A state-space Bayesian framework for estimating biogeochemical transformations using time-lapse geophysical data

Jinsong Chen,[1] Susan S. Hubbard,[1] Kenneth H. Williams,[1] Steve Pride,[1] Li Li,[1] Carl Steefel,[1] and Lee Slater[2]

[1] We develop a state-space Bayesian framework to combine time-lapse geophysical data with other types of information for quantitative estimation of biogeochemical parameters during bioremediation. We consider characteristics of end products of biogeochemical transformations as state vectors, which evolve under constraints of local environments through evolution equations, and consider time-lapse geophysical data as available observations, which could be linked to the state vectors through petrophysical models. We estimate the state vectors and their associated unknown parameters over time using Markov chain Monte Carlo sampling methods. To demonstrate the use of the state-space approach, we apply it to complex resistivity data collected during laboratory column biostimulation experiments that were poised to precipitate iron and zinc sulfides during sulfate reduction. We develop a petrophysical model based on sphere-shaped cells to link the sulfide precipitate properties to the time-lapse geophysical attributes and estimate volume fraction of the sulfide precipitates, fraction of the dispersed, sulfide-encrusted cells, mean radius of the aggregated clusters, and permeability over the course of the experiments. Results of the case study suggest that the developed state-space approach permits the use of geophysical data sets for providing quantitative estimates of end-product characteristics and hydrological feedbacks associated with biogeochemical transformations. Although tested here on laboratory column experiment data sets, the developed framework provides the foundation needed for quantitative field-scale estimation of biogeochemical parameters over space and time using direct, but often sparse wellbore data with indirect, but more spatially extensive geophysical data sets.

1. Introduction

[2] In-situ contaminant remediation treatments are being used to facilitate reactions that degrade or immobilize contaminants in the subsurface, rendering them less hazardous to human and ecological health [e.g., *Hazen and Tabak*, 2005]. These remediation treatments induce various biogeochemical reactions, such as the dissolution and precipitation of minerals, gas evolution, changes in total dissolved solids, and biofilm generation. Direct aqueous geochemical measurements obtained using wellbore groundwater samples are typically used to assess the efficacy of the remedial treatments [e.g., *Lovley et al.*, 1994; *Chapelle*, 2001]. However, given the spatially variable distribution of remediation treatments introduced into the subsurface and the complexity of the subsequent biogeochemical reactions [*Scheibe et al.*, 2006], it is often difficult to assess the efficacy of remediation treatments over time and space with reasonable confidence using wellbore measurements alone [*Hubbard et al.*, 2008]. In addition, it is challenging to directly measure the evolution of solid phase transformations (such as the generation of precipitates) using conventional wellbore-based sampling approaches.

[3] Time-lapse geophysical methods hold potential for providing information about remediation-induced biogeochemical changes in a cost-effective and minimally invasive manner because they are often sensitive to changes in pore fluid and matrix properties associated with the induced biogeochemical transformations. Several biogeophysical studies have been performed in recent years to test this hypothesis [*Atekwana et al.*, 2006]. For example, *Williams et al.* [2005] performed a laboratory-scale biostimulation experiment where time-lapse complex resistivity, seismic, and various geochemical measurements were measured over the length of the experimental columns during the experiments. They showed that changes in complex resistivity and seismic amplitude measurements corresponded to the onset and spatial distribution of microbial-mediated iron and zinc sulfide precipitation. High-frequency seismic wave amplitudes were reduced by nearly 84%; within the context of a double porosity model [*Pride et al.*, 2004], the attenuation was interpreted to be caused by the wave-induced flow

---

[1]Earth Science Division, Lawrence Berkeley Laboratory, Berkeley, California, USA.
[2]Department of Earth and Environmental Sciences, Rutgers University, Newark, New Jersey, USA.
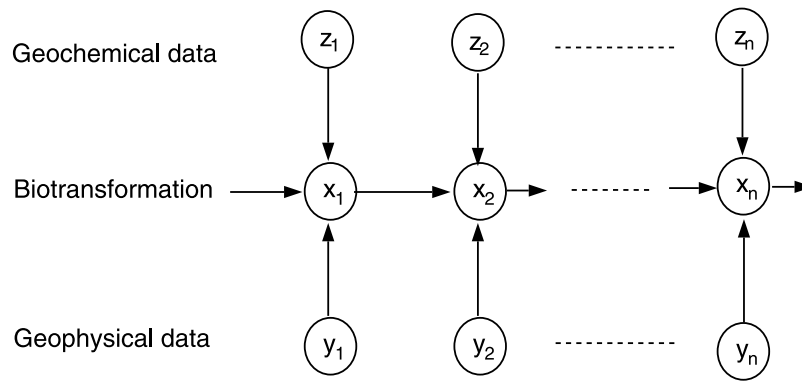
Figure 1. Schematic map of the state-space Bayesian estimation framework.

resulting from the heterogeneous formation of high bulk modulus sulfide precipitates within formerly fluid-filled pore spaces. The phase response of the complex resistivity data also tracked the spatiotemporal development of the precipitates. In the frequency range used to collect the complex resistivity measurements $(0.1-1000 \text{ Hz})$, the energy storage reflected by the phase response results primarily from the polarization of the ions in the electrical double layer at the mineral-fluid interface and from the formation of electrically conductive pathways accompanying the precipitation of (semi) conductive minerals. As such, changes in the complex resistivity response were attributed to alterations in subsurface mineralogy arising from stimulated microbial activity within the pore space, including precipitation reactions, aggregation dynamics, and solid-state mineral transformations.

[4] More recent studies have shown that time-lapse geophysical methods can be useful for tracking remediation processes at the field scale. *Lane et al.* [2006] used time-lapse, crosshole, zero-offset radar data and electrical logs to indicate subsurface regions impacted by injection of emulsified vegetable oil during a biostimulation experiment. *Hubbard et al.* [2008] explored the use of geophysical data sets for monitoring the distribution of electron donor and subsequent transformations associated with a Cr(VI) bioremediation treatment. Using the constraints provided by laboratory biogeochemical experiments and field geochemical data sets, *Hubbard et al.* [2008] interpreted field-scale, time-lapse seismic and radar tomographic data sets in terms of hydrological and biogeochemical transformations associated with the remedial treatment over approximately 3-year monitoring period, including the spatial distribution of injected electron donor, gas bubble formation, variations in total dissolved solids, and the formation of precipitates. The integrated interpretation revealed how geophysical techniques can provide information about coupled hydrobiogeochemical responses to remedial treatments.

[5] Although both laboratory- and field-scale studies have illustrated the potential of geophysical methods for providing information about biogeochemical end products, the use of geophysical data for this objective has to date been primarily qualitative in nature. In this study, we develop a state-space Bayesian estimation framework that permits rigorous integration of multiple types of time-lapse data sets (e.g., geophysical and geochemical) for quantitative estimation of biogeochemical end products. The developed method is subsequently applied to the laboratory biostimulation data

sets of *Williams et al.* [2005] to demonstrate the utility of time-lapse complex resistivity data for remotely estimating the evolution of volume fraction of metal sulfides and their associated parameters. Although we test the developed estimation framework by applying it to measurements collected over time at a single location within an experimental column, the methodology can be extended to larger, multidimensional data sets and regions.

[6] The remainder of this paper is organized as follows. Section 2 describes the state-space Bayesian framework for estimation of biogeochemical transformations and methods for obtaining solutions from the Bayesian model. In section 3, we apply the developed method to laboratory column experiments. The estimation results are given in section 4 and discussion and conclusions are provided in section 5.

2. State-Space Bayesian Framework

[7] In this section, we describe a general state-space approach for estimation of end products associated with biogeochemical transformations using time-lapse geophysical data and other types of information, such as direct and indirect measurements of geochemical or biogeochemical parameters.

2.1. Dynamic System

[8] We consider a typical bioremediation system as a dynamic system in which numerous geochemical reactions and biogeochemical processes may take place that are controlled or affected by the local environment. As shown in Figure 1, the dynamic system is described by a state vector $\mathbf{x}_i$, which consists of some characteristics of biogeochemical transformations at time t_i. This state vector can include properties that are helpful for ascertaining the system response to the remedial treatments, such as concentrations of electron donors or acceptors, or solid phase transformations such as the volume fraction of precipitates resulting from microbial activity. The biogeochemical state vector changes over time as the system evolves in response to the remediation; the change can be described by the following evolution equation.

$$\mathbf{x}_i = F(\mathbf{x}_{i-1}, \boldsymbol{\theta}_1) + \mathbf{w}_i, \tag{1}$$

where F represents a biogeochemical process forward model as a function of previous states, available geochemical measurements at time t_i, and an unknown time-invariant

parameter vector $\boldsymbol{\theta}_1$. Vector $\mathbf{w}_i$ represents random errors associated with the forward modeling.

[9] We can numerically obtain a series of state vectors $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$, by using equation (1) and starting from an initial state of the system $\mathbf{x}_0$, such as the initial biogeochemical conditions prior to bioremediation. Those state vectors form a Markov chain because the state vector $\mathbf{x}_i$ is conditionally independent of the state vector $\mathbf{x}_{i-2}$, given the state vector $\mathbf{x}_{i-1}$. Let f_0 and f_w be the probability distribution functions of the initial state $\mathbf{x}_0$ and the error vector $\mathbf{w}_i$, respectively. We can obtain the joint distribution of the Markov chain as follows [*Shumway and Stoffer*, 2000].

$$f(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n | \boldsymbol{\theta}_1) = f_0(\mathbf{x}_0) \prod_{i=1}^{n} f_w(\mathbf{x}_i - F(\mathbf{x}_{i-1}, \boldsymbol{\theta}_1)). \quad (2)$$

2.2. Time-Lapse Geophysical Data and Petrophysical Models

[10] Perhaps one of the most powerful aspects of environmental geophysics is the use of geophysical data for monitoring dynamic processes. Observing the data in a time-lapse mode (i.e., measurements collected at an earlier time subtracted from those collected at a later time) enhances the imaging of subtle changes in geophysical attributes caused by system perturbations and reduces the correlated errors and the dependence of geophysical measurements on the static geological heterogeneities [*Day-Lewis et al.*, 2002; *Vasco et al.*, 2004].

[11] Here, rather than differentiating the time-lapse data sets, we incorporate the geophysical data $\mathbf{y}_i$ (such as complex resistivity and seismic measurements) collected at each time t_i within the estimation framework (Figure 1). If we let: G be the petrophysical model that relates geophysical data observed at time t_i to the biogeochemical state vector at the same time; $\boldsymbol{\theta}_2$ be the unknown time-invariant parameter vector associated with the petrophysical model; $\mathbf{v}_i$ be the random error vector in the petrophysical model, we can obtain

$$\mathbf{y}_i = G(\mathbf{x}_i, \boldsymbol{\theta}_2) + \mathbf{v}_i. \quad (3)$$

Let f_v be a probability distribution function of the error vector $\mathbf{v}_i$. It is common to assume that errors in geophysical data collected at different times are independent of each other. We thus obtain the following likelihood function that relates the geophysical measurements to the biogeochemical parameters that we desire to estimate.

$$f(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n | \mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \prod_{i=1}^{n} f_v(\mathbf{y}_i - G(\mathbf{x}_i, \boldsymbol{\theta}_2)). \quad (4)$$

2.3. Bayesian Estimation Framework

[12] Our goal is to quantitatively estimate the end-product evolution associated with remediation-induced biogeochemical transformations using direct borehole geochemical and indirect geophysical data sets. As shown in Figure 1, we specifically strive to estimate state vectors $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$, and time-invariant parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, given geophysical data $\mathbf{y}_1$, $\mathbf{y}_2$, ..., $\mathbf{y}_n$. We formulate the above problem within the Bayesian framework. Using the Bayes' theorem and equations (2) and (4), we obtain the following joint posterior distribution function.

$$
\begin{aligned}
&f(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n) \\
&\propto f_0(\mathbf{x}_0) f_1(\boldsymbol{\theta}_1) f_2(\boldsymbol{\theta}_2) \prod_{i=1}^{n} f_w(\mathbf{x}_i - F(\mathbf{x}_{i-1}, \boldsymbol{\theta}_1)) f_v(\mathbf{y}_i - G(\mathbf{x}_i, \boldsymbol{\theta}_2)),
\end{aligned}
$$
$$(5)$$

where symbol "$\propto$" represents "is proportional to", which negates the use of a normalizing constant that does not affect the solution to equation (5), and f_1 and f_2 represent probability distributions of parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. As will be discussed in section 3.4, we can parameterize the general Bayesian formulation given in equation (5) for a specific biogeophysical estimation problem through specifying: the prior probability distributions (f_0, f_1, and f_2) of the initial state vector $\mathbf{x}_0$, the time-invariant parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, the error probability distributions (f_w and f_v), and the forward biogeochemical and petrophysical models (F and G).

2.4. Markov Chain Monte Carlo Sampling Methods

[13] The key to estimating the evolution of biogeochemical parameters using geophysical data sets and the Bayesian model defined in equation (5) is to obtain state vectors and unknown time-invariant parameters. Since the forward and petrophysical models F and G are often nonlinear, it is very challenging to analytically solve the inverse problem. Instead, we use Markov chain Monte Carlo (MCMC) sampling methods to draw many samples from the posterior joint probability distribution function following the procedure outlined by *Chen et al.* [2006]. With this approach, we obtain many samples of the biogeochemical parameters of interest, from which we can calculate statistics such as the medians, mean values, and variances of those parameters.

3. Application to Laboratory Column Experiments

[14] In this section, we formulate the developed state-space estimation framework (equation (5)) to specifically estimate parameters in connection with FeS and ZnS precipitates formed as a result of stimulated microbial activity using time-lapse complex resistivity measurements. Although our goal in this section is to demonstrate the use of the framework for solving a specific biogeochemical transformation estimation problem and to provide some insights into the utility of time-lapse complex resistivity data in the estimation of precipitation processes, we emphasize that the framework developed here is general in nature and could be applied to a variety of estimation problems.

[15] To test this framework, we use the data sets collected during the biostimulation column experiments described by *Williams et al.* [2005] and *Ntarlagiannis et al.* [2005]. We first introduce the column experiment setup and the collected time-lapse geochemical and geophysical data. We then develop a petrophysical model to link the geochemical and geophysical data sets. We finally use this information to parameterize the general Bayesian framework (equation (5)) for this specific estimation problem.

3.1. Laboratory Column Experiments

[16] The column experiments of *Williams et al.* [2005] were designed to examine the geophysical response to microbe-induced ZnS and FeS precipitation during a bio-stimulation experiment performed using sulfate-reducing bacteria. The experimental columns were instrumented along their length with geophysical sensors, as well as with biogeochemical fluid sampling ports. The experiments were conducted under temperature-controlled conditions over a period of 78 days using five polycarbonate columns having inner diameters of 5.08 cm and lengths of 30.5 cm. Although different columns were used to collect seismic, complex resistivity and biogeochemical data sets and to serve as abiotic control columns, care was taken to ensure that the column packing, flow rates, and other experimental parameters were similar across the columns.

[17] Several pore volumes of lactate were flushed through the water-saturated, sand packed system before the experiment started, at which time the sulfate-reducing bacteria *Desulfovibrio vulgaris* were introduced into the middle and the nutrients were introduced into the bottom of the upward-flowing column. From the multilevel sampling ports, spaced 3.8 cm along column length, sulfate reduction was monitored over seven weeks, as indicated by decreasing substrate and metals concentrations, increasing biomass, and visually discernable regions of metal sulfide accumulation. The region of sulfide mineral precipitation showed a shift toward the influent (bottom) portion of the column over time as a result of microbial chemotaxis toward elevated substrate concentrations at the base of the column [*Williams et al.*, 2005]. Upon termination, the fluid sampling and geophysical measurement columns were destructively evaluated; the sediment samples were collected to determine grain-affixed biomass, extractable metals, and to provide materials for electron microscopy.

[18] *Williams et al.* [2005] showed that changes in seismic and complex resistivity measurements tracked the onset, spatial distribution, and aging of FeS and ZnS accumulation. In addition, the scanning electron microscope (SEM) images indicated that the biostimulation led to the aggregation of sulfide-encrusted bacterial cells. In this study, we extend this effort from a qualitative tracking of the system response using geophysical measurements to a quantitative estimation of the bioaggregated precipitate characteristics over time.

3.2. Geochemical Data and Evolution Model

[19] Several types of aqueous geochemical measurements were collected over time during the course of the experiments. The principal reaction taking place in the column involves the microbially mediated oxidation of lactate to acetate while reducing sulfate according to $CH_3CH(OH)COO^- + \frac{1}{2}SO_4^{2-} \rightarrow CH_3COO^- + \frac{1}{2}HS^- + HCO_3^- + \frac{1}{2}H^+$. Since the lactate and sulfate concentrations are strongly correlated to the acetate concentrations through the reaction stoichiometry, we only show in Figure 2 the acetate concentrations (a byproduct of lactate oxidation) measured at the sampling port located 3.8 cm from the column base in the experiments as a function of time. The production rate of acetate according to the stoichiometry shown above is in the proportion of 2:1 to the sulfide generated, the dissolved species that drives the precipitation of both FeS and ZnS. In theory and ideally, we could simulate FeS and ZnS precipitates
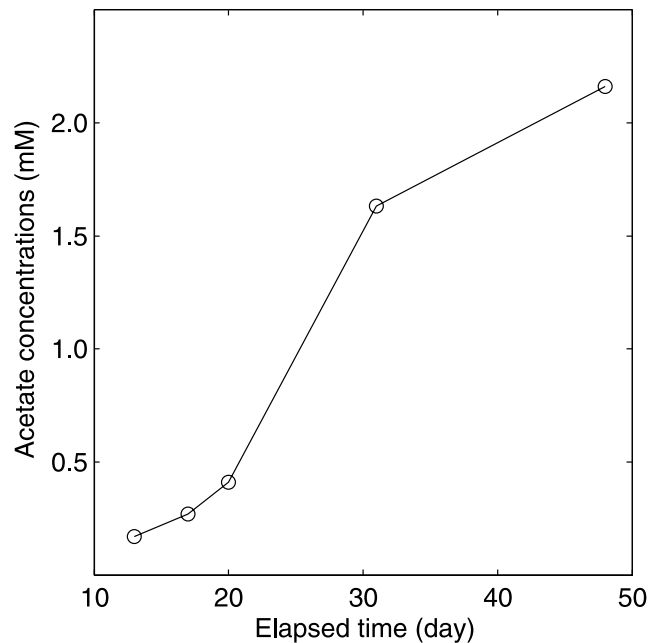

Figure 2. Measured acetate concentrations over time from the column experiments.

rigorously through numerical reactive transport modeling of bioremediation processes based on those measured aqueous geochemical data, but the chemotaxis of the bacteria was a process that was beyond the capabilities of the software at the time.

[20] For the column experiments of *Williams et al.* [2005], we can estimate the volume fraction of FeS and ZnS precipitates from the profiles of the measured total dissolved Fe^{2+} and Zn^{2+} concentrations using a mass balance method. For every mole of acetate produced, one half mole of sulfide is generated, which then results in the precipitation of sulfides according to the reactions $Fe^{2+} + S^{2-} \rightarrow FeS_s$, and $Zn^{2+} + S^{2-} \rightarrow ZnS_s$. For a column having a steady flow, the mass change of an aqueous species (after ignoring dispersion process) can be described by $R = -\phi v(\frac{\partial C}{\partial x})$ at steady state ($\frac{\partial C}{\partial t} = 0$), where ϕ is porosity, v is flow velocity, R is the precipitation rate of the sulfide mineral phase, C is the concentration of Fe or Zn in solution, and x is the distance along the column from the base. According to this equation, the loss rates of Fe(II) and Zn in the aqueous phase were computed by dividing their corresponding concentration differences by the distance between two consecutive sampling ports.

$$R_{j-1/2} = -\phi v \left[\frac{C_j - C_{j-1}}{x_j - x_{j-1}} \right]. \tag{6}$$

Here $R_{j-1/2}$ is the reaction rate defined in the interval between two discrete data points in space x_j and x_{j-1}, where the aqueous concentrations C_j and C_{j-1} are measured [*Steefel and Maher*, 2009]. The FeS and ZnS accumulated during given time intervals were calculated by multiplying equation (6) by the time interval during the sampling process. The accumulated FeS and ZnS calculated using equation (6) matches well the amount of extractable FeS and ZnS measured at the end of the experiment (see Figure 3).
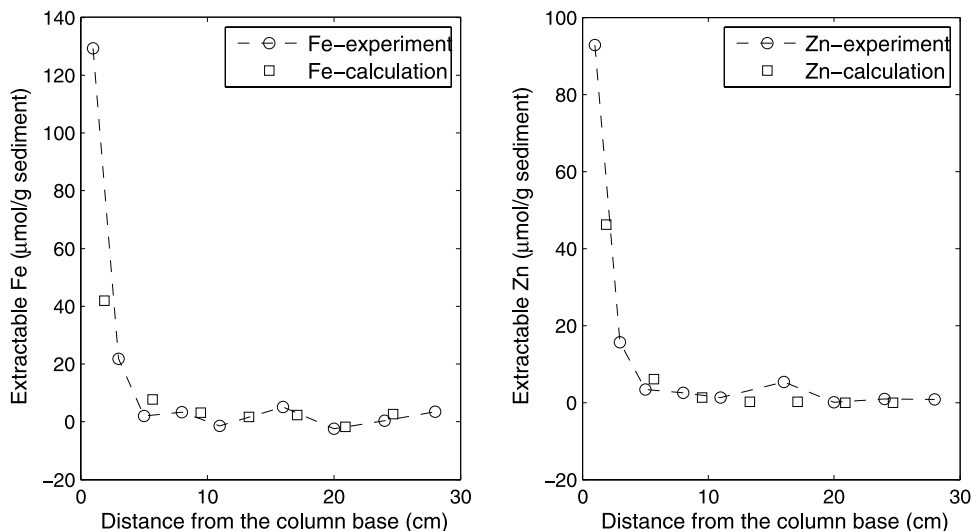
Figure 3. Comparison between the measured and calculated extractable (left) Fe and (right) Zn.

The overall reaction stoichiometry outlined above is also supported by the measurements of other redox-active species (lactate, acetate, and sulfate) in the column, which are in the proper proportions for the electron balance. This further supports the validity of the use of the aqueous concentrations to calculate mineral precipitation rates.

[21] The mass-balance-based estimation, however, is practically impossible under field conditions because many more processes are involved in the mass balance of Fe(II) and Zn, and it is typically challenging to decouple these different processes. For example, in addition to the process of FeS precipitation, minerals such as iron oxide can absorb Fe(II) on their surfaces. Since our ultimate goal of developing the estimation framework is to apply it to field data sets, we assume that the direct estimates of FeS and ZnS will not be available through this simple procedure, but that we may be able to approximate the expected amount and distribution of these mineral phases using more sophisticated geochemical models, normally multicomponent reactive transport models [*Steefel and Maher*, 2009]. The accuracy of the approximation might range from simple qualitative relationships to more sophisticated numerical reactive transport modeling platforms, such as CrunchFlow [*Steefel*, 2008] and TOUGH-React [*Xu et al.*, 2003], depending on available information. As a result, for the purposes of this study, we use the results obtained from the above mass balance method as the ground truth for evaluating the applicability and effectiveness of our state-space estimation framework.

[22] For our application example, we use a simple qualitative relationship with a statistical model for describing possible uncertainty to represent the geochemical evolution. On the basis of the observation from the column experiments, we assume that the increment in volume fraction of metal precipitates is nonlinearly proportional to the concentrations of acetate. This assumption is not important and the relationship can be replaced with a more sophisticated numerical model as it becomes available. For now, this approach is sufficient for testing the developed framework. Let z_t represent the increment of total acetate concentrations from time $t - 1$ to time t and let p_t and p_{t-1} represent

volume fraction of metal precipitates at time t and $t - 1$, respectively. The increment of volume fraction thus can be modeled using function $B(z_t, \theta_1, \theta_2) = \theta_1(1 - \exp(-\theta_2 z_t))$, where θ_1 and θ_2 are parameters associated with the model. This empirical model is intuitively plausible because it is consistent with the fact that the increment of volume fraction increases with increasing of acetate concentrations and the rate of increase in volume fraction decreases. Parameter θ_1 is the limit of the increment of volume fraction, whereas parameter θ_2 depends on the unit of acetate concentrations and the increasing speed of volume fraction. To account for uncertainty in the model, we assume that the two parameters are known within some ranges and the output of the model is subject to Gaussian relative random noise with standard deviation of β_1. Consequently, we obtain the following statistical model that we use for this example to describe the evolution of the precipitate volume fraction from time $t - 1$ to time t.

$$\varepsilon_1 = \frac{(p_t - p_{t-1}) - B(z_t, \theta_1, \theta_2)}{B(z_t, \theta_1, \theta_2)} \sim N(0, \beta_1). \qquad (7)$$

### 3.3. Complex Resistivity Data and Petrophysical Model

[23] The complex resistivity data were collected from several locations along the length of the column and over time by using frequencies from 0.01 Hz to 1000 Hz. In this example, we focus only on the complex resistivity data collected between ports 1 and 2, which correspond to the length interval between 3.5 cm and 7.0 cm away from the column base. Theoretically based models for predicting spectral induced polarization (SIP) signatures in metal containing soils are lacking, despite recent advances in semi-theoretical modeling of SIP signatures in nonmetallic soils [*Leroy et al.*, 2008]. The one exception is the classic electrochemical model of *Wong* [1979]. He attributed the polarization in metallic soils when the metal is less than 10% of the soil volume to diffusion of redox active and inactive ions that are predominantly perpendicular to the metal surface under an applied electric field and to an electrochemical mechanism
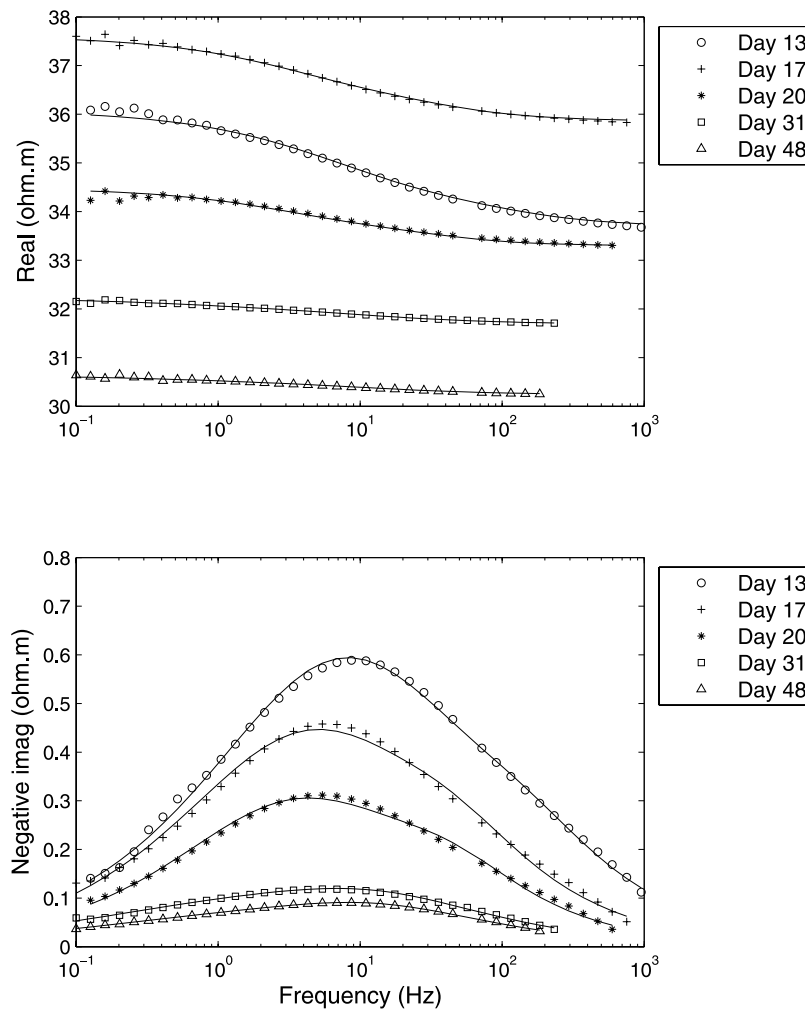
Figure 4. Complex resistivity data (symbols) and their corresponding fits (solid curves) for Cole-Cole models using the stochastic inversion method developed by *Chen et al.* [2008].

associated with the redox active ions that facilitate transport of charge between ionic and electronic conduction. In the model, he also assumed no interaction between the electric fields of the individual polarizable particles (i.e., the metallic minerals), a condition that *Wong* [1979] stated was reasonable for metal concentrations up to 16%. However, since the theoretical model requires the definition of several (more than eight) electrochemical parameters that are typically poorly determined, no practical applications have been presented in the peer-reviewed literature.

[24] Given the lack of easily applied theoretical models to adequately describe the SIP response of soils containing metallic minerals, phenomenological formulations, such as the Cole-Cole relaxation model [*Cole and Cole*, 1941], are often invoked [*Pelton et al.*, 1978, 1983; *Binley et al.*, 2005; *Slater et al.*, 2006]. Similar to those studies, the complex resistivity data are first inverted for Cole-Cole model parameters (e.g., chargeability and time constant) using the stochastic inversion method developed by *Chen et al.* [2008]. Figure 4 shows the real and imaginary components of the measured complex resistivity data after inoculation as well as their corresponding fits to Cole-Cole models. Figures 5 and 6 give the medians and 95% predictive intervals of the inverted chargeability normalized by zero-
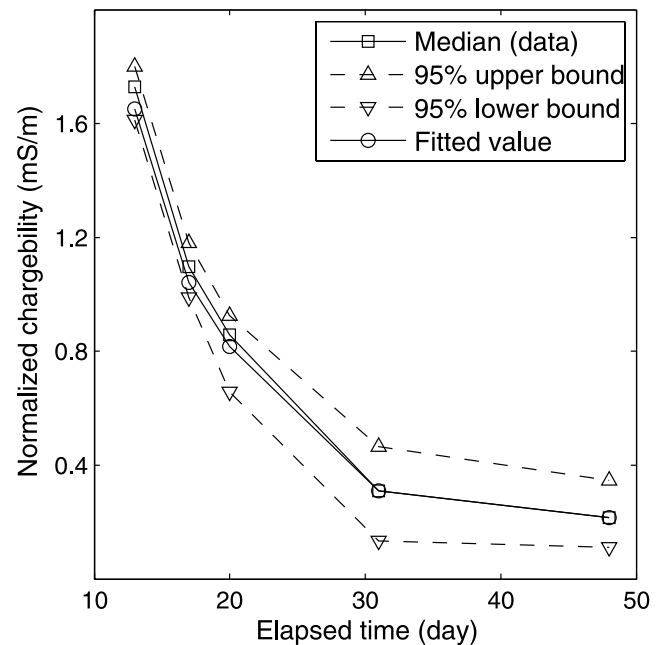
Figure 5. Time-lapse normalized chargeability data (squares) and their corresponding fits (circles).
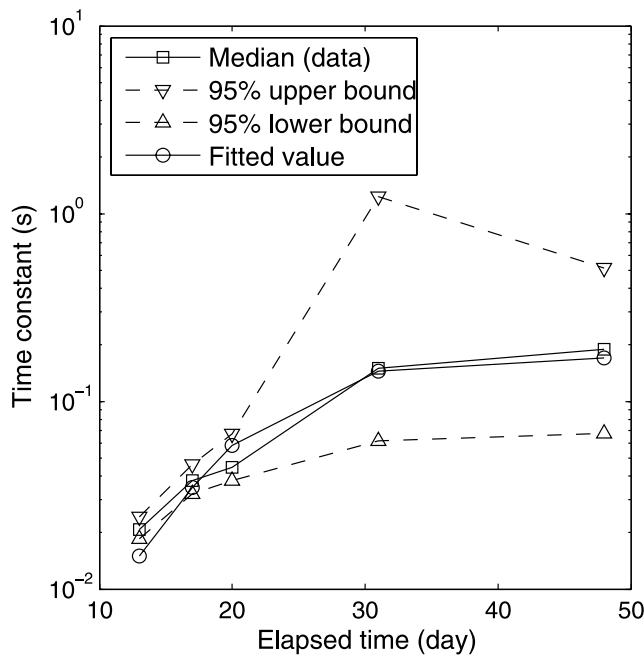
Figure 6. Time-lapse time constant data (squares) and their corresponding fits (circles).

frequency resistivity (referred to as normalized chargeability) and time constant parameters from day 13 to day 48. We did not get reliable estimates of Cole-Cole parameters from the complex resistivity data collected on the date earlier than day 13. We speculate that under the conditions where geochemical (i.e., aqueous chemistry) conditions are changing rapidly, Cole-Cole parameters may not adequately capture changes in the complete spectral response.

[25] We develop a petrophysical model to link the inverted Cole-Cole parameters to the properties of metal precipitates based on the observations of the column experiments on the date after day 13. From Figures 5 and 6, we can see that the normalized chargeability, a nearly linear function of the surface area of sulfide minerals in contact with water, is decreasing through time while volume fraction of the precipitates suggested by the geochemical data (Figure 2) is increasing through time. These observations perhaps are different from the response of complex resistivity obtained at early time because at early time, a single cell has an increasing layer of sulfide on it and both surface area and volume fraction increase over time. To explain the observations at the later time, we develop a rock-physics model of cells aggregating into clusters, which provides the key geometric parameters involved in modeling both permeability and induced polarization (IP) responses of the sand column. In the following, we conceptually describe the petrophysical model and present the results that are directly related to the inverted Cole-Cole parameters (i.e., normalized chargeability and time constant). The detailed derivations are given in Appendix A. The developed model involves many parameters, some of which can be approximately determined from SEM images and some need to be estimated during the inversion, which are also explicitly given in the following description.

[26] We assume that the formation of metal precipitates includes two main phases based on our observations from the column experiments. Similar processes were also observed by *Moreau et al.* [2004] under the natural conditions where the concentration of aqueous metals (e.g., zinc) was much lower. The early phase involves the coating of an individual cell, that is, the bacterial cells in the system produce sulfide mineral to the point that they become entirely covered in a sulfide layer and ultimately die [*Williams et al.*, 2005]. The subsequent second phase involves the aggregation of individual coated biominerals, in which the dispersed individual coated cells form clusters. Since we only have data after 13 days of bioremediation, we assume the dominant process involved in this example is the cell aggregation. For ease of description, we assume that both cells and metal sulfides are spherical, and the effects of deviations between the actual shape and that of a sphere will be addressed by some coefficients. As shown in Figure 7, all cells with a sulfide coating are assumed to be initially dispersed (i.e., widely separated from one another). Over time, the dispersed cells gradually aggregate into clusters, in the present simple model, taking the form of spheres. We employ a face-centered sphere packing approach to represent the aggregation, as is described in Appendix A. These spherical clusters grow through the attachment of additional dispersed cells. Since an isolated coated cell has a larger mineral-fluid surface area than a cell attached to a cluster, the surface area of sulfide will decline as long as the rate of cells attaching to clusters is greater than the rate at which new dispersed cells are being formed. This is the case in the column experiments as shown by *Williams et al.* [2005]. Given the near complete consumption of lactate within the first 1.9 cm of the column by day 12, this loss of the primary electron donor (lactate) severely
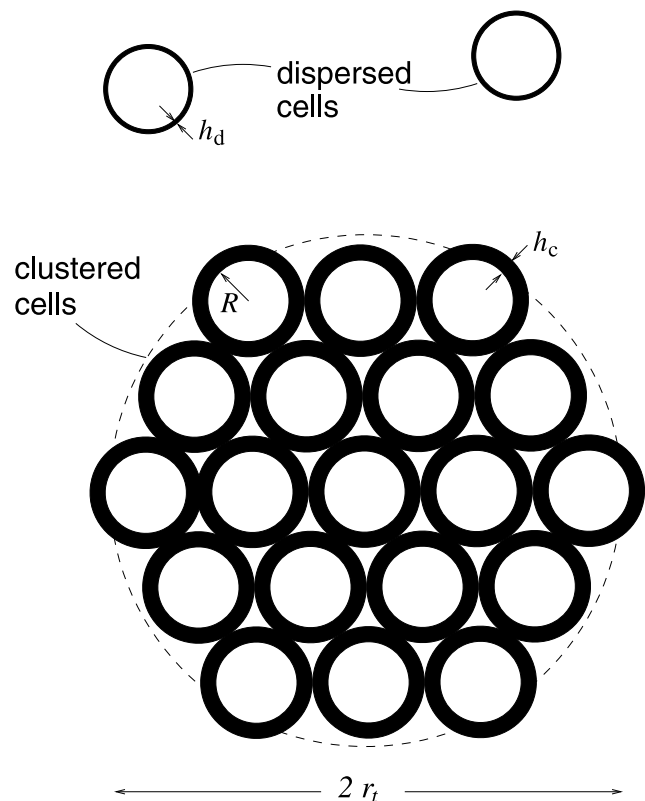
Figure 7. Schematic representation of FeS and ZnS precipitation for the induced polarization (IP) data inversion.

limits subsequent microbial growth and cell division, thereby minimizing the rate at which new dispersed cells are formed. To account for the observation that total sulfide volume in the pores increases over time, we also assume that the cells in a cluster have a thicker layer of sulfide on them than do the dispersed cells (i.e., $h_c \gg h_d$ in Figure 7). To describe the process, we define two key parameters: One is the volume fraction of metal precipitates (p_t) and the other is the fraction of dispersed coated biominerals (w_t). Both are functions of time and will be estimated in the inversion.

[27] We can obtain an analytical relationship between normalized chargeability (m_t) and parameters p_t, w_t, and θ_3, the latter of which is a coefficient that accounts for incomplete knowledge about the thickness of encrusted cells. Within the model and under certain assumptions, we can obtain the specific area $S_t = G_0(p_t, w_t, \theta_3)$ (see equation (A4)). Additionally, normalized chargeability (i.e., polarization magnitude) has been repeatedly shown to scale with S_t in laboratory studies conducted on both metallic soils [e.g., *Slater et al.*, 2006] and nonmetallic soils [*Scott and Barker*, 2005; *Slater et al.*, 2006]. Therefore we can assume that normalized chargeability is proportional to specific surface area, i.e., $m_t = \theta_4 S_t = G_1(p_t, w_t, \theta_3, \theta_4)$, where θ_4 is a parameter that may partially account for disparity in the shapes between spheres and actual ones and partially explain the ratios between the specific area and chargeability. This is an empirical based model, which is critical for the success of our estimation because it links the IP responses to the physical properties of geochemical precipitation. To consider uncertainty in the model, we also assume the empirical relationship is subject to relative Gaussian random errors with the standard deviation of β_2. This is a common assumption for likelihood functions because the Gaussian distribution is the most robust probability distribution for characterizing errors, even the errors are non-Gaussian [*Stone*, 1996]. Thus we obtain the following model.

$$\varepsilon_2 = \frac{m_t^{obs} - G_1(p_t, w_t, \theta_3, \theta_4)}{m_t^{obs}} \sim N(0, \beta_2). \tag{8}$$

[28] We can also obtain an analytical formula to link time constant (τ_t) to the fraction of dispersed biominerals (w_t). Time constant, describing the length scale of the relaxation in IP responses, has been widely recognized as a function of the pore or grain size characteristics of soils [e.g., *Olhoeft*, 1985; *Chelidze and Gueguen*, 1999] and therefore can be linked to the mean radius of clusters formed from metal precipitates. *Schwartz* [1962] showed that the function is consistent with electrochemical theory for colloidal suspensions, whereby we can tie time constant τ_t at time t to the mean radius of aggregated clusters (r_t) using the following formula: $\tau_t = r_t^2/(2D)$, where D is referred to as the surface ionic diffusion parameter and its value is given by 3×10^{-9} m^2/s as used by *Tarasov and Titov* [2007] and *Slater et al.* [2007]. In addition, we can derive the mean radius as $r_t = \theta_5 l_0 \sqrt{1 - w_t}$ (see equation (A8)), where l_0 is the characteristic pore-throat radius of the system and has a value of 1.3×10^{-4} m as determined from *Thompson et al.* [1987] permeability model prior to precipitation, and θ_5 is a parameter that explains the effects of differences between the actual shape and the used sphere and the effects of uncertainty in the values of the surface ionic parameter and

the characteristic pore-throat radius. This parameter will be determined in the inversion with a value between 0.2 and 0.9. By combining the above two relationships, we obtain $\tau_t = \theta_5^2 l_0^2 (1 - w_t)/(2D) = G_2(w_t, \theta_5)$. This is an important relation for the estimation because it provides a linkage between time constant and the fraction of dispersed cells. Again, to account for uncertainty in the model, we assume the empirical relationship is subject to relative Gaussian random errors with the standard deviation of β_3. Thus we obtain the following model.

$$\varepsilon_3 = \frac{\tau_t^{obs} - G_2(w_t, \theta_5)}{\tau_t^{obs}} \sim N(0, \beta_3). \tag{9}$$

3.4. Bayesian Model

[29] We apply the estimation framework given in section 2 to the column experimental data described by *Williams et al.* [2005]. We consider volume fraction ($p_1, p_2, \ldots, p_n$) as state variables and time-lapse normalized chargeability ($m_1^{obs}, m_2^{obs}, \ldots, m_n^{obs}$) and time constant ($\tau_1^{obs}, \tau_2^{obs}, \ldots, \tau_n^{obs}$) as measurements with Gaussian relative random errors. We also consider the fraction of dispersed biominerals ($w_1, w_2, \ldots, w_n$) and five time-independent parameters ($\theta_1, \theta_2, \ldots, \theta_5$) as unknowns. We jointly estimate those state variables and time-dependent and independent parameters by conditioning on the inverted Cole-Cole parameters.

[30] We can specify the general Bayesian framework given in equation (5) with the geochemical evolution model described in section 3.2, and the complex resistivity rock-physics model conceptually summarized in section 3.3 (and described in detail in Appendix A) to obtain the following specific Bayesian model for estimation of precipitate related parameters (see Appendix B).

$$f(p_1, p_2, \cdots, p_n, w_1, w_2, \cdots, w_n, \theta_1, \theta_2, \cdots,$$
$$\theta_5 | m_1^{obs}, m_2^{obs}, \cdots, m_n^{obs}, \tau_1^{obs}, \tau_2^{obs}, \cdots, \tau_n^{obs})$$
$$\propto f(\theta_1, \theta_2, \cdots, \theta_5) f(w_1, w_2, \cdots, w_n) \prod_{t=1}^{n} \left(\frac{1}{B(z_t, \theta_1, \theta_2)} \right)$$
$$\cdot \exp\left\{ -\frac{1}{2\beta_1^2} \sum_{t=1}^{n} \left(\frac{(p_t - p_{t-1}) - B(z_t, \theta_1, \theta_2)}{B(z_t, \theta_1, \theta_2)} \right)^2 \right\}$$
$$\cdot \exp\left\{ -\frac{1}{2\beta_2^2} \sum_{t=1}^{n} \left(\frac{m_t^{obs} - G_1(p_t, w_t, \theta_3, \theta_4)}{m_t^{obs}} \right)^2 \right\}$$
$$\cdot \exp\left\{ -\frac{1}{2\beta_3^2} \sum_{t=1}^{n} \left(\frac{\tau_t^{obs} - G_2(w_t, \theta_5)}{\tau_t^{obs}} \right)^2 \right\}. \tag{10}$$

[31] In equation 10, we assume $p_0 = 0$ (i.e., no precipitates at time t_0), $\beta_1 = 5\%$, $\beta_2 = 1\%$, and $\beta_3 = 10\%$. In this model, we only take account for random measurement errors, and systematic errors in data, model assumptions, and parameterization cannot be resolved. However, given the flexibility of our estimation framework, we can certainly combine them into the model if we know the structures of those systematic errors. To obtain samples from the joint posterior distribution given in equation (10), we first derive conditional distributions for unknown variables and then use the MCMC sampling methods to obtain many samples of the unknowns. Details about the MCMC sampling
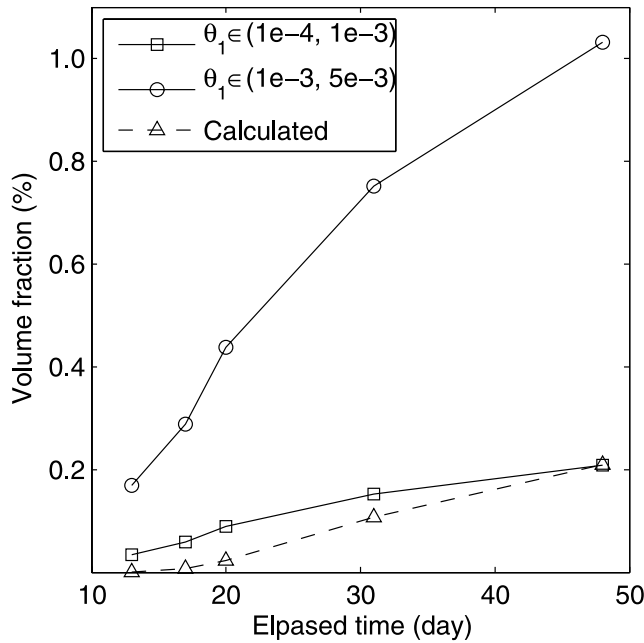
Figure 8. Effects of parameters in the geochemical model on the estimates of volume fraction.

methods are provided by *Chen et al.* [2006] and in Appendix C.

4. Estimation Using Laboratory Column Experimental Data

4.1. Estimation of Volume Fraction of FeS and ZnS Precipitates

[32] We first estimate volume fraction of FeS and ZnS precipitates using only the measured acetate concentrations. By dropping the last two terms on the right side of equation (10), we can obtain the joint distribution of evolved precipitate volume fraction as functions of the measured acetate concentrations and the evolution model given in equation (7). For the given evolution model $B(z_t, \theta_1, \theta_2)$, we choose the prior ranges of parameters θ_1 and θ_2 on which the estimated medians of the volume fraction have a similar range to the values calculated from direct measurements of metal sulfide precipitates. The estimates of volume fraction are very sensitive to the choice of parameter θ_1, which is the limit or maximum increment of volume fraction for given acetate concentrations. Figure 8 shows the effect of its prior range on the estimates of volume fraction. If we assume the maximum increment of volume fraction is in the range of $(1e-3, 5e-3)$, the estimated medians of volume fraction (circles in Figure 8) are one order larger than those calculated from direct measurements of dissolved metal concentrations (triangles in Figure 8). However, if we choose a prior range of $(1e-4, 1e-3)$ for parameter θ_1, we can obtain the medians of volume fraction (squares in Figure 8) that are in the same order as those calculated from dissolved metal concentrations. Therefore, in this example, we assume that parameter θ_1 is uniformly distributed over $(1e-4, 1e-3)$. The estimates of volume fraction are less sensitive to the value of parameter θ_2, the increment rate of volume fraction for given acetate concentrations. We

assume that the parameter θ_2 is uniformly distributed between 1 and 10.

[33] We combine information from complex resistivity data into the estimation using all the terms in equation (10). The added data are normalized chargeability and time constant, both of which are obtained from fitting complex resistivity data with Cole-Cole models following *Chen et al.* [2008]. In this case, we must also invert the fraction of dispersed cells over time and three additional time-independent parameters θ_3, θ_4, and θ_5. As will be subsequently discussed, all those parameters can be estimated well from the joint inversion. Figure 9a shows the estimated medians of volume

(a)

(b)
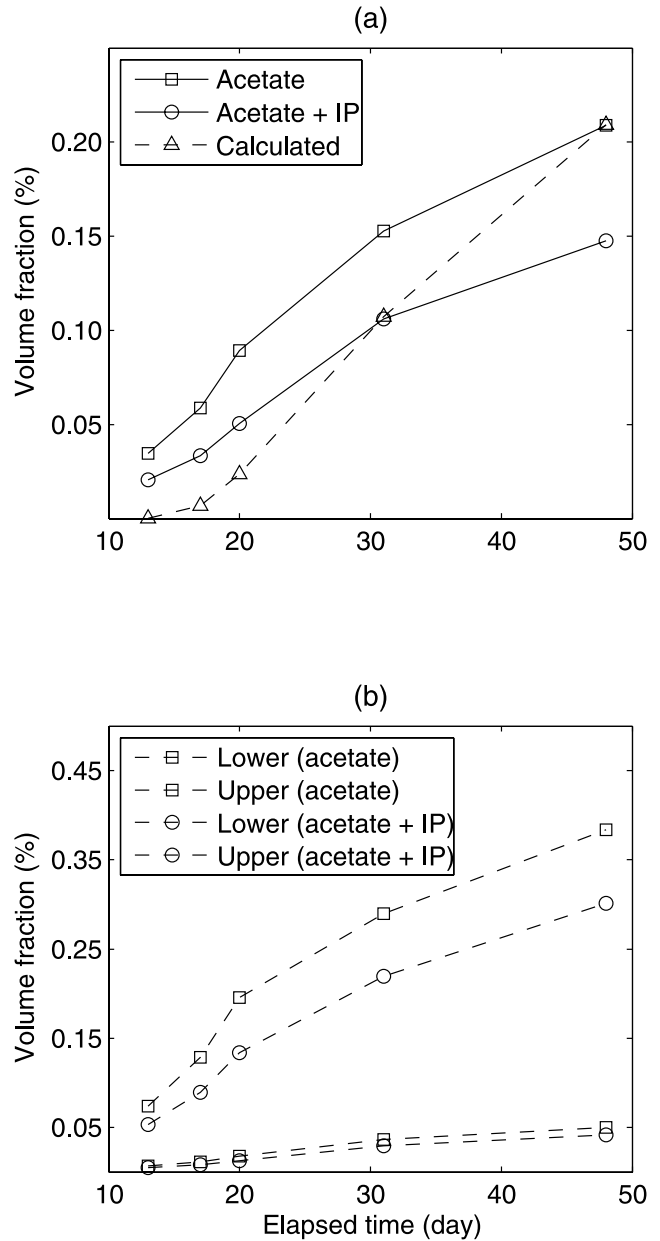
Figure 9. (a) Estimates of volume fraction obtained using acetate data only (squares) and using both acetate and IP data (circles), and those calculated from the dissolved iron and zinc concentrations (triangles); (b) 95% highest probability domains (HPDs) obtained using acetate data only (squares) and using both acetate and IP data (circles).
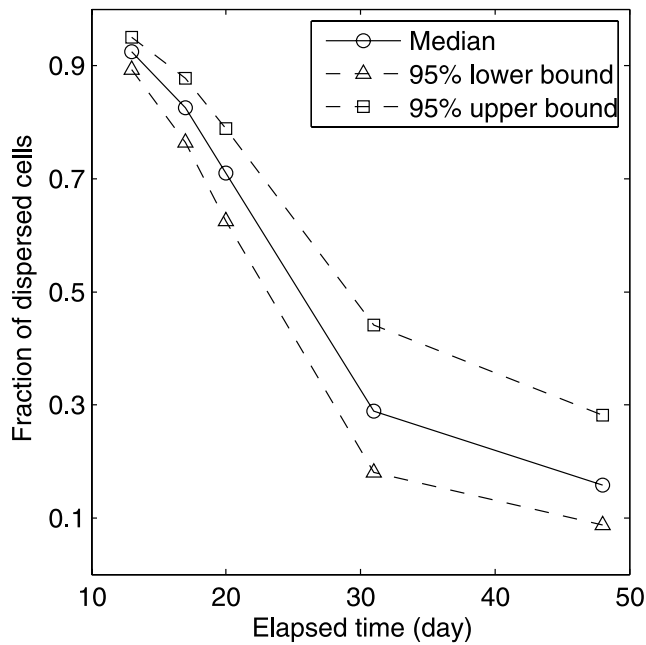
Figure 10. Estimated fraction of dispersed biominerals over time.

Figure 11. Estimated mean radius of aggregated clusters over time.

fraction of FeS and ZnS precipitates obtained using the acetate concentrations only (squares) and using both acetate concentrations and complex resistivity data (circles). The calculated volume fraction using equation (6) from the dissolved Fe^{2+} and Zn^{2+} concentrations are also shown in Figure 9a as triangles. Comparing the estimated and calculated volume fraction, we find that the estimates of volume fraction of FeS and ZnS precipitates obtained using both acetate concentration and complex resistivity data are much better, having a root-mean-square (RMS) difference of 0.0337, relative to those using acetate concentration data only, having a RMS value of 0.0453. Figure 9b shows the 95% highest probability domains (HPDs) of the estimated volume fraction. We can see that combination of complex resistivity and acetate concentration data yields only slightly smaller uncertainty bounds. Although we include more data in the procedure, we have also added more unknown parameters.

4.2. Estimation of Fraction of Dispersed Cells and Mean Radius of Aggregated Clusters

[34] We can directly estimate the fraction of dispersed cells as a function of time as shown in equation (10) through incorporating complex resistivity data into the inversion. Figure 10 shows the medians of the marginal posterior probability distribution of the fraction of the dispersed cells and their corresponding 95% HPDs. The fraction of dispersed, coated cells decreases from about 90% to about 10% from day 13 to day 48 because of aggregation of dispersed cells into large clusters.

[35] Although we did not directly estimate the mean radius of aggregated clusters in equation (10), we can calculate it through the formula: $r_t = \theta_5 l_0 \sqrt{1 - w_t}$ from the fraction of dispersed cells and time-independent parameter θ_5. Figure 11 shows the estimated medians of the mean radius of aggregated clusters, together with their corresponding 95% HPDs. From Figure 11, we can see that the
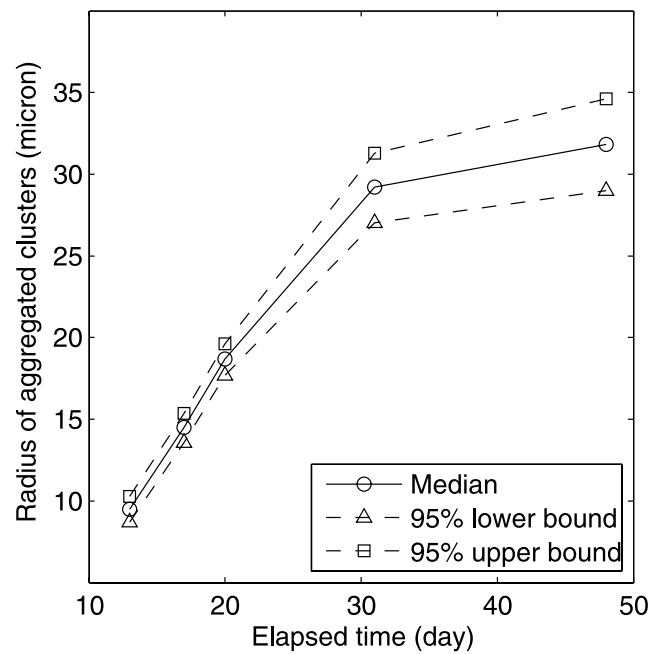
mean radius of aggregated clusters increases as we expected from about 10 microns to about 30 microns, which is reasonable according to the observations from SEM images of samples from the destructed experimental columns (10–20 microns) [*Williams et al.*, 2005].

4.3. Estimation of Permeability

[36] We can also estimate effective permeability and its change over time in the zone impacted most significantly by
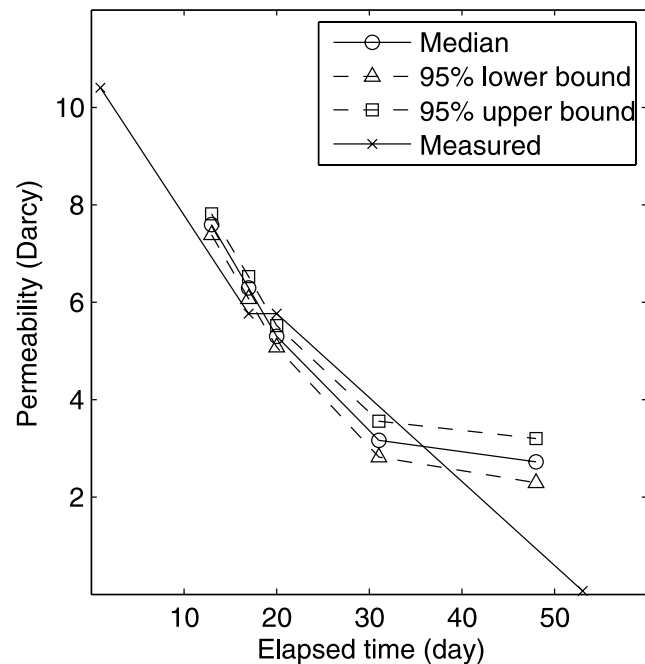
Figure 12. Comparison between the estimated (circles) and measured (crosses) permeability.
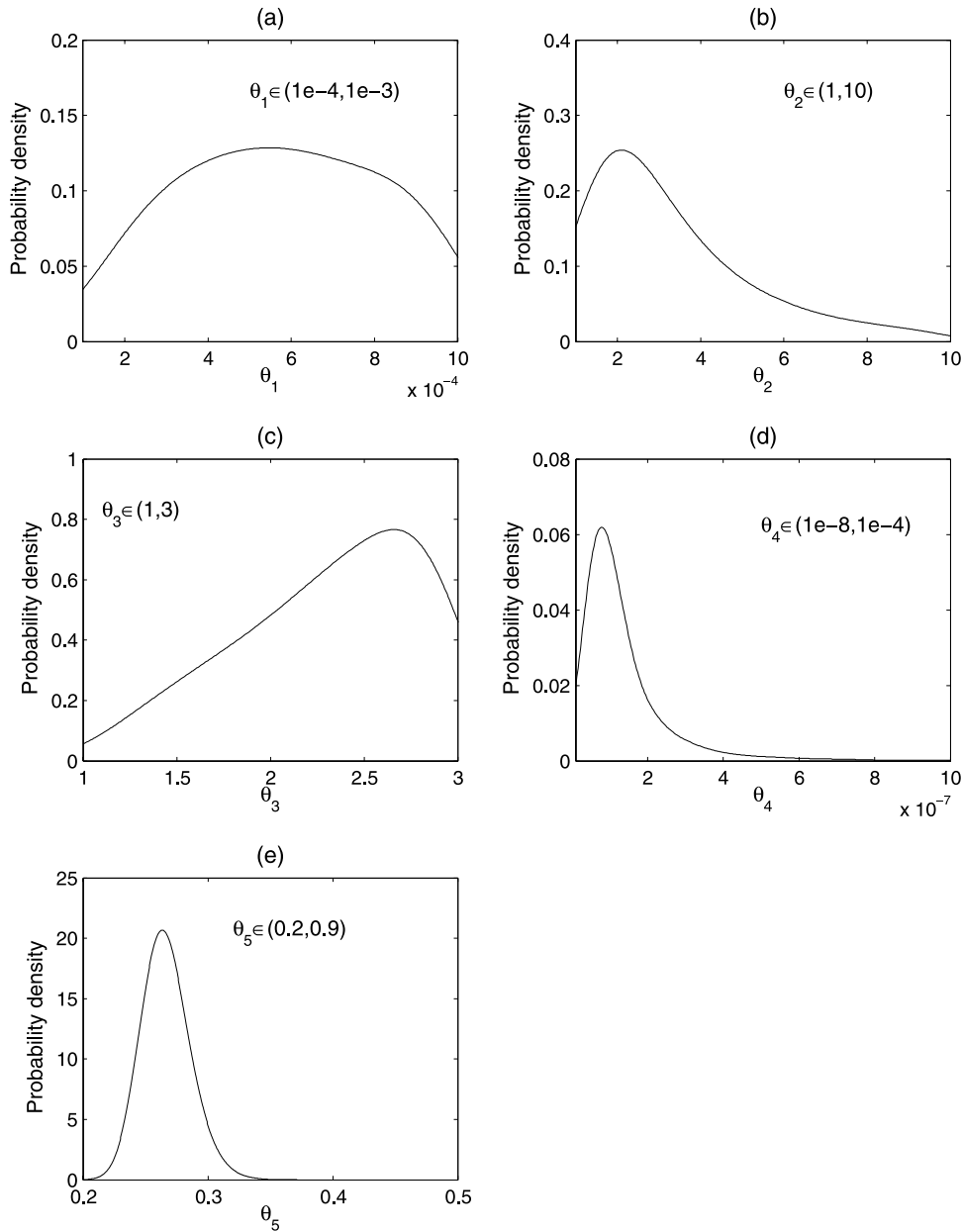
Figure 13. Estimated posterior probability density functions (pdfs) of time-independent parameters, where priors are uniform distributions on given ranges shown on Figures 13a–13e.

the biostimulation using the developed petrophysical model and the complex resistivity data. Permeability is a key parameter for flow transport and is difficult to measure in hydrogeology. Following *Thompson et al.* [1987], we can obtain permeability at time t as below:

$$k_t = \frac{1}{226}(\phi_0(1 - \Pi_t)(l_0 - 2r_t))^2, \tag{11}$$

where Π_t is the volume fraction of the pores occupied by clusters, which is a function of both p_t and w_t (see Appendix A) and typically is much larger than the fraction of FeS and ZnS precipitates p_t. Symbol $\phi_0 = 0.37$ is the initial porosity of sand grain prior to precipitation. Figure 12 shows the medians (solid lines with circles) of the estimated permeability over time, together with their corresponding 95%

HPDs (dashed lines with triangles or squares). The effects of the evolved precipitates on the effective permeability are evident; the formation of the aggregated clusters reduces permeability at the location from about 8 darcies to 2 darcies.

[37] To justify the estimated permeability, we compare these results with those calculated from the measured permeability of sand column by *Williams et al.* [2005]. In the column experiments, after the initial migration of the precipitation front toward the column base (influent), the microbially mediated sulfide precipitation mainly occurred in the first several centimeters of the soil column. Let k_{sc} be the permeability of the sand column, which is calculated from the measured hydraulic conductivity data by *Williams et al.* [2005] and has a value of 10.4 darcies before precipitation, 9.15 darcies on days 17 and 20, and 0.4 darcy on day 53. Let v_c be the volume fraction of the location where cluster

development is occurring, which is about 0.17 in the case. Thus we have

$$\frac{1}{k_{sc}} = \frac{v_c}{k_t} + \frac{1 - v_c}{k_0}, \qquad (12)$$

where k_0 is effective permeability prior to biostimulation and has a value of 10.4 darcies. Using the available information and equation (12), we can calculate the effective permeability as a function of time, which is shown as solid lines with crosses in Figure 12. Comparison between the estimated and calculated effective permeability suggests that the developed estimation framework and petrophysical model permit a reasonable estimation of changes in permeability conditioned on complex resistivity data.

4.4. Estimation of Time-Independent Parameters

[38] Figure 13 shows the marginal posterior probability distribution functions (pdfs) of five time-independent parameters. We show these results in order to demonstrate an important benefit provided by Bayesian estimation approaches. This is that they allow us to consider those parameters in the model that we do not have enough information as unknowns with prior ranges. The posterior results of those parameters may or may not get information from the data that we are conditioning to, depending on relationships between the data and those parameters. For a parameter, such as θ_1, the limit of increment of volume fraction, whose posterior pdf is almost the same as its prior pdf, we should generally be careful in choosing its prior range and analyzing its effects on the estimated results. However, in the current study, since our goal is to show the increasing values of complex resistivity data for a given prior model, we pick such a prior range.

[39] For parameters, such as θ_2, θ_3, θ_4, and θ_5, the choice of prior distributions is not crucial. For example, from Figure 13, we can see that parameter θ_5 is well resolved. Even if we start from a wider prior range, we still get a similar posterior pdf. We can use the estimated results as calibration of the rock-physics model and apply them for prediction. We can also gain insights from the results that justify the developed petrophysical model.

5. Discussion and Conclusions

[40] We developed a general Bayesian framework based on a state-space approach to estimate biogeochemical end products using time-lapse geochemical and geophysical data. The developed framework is very flexible, as it allows for systematic incorporation of multisource and multiscale information and permits use of different forms of forward geochemical and petrophysical models.

[41] We demonstrated the utility of the developed estimation framework for quantitative estimation of biogeochemical parameters by applying it to geophysical and geochemical data sets collected during laboratory column biostimulation experiments. In the case study, we estimated the evolution of several parameters in connection with biostimulation-induced metal sulfide precipitates. We used empirical relationships to link the total concentrations of acetate to volume fraction of FeS and ZnS precipitates and developed a novel rock-physics model based on face-centered sphere packing to link normalized chargeability and time constant obtained from complex resistivity data to various time- dependent and independent parameters related to the aggregated precipitates. We note that the petrophysical model included within the estimation framework is expected to be refined as our understanding of the evolution of biogeochemical end products and their impact on pore structures become available; this topic is a subject of ongoing research by the authors. For testing of the developed estimation framework, we have developed a model that is conceptually simple and consistent with all available observations made by *Williams et al.* [2005].

[42] Our results show that we can obtain quantitative estimates of the evolution of volume fraction and several other types of information related to the precipitation from the time-lapse complex resistivity data using the developed Bayesian framework and the assumed petrophysical model. The incorporation of time-lapse complex resistivity data improves the estimates of volume fraction over the estimates obtained using measured geochemical data alone, and provides the estimates of dispersed cell fraction, mean radius of aggregated clusters, and permeability, which geochemical data alone could not provide.

[43] Estimation of biogeochemical parameters using time-lapse geochemical and geophysical data is subject to uncertainty. This may come from the choice of models for linking geochemical and geophysical properties to parameters related to biogeochemical end products, from the choice of prior distributions of unknown parameters, and from the estimation of parameters associated with the petrophysical model. To address those uncertainties, we assume the output of models includes Gaussian relative random errors and the associated model parameters are uniformly distributed on given prior ranges. The uncertainty can be reduced through two different approaches. The first one is to incorporate multiple types and multiple scales of information using the Bayesian integrated approach; the other approach is to improve our understanding of the petrophysics of precipitation through additional laboratory, theoretical, and numerical experiments. Ongoing efforts within the environmental community to advance our understanding of petrophysical models and to incorporate a variety of data sets for exploring system behavior are expected to lead to improved quantitative estimates of biogeochemical end-product characteristics.

[44] The obvious potential of the developed framework is its use for quantitative estimation of biogeochemical parameters at the field scale, using time-lapse direct borehole and indirect geophysical data sets. Application of the developed procedure with time-lapse geophysical data sets has the potential to provide a wealth of information about the spatiotemporal evolution of biogeochemical processes associated with remedial treatments that are difficult to obtain using borehole data alone. However, for use at the field scale, we may need to consider state vectors and time-lapse geophysical data as functions of the spatial variability associated with natural heterogeneity and its controls on geophysical and geochemical responses. We need to develop models to characterize spatial patterns of biogeochemical properties and geochemical and geophysical data as functions of time.

[45] Extension of the estimation framework to the field-scale presents other challenges as well. Different types of

geochemical and geophysical data typically have different measurement support scales. For example, geochemical data are typically collected from borehole fluid samples and are often considered to provide high-resolution "point measurements", whereas geophysical data often are collected from crosshole or surface surveys at relatively lower resolution but with larger spatial coverage. To use those data together, we need to find ways to bridge the scale discrepancies for integration and to permit development and validation of petrophysical models. Additionally, in-situ remediation treatments often lead to multiple and competing biogeochemical reactions in the subsurface. In our case study, the column experiments only involved the stimulation of sulfate reducing bacteria through a use of a pure culture, which led to the controlled precipitation of metal sulfide minerals following the introduction of dissolved metal ions at a known concentration. However, in nature many biogeochemical processes often exist that can occur within the footprint of the geophysical measurements, such as dissolution, precipitation, gas generation, and biofilm formation. To apply the developed approach to natural field conditions, we will likely need to augment the Bayesian framework to distinguish the dominant process and associated end products.

[46] Our study focused on developing and testing a stochastic approach for estimating biogeochemical end products associated with bioremediation treatments using time-lapse geophysical laboratory data sets. This approach builds upon recent biogeophysical research that indicated that geophysical data can track system responses over time; it now allows for quantitative estimation of transformational end products in a minimally invasive manner. Further development and application of the estimation framework is expected to significantly improve our understanding of complex biogeochemical processes in naturally heterogeneous subsurface systems and our ability to monitor processes remotely. An improved understanding and ability to monitor in-situ biogeochemical processes is expected to lead to an improved ability to design, guide, predict, and assess in-situ remediation approaches at the field scale.

Appendix A: Rock-Physics Model for Cells Aggregating Into Clusters

[47] We develop a rock-physics model in this section to link Cole-Cole parameters (i.e., normalized chargeability and time constant) to the properties of metal precipitates. The derivation is mainly based on observations from the column experiments performed by *Williams et al.* [2005].

[48] As shown in Figure 7, we assume that the bacterial cells in the system produce sulfide minerals to the point that they become covered in a sulfide layer and ultimately die. In a highly simplified model, we assume that all sulfides in the system reside as spherical shells around the cells; we distinguish between dispersed cells and clustered cells. Initially, all cells with a sulfide coating are dispersed (i.e., widely separated from one another); through time, the dispersed cells aggregate into clusters with the form of spheres in the present simple model. These spherical clusters grow through the attachment of additional dispersed cells. Since an isolated cell has a larger mineral-fluid surface area than a cell attached to a cluster, the mineral-fluid surface area of sulfides will decline as long as the rate of cells attaching to clusters is

greater than the rate at which new dispersed cells are being formed. We assume this is the case in the column experiments. To account for the observation that total sulfide volume in the pores is increasing through time, we assume that the cells in a cluster have a thicker layer of sulfides on them than do the dispersed cells.

[49] We assume that there are N_t cells with sulfides on them in every unit volume of pore space, which are partitioned into N_d dispersed cells and N_c clustered cells such that $N_t = N_d + N_c$. The dispersed cells are coated with a sulfide layer of thickness h_d while the clustered cells have a sulfide layer with thickness of h_c (see Figure 7). For ease of description, we define two time-invariant dimensionless parameters $\chi_d = h_d/R$ and $\chi_c = h_c/R$, where R is the radius of a cell without sulfides on it. Thus we can obtain the volume of sulfides surrounding a single dispersed cell by $4\pi R^2 h_d(1 + \chi_d + \chi_d^2/3)$ and that surrounding a coated cell within the cluster approximately by $4\pi R^2 h_c(1 + \chi_c + \chi_c^2/3)$. Let $w_t = N_d/N_T$, which varies over time and has a value between 0 and 1. Let $g_d = \chi_d(1 + \chi_d + \chi_d^2/3)$ and $g_c = \chi_c(1 + \chi_c + \chi_c^2/3)$. Consequently, the fraction of pore volume p_t occupied by sulfide precipitate is given by

$$p_t = 4\pi N_T R^3 (g_d w_t + g_c(1 - w_t)). \qquad (A1)$$

The relationship between p_t and w_t given by equation (A1) requires knowledge of the total number N_T of cells per unit pore volume covered with precipitates. Since this generally is unknown, we consider both p_t and w_t as unknown parameters that are determined by the inversion at each time step. We subsequently express all other time-varying petrophysical parameters required within the modeling as functions of p_t and w_t.

[50] We first derive the specific surface area S_t, defined as the area of sulfides in contact with water per unit pore volume, in terms of p_t and w_t. Note that the number of coated cells n_c in a cluster is given by $n_c = (r_t/R)^3 (1 - \phi_c)/(1 + \chi_c)^3$, where ϕ_c is the porosity in a cluster. Therefore the total number of clusters per unit volume of pore space M_c is given by

$$M_c = \frac{N_T - N_d}{n_c} = N_T(1 - w_t)\left(\frac{R}{r_t}\right)^3 \frac{(1 + \chi_c)^3}{(1 - \phi_c)}, \qquad (A2)$$

and the volume fraction (Π_t) of pores occupied by aggregated clusters is given by

$$\Pi_t = \frac{4}{3}\pi r_t^3 M_c = \frac{(1 + \chi_c)^3 (1 - w_t) p_t}{3(1 - \phi_c)(g_d w_t + g_c(1 - w_t))}. \qquad (A3)$$

The specific surface area S_t can be modeled as $N_d 4\pi R^2(1 + \chi_d)^2 + M_c(1 - \phi_c)4\pi r_t^2$, where the first term is the surface area associated with the individual dispersed cells and the second term is the surface area of the clusters. Since a cluster is an electronically conducting object, only its exterior surface contributes to the IP effect. Thus we obtain

$$S_t = \frac{p_t(R^{-1}w_t(1 + \chi_d)^2 + r_t^{-1}(1 - w_t)(1 + \chi_c)^3)}{g_d w_t + g_c(1 - w_t)}, \qquad (A4)$$

where $R \approx 0.3 \times 10^{-6}$ m is the radius of a cell based on SEM imagery described by *Williams et al.* [2005]. The time-

invariant parameters χ_d and χ_c that represent the fraction of a cell radius occupied by sulfide are not well known. We model them as $\chi_d = 10^{-3}\theta_3$ and $\chi_c = 10^{-1}\theta_3$, where θ_3 is a parameter that must be determined from the inversion with allowed values in the range between 1 and 3. The observation that $\chi_c \approx 10^{-1}$ is consistent with the SEM images of cells from clusters obtained by *Williams et al.* [2005].

[51] We can derive the mean radius of the cluster within the rock-physics model in terms of p_t and w_t. Let n_l be the number of cells, with the radius of $R(1 + \chi_c)$, needed to uniformly coat a cluster with radius of r_t, where $n_l = 4\pi r_t^2(1 - \phi_c)/(\pi R^2(1 + \chi_c)^2)$, we can obtain

$$\frac{1}{2R(1 + \chi_c)}\frac{dr_t}{dt} = \frac{1}{n_l M_c}\frac{dN_c}{dt}. \tag{A5}$$

To understand the rate dN_c/dt at which cells in clusters are increasing, we consider the rate at which $w_t = N_d/N_T$ is changing. A change dN_T occurs whenever new dispersed cells are created (presumably, this is occurring to some degree); a change dN_d occurs both as a loss $-dN_c$ to clusters and as a gain dN_T from the newly created cells. From the definition of the derivative, we have

$$d\left(\frac{N_d}{N_T}\right) = \frac{N_d - dN_c + dN_T}{N_T + dN_T} - \frac{N_d}{N_T}. \tag{A6}$$

After rearranging and ignoring products of infinitesimals, we obtain

$$-dN_c = N_T\left(dw_t - (1 - w_t)\frac{dN_T}{N_T}\right) \approx N_T dw_t. \tag{A7}$$

The second term in equation (A7) can be neglected at early times where w_t is close to one. At later time, it is expected that the rate at which new dispersed cells is forming is much smaller than that at which dispersed cells are attaching to clusters. Equation (A7) should thus be a reasonable approximation at all the time.

[52] We can obtain the following differential equation that relates r_t to w_t from equations (A5)–(A7).

$$\frac{1}{r_t}\frac{dr_t}{dt} = \frac{1}{2(1 - w_t)}\frac{d(1 - w_t)}{dt}. \tag{A8}$$

By solving the equation, we have $r_t = r_{max}\sqrt{1 - w_t}$, where r_{max} is the maximum cluster size that occurs when all dispersed cells have been deposited on clusters ($w_t = 0$) and it depends on N_T and other parameters that are not precisely known. Our final result is given by $r_t = \theta_5 l_0 \sqrt{1 - w_t}$, where l_0 is the characteristic pore-throat radius in the system and has a value of 1.3×10^{-4} m and θ_5 is a time-independent parameter that will be determined as part of the inversion with allowed values in the range between 0.2 and 0.9.

Appendix B: Bayesian Model for the Column Experiment Data

[53] The joint posterior distribution function in equation (10) combines information from the evolution model in equation (7) and petrophysical models in equations (8)

and (9). On the basis of equation (5), we can write the joint pdf as follows.

$$f(p_1, p_2, \cdots p_t, w_1, w_2, \cdots, w_n, \theta_1, \theta_2, \cdots,$$
$$\theta_5 | m_1^{obs}, m_2^{obs}, \cdots, m_n^{obs}, \tau_1^{obs}, \tau_2^{obs}, \cdots, \tau_n^{obs})$$
$$\propto f(\theta_1, \theta_2, \cdots, \theta_5)f(w_1, w_2, \cdots, w_n)\prod_{t=1}^{n} f(p_t | p_{t-1}, \theta_1, \theta_2)$$
$$\cdot \prod_{t=1}^{n} f(m_t^{obs} | p_t, w_t, \theta_3, \theta_4)\prod_{t=1}^{n} f(\tau_t^{obs} | w_t, \theta_5), \tag{B1}$$

where $p_0 = 0$ and $\theta_1, \theta_2, \ldots, \theta_5$ are assumed to be uniformly distributed on given ranges.

[54] We first derive conditional probability distribution $f(p_t | p_{t-1}, \theta_1, \theta_2)$ from the normal distribution $N(0, \beta_1)$ using variable transformations, which is given by

$$f(p_t | p_{t-1}, \theta_1, \theta_2)$$
$$\propto \left(\frac{1}{B(z_t, \theta_1, \theta_2)}\right)$$
$$\cdot \left(\frac{1}{\beta_1} \exp\left\{-\frac{1}{2\beta_1^2}\left(\frac{(p_t - p_{t-1}) - B(z_t, \theta_1, \theta_2)}{B(z_t, \theta_1, \theta_2)}\right)^2\right\}\right) \tag{B2}$$

Similarly, we can obtain likelihood functions of chargeability using variable transformations from the normal distribution as below.

$$f(m_t^{obs} | p_t, w_t, \theta_3, \theta_4)$$
$$\propto \frac{1}{\beta_2} \exp\left\{-\frac{1}{2\beta_2^2}\left(\frac{m_t^{obs} - G_1(p_t, w_t, \theta_3, \theta_4)}{m_t^{obs}}\right)^2\right\}, \tag{B3}$$

and

$$f(\tau_t^{obs} | w_t, \theta_5) \propto \frac{1}{\beta_3} \exp\left\{-\frac{1}{2\beta_3^2}\left(\frac{\tau_t^{obs} - G_2(w_t, \theta_5)}{\tau_t^{obs}}\right)^2\right\}. \tag{B4}$$

Combing equations (B1) to (B4), we can obtain the joint posterior distribution given in equation (10).

Appendix C: Sampling Methods

[55] We group unknown parameters in equation (10) into five subsets: (1) $\{\theta_1, \theta_2\}$, parameters related to the geochemical model $B(z_t)$, (2) $\{\theta_3, \theta_4\}$, parameters related to normalized chargeability, (3) θ_5, a parameter related to time constant, (4) $\{p_1, p_2, \ldots, p_n\}$, the volume fraction of metal precipitates, and (5) $\{w_1, w_2, \ldots, w_n\}$, the partition factors of dispersed biominerals. We use block-sampling methods [*Chen et al.*, 2006] to obtain many samples from the joint posterior distribution function given in equation (10). The conditionals for those subsets are given below:

$$f(\theta_1, \theta_2 | \cdot) \propto Ind((\theta_1, \theta_2) \in D_1)\prod_{t=1}^{n}\left(\frac{1}{B(z_t, \theta_1, \theta_2)}\right)$$
$$\cdot \exp\left\{-\frac{1}{2\beta_1^2}\sum_{t=1}^{n}\left(\frac{(p_t - p_{t-1}) - B(z_t, \theta_1, \theta_2)}{B(z_t, \theta_1, \theta_2)}\right)^2\right\}, \tag{C1}$$

$$f(\theta_3, \theta_4|\cdot) \propto Ind((\theta_3, \theta_4) \in D_2)$$

$$\cdot \exp\left\{-\frac{1}{2\beta_2^2}\sum_{t=1}^{n}\left(\frac{m_t^{obs} - G_1(p_t, w_t, \theta_3, \theta_4)}{m_t^{obs}}\right)^2\right\}, \tag{C2}$$

$$f(\theta_5|\cdot) \propto Ind(\theta_5 \in D_3)\exp\left\{-\frac{1}{2\beta_3^2}\sum_{t=1}^{n}\left(\frac{\tau_t^{obs} - G_2(w_t, \theta_5)}{\tau_t^{obs}}\right)^2\right\}, \tag{C3}$$

$$f(p_1, p_2\cdots, p_n|\cdot)$$

$$\propto \exp\left\{-\frac{1}{2\beta_1^2}\sum_{t=1}^{n}\left(\frac{(p_t - p_{t-1}) - B(z_t, \theta_1, \theta_2)}{B(z_t, \theta_1, \theta_2)}\right)^2\right\}$$

$$\cdot \exp\left\{-\frac{1}{2\beta_2^2}\sum_{t=1}^{n}\left(\frac{m_t^{obs} - G_1(p_t, w_t, \theta_3, \theta_4)}{m_t^{obs}}\right)^2\right\}, \tag{C4}$$

and

$$f(w_1, w_2\cdots, w_n|\cdot) \propto \prod_{t=1}^{n}Ind(w_t \in D_4)$$

$$\cdot \exp\left\{-\frac{1}{2\beta_2^2}\sum_{t=1}^{n}\left(\frac{m_t^{obs} - G_1(p_t, w_t, \theta_3, \theta_4)}{m_t^{obs}}\right)^2\right\}$$

$$\cdot \exp\left\{-\frac{1}{2\beta_3^2}\sum_{t=1}^{n}\left(\frac{\tau_t^{obs} - G_2(w_t, \theta_5)}{\tau_t^{obs}}\right)^2\right\}. \tag{C5}$$

References

Atekwana, E. A., D. D. Werkema, and E. A. Atekwana (2006), Biogeophysics: The effects of microbial processes on geophysical properties of the shallow subsurface, in *Applied Geophysics, NATO Science Series: IV. Earth and Environmental Sciences*, vol. 71, edited by H. Vereecken et al., pp. 161–194, Springer, New York.

Binley, A., L. Slater, M. Fukes, and G. Cassiani (2005), The relationship between frequency dependent electrical conductivity and hydraulic properties of saturated and unsaturated sandstone, *Water Resour. Res.*, *41*, W12417, doi:10.1029/2005WR004202.

Chapelle, F. H. (2001), *Ground-Water Microbiology and Geochemistry*, John Wiley, Hoboken, N. J.

Chelidze, T. L., and Y. Gueguen (1999), Electrical spectroscopy of porous rocks: A review. I: Theoretical models, *Geophys. J. Int.*, *137*, 1–15.

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*(6), W06425, doi:10.1029/2005WR004694.

Chen, J., A. Kemna, and S. S. Hubbard (2008), A comparison between Gauss-Newton and Markov Chain Monte Carlo based methods for inverting spectral induced polarization data for Cole-Cole parameters, *Geophysics*, *73*(6), F247–F259.

Cole, K. S., and R. H. Cole (1941), Dispersion and absorption in dielectrics. I: Alternating current field, *J. Chem. Phys.*, *1*, 341–351.

Day-Lewis, F. D., J. M. Harris, and S. M. Gorelick (2002), Time-lapse inversion of crosswell radar data, *Geophysics*, *67*(6), 1740–1752.

Hazen, T. C., and H. H. Tabak (2005), Developments in bioremediation of soils and sediments polluted with metals and radionuclides: 2. Field research on bioremediation of metals and radionuclides, *Rev. Environ. Sci. Biotechnol.*, *4*(3), 157–183.

Hubbard, S., K. Williams, M. Conrad, B. Faybishenko, J. Peterson, J. Chen, P. Long, and T. Hazen (2008), Geophysical monitoring of hydrological and biogeochemical transformations associated with Cr (VI) bioremediation, *Environ. Sci. Technol.*, *42*, 3757–3765.

Lane, J. W., F. D. Day-Lewis, and C. C. Casey (2006), Geophysical monitoring of a field-scale biostimulation pilot project, *Ground Water*, *44*(3), 430–443.

Leroy, P., A. Revil, A. Kemna, P. Cosenza, and A. Ghorbani (2008), Complex conductivity of water-saturated packs of glass beads, *J. Colloid Interface Sci.*, *321*(1), 103–117.

Lovley, D. R., F. H. Chapelle, and J. C. Woodward (1994), Use of dissolved H_2 concentrations to determine distribution of microbially catalyzed redox reactions in anoxic groundwater, *Environ. Sci. Technol.*, *28*, 1205–1210.

Moreau, J. W., R. I. Webb, and J. F. Banfield (2004), Ultra-structure, aggregation-state, and crystal growth of biogenic nano-crystalline sphalerite and wurtzite, *Am. Mineral.*, *89*, 950–960.

Ntarlagiannis, D., K. H. Williams, L. Slater, and S. Hubbard (2005), Low-frequency electrical response to microbial induced sulfide precipitation, *J. Geophys. Res.*, *110*, G02009, doi:10.1029/2005JG000024.

Olhoeft, G. R. (1985), Low-frequency electrical properties, *Geophysics*, *50*, 2492–2503.

Pelton, W. H., S. H. Ward, P. G. Hallof, W. R. Sill, and P. H. Nelson (1978), Mineral discrimination and removal of inductive coupling with multi-frequency IP, *Geophysics*, *43*, 588–609.

Pelton, W. H., W. R. Sill, and B. D. Smith (1983), Interpretation of complex resistivity and dielectric data: Part I, *Geophys. Trans.*, *29*, 297–330.

Pride, S. R., J. G. Berryman, and J. M. Harris (2004), Seismic attenuation due to wave-induced flow, *J. Geophys. Res.*, *109*, B01201, doi:10.1029/2003JB002639.

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

Schwartz, G. (1962), A theory of the low-frequency dielectric dispersion of colloidal particles in electrolyte solution, *J. Phys. Chem.*, *66*, 2636–2642.

Scott, J. B. T., and R. D. Barker (2005), Characterization of sandstone by electrical spectroscopy for stratigraphical and hydrogeological investigations, *Q. J. Eng. Geol. Hydrogeol.*, *38*, 143–154.

Shumway, R. H., and D. S. Stoffer (2000), *Time Series Analysis and Its Applications*, Springer, New York.

Slater, L., D. Ntarlagiannis, and D. Wishart (2006), On the relationship between induced polarization and surface area in metal-sand and clay-sand mixture, *Geophysics*, *71*(2), A1–A5.

Slater, L., D. Ntarlagiannis, Y. R. Personna, and S. Hubbard (2007), Pore-scale spectral induced polarization signatures associated with FeS bio-mineral transformations, *Geophys. Res. Lett.*, *34*, L21404, doi:10.1029/2007GL031840.

Steefel, C. I. (2008), CrunchFlow: Software for Modeling Multicomponent Reactive Flow and Transport, User's Manual, Lawrence Berkeley Nat. Lab., Berkeley, Calif.

Steefel, C. I., and K. Maher (2009), Fluid-rock interaction: A reactive transport approach, *Rev. Mineral. Geochem.*, *70*, 485–532.

Stone, C. J. (1996), *A Course in Probability and Statistics*, Duxbury, New York.

Tarasov, A., and T. Titov (2007), Relaxation time distribution from time domain induced polarization measurements, *Geophys. J. Int.*, *170*, 31–34.

Thompson, A. H., A. J. Katz, and C. E. Krohn (1987), The micro-geometry and transport properties of sedimentary rock, *Adv. Phys.*, *36*, 625–694.

Vasco, D. W., A. Datta-Gupta, R. Behrens, P. Condon, and J. Rickett (2004), Seismic imaging of reservoir flow properties: Time-lapse amplitude changes, *Geophysics*, *69*(6), 1425–1442.

Williams, K., D. Ntarlagiannis, L. Slater, A. Dohnalkova, S. Hubbard, and J. Banfield (2005), Geophysical imaging of stimulated microbial bioremediation, *Environ. Sci. Technol.*, *39*(19), 7592–7600.

Wong, J. (1979), An electrochemical model of the induced-polarization phenomenon in disseminated sulfide ores, *Geophysics*, *44*(7), 1245–1265.

Xu, T., E. Sonnenthal, and G. Bodvarsson (2003), A reaction-transport model for calcite precipitation and evaluation of infiltration-percolation fluxes in unsaturated fractured rock, *J. Contam. Hydrol.*, *64*, 113–127.

––––––––––––––––––

J. Chen, S. S. Hubbard, L. Li, S. Pride, C. Steefel, and K. H. Williams, Earth Science Division, Lawrence Berkeley Laboratory, MS 90-1116, One Cyclotron Road, Berkeley, CA 94720, USA. (jchen@lbl.gov)

L. Slater, Department of Earth and Environmental Sciences, Rutgers University, 101 Warren Street, Smith Hall, Room 136, Newark, NJ 07102, USA.