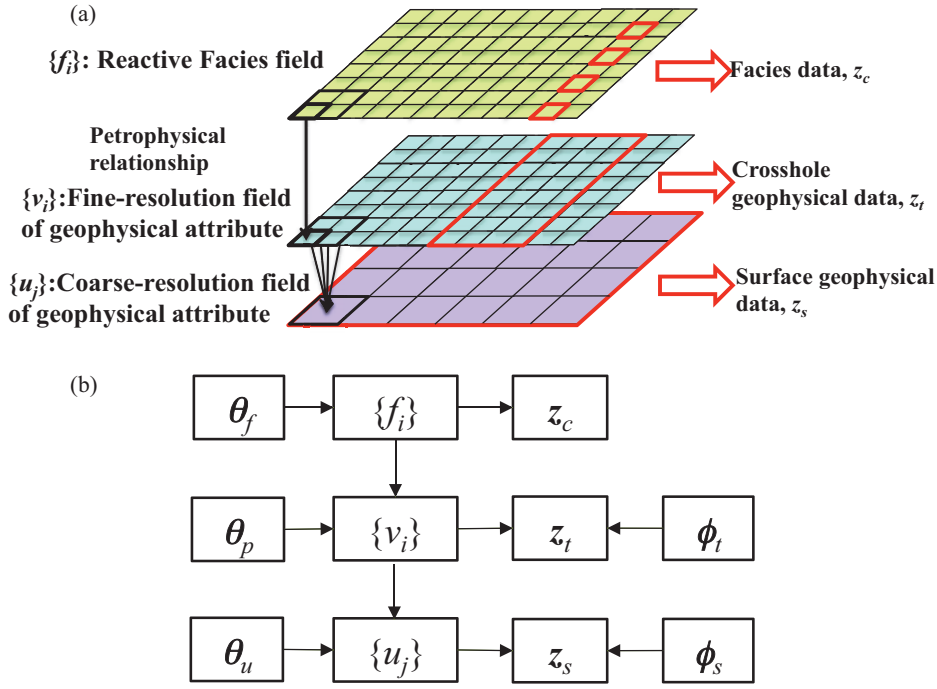

Figure 2. Multiscale estimation diagram; (a) conceptual model and (b) graphical model. In Figure 2a, the areas outlined by black represent the pixels in each field, and the areas outlined by red represent the pixels that are associated with each measurement type (i.e., the pixels are at the point measurement locations or within the coverage of geophysical data sets).

The data model parameters ϕ include (1) the crosshole data parameters ϕ_t, which describe the distributions of crosshole data $\boldsymbol{z}_t$ conditioned on a given $\{v_i\}$, and (2) the surface data parameters ϕ_s, which describe the distributions of surface geophysical data $\boldsymbol{z}_s$ conditioned on given $\{u_j\}$. Each component of θ and ϕ can have a probabilistic distribution to represent parameter uncertainty, or they can be fixed (equivalent to having a Dirac delta function for a distribution), depending on information available at the site. In sections 2.2 and 2.3, we define conditional models of those connections shown in Figure 2b.

2.2. Data Models

The data model connects geophysical data sets to the underlying geophysical attributes. We denote the surface geophysical data $\boldsymbol{z}_s$ as a function of the coarse-resolution geophysical-attribute field $\{u_j\}$, surface data parameters ϕ_s, and measurement-error vector ε_s:

$$z_s = f_s(\{u_j\}, \phi_s) + \varepsilon_s. \tag{1}$$

Here we assume that ε_s has a zero-mean multivariate normal (MVN) distribution with a given covariance matrix D_s: $\varepsilon_s \sim MVN(0, D_s)$. For independent and identically distributed errors, D_s is a diagonal matrix with the measurement error variance of σ^2_s. The function $f_s(\bullet)$ represents a forward geophysical model. For example, $f_s(\bullet)$ may describe seismic amplitude as a function of acoustic impedance and incident waveform obtained at each trace using a convolution model [*Aki and Richards*, 1980; *Bosch et al.*, 2007]. According to the normality assumption of the measurement errors, we have the following conditional distribution for $\boldsymbol{z}_s$:

$$p(z_s|\{u_j\}, \phi_s) = MVN(f_s(\{u_j\}, \phi_s), D_s). \tag{2}$$

Similarly, we can define the crosshole data in terms of the fine-resolution geophysical attributes $\{v_i\}$, crosshole model parameters ϕ_t, and measurement error vector ε_t:

© 2014. American Geophysical Union. All Rights Reserved.

$$z_t = f_t(\{v_i\}, \boldsymbol{\phi}_t) + \boldsymbol{\varepsilon}_t, \tag{3}$$

where $f_t(\bullet)$ represents the forward model (such as a straight-ray model [*Peterson et al*., 1985] or Fresnel zone-based model [*Johnson et al*., 2005]). Again, we assume that $\boldsymbol{\varepsilon}_t$ has a zero-mean multivariate normal distribution with a given covariance matrix: $\boldsymbol{\varepsilon}_t \sim N(0, \mathrm{D}_t)$. For independent and identically distributed errors, D_t is a diagonal matrix with the measurement error variance of σ_t^2. The conditional distribution for $\boldsymbol{z}_t$ given the fine-resolution geophysical attributes thus becomes:

$$p(z_t|\{v_i\}, \boldsymbol{\phi}_t) = MVN(f_t(\{v_i\}, \boldsymbol{\phi}_t), \mathrm{D}_t). \tag{4}$$

The facies data, denoted by $\boldsymbol{z}_c$, are the direct point measurements of facies, which are a subset of $\{f_i\}$.

2.3. Process Models

2.3.1. Reactive Facies Field

We follow the approach developed by *Chen et al*. [2006] to formulate the reactive facies field $\{f_i\}$. Here $\{f_i\}$ is an indicator field such that the facies at each element f_i has a Bernoulli distribution given by:

$$p(f_i = 1|\{f_k, k \neq i\}, \boldsymbol{\theta}_f) = \text{Bernoulli}\,(p_i), \tag{5}$$

where p_i can be determined by indicator kriging:

$$p_i = \mu + \sum\nolimits_{k \neq i} \lambda_k (f_k - \mu), \tag{6}$$

where μ is the overall mean. Note that p_i is truncated within [0, 1]. The kriging coefficients λ_k are a function of the correlation length and sill, and the distance between Pixel k and Pixel i. The model parameter vector $\boldsymbol{\theta}_f$ includes the overall mean μ, and correlation length and sill.

2.3.2. Petrophysical Relationship

The petrophysical model connects the fine-resolution geophysical-attribute field $\{v_i\}$ with the reactive facies field $\{f_i\}$. Following *Sassen et al*. [2012] and *Chen et al*. [2006], we assume that the fine-resolution geophysical attribute at Pixel i v_i has a distribution in each facies described as:

$$v_i = a_1 + a_2 f_i + \varepsilon_v, \tag{7}$$

where a_1 and a_2 are the coefficients that can be obtained from fitting colocated borehole data, and ε_v represents the variability within each facies defined by the variance τ_v ($\varepsilon_v \sim N(0, \tau_v)$). Although we may consider a nonlinear function within the petrophysical model, equation (7) is fairly general for a two facies case, since equation (7) signifies that the two facies have two distributions of geophysical attributes with mean a_1 and $a_1 + a_2$, and variance τ_v. We denote all the petrophysical parameters by $\boldsymbol{\theta}_p = \{a_1, a_2, \tau_v\}$. Although the variance τ_v is currently the same for both facies, the model can be modified to have two different variances [*Chen et al*., 2006]. We write the conditional distribution for v_i as follows:

$$p(v_i|f_i, \boldsymbol{\theta}_p) = N(a_1 + a_2 f_i, \tau_v). \tag{8}$$

2.3.3. Upscaling of Geophysical Attributes

To integrate data sets with different resolutions, we must aggregate fine-resolution geophysical attributes to a coarse grid. Geophysical data analysis routinely uses filtering to combine small-scale and large-scale data (for example, simple or Backus averaging in the seismic data analysis [*Bosch et al*., 2007; *Tiwary et al*., 2009]). Although we follow this approach, we include this averaging process directly into the inversion, so that we can address upscaling uncertainty associated with the estimation process. We consider the coarse-resolution geophysical attributes at each pixel u_j as a function of the fine-resolution geophysical attributes at their child pixels $\{v_k, k \in \boldsymbol{\alpha}(j)\}$ and an additive random variability ε_u:

© 2014. American Geophysical Union. All Rights Reserved.

$$u_j = g(\{v_k, k \in \boldsymbol{\alpha}(j)\}) + \varepsilon_u, \tag{9}$$

where we assume that $\varepsilon_u \sim N(0, \tau_u)$ and g(•) is an upscaling function, such as simple averaging or Backus averaging. The variance τ_u represents random variability. We denote the upscaling parameter vector by $\boldsymbol{\theta}_u = \{\tau_v\}$. We can then write the conditional distribution as:

$$p(u_j|\{v_i\}, \boldsymbol{\theta}_u) = N(g(\{v_k, k \in \alpha(j)\}), \tau_u). \tag{10}$$

2.4. Joint Probability Distribution

The goal is to determine the posterior distribution of the reactive-facies field $\{f_i\}$ given the available data sets ($\boldsymbol{z}_s$, $\boldsymbol{z}_t$, $\boldsymbol{z}_c$) at the two different scales $p(\{f_i\}|\boldsymbol{z}_s, \boldsymbol{z}_t, \boldsymbol{z}_c)$, which is the marginal distribution of the joint posterior distribution $p(\{f_i\},\{u_j\},\{v_i\}, \boldsymbol{\phi}, \boldsymbol{\theta}|\boldsymbol{z}_s, \boldsymbol{z}_t, \boldsymbol{z}_c)$. The data-model parameters are $\boldsymbol{\phi} = \{\boldsymbol{\phi}_s, \boldsymbol{\phi}_t\}$ (defined in section 2.2) and the process-model parameters are $\boldsymbol{\theta} = \{\boldsymbol{\theta}_f, \boldsymbol{\theta}_p, \boldsymbol{\theta}_u\}$ (defined in section 2.3). Bayes' rule yields:

$$\begin{aligned} &p(\{f_i\}, \{u_j\}, \{v_i\}, \boldsymbol{\phi}, \boldsymbol{\theta}|z_s, z_t, z_c) \\ &\alpha p(z_s|\{u_j\}, \boldsymbol{\phi}_s)p(z_t|\{v_i\}, \boldsymbol{\phi}_t)p(\{u_j\}|\{v_i\}, \boldsymbol{\theta}_u)p(\{v_i\}|\{f_i\}, \boldsymbol{\theta}_p)p(\{f_i\}|\boldsymbol{\theta}_f, \ \mathbf{z}_c)p(\boldsymbol{\phi}, \boldsymbol{\theta}). \\ &\alpha p(z_s|\{u_j\}, \boldsymbol{\phi}_s)p(z_t|\{v_i\}, \boldsymbol{\phi}_t)\prod_{j=1}^{m} p(u_j|\{v_i\}, \boldsymbol{\theta}_u)\prod_{i=1}^{n} p(v_i|f_i, \boldsymbol{\theta}_p)p(\{f_i\}|\boldsymbol{\theta}_f, \ \mathbf{z}_c)p(\boldsymbol{\phi}, \boldsymbol{\theta}). \end{aligned} \tag{11}$$

In the previous sections, we have defined all the conditional distributions on the right-hand sides of equations (2), (4), (5), (8), and (10) except for the prior distribution $p(\boldsymbol{\phi}, \boldsymbol{\theta})$.

Sampling methods are invoked to sample from the joint probability distribution in equation (11), which can then be used to obtain posterior probability distributions and associated confidence intervals of the desired properties (in this case, reactive facies). MCMC sampling methods [e.g., *Andreu et al.*, 2003; *Gamerman and Lopes*, 2006] have been used in Bayesian estimation of hydrological properties [e.g., *Michalak and Kitanidis*, 2003; *Chen et al.*, 2004, 2006; *Vrugt et al.*, 2005, 2006]. MCMC sampling requires formulation of the probability distribution of each parameter conditioned on the other parameters and all data sets. For this study, we use two MCMC sampling methods: Gibbs sampling for the parameters with distributions from which sampling is straightforward (e.g., normal and multivariate normal), and Metropolis-Hasting (MH) sampling for all other parameters. When the forward model is linear and the prior model is normal or multivariate, for example, the posterior distribution is multivariate normal, which allows us to use the Gibbs sampler. We sequentially sample each parameter defined in equation (11) (i.e., $\{f_i\}$,$\{u_j\}$,$\{v_i\}$, $\boldsymbol{\phi}$, $\boldsymbol{\theta}$) according to their corresponding conditional distributions given below.

2.4.1. Sampling of Coarse-Resolution Geophysical Attributes

Let $p(\{u_j\}|\ \bullet)$ denote the conditional distribution of the coarse-resolution geophysical-attribute field $\{u_j\}$ conditioned on all the other variables and data sets. Based on the conditional independence shown in the diagram in Figure 2b and equation (11), we have:

$$p(\{u_j\}|\bullet)\alpha p(z_s|\{u_j\}, \boldsymbol{\phi}_s)p(\{u_j\}|\{v_i\}, \boldsymbol{\theta}_u). \tag{12}$$

Equation (12) implies that sampling $\{u_j\}$ only involves the likelihood function of the surface geophysical data $p(\boldsymbol{z}_s|\{u_j\}, \boldsymbol{\phi}_s)$, and the upscaling model $p(\{u_j\}|\{v_i\}, \boldsymbol{\theta}_u)$ that is the probability of $\{u_j\}$ given the fine-resolution geophysical attributes at the child pixels.

2.4.2. Sampling of Fine-Resolution Geophysical Attributes

To sample the fine-resolution geophysical attribute from $p(\{v_i\}|\ \bullet)$, we need to separate the locations containing crosshole data from other locations. We define crosshole-data locations as the pixels involved in a forward geophysical process (e.g., pixels along the ray-paths or within Fresnel zones). At the crosshole-data locations, we have the following conditional distribution:

$$p(\{v_i\}|\bullet)\alpha p(z_t|\{v_i\}, \boldsymbol{\phi}_t)p(\{u_j\}|\{v_i\}, \boldsymbol{\theta}_u)p(\{v_i\}|\{f_i\}, \theta_p). \tag{13}$$

Sampling $\{v_i\}$ requires the likelihood function of crosshole data $p(\boldsymbol{z}_t|\{v_i\}, \boldsymbol{\phi}_t)$, the upscaling model of the geophysical attributes $p(\{u_j\}|\{v_i\}, \boldsymbol{\theta}_u)$, and the petrophysical relationship $p(\{v_i\}|\{f_i\}, \boldsymbol{\theta}_p)$.

© 2014. American Geophysical Union. All Rights Reserved.

At the locations where crosshole data are not available, the data model is not needed, and the conditional distribution includes only two components:

$$p(\{v_i\}|\bullet)\alpha p(\{u_j\}|\{v_i\},\boldsymbol{\theta}_u)p(\{v_i\}|\{f_i\},\boldsymbol{\theta}_p). \tag{14}$$

2.4.3. Sampling of Reactive Facies Field

The conditional distribution of the reactive facies field is written by $p(\{f_i\}|\bullet) = p(\{f_i\}|\{v_i\}, \boldsymbol{\theta}_p, \theta_f, z_c)$. Following *Chen et al*. [2006], we use the Gibbs sampler to draw samples at each pixel from $p(f_i|\bullet)$. Bayes' rule yields:

$$p(f_i|\bullet)\alpha p(v_i|f_i,\boldsymbol{\theta}_p)p(f_i|\{f_k,k\neq i\},\boldsymbol{\theta}_f,z_c). \tag{15}$$

The petrophysical relationship $p(v_i|f_i, \boldsymbol{\theta}_p)$ and the conditional facies at each pixel $p(f_i|\{f_k, k\neq i\}, \boldsymbol{\theta}_f, \boldsymbol{z}_c)$ are defined by equations (8) and (5), respectively. The conditional distribution $p(f_i|\bullet)$ is the Bernoulli distribution with probability p_i^*:

$$p_i^* = \frac{p_i q_i}{1-p_i+p_i q_i}, \tag{16}$$

where $q_i = \exp[\tau_v^{-1} a_2(v_i - a_1 - 0.5a_2)]$.

3. Site and Data Descriptions

3.1. Site Description

The hydrologeology of the F-Area, described in several previous studies [*Jean et al*., 2002, 2004; *Flach*, 2004; *Phifer et al*., 2006; *Sassen et al*., 2012], is briefly summarized here. The Savannah River Site (SRS) is located within the Atlantic Coastal Plain physiographic province, which is characterized by subhorizontal sedimentary layers. The water table is ~5–20 m below ground surface, depending on location and elevation. The uranium plume from the F-Area basins mostly exists in one geologic unit called the Barnwell Group [*Jean et al*., 2002, 2004]. The Barnwell Group was deposited in a shallow marginal marine environment and consists of two depositional facies: a sandy Barrier Beach facies and a silty Lagoonal facies. *Sassen et al*. [2012] established that these depositional facies were equivalent to reactive facies, and also established the geophysical identifiability of these facies using wellbore and crosshole local-scale data sets. From a hydrostratigraphic perspective, the Barnwell Group is divided by the Tan Clay Zone (TCZ) into two groundwater aquifers: the Upper and Lower Upper Three Runs Aquifers. The TCZ has a high occurrence of Lagoonal facies, although it is discontinuous and multilayered at some locations rather than a single continuous layer. The site data suggest that the two aquifers are hydrologically connected, because there is little difference in the piezometric data and the uranium plume exists in both aquifers. Characterizing the continuity of the Lagoonal Facies in the Tan Clay Zone is of great interest for simulating the long-term behavior of the subsurface plumes.

3.2. Point-Scale Data Sets

Several decades of intensive site characterization has resulted in the accumulation of rich historical data sets, including core sample analysis, cone penetrometer testing (CPT), and groundwater monitoring data sets at many locations. The analysis of cores (every 0.3048 m) along more than 10 wells (Figure 1b) includes soil texture analysis, visual inspection of color, and classification of depositional facies [*Jean et al*., 2002, 2004]. The abundance of analyzed cores provides a unique opportunity for characterizing subsurface variations in lithology and depositional facies.

CPT, an in situ soil exploration tool routinely used for environmental and geotechnical applications in shallow unconsolidated environments [*Lunne et al*., 1997], is also available for characterizing lithological (or depositional facies) variability. With CPT, a cone on the end of a series of rods is pushed into the ground at a constant rate, which provides continuous measurements of resistance and friction against the cone penetrating through the soil. The piezocone penetrometer can also simultaneously measure pore pressure behind the cone. The pore pressure and friction ratio (i.e., the ratio between the cone resistance and sleeve friction) can be used to characterize geologic interfaces and soil types. Empirical relations are often used to

© 2014. American Geophysical Union. All Rights Reserved.

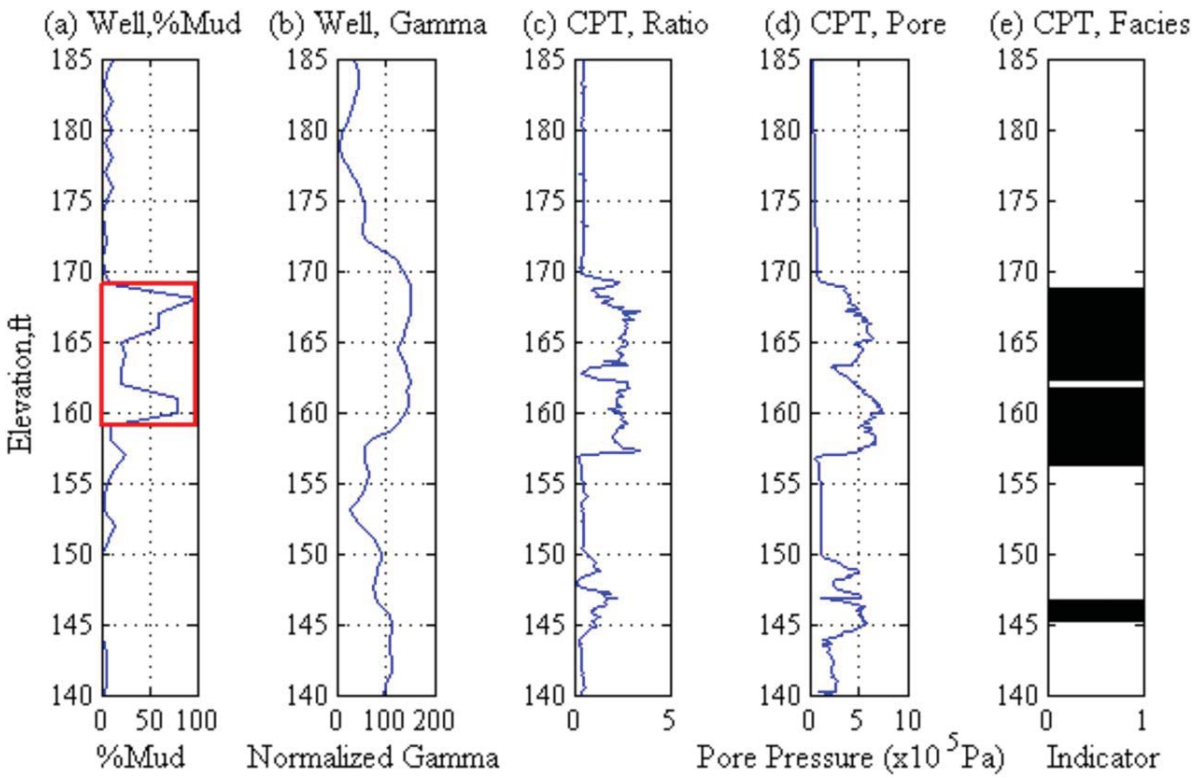

Figure 3. Comparison between the well data (FSB79A) and CPT data (FCPT14) located 6.7 m away; (a) % mud at the well, with the Lagoonal Facies identified by the red box (interpreted by *Jean et al*. [2002, 2004]), (b) Well-based gamma-ray log data, (c) CPT friction ratio, (d) CPT pore pressure, and (e) depositional facies based on the CPT data cluster analysis shown in Figure 4 (the white region is the Barrier Beach facies, and the black region is the Lagoonal facies).

link CPT and other properties; one example of this is the relationship between cone resistance and acoustic impedance [e.g., *Lunne et al*., 1997]. Site-specific CPT-based petrophysical relationships can also be developed. At the F-Area, most of the CPT data are located at the southern portion of the site (Figure 1b), where fewer cores are available. Consequently, in this study, we use the CPT data to augment the point-scale data sets for characterizing the subsurface environment. Although the CPT measurements were done during several campaigns, we assume that all the CPT data sets have the same quality. We note that all the CPT sites were surveyed and all the data sets have been carefully archived in the data management system at the SRS.

To relate the CPT and wellbore information, we compared nearby data sets and performed a cluster analysis of CPT-associated facies interpretation. Figure 3 shows a comparison of the CPT data with the adjacent well data (6.7 m separation distance) that had been interpreted in terms of depositional facies [*Jean et al*., 2002, 2004]. It suggests that CPT can identify interfaces between Lagoonal and Barrier Beach facies better than gamma-ray logs, which are traditionally used to characterize clay content [*Keys*, 1989]. In Figure 4, the CPT friction ratio and pore pressure has two Gaussian-like distributions in the Barnwell Group, which correspond to the two depositional facies. To estimate depositional facies using the CPT friction ratio and pore pressure without subjective picking of the interfaces between the two facies, we used the expectation maximization (EM) algorithm [*Hastie et al*., 2001] to cluster the CPT responses into the Lagoonal and Barrier Beach facies. Figure 3e shows a reasonable agreement between the facies based on the CPT data and the actual depositional facies in Figure 3a, although an exact match is not expected due to the separation distance between the CPT and wellbore. In this study, we assume that the errors in the CPT-determined facies are insignificant.

3.3. Geophysical Data Sets

Surface seismic reflection data were acquired at the SRS F-Area in 2010. Common midpoint (CMP) data were acquired by Rick Miller (Kansas Geological Survey) using a minivibrator source and a 392 channel Geometric Geode distributed seismograph. Vertical seismic profiling along one borehole and normal moveout

 © 2014. American Geophysical Union. All Rights Reserved.

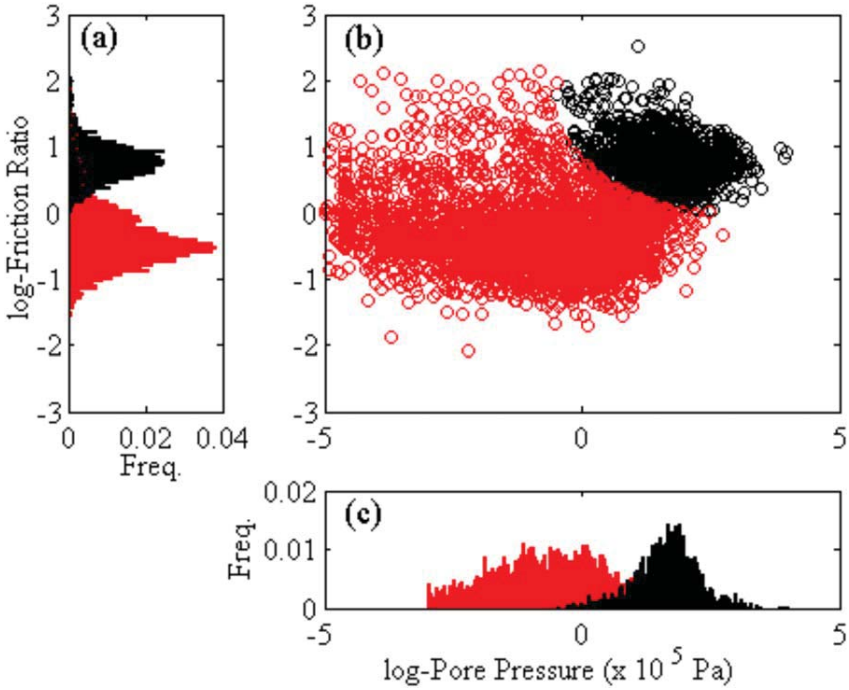

Figure 4. Clustered CPT responses: (a) histogram of friction ratio, (b) cross plot of pore pressure and friction ratio, and (c) histogram of pore pressure. The red color is the interpreted Barrier Beach facies and the black color is the Lagoonal facies.

velocities at 16 locations provided estimates of depth-discrete seismic velocities. These normal moveout velocities did not show significant lateral change along this line. The acquired tenfold CMP data were filtered, migrated, and stacked [*Miller et al.*, 1990; *Steeples and Miller*, 1990; *Miller*, 1992; *Miller et al.*, 2010]. As is

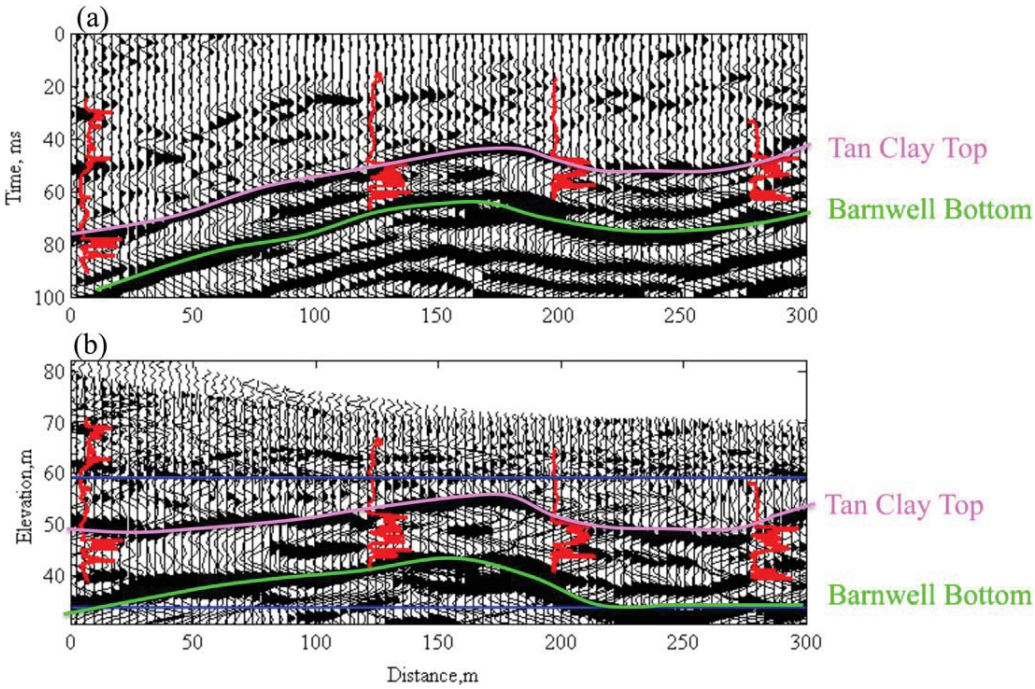

Figure 5. Stacked and migrated surface seismic data displayed as function of (a) two-way travel time and (b) elevation. The red lines are the CPT data used in this study, and the blue rectangle represents the estimation domain. The NMO velocity estimates were used to convert the two-way time section shown in Figure 5a to the depth section shown in Figure 5b.

© 2014. American Geophysical Union. All Rights Reserved.

Table 1. Parameters Used in the Estimation

Domain and pixel sizes	
Domain size {horizontal, vertical}, m	{300, 25.6}
Fine pixel size {horizontal, vertical}, m	{1.20, 0.305}
Coarse pixel size {horizontal, vertical}, m	{6.00, 0.915}
Surface seismic data model	
Data error variance σ_s	
Standard deviation	5% of the amplitude range
Data parameters $\boldsymbol{\phi}_s$	
Vadose zone seismic velocity, m/s	7.01×10^2
Saturated zone homogeneous velocity, m/s	1.25×10^3
Mean frequency, Hz {mean, standard deviation}	{69.2, 11.5}
Source amplitude {mean, standard deviation}	{234.0, 87.5}
Crosshole seismic data model	
Data error variance σ_t	
Standard deviation	5% of the arrival time range
Data parameters $\boldsymbol{\phi}_t$	
None	
Process models	
Geostatistical model parameters $\boldsymbol{\theta}_f$	
Overall mean μ	0.322
Sill	0.201
Horizontal correlation length, m	15.1
Vertical correlation length, m	1.5
Petrophysical model parameters $\boldsymbol{\theta}_p$	
a_1, s/m	5.78×10^{-1}
a_2, s/m	1.13×10^{-2}
τ_v, (s/m)^2	4.50×10^{-5}
Upscaling model parameters $\boldsymbol{\theta}_u$	
τ_u, (s/m)^2	3.00×10^{-6}

shown in Figure 5, the migrated section displays good reflection coherency and high signal-to-noise ratio over our domain of interest (blue rectangle in Figure 5). Figure 5 also shows the interface of the Tan Clay Zone top and the bottom of the Barnwell Group, interpolated based on the deep wellbore data nearby, although it is difficult to tell the thickness and continuity of the Tan Clay Zone.

Multioffset seismic crosshole data were collected in the saturated section of Upper Three Runs Aquifers using an LBNL piezoelectric source, and a hydrophone sensor string. The central frequency of the pulse was 4000 Hz, with a bandwidth from ~1000 to 7000 Hz. Source and geophone spacing in the boreholes was 0.25 m, and travel times were picked for all source-receiver pairs. Picking and inversion of the crosshole seismic data were described by *Sassen et al.* [2012].

4. Site Specific Model

Our objective is to estimate the spatial distribution of the two reactive facies and associated transport parameters (percent fines, hydraulic conductivity, and Al:Fe ratio) along a 300 m traverse that is approximately aligned with the plume centerline. We demonstrate our approach along the two-dimensional (2-D) domain transect shown in Figure 1b, which corresponds both to the surface seismic data line (Figure 5) and to the 2-D domain used for reactive transport simulations [*Bea et al.*, 2013]. The domain and pixel sizes are shown in Table 1. We used four CPT profiles and one set of crosshole seismic data along this line for conditioning. We focus on characterizing a vertical rectangular domain in the saturated and contaminated portion of the Barnwell Group, where we are confident in the seismic data coverage and quality.

Following the model structure described in section 2.1, we used the reactive-facies field $\{f_i\}$, the fine-resolution seismic-slowness field $\{v_i\}$, and coarse-resolution seismic-slowness field $\{u_j\}$, where slowness is the inverse of velocity in the estimation procedure. As described in Figure 2a and section 2.1, the crosshole data $\boldsymbol{z}_t$ is a function of $\{v_i\}$, and the surface seismic data $\boldsymbol{z}_s$ is a function of $\{u_j\}$. The point measurements of reactive facies $\boldsymbol{z}_c$ are the CPT-based facies data. We chose the fine-resolution pixel size as 1.20 m and 0.305 in the horizontal and vertical directions, respectively, which is approximately the scale of cores as well as the resolution of the crosshole tomography data. For the coarse-resolution pixel, we averaged every five pixels in the horizontal direction and three pixels in the vertical direction, since the surface seismic data resolution is approximately several meters in the horizontal direction and 1 m in the vertical direction.

The site-specific geophysical data models (following section 2.2) were formulated as follows. The data model parameters ($\boldsymbol{\phi}_s$, $\boldsymbol{\phi}_t$) and measurement error variance (σ_s, σ_t) are shown in Table 1. For the surface seismic data, we used the same forward model as *Bosch et al.* [2007]. Each trace was simulated as one-dimensional zero-offset seismic data reflected in a horizontally layered medium, assuming horizontal homogeneity within the region surrounding each trace. The forward model for the likelihood function $p(\boldsymbol{z}_s|\boldsymbol{\phi}_s,\{u_j\})$ is a convolution model as a function of acoustic impedance along each trace, which is the coarse-resolution seismic velocity (i.e., the inverse of $\{u_j\}$) multiplied by the bulk density. We assumed that the noise had an independent Gaussian distribution with a standard deviation of 5% of the amplitude range, following *Bosch et al.* [2007]. In addition, the surface geophysical model parameters $\boldsymbol{\phi}_s$

© 2014. American Geophysical Union. All Rights Reserved.

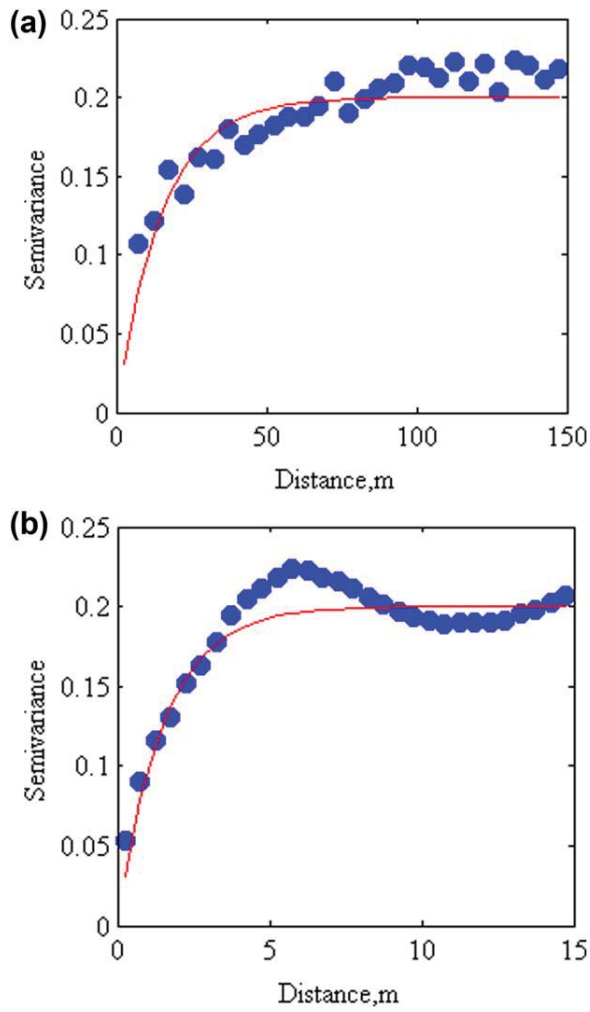

Figure 6. Indicator variogram of Lagoonal Facies (a) in the horizontal direction and (b) in the vertical direction.

include the vadose-zone and saturated-zone (above the estimation domain) seismic velocity, and the source amplitude and frequency. We also assumed that the seismic velocity in the unsaturated zone and saturated zone above the domain is constant and laterally homogeneous, based on the NMO velocity profiles. We estimated the source amplitude and frequency, assuming the Ricker waveform, based on the known facies at the traces corresponding to the CPT locations by the least-square fitting. The estimates were quite uncertain, so we assumed a normal distribution for both the amplitude and frequency, with the mean and standard deviation shown in Table 1. We used the Hann window function to create a window for our domain of interest within the Barnwell Group. We also applied Metropolis-Hastings methods to draw samples from the probability distribution function (Appendix A), since surface seismic data are nonlinear functions of seismic slowness $\{u_j\}$, and we cannot obtain analytical forms for the likelihood function $p(\boldsymbol{z}_s|\phi_s,\{u_j\})$.

The crosshole seismic data $\boldsymbol{z}_t$ is the first-arrival time at every transmitter-receiver combination, which is a function of fine-resolution seismic slowness $\{v_i\}$. We used a straight-ray approximation [*Peterson et al*., 1985], so that the model is linear with respect to slowness. Since the arrival time is a function of the grid size and seismic slowness only, we do not need any other parameters (ϕ_t is an empty vector). The data noise is assumed to be an independent Gaussian distribution, with a standard deviation of 5% of the range for the first arrival time. For this MCMC estimation, we used the same procedure for sampling fine-resolution seismic velocities as the one described in *Chen et al*. [2006] (Appendix B), who inverted crosshole seismic data sets jointly with wellbore data sets.

The process models (described in section 2.3) include the geostatistical model for the reactive-facies field, petrophysical relationship, and upscaling model. Table 1 shows all the parameters (θ_f, θ_p, θ_u). The geostatistical parameters θ_f were estimated based on the variogram analysis of core-based facies and CPT-based facies data. Figure 6 shows the empirical indicator variogram fitted by the exponential covariance model based on the facies data from the 105 CPT and 12 wells at the site. Although *Jean et al*. [2002, 2004] used nested variograms, we used a variogram with single correlation length, since it was found to be sufficient for this inversion. We fixed the geostatistical parameters during the multiscale estimation to avoid numerical instability, although theoretically we could estimate the geostatistical parameters simultaneously if we have sufficient information. The petrophysical relationship θ_p is identified by the same data sets used in *Sassen et al*. [2012]. For upscaling from the fine-resolution to coarse-

 © 2014. American Geophysical Union. All Rights Reserved.

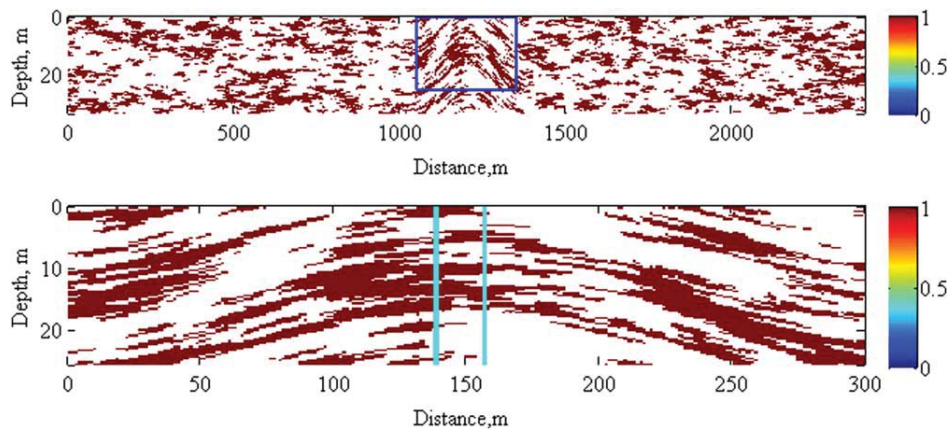

Figure 7. (a) Reactive-facies field with the local curvature generated for the synthetic study (site-scale) and (b) reactive-facies field in the geophysical-inversion domain. In these figures, the white regions are the Barrier Beach facies, and the red regions are the Lagoonal facies. In Figure 7b, the blue lines represent the crosshole geophysical data wellbore locations, and the black lines are the CPT data locations. The image shown in Figure 7b is an enlargement of the area shown by the blue rectangle in Figure 7a.

resolution seismic slowness, we used arithmetic averaging [*Tiwary et al.* 2009], given by $u_j = n(j)^{-1} \sum_{k\in\alpha(j)} v_k + \varepsilon_u$ following equation (10), where $n(j)$ is the number of child pixels of Pixel j in the coarse-resolution field. We assumed that the variance of the upscaling process (τ_u) in equation (10) represents the uncertainty in averaging, so that the variance is that of fine-resolution seismic slowness (τ_v) divided by the number of averaged cells (15).

5. Results and Discussion

We first perform a synthetic study using a domain and data sets that mimic those of the F-Area, followed by an application of our new methodology to real F-Area field data sets.

5.1. Synthetic Study

For the synthetic study, we used the domain, data locations, and data types representative of the actual F-Area field data sets—except for the domain of the crosshole seismic data set, which we expanded for this synthetic study to facilitate visualization and understanding of the estimation results. We used the same geophysical forward models for the synthetic data sets (including the noise) described in section 4 and Appendices.

Figure 7 shows a synthetic "true" reactive-facies field generated using the GSTAT package [*Pebesma*, 2004] in R [*R Development Core Team*, 2007]. We used the geostatistical parameters identified at the site shown in Table 1 and Figure 6. The surface seismic data (Figure 5) shows that the large Lagoonal Facies within the TCZ is likely to have a curved feature in the geophysical-inversion domain. Such local feature is difficult to capture in the variogram analysis, since there are few point measurements within this domain (only four CPT logs). The previous study at the site also used a variogram in the horizontal and vertical directions [*Jean et al.*, 2002, 2004]. In the hydrogeophysical inversion, a geostatistical model—developed based on the data sets from a much wider region than the geophysical-inversion domain—serves as prior information before integrating geophysical data sets. Previous studies have shown that geophysical inversion can capture such curved features despite using horizontal and vertical variograms thanks to the strong conditioning effects of geophysical data sets [*Chen et al.*, 2004, 2006]. In this synthetic example, we mimic such a realistic situation by generating a large field with the horizontal and vertical variogram (Figure 7a), and including a local curved feature (around a blue rectangle in Figure 7a). We extracted a small domain (a blue rectangle in Figure 7a) and used it as a domain for synthetic geophysical inversion (Figure 7b).

Using the MCMC approach, we generated 80,000 samples of the reactive-facies field, the convergence of which was confirmed by Geweke's convergence diagnostic [*Geweke*, 1992]. The first 10,000 samples were

 © 2014. American Geophysical Union. All Rights Reserved.

Table 2. Average Misclassified Fraction in the Entire Domain and in the Crosshole Domain for the Prior Field ("Prior": Corresponding to Figure 8a), the Posterior Field ("Posterior": Figure 8b), and the Result Without Crosshole Data ("No Crosshole": Figure 10)

Cases	Prior	Posterior	No Crosshole
Entire domain	0.427	0.247	0.257
Crosshole domain	0.378	0.0953	0.226

discarded, since the data misfit of the surface seismic data (the square sum of the difference between the observed and simulated data) were reduced in the first several thousand samples. For each sampled field, we defined the number of misidentified pixels divided by the total number of pixels as the misclassified fraction. To quantify the performance of the estimation, we calculated the average of the misclassified fraction of the resulting fields sampled in MCMC (Table 2). The average misclassified fraction was calculated for the entire domain and for the crosshole data domain (the blue rectangle in Figure 7a).

Figure 8a shows the prior probability field of reactive facies conditioned only on the point measurements (CPT-based facies data) created by simple kriging. Although the prior field captures the true field at and near the CPT locations, the regions far from the CPT locations have a probability close to the overall mean (0.322). The average misclassified fraction is about 42.7% in the entire domain and 37.8% in the crosshole domain.

Figure 8b shows the posterior probability field of facies obtained using the new method and the multiscale data. This is the mean field of the all the sampled fields from MCMC. Compared to the true field in Figure 7a, the posterior probability field shows good agreement with the true field. The average misclassified fraction (Table 2) was reduced to about a half of the prior field (24.7%) in the entire domain, and to about a forth in the crosshole domain (9.53%). The posterior field in Figure 8b captures the curved structure despite the fact that the horizontal and vertical correlation structure was used thanks to the information from the geophysical data sets. Near the crosshole region especially, the interfaces are sharp and the small-scale heterogeneity is well defined. Even in the regions away from the crosshole data and point measurements, the surface seismic data provide fairly detailed information about facies distribution. The estimates obtained using the new method are greatly improved over the ones obtained using standard geostatistical approaches (Figure 8a) and are expected to be greatly improved over what one could qualitatively interpret from conventional seismic data displays (such as Figure 5).

Figure 9 shows the confidence interval of the facies estimation associated with the borehole locations indicated by the two green lines in Figure 8b. Figure 9a shows the location far from the point or crosshole data sets; Figure 9b shows the location within the crosshole domain. In both figures (Figures 9a and 9b), the true

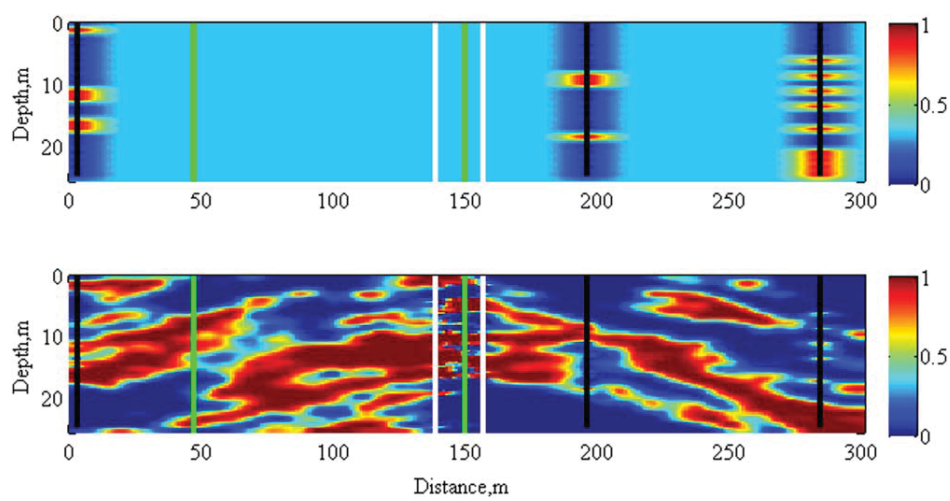

Figure 8. Synthetic study results: (a) prior probability field of Lagoonal facies and (b) posterior probability field of Lagoonal facies. The white lines represent the crosshole geophysical data wellbore locations, and the black lines are the CPT data locations. The green lines are wellbores where confidence intervals are calculated (Figure 9). A probability of 0 indicates the Barrier Beach facies, while a probability of 1 indicates the Lagoonal facies.

© 2014. American Geophysical Union. All Rights Reserved.

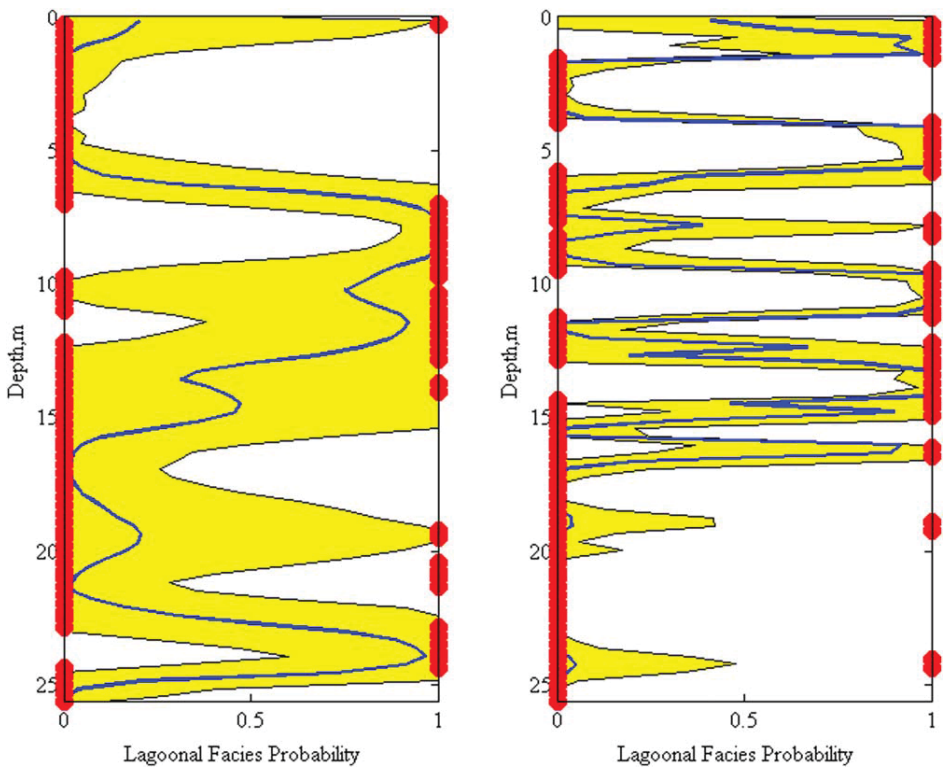

Figure 9. Synthetic study results: Estimated mean and confidence interval obtained using the estimation methodology compared with the true field at Distance (a) 47 m and (b) 150 m in Figure 8b. The red dots represent the true facies, the blue line is the mean probability, and the yellow region represents the 95% confidence interval based on the standard deviation (STD) multiplied by two. A probability of 0 indicates the Barrier Beach facies, while a probability of 1 indicates the Lagoonal facies.

facies is mostly contained by the confidence intervals, and the mean curves follow the true field. The confidence interval is narrower in the crosshole location in general, which is consistent with the sharp interfaces observed in Figure 8b. Figure 9b indicates that there are some locations that miss the true facies even in the crosshole region (depth approximately 19 and 24 m), which is caused by the presence of the small Lagoonal Facies (see Figure 8b) near the edge of the domain.

We examined a scenario without the crosshole data to assess the value of crosshole data to the estimation result. Figure 10 shows that when crosshole data are not available, the region close to the crosshole data becomes smoother, but that the lack of crosshole data has little impact on the estimation in other regions. Table 2 indicates that, compared to the full data set, the average misclassified fraction does not increase significantly over the entire domain (from 24.7% to 25.7%), although it is more than double in the crosshole domain (from 9.53% to 22.6%). Based on this result, one may conclude that the crosshole data have only a local impact on the large-scale subsurface characterization using this framework. However, we note that the crosshole data were essential in this framework for determining the distribution of seismic velocity in each facies [*Sassen et al.*, 2012]. The crosshole data can also potentially be used to identify the spatial correlation [*Hubbard et al.*, 1999], although smearing and inversion artifacts may hinder such estimation [*Day-Lewis and Lane*, 2004].

5.2. Site Application

After testing the developed methodology through a synthetic study, we applied it to the actual data sets from the F-Area. Since the true reactive-facies field is unknown, we evaluated the performance of the inversion by predicting the facies profile at one CPT location that was not used in the estimation, and then comparing the estimation results with the CPT-based facies data at that location.

Figure 11 shows the prior and posterior probability field of the Lagoonal facies obtained using the estimation methodology developed here. Figure 11a shows that the prior field has nearly horizontal and smooth

 © 2014. American Geophysical Union. All Rights Reserved.

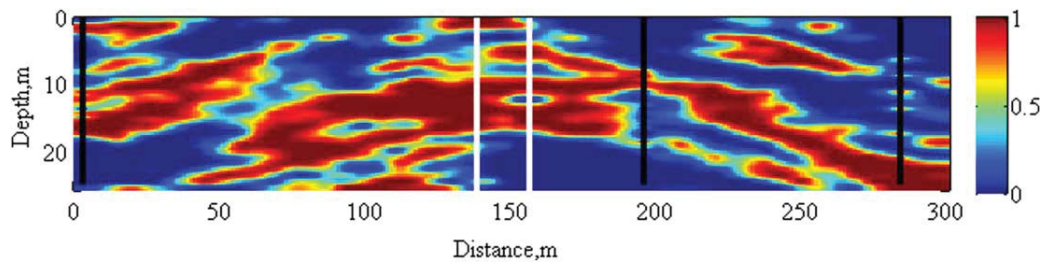

Figure 10. Posterior reactive-facies probability field obtained without using the crosshole data set. The white rectangle represents the crosshole geophysical data location, and the black lines are the CPT data locations. A probability of 0 indicates the Barrier Beach facies, while a probability of 1 indicates the Lagoonal facies.

layers of the Lagoonal facies in the Tan Clay Zone, although the multilayered Lagoonal facies in the Tan Clay Zone was captured locally at the CPT locations. In Figure 11b, the posterior facies probability field shows the large-scale curved structure of the Lagoonal facies. It is nearly continuous in this domain, although some bifurcations and isolated lenses exist. There are several discontinuous lenses of the Lagoonal facies in the lower aquifer, which is consistent with the findings of *Sassen et al*. [2012] and borehole data sets at the sites. Although the low-permeability Lagoonal facies in the Tan Clay zone is continuous in our domain, changing the depth and thickness of the Lagoonal facies in the Tan Clay zone would have a significant impact on the plume mobility and distribution, since it controls the groundwater flow rate into the lower aquifer. The discontinuous lenses of the Lagoonal facies would affect sorption as well as other geochemical reactions.

The continuity and curved structure of the Lagoonal facies in the Tan Clay zone (TCZ) found in Figure 11b might raise a question of the prior information such that we could have included stronger prior information imposing the continuous and curved layer of the Lagoonal facies in the TCZ based on the surface seismic data set (Figure 5). The surface seismic data, however, contains uncertainty, and there exist multiple layers of Lagoonal facies at some locations in TCZ. In addition, as is mentioned in section 3.1, it is known that the Lagoonal facies in the TCZ is not perfectly continuous over the site, and there exist discontinuous features of Lagoonal facies. Having a stronger prior would have caused a strong bias in our estimation.

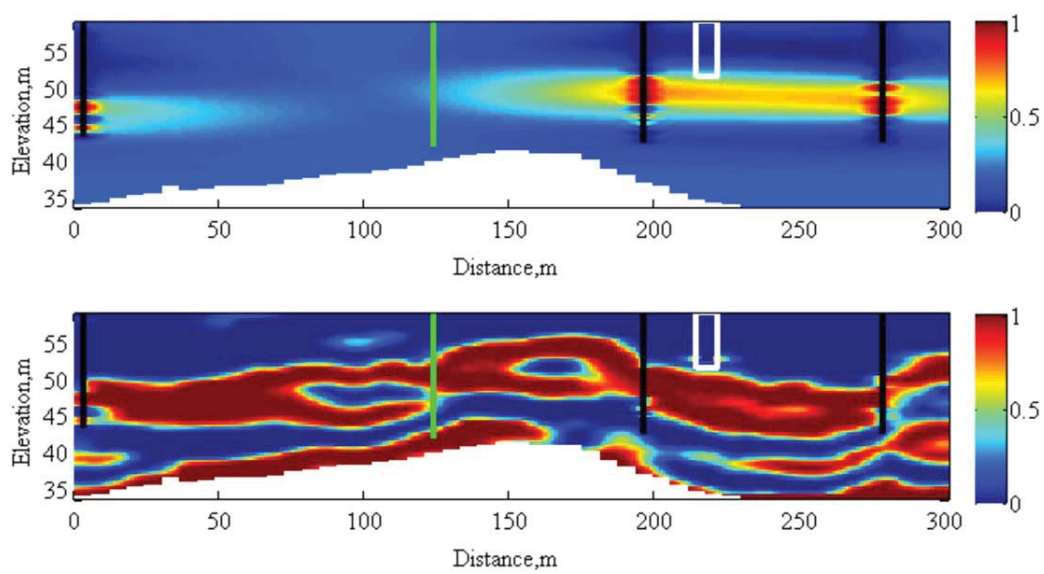

Figure 11. F-Area estimation results: (a) prior reactive-facies probability field and (b) posterior reactive-facies probability field. The black vertical lines represent the CPT data locations, the green line represents the CPT data used for validation, and the white box represents the crosshole data location. The underlying formation (lower white region) was excluded in the estimation. A probability of 0 indicates the Barrier Beach facies, while a probability of 1 indicates the Lagoonal facies.

© 2014. American Geophysical Union. All Rights Reserved.

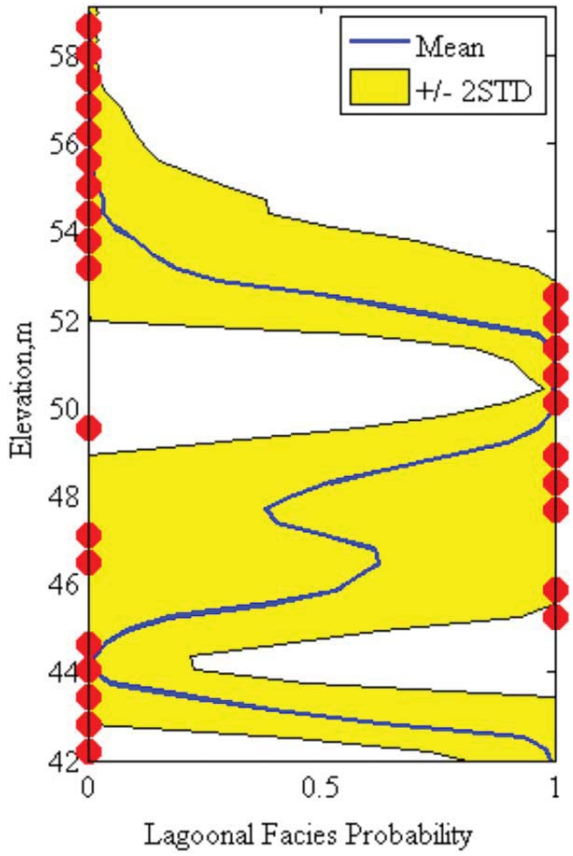

Figure 12. Estimated mean and confidence interval obtained using the estimation methodology compared with the interpretations from the CPT data not used in the estimation. The red dots represent CPT-based facies, the blue line is the mean probability, and the yellow region represents the 95% confidence interval based on the standard deviation (STD) multiplied by two. A probability of 0 indicates the Barrier Beach facies, while a probability of 1 indicates the Lagoonal facies.

In the Bayesian method to invert geophysical data [e.g., *Chen et al*., 2004, 2006; *Bosch et al*., 2007], the geostatistical (or variogram) model serves as prior information before conditioning on geophysical data sets. The geostatistical model is developed based on the data sets not only in the geophysical-data domain but also in the surrounding larger domain, which provides general information on heterogeneity at the site. The geophysical data provide the domain-specific conditional information such as the location and extent of particular geological units. The Bayesian method combines both geostatistical and geophysical information to identify the location and the extent of particular geological units with uncertainty quantified. Our method successfully identified both the continuous layer and curved feature without explicitly assuming them in the prior information, which suggests that our method can successfully honor the information of the surface seismic data.

For validation, Figure 12 shows a comparison of the estimated facies mean value and confidence interval, with facies interpretation from the CPT that was not used in the estimation procedure. The figure shows that the true facies profile at the validation points are close to the mean field curves and mostly within the 95% confidence interval. The comparison suggests that our method has estimated the facies variability in the vertical direction successfully, although it was not successful at estimating thin zones that were below the resolution of the seismic data and thus smaller than the estimation discretization (i.e., the thin Barrier Beach facies layer embedded within the Lagoonal facies at elevation around 50 m).

As a final step in our estimation, we generated reactive transport parameters based on the estimated reactive-facies field and petrophysical relationships developed by *Sassen et al*. [2012] Within each reactive facies, three properties were randomly generated without considering spatial correlation (i.e., each property at each pixel was sampled independently from the distribution shown in *Sassen et al*. [2012, Figure 8]), depending on the reactive facies present at the pixel. We did not consider the spatial correlation and cross-correlation among the three properties in this sampling process, since we assume that the heterogeneity and cross-correlation is captured by the reactive facies.

Figure 13 shows the fields of percent of fine, hydraulic conductivity, and Al:Fe ratio, based on one realization of the reactive facies field. The figure highlights the coupled physiochemical nature of the reactive facies: the Lagoonal facies always has a higher percent of fine, lower hydraulic conductivity, and higher Al:Fe ratio. The figure also still clearly identifies a two-facies structure, which suggests that the transport-property heterogeneity represented by the reactive facies could potentially be sufficient information for informing reactive transport simulators [e.g., *Sassen et al*., 2012; *Bea et al*., 2013]. It is worth emphasizing that in this 2-D domain, there are no direct measurements of the reactive transport properties; only CPT and geophysical data sets along with petrophysical relationships were used.

© 2014. American Geophysical Union. All Rights Reserved.

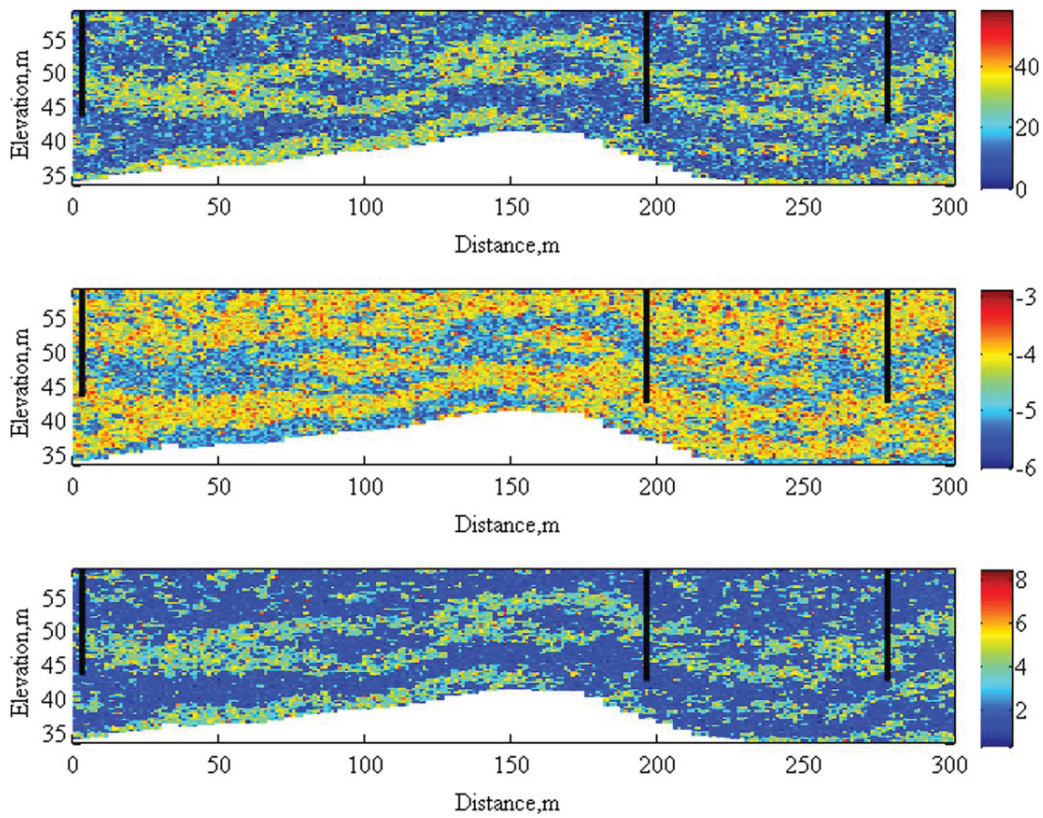

Figure 13. Reactive facies-based transport properties estimated along the plume centerline (refer to Figure 1b for location): (a) percent of fines, (b) hydraulic conductivity (log10) in cm/s, and (c) Al:Fe ratio. The black lines are the point measurement locations at the CPT locations.

6. Summary and Conclusion

In this study, we have developed a multiscale Bayesian method for integrating multiple geophysical data sets and wellbore/point data sets over large spatial extents to estimate the spatial distribution of reactive facies and associated flow and reactive transport properties. Our approach is based on the previous local-scale study by *Sassen et al*. [2012], which identified two reactive facies at the Savannah River F-Area that have unique distributions of reactive transport properties and geophysical attributes. Through synthetic and real data studies associated with the same site, we found that the hierarchical Bayesian model can effectively integrate multiple types and scales of data, providing an approach for distributing heterogeneous properties that control flow and transport over the scales needed for simulating plume migration and guiding the development of remediation strategies. Importantly, our approach permits the use of surface seismic-reflection data, which provides large spatial coverage in a noninvasive manner, but is often difficult to interpret in terms of transport parameters.

The developed approach was tested using both synthetic and real data sets. Synthetic studies demonstrated that the estimation method worked well, providing high-resolution estimates of reactive facies in the vicinity of point-scale and crosshole data, and providing lower resolution but acceptable estimates in the region where only surface seismic data were available. Application of the new methodology to real site data (including CPT, crosshole seismic, and surface seismic data) highlighted the use of this approach for characterizing reactive facies and associated transport parameters along a plume centerline. The surface seismic data were useful for estimating the continuity and multiple layers of Lagoonal facies in the Tan Clay Zone, as well as discontinuous lenses in the upper and lower aquifer. Although the crosshole seismic data set had a small spatial coverage in this site, such data were useful in identifying the differences in seismic velocity between the two facies. The newly developed approach yielded multiple random fields of reactive

© 2014. American Geophysical Union. All Rights Reserved.

transport parameters along a 2-D, 300 m traverse and with a pixel resolution of 0.305 m (vertical) by 1.2 m (horizontal). These estimates can be directly used to parameterize stochastic simulations of reactive transport models, which can in turn be used to simulate long-term plume behavior and to develop science-based remediation strategies.

Although demonstrated here along a 2-D transect, the developed framework can be extended to the 3-D domain to provide input to 3-D reactive transport models. The computational cost associated with the increased number of pixels will be a challenge, but parallel computing can be used. In this framework, sampling of each field (except reactive facies) can be parallelized, since each pixel is independent of all other pixels. Another challenge of extending the method to 3-D is the availability of data appropriate for a 3-D estimation, including geophysical data collected using 3-D or pseudo-3-D geometries as well as conventional (borehole) data sets having a reasonably dense and azimuthal distribution [e.g., *Hyndman and Gorelick*, 1996; *Doetsch et al*., 2010; *Dafflon et al*., 2011; *Dafflon and Barrash*, 2012]. Although our study was conducted using 2-D geophysical data sets, it would be possible to generate 3-D reactive facies field using conditional realizations by regarding all the pixels in the vicinity of the transects as point measurements.

Although facies-based approaches have been used in the past, the multiscale-estimation reactive-facies approach developed here is a new formal methodology that quantifies the uncertainty of reactive-facies properties, as well as the uncertainty in the spatial distribution of reactive facies, using multiscale data sets. Also in contrast to previous methods, our approach includes both the joint inversion of geophysical data sets and random field generation. This study further suggests that the reactive facies concept and estimation approach used here are effective for transferring information on reactive transport parameters from direct measurements to indirect measurements (including geophysical data, lithology data and CPT), and for maximizing the value of data sets for parameterizing reactive transport models at large scales.

Appendix A: Sampling of Coarse-Resolution Seismic Slowness $\{u_j\}$ in the Site-Specific Model

According to equation (12), sampling of $\{u_j\}$ requires the surface seismic forward model $p(\boldsymbol{z}_s|\{u_j\}\ \boldsymbol{\phi}_s)$, and upscaling model $p(\{u_j\}|\{v_i\}, \boldsymbol{\theta}_u)$. The forward model is a convolution model specified in section 4. Since upscaling involves simple averaging of the child elements such that $u_j = n(j)^{-1}\sum_{k\in\alpha(j)} v_k + \varepsilon_u$, the distribution $p(u_j|\{v_{\alpha(j)}\}, \boldsymbol{\theta}_u)$ is a Gaussian distribution with the mean $n(j)^{-1}\sum_{k\in\alpha(j)} v_k$ and variance τ_u. According to equation (12), we sample each element by:

$$p\left(u_j|\bullet\right)\alpha p(z_s|u_j, \boldsymbol{\phi}_s)N(n(j)^{-1}\sum_{k\in\alpha(j)} v_k, \tau_u). \tag{A1}$$

We use the MH sampler with the proposal distribution $N(n(j)^{-1}\sum_{k\in\alpha(j)} v_k, \tau_u)$, and the decision criteria based on the likelihood $p(\boldsymbol{z}_s|u_j, \boldsymbol{\phi}_s)$.

Appendix B: Sampling of Fine-Resolution Seismic Slowness $\{v_i\}$ in the Site-Specific Model

We assume that the first arrival time in the crosshole data is the sum of seismic slowness along the straight ray path, multiplied by the ray path length in each pixel, following *Chen et al*. [2006]. For Pixel i, the first arrival time passing through Pixel i ($z_{t,i}$) has a Gaussian distribution with the mean $\sum_{k\in r(i)} d_k v_k$ and variance σ_t, where $r(i)$ represents the pixels along the ray path passing through Pixel i, d_k is the ray path length in Pixel i, and σ_t is the measurement error variance. At the crosshole data location, we sample the seismic slowness at each pixel based on equation (13):

$$p(v_i|\bullet)\alpha N(\sum_{k\in r(i)} d_k v_k, \sigma_t)N(n(\gamma(i))^{-1}\sum_{k\in\alpha(\gamma(i))} v_k, \tau_u)N(a_1 + a_2 f_i, \tau_v). \tag{B1}$$

Since the three distributions on the right-hand side of equation (18) are all Gaussian, we can use a conjugate prior such that the posterior distribution $p(v_i|\ \bullet)$ is also a Gaussian distribution with mean A_1 and variance B_1, where:

© 2014. American Geophysical Union. All Rights Reserved.

$$A_1 = B_1^{-1}\left\{\frac{z_{t,i} - \sum_{k\in r(i),k\neq i} d_k v_k}{\sigma_t} + \frac{n(\gamma(i))u_{\gamma(i)} - \sum_{k\in\alpha(\gamma(i)),k\neq i} v_k}{n(\gamma(i))^2\tau_u} + \frac{a_1 + a_2 f_i}{\tau_v}\right\} B_1 = \left\{\frac{1}{\sigma_t} + \frac{1}{n(\gamma(i))^2\tau_u} + \frac{1}{\tau_v}\right\}^{-1} \tag{B2}$$

At the noncrosshole locations, we sample each element based on equation (14):

$$p(v_i|\bullet) \propto N\left(n(\gamma(i))^{-1}\sum_{k\in\alpha(\gamma(i))} v_k, \tau_u\right) N(a_1 + a_2 f_i, \tau_v). \tag{B3}$$

In the same manner as the crosshole locations, the posterior distribution $p(v_i|\bullet)$ is also a Gaussian distribution with mean A_2 and variance B_2, where:

$$A_2 = B_2^{-1}\left\{\frac{n(\gamma(i))u_{\gamma(i)} - \sum_{k\in\alpha(\gamma(i)),k\neq i} v_k}{n(\gamma(i))^2\tau_u} + \frac{a_1 + a_2 f_i}{\tau_v}\right\}, B_2 = \left\{\frac{1}{n(\gamma(i))^2\tau_u} + \frac{1}{\tau_v}\right\}^{-1}. \tag{B4}$$

Acknowledgments
Our study was supported as part of the Subsurface Science Scientific Focus Area funded by the U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research under Award DE-AC02–05CH11231 to the LBNL Sustainable Systems SFA. The surface seismic data acquisition was supported by the DOE EM-32 for field support at the F-Area Applied Field Study Site. We thank Miles Denham (SRNL) for providing site access as well site data sets information, Deb Agarwal and the ASCEM data management team for providing and managing the data sets, John E. Peterson (LBNL) for collecting and processing field geophysical data, Rick Miller (KGS) for collecting and processing the surface seismic data, and Maggie Millings (SRNL) and others for assisting with the field campaigns. We also appreciate constructive comments from the Associate Editor Andrew Binley as well as Chris Murray (PNNL) and two other anonymous reviewers. Data sets will be available upon request (contact the corresponding author). The permission will be granted after the review by the Savannah River National Laboratory.

References

Aki, K., and P. Richards (1980), *Quantitative Seismic Theory and Methods*, W. H. Freeman, San Francisco, Calif.
Andreu, C., N. de Freitas, A. Doucet, and M. I. Jordan, (2003), An introduction to MCMC for machine learning, *Mach. Learning.*, *50*, 5–43.
Bea, S. A., H. Wainwright, N. Spycher, B. Faybishenko, S. Hubbard, and M. E. Denham (2013), Identifying key controls on acidic-U(VI) plume behavior at the Savannah River Site, using reactive transport modeling, *J. Contam. Hydrol.*, *151*, 34–54.
Bosch, M., L. Cara, J. Rodrigues, A. Navarro, and M. Diaz (2007), A Monte Carlo approach to the joint estimation of reservoir and elastic parameters from seismic amplitudes, *Geophysics*, *72*, O29–O39, doi:10.1190/1.2783766.
Chen, J., and M. Hoversten (2012), Joint inversion of marine seismic AVA and CSEM data using statistical rock-physics models and Markov random fields, *Geophysics*, *77*, R65–R80, doi:10.1190/GEO2011-0219.1.
Chen, J., S. Hubbard, Y. Rubin, C. Murray, E. Roden, and E. Majer (2004), Geochemical characterization using geophysical data and Markov chain Monte Carlo methods: A case study at the South Oyster Bacterial Transport Site in Virginia, *Water Resour. Res.*, *40*, W12412, doi:10.1029/2003WR002883.
Chen, J., S. Hubbard, J. Peterson, K. Williams, M. Fienen, P. Jardine, and D. Watson (2006), Development of a joint hydrogeophysical inversion approach and application to a contaminated fractured aquifer, *Water Resour. Res.*, *42*, W06425, doi:10.1029/2005WR004694.
Dafflon, B., and W. Barrash (2012), Three-dimensional stochastic estimation of porosity distribution: Benefits of using ground-penetrating radar velocity tomograms in simulated-annealing-based or Bayesian sequential simulation approaches, *Water Resour. Res.*, *48*, W05553, doi:10.1029/2011WR010916.
Dafflon, B., J. Irving, and W. Barrash (2011), Inversion of multiple intersecting high-resolution crosshole GPR profiles for hydrological characterization at the Boise Hydrogeophysical Research Site, *J. Appl. Geophys.*, *73*, 305–314.
Day-Lewis, F. D., and J. W. Lane Jr. (2004), Assessing the resolution-dependent utility of tomograms for geostatistics, *Geophys. Res. Lett.*, *31*, L07503, doi:10.1029/2004GL019617.
Denham, M., and K. M. Vangelas (2008), Biogeochemical gradients as a framework for understanding waste-site evolution, *Remediation*, *19*, 5–17, doi:10.1002/rem.20188.
Doetsch, J., N. Linde, I. Coscia, S. A. Greenhalgh, and A. G. Green (2010), Zonation for 3D aquifer characterization based on joint inversions of multimethod crosshole geophysical data, *Geophysics*, *75*(6), G53–G64.
Dong, W., T. K. Tokunaga, J. A. Davis, and J. Wan (2012), Uranium (VI) adsorption and surface complexation modeling on background sediments from the F-Area, Savannah River Site, *Environ. Sci. Technol.*, *46*, 1565–1571.
Ferreira, M. A., and H. K. Lee (2007), *Multiscale Modeling: A Bayesian Perspective*, Springer, N. Y.
Flach, G. (2004), Groundwater Flow Model of the General Separations Area Using Porflow (U), WSRC-TR-2004-00106, Westinghouse Savannah River Co. LLC, Savannah River Site, Aiken, S. C.
Fogg, G. E., C. D. Noyes, and S. F. Carle (1998), Geologically based model of heterogeneous hydraulic conductivity in an alluvial setting, *Hydrogeol. J.*, *6*, 131–143.
Gamerman, D., and H. F. Lopes (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed., Chapman and Hall, Boca Raton, Fla.
Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to calculating posterior moments, in *Bayesian Statistics*, vol. 4, edited by J. M. Bernado et al., Clarendon, Oxford, U. K.
Gotway, C. A., and L. J. Young (2002), Combining incompatible spatial data, *J. Am. Stat. Assoc.*, *97*(458), 632–648.
Hammond, G. E., P. C. Lichtner, and M. L. Rockhold (2011), Stochastic simulation of uranium migration at the Hanford 300 Area, *J. Contam. Hydrol.*, *120–121*, 115–128, doi:10.1016/j.jconhyd.2010.04.005.
Hastie, T., R. Tibshirani, and J. H. Friedman (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, N. Y.
Heinz, J., S. Kleineidam, G. Teutsch, and T. Aigner (2003), Heterogeneity patterns of Quaternary glaciofluvial gravel bodies (SW-Germany): Application to hydrogeology, *Sediment. Geol.*, *158*(1–2), 1–23.
Hubbard, S. S., and N. Linde (2011), Hydrogeophysics, in *Treatise on Water Science*, vol. 2, edited by P. Wilderer, Academic, Oxford, U. K.
Hubbard, S. S., and Y. Rubin (2005), Hydrogeophysics, in *Hydrogeophysics*, edited by Y. Rubin and S. Hubbard, chap. 1, Elsevier, Dordrecht, Netherlands.
Hubbard, S. S., Y. Rubin, and E. L. Majer (1999), Spatial correlation structure estimation using geophysical data, *Water Resour. Res.*, *35*(6), 1809–1825.
Hyndman, D. W., and S. M. Gorelick (1996), Estimating lithologic and transport properties in three dimensions using seismic and tracer data: The Kesterson aquifer, *Water Resour. Res.*, *32*(9), 2659–2670, doi:10.1029/96WR01269.

 © 2014. American Geophysical Union. All Rights Reserved.

Ines, A. V. M., and B. P. Mohanty (2009), Near-surface soil moisture assimilation for quantifying effective soil hydraulic properties using genetic algorithms: 2. Using airborne remote sensing during SGP97 and SMEX02, *Water Resour. Res.*, *45*, W01408, doi:10.1029/2008WR007022.

Jean, G. A., J. M. Yarus, G. P. Flach, M. R. Millings, M. K. Harris, R. L. Chambers, and F. H. Syms (2002), Geological 3D Modeling of Southeastern Tertiary Coastal-Plain Sediments, Savannah River Site, SC: An Applied Geostatistical Approach, WSRC-OS-2002-00052, Westinghouse Savannah River Co. LLC, Savannah River Site, Aiken, S. C.

Jean, G. A., J. M. Yarus, G. P. Flach, M. R. Millings, M. K. Harris, R. L. Chambers, and F. H. Syms (2004), Three-dimensional geologic model of southeastern Tertiary coastal-plain sediments, Savannah River Site, South Carolina: An applied geostatistical approach for environmental applications, *Environ. Geosci.*, *11*(4), 205–220.

Johnson, T., P. Routh, and M. Knoll (2005), Fresnel volume georadar attenuation-difference tomography, *Geophys. J. Int.*, *162*, 9–24.

Keys, W. S. (1989), *Borehole Geophysics Applied to Ground-Water Investigations*, p. 313, Natl. Water Well Assoc., Worthington, OH.

Klingbeil, R., S. Kleineidam, U. Asprion, T. Aigner, and G. Teutsch (1999), Relating lithofacies to hydrofacies: Outcrop-based hydrogeological characterization of Quaternary gravel deposits, *Sediment. Geol.*, *129*, 299–310.

Li, L., C. I. Steefel, M. B. Kowalsky, A. Englert, and S. Hubbard (2010), Effects of physical and geochemical heterogeneities on mineral transformation and biomass accumulation during biostimulation experiments at Rifle, Colorado, *J. Contam. Hydrol.*, *112*, 45–63.

Lunne, T., P. K. Robertson, and J. J. M. Powell (1997), *Cone Penetration Testing in Geotechnical Practice*, E&FN Spon, London.

Michalak, A. M., and P. K. Kitanidis (2003), A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification, *Water Resour. Res.*, *39*(2), 1033, doi:10.1029/2002WR001480.

Miller, R. (1992), Normal moveout stretch mute on shallow: Reflection data, *Geophysics*, *57*(11), 1502–1507, doi:10.1190/1.1443217.

Miller, R., D. Steeples, R. Hill Jr., and B. Gaddis (1990), Identifying intra-alluvial and bedrock structures shallower than 30 meters using seismic reflection techniques, *Geotech. Environ. Geophys.*, pp. *89–98*.

Miller, R., S. L. Peterie, B. A. Wedel, A. R. Wedel, O. M. Metheny, B. T. Rickards, J. T. Schwenk, B. C. Bennett, and M. C. Brohammer (2010), Surface seismic characterization at F Area Seepage Basin SRNL, internal report, Savannah River National Laboratory, Aiken, SC, USA.

Pebesma, E. J. (2004), Multivariable geostatistics in S: The GSTAT package, *Comput. Geosci.*, *30*, 683–691.

Peterson, J. E., B. N. P. Paulson, and T. V. McEvilly (1985), Applications of algebraic reconstruction to crosshole seismic data, *Geophysics*, *50*, 1556–1580.

Phifer, M. A., M. R. Milling, and G. P. Flach (2006), Hydraulic Property Data Package for the E-Area and Z-Area Soils, Cementitious Materials, and Waste Zone, WSRC-STI-2006-00198, Rev 0, Washington Savannah River Co., Aiken, S. C.

R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna.

Sassen, D. S., S. S. Hubbard, S. A. Bea, J. Chen, N. Spycher, and M. E. Denham (2012), Reactive facies: An approach for parameterizing field-scale reactive transport models using geophysical methods, *Water Resour. Res.*, *48*, W10526, doi:10.1029/2011WR011047.

Scheibe, T. D., and Y. J. Chien (2003), An evaluation of conditioning data for solute transport prediction, *Ground Water*, *41*(2), 128–141.

Scheibe, T. D., Y. Fang, C. J. Murray, E. E. Roden, J. Chen, Y. J. Chien, S. C. Brooks, and S. S. Hubbard (2006), Transport and biogeochemical reaction of metals in a physically and chemically heterogeneous aquifer, *Geosphere*, *2*(4), 220–235.

Steeples, D. W., and R. D. Miller (1990), Seismic reflection methods applied to engineering, environmental and groundwater problems, *Geotech. Environ. Geophys.*, *1*, 1–30.

Tiwary, D. K., I. O. Bayuk, A. A. Vikhorev, and E. M. Chesnokov (2009), Comparison of seismic upscaling methods: From sonic to seismic, *Geophysics*, *74*, WA3–WA14, doi:10.1190/1.3054144.

Vereecken, H., A. Binley, G. Cassiani, A. Revil, and K. Titov (2006), *Applied Hydrogeophysics*, Springer, N. Y.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, *41*, W01017, doi:10.1029/2004WR003059.

Vrugt, J. A., H. V. Gupta, S. Sorooshian, T. Wagener, and W. Bouten (2006), Application of stochastic parameter optimization to the Sacramento Soil Moisture Accounting model, *J. Hydrol.*, *325*(1–4), 288–307, doi:10.1016/j.hydrol.2005.10.041.

Wan, J., T. K. Tokunaga, W. Dong, M. E. Denham, and S. S. Hubbard (2012), Persistent source influences on the trailing edge of a groundwater plume, and natural attenuation timeframes: The F-Area Savannah River Site, *Environ. Sci. Technol.*, *46*, 4490−4497.

Weissmann, G. S., Y. Zhang, E. M. LaBolle, and G. E. Fogg (2002), Dispersion of groundwater age in an alluvial aquifer system, *Water Resour. Res.*, *38*(10), 1198, doi:10.1029/2001WR000907.

Wikle, C. K., R. F. Milliff, D. Nychka, and L. M. Berliner (2001), Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds, *J. Am. Stat. Assoc.*, *96*(454), 382–397.

Willsky, A. S. (2002), Multiresolution Markov models for signal and image processing, *Proc. IEEE*, *90*(8), 1396–1458.

Yabusaki, S. B., Y. Fang, K. H. Williams, C. J. Murray, A. L. Ward, R. Dayvault, S. R. Waichler, D. R. Newcomer, F. A. Spane, and P. E. Long (2011), Variably saturated flow and multicomponent biogeochemical reactive transport modeling of a uranium bioremediation field experiment, *J. Contam. Hydrol.*, *126*(3–4), 271–290, doi:10.1016/j.jconhyd.2011.09.002.

 © 2014. American Geophysical Union. All Rights Reserved.