Joint Inversion of Marine Seismic AVA and CSEM Data Using Statistical Rock-physics Models and Markov Random Fields

Stochastic inversion of AVA and CSEM data

Jinsong Chen[1] and Michael Hoversten[2]

1: Earth Sciences Division, Lawrence Berkeley National Laboratory, USA;

Email: jchen@lbl.gov

2: Chevron Energy Technology Company, San Ramon, California, USA

Email: hovg@chevron.com

ABSTRACT

Joint inversion of seismic AVA and CSEM data requires rock-physics relationships to link seismic attributes to electrical properties. Ideally, we can connect them through reservoir parameters (e.g., porosity and water saturation) by developing physical-based models, such as Gassmann's equations and Archie's law, using nearby borehole logs. This could be difficult in the exploration stage because information available is typically insufficient for choosing suitable rock-physics models and for subsequently obtaining reliable estimates of the associated parameters. The use of improper rock-physics models and the inaccuracy of the estimates of model parameters may cause misleading inversion results. Conversely, it is easy to derive statistical relationships among seismic and electrical attributes and reservoir parameters from distant borehole logs. In this study, we develop a Bayesian model to jointly invert seismic AVA and CSEM data for reservoir parameter estimation using statistical rock-physics models; the spatial dependence of geophysical and reservoir parameters are carried out by lithotypes through Markov random fields. We apply the developed model to a synthetic case, which simulates a CO2 monitoring application. We derive statistical rock-physics relations from borehole logs at one location and estimate seismic P- and S-wave velocity ratio, acoustic impedance, density, electrical resistivity, lithotypes, porosity, and water saturation at three different locations by conditioning to seismic AVA and CSEM data. Comparison of the inversion results with their corresponding true values shows that the correlation-based statistical rock-physics models provide significant information for improving the joint inversion results.

INTRODUCTION

Joint inversion of seismic AVA and CSEM data has been demonstrated to be beneficial for reservoir parameter estimation (Hoversten et al., 2006; Chen et al., 2007; Chen and Dickens, 2009) because they provide complementary information on reservoir parameters. This estimation requires rock-physics relationships to connect geophysical attributes to reservoir parameters, which are often derived from suitable nearby borehole logs in practice. An appropriate family of rock-physics relationships is first chosen and the parameters associated with the rock-physics relationships are then estimated by fitting them to the selected borehole logs.

However, the derivation of detailed rock-physics relationships in the exploration stage could be very difficult due to the lack of nearby wells. On the other hand, we can often find logging data at more distant locations where regression-based relations can be derived sufficiently. The statistical-based relationships, with associated uncertainty, may improve the estimates of geophysical attributes and thus reservoir parameters even though they are less precise than the physical-based rock-physics models.

Conventional methods for reservoir exploration follow a three-step approach. Firstly, seismic AVA data are inverted for seismic attributes (e.g., seismic P-wave and S-wave velocity and density), and CSEM data are inverted for electrical resistivity. Secondly, physical-based rock-physics models are derived from borehole logs, in which CO2, oil or gas reservoirs are classified as clusters or lithotypes. Finally, the estimated geophysical attributes are transferred into the lithotype images based on the rock-physics models. The main limitation of the approach is that the joint use of seismic AVA and CSEM data does not allow for information sharing across the two different types of datasets. Joint inversion approaches have been developed by many researchers, for example, cross-gradient methods as used by Gallardo and Meju (2004),

and Fregoso and Gallardo (2009). One limitation of the cross-gradient based methods is that it does not make use of information from nearby boreholes.

In this study, we develop a Bayesian method to jointly invert seismic AVA and CSEM data by using correlation-based rock-physics relationships, referred to as 'statistical rock-physics models'. Unlike conventional methods, where seismic and electrical properties are linked through physical-based rock-physics models such as Gassmann's equations (Gassmann, 1951), statistical rock-physics models are regression-based models derived from borehole logs. We assume that such rock-physics models are dependent on lithotypes or clusters obtained from borehole logs and use Markov random fields (Swendsen and Wang, 1987) to describe the spatial correlation of lithotypes. We use Markov chain Monte Carlo (MCMC) (Gilks et al., 1996) methods to draw many samples from the joint posterior distribution for the estimation.

We apply the developed Bayesian model to a synthetic dataset generated from a CO2 monitoring application. We use borehole logs at one location to derive statistical rock-physics relations and invert seismic AVA and CSEM data at three different locations: one is close to the training borehole and others are far away from the borehole. We compare the inversion results with the corresponding true values to demonstrate the effectiveness of the developed Bayesian models and the usefulness of the derived statistical rock-physics models.

The remaining sections are arranged as follows. Section 2 describes the inverse problem and Bayesian model for joint inversion of seismic AVA and CSEM data. Section 3 provides basic information on the case study, including the 2D estimation cross-section, statistical rock-physics models, and synthetic data. Section 4 gives inversion results and discussion. Conclusions are given in Section 5.

METHODOLOGY

In this section, we describe a Bayesian model to invert seismic AVA and CSEM data for reservoir parameter estimation based on typical situations in geophysical exploration, where derivation of physical-based rock-physics models for linking seismic and electrical properties to reservoir parameters is difficult. Instead, statistical relationships among geophysical attributes and reservoir parameters can be obtained rapidly from borehole data.

Parameterization and Bayesian model

Although we develop a Bayesian model for estimating geophysical attributes, reservoir parameters, and lithotypes along depth (or 1D) using seismic AVA and CSEM data in the current study, the model can be extended to a 2D cross-section later. The 1D model is similar to the layered reservoir as used by Hoversten et al. (2006) and Chen et al. (2007). The main difference between this and the previous studies are (1) that we use statistical rock-physics models derived from borehole logs, (2) that we consider lithotypes as unknown variables, and (3) that for seismic attributes, we estimate the ratio of P-wave to S-wave velocity, acoustic impedance, and density, rather than seismic P-wave and S-wave velocity, and density.

Let vectors $\boldsymbol{\alpha}$, $\mathbf{Z}$, $\boldsymbol{\rho}$, $\mathbf{r}$, and $\mathbf{L}$ represent the ratio of seismic P-wave to S-wave velocity, acoustic impedance, density, logarithmic resistivity, and lithotypes in the target layers. The lithotypes are categorical variables and can be two (e.g., CO2 sand and non-CO2 sand and shale) or three (e.g., CO2 sand, shale and brine sand). Let vectors $\boldsymbol{\varphi}$ and $\mathbf{S}$ be the logarithmic porosity and water saturation in those layers. Similar to the model given by Chen et al. (2007), we add several layers above and below the target layers to account for uncertainty in selecting the time window for seismic AVA data inversion. For those layers, we directly invert for the P- and S-

wave velocity ratio ($\boldsymbol{\alpha}_0$), acoustic impedance ($\mathbf{Z}_0$), and density ($\boldsymbol{\rho}_0$). Because the resistivity in the seawater and in the overburden and bedrock affects the estimates of reservoir parameters, we also consider them as unknown parameters in this model and denoted by the vector $\mathbf{r}_0$.

Since seismic and EM data are often subject to unknown random errors, we consider them as unknown parameters. Let $\boldsymbol{\theta}_s = (\theta_{s1}, \theta_{s2})^T$, where θ_{s1} and θ_{s2} are the signal-to-noise ratios (S/N) of near and far offsets of seismic data. Let $\boldsymbol{\theta}_e = (\theta_{e1}, \theta_{e2})^T$, where θ_{e1} and θ_{e2} are the relative errors of EM data from near to far offsets. For those offsets between the nearest and furthest offsets, the signal-to-noise ratios of seismic data and the relative errors of EM data are linear interpolation of the values at the two ends.

Let matrix $\mathbf{A}$ represent seismic AVA data, which are some explicit functions of seismic P- and S-wave velocity and density within and outside the target zone. Let matrix $\mathbf{E}$ represent CSEM data, which are some functions of resistivity in the reservoir, seawater, overburden, and bedrock. Since seismic AVA and CSEM data are two different types of geophysical measurements, we assume their random errors are independent of each other. Consequently, we obtain the following Bayesian model:

$$\begin{aligned} & f(\boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \boldsymbol{\alpha}_0, \mathbf{Z}_0, \boldsymbol{\rho}_0, \mathbf{r}, \mathbf{r}_0, \boldsymbol{\varphi}, \mathbf{S}, \mathbf{L}, \boldsymbol{\theta}_s, \boldsymbol{\theta}_e \mid \mathbf{A}, \mathbf{E}) \\ & \quad \propto f(\mathbf{A} \mid \boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \boldsymbol{\alpha}_0, \mathbf{Z}_0, \boldsymbol{\rho}_0, \boldsymbol{\theta}_s) \times f(\mathbf{E} \mid \mathbf{r}, \mathbf{r}_0, \boldsymbol{\theta}_e) \\ & \quad \times f(\boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \boldsymbol{\alpha}_0, \mathbf{Z}_0, \boldsymbol{\rho}_0, \mathbf{r}, \mathbf{r}_0, \boldsymbol{\varphi}, \mathbf{S}, \mathbf{L}, \boldsymbol{\theta}_s, \boldsymbol{\theta}_e) \end{aligned} \tag{1}$$

Equation 1 defines a joint posterior probability distribution function of all unknown parameters and it is known up to a normalizing constant. The first and second terms on the right side of the equation are the likelihood functions of seismic AVA and CSEM data, respectively. The third term on the right side of the equation is the prior distribution of unknown parameters.

Likelihood model of seismic AVA and CSEM data

Seismic AVA data are post-stacked and NMO-removed full waveforms for different incident angles (Yilmaz, 1987). Let $\mathbf{A} = (a_{11}, a_{21}, \cdots, a_{m1}, a_{12}, a_{22}, \cdots, a_{m2}, \cdots, a_{1n}, a_{2n}, \cdots, a_{mn})^T$, where m denotes the total number of time samples, n denotes the total number of incident angles, and T denotes the transpose of vectors or matrices. Let M_{ij}^s be the forward model of seismic data at the *ith* time and the *jth* incident angle and ε_{ij}^a be the corresponding random error. We have

$$a_{ij} = M_{ij}^a(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{\rho}, \boldsymbol{\alpha}_0, \mathbf{z}_0, \boldsymbol{\rho}_0) + \varepsilon_{ij}^a. \tag{2}$$

Let $\boldsymbol{\varepsilon} = (\varepsilon_{11}^a, \varepsilon_{21}^a, \cdots, \varepsilon_{m1}^a, \varepsilon_{12}^a, \varepsilon_{22}^a \cdots, \varepsilon_{m2}^a, \cdots, \varepsilon_{mn}^a)^T$ be a vector representing all the random errors. To consider correlation of those errors in time and across incident angles, we assume that they have the multivariate Gaussian distribution with zero mean and covariance matrix $\boldsymbol{\Sigma}$, as used by Buland and Omre (2003). Thus, we obtain the following likelihood function for the seismic data:

$$f(\mathbf{A} \mid \boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{\rho}, \boldsymbol{\alpha}_0, \mathbf{z}_0, \boldsymbol{\rho}_0, \boldsymbol{\theta}_s) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \right), \tag{3}$$

where $|\boldsymbol{\Sigma}|$ and $\boldsymbol{\Sigma}^{-1}$ denote the determinant and the inverse of the covariance matrix, respectively, which are functions of $\boldsymbol{\theta}_s$. If the error structure is non-Gaussian and can be modeled, then other appropriate likelihood functions should be used.

We can model the joint likelihood function of random errors in CSEM data using the approach in Chen et al. (2007). Let matrix $\mathbf{E} = \{e_{ijk}\}$, where $i = 1, 2, \cdots, n_f$, representing different frequencies of CSEM sources, $j = 1, 2, \cdots, n_0$, representing different offsets, and $k = 1, 2$, representing real and quadrature components of the recorded electrical fields. Unlike seismic

AVA data, we use relative errors for CSEM data because the amplitudes of EM data differ by several orders of magnitude for different frequencies and offsets. We assume that the relative ratio ε_{ijk}^{r} has the Gaussian distribution with zero mean and the standard deviation γ_j that depends on offsets and is linearly interpolated from the values at the near (θ_{e1}) and the far (θ_{e2}) offsets. As a result, we obtain the following likelihood function:

$$f(\mathbf{E}\,|\,\mathbf{r},\mathbf{r}_0,\boldsymbol{\theta}_e)=\prod_{i=1}^{n_f}\prod_{j=1}^{n_o}\prod_{k=1}^{2}\frac{1}{\sqrt{2\pi\gamma_j^2}}\exp\left\{-\frac{1}{2\gamma_j^2}\left(\frac{e_{ijk}-M_{ijk}^{e}(\mathbf{r},\mathbf{r}_0)}{e_{ijk}}\right)^2\right\}, \tag{4}$$

where M_{ijk}^{e} represents the real and quadrature components of the simulated electrical fields using a forward numerical model. Like the seismic AVA data, we should use other appropriate likelihood functions if the error structure of CSEM data is non-Gaussian.

Prior models, statistical rock-physics models, and Markov random fields

The prior probability distribution in Equation 1 should be determined from information that is not included in seismic and EM data, for example, borehole logs. It is reasonable to assume that the unknown parameters outside the target zone are independent of the parameters within the zone. Therefore, we can write the prior distribution as follows:

$$\begin{aligned} f(\boldsymbol{\alpha},\mathbf{Z},\boldsymbol{\rho},\boldsymbol{\alpha}_0,\mathbf{Z}_0,\boldsymbol{\rho}_0,\mathbf{r},\mathbf{r}_0,\boldsymbol{\varphi},\mathbf{S},\mathbf{L},\boldsymbol{\theta}_s,\boldsymbol{\theta}_e) &= f(\boldsymbol{\theta}_s)\times f(\boldsymbol{\theta}_e) \\ &\times f(\boldsymbol{\alpha}_0,\mathbf{Z}_0,\boldsymbol{\rho}_0,\mathbf{r}_0)\times f(\boldsymbol{\alpha},\mathbf{Z},\boldsymbol{\rho},\mathbf{r},\boldsymbol{\varphi},\mathbf{S}\,|\,\mathbf{L})\times f(\mathbf{L}). \end{aligned} \tag{5}$$

We assume that the first three terms on the right side of Equation 5 have uniform distribution on given ranges, as they are not our focuses. We can derive the fourth term from borehole logs and describe lithotypes using Markov random fields.

We simplify $f(\boldsymbol{\alpha},\mathbf{Z},\boldsymbol{\rho},\mathbf{r},\boldsymbol{\varphi},\mathbf{S}\,|\,\mathbf{L})$ by assuming seismic and electrical properties and reservoir parameters at each layer are conditionally independent given lithotypes. This means

that given lithotypes at two adjacent layers, we can assume the unknown parameters in the two layers are independent of each other because both geophysical attributes and reservoir parameters depend on lithotypes. Let α_i, Z_i, ρ_i, r_i, φ_i, S_i, and L_i be the P- and S-wave velocity ratio, acoustic impedance, density, resistivity, porosity, water saturation, and lithotype at the *ith* layer, we have

$$f(\boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \mathbf{r}, \boldsymbol{\varphi}, \mathbf{S} \mid \mathbf{L}) = \prod_{i=1}^{n} f(\alpha_i, Z_i, \rho_i, r_i, \varphi_i, S_i \mid L_i). \tag{6}$$

We assume that the conditional distribution on the right side of Equation 6 has a multivariate Gaussian distribution, i.e., $f(\alpha_i, Z_i, \rho_i, r_i, \varphi_i, S_i \mid L_i) \square N(\boldsymbol{\mu}(L_i), \boldsymbol{\Sigma}(L_i))$, where both mean vector $\boldsymbol{\mu}(L_i)$ and covariance matrix $\boldsymbol{\Sigma}(L_i)$ are functions of lithotypes and are derived from borehole logs.

We use the first-order Markov random field to simulate the spatial dependence of lithotypes as done by Eidsvik et al. (2004). Although in this study, we only estimate seismic and electrical parameters along a 1D profile, the developed model can be extended to 2D cross-sections. The Markov random field is the Ising and Potts model. Let i and j be the indices at two different reservoir layers. As shown by Higdon (1998), we can write the joint conditional distribution of lithotypes given the spatial correlation parameter as:

$$f(\mathbf{L}) \propto \exp\left\{ \sum_{i \square j} \beta \ \mathrm{Ind}[L_i = L_j] \right\}. \tag{7}$$

In Equation 7, Ind[ ] denotes the indicator function, taking the value of 1 if $L_i = L_j$ and the value of 0, otherwise. Parameter β is a coefficient that determines the spatial dependence of lithotypes. Since we only use data along a 1D profile, we cannot get a good estimate of the

spatial dependence parameter. Therefore, we consider it as a fixed number and set $\beta = 1.0$, which is reasonable for the current synthetic study after several tries.

Markov chain Monte Carlo sampling methods

We use Markov Chain Monte Carlo methods to obtain many samples of the unknown parameters. We adapt different sampling strategies for geophysical attributes ($\boldsymbol{\alpha}$, $\mathbf{Z}$, $\boldsymbol{\rho}$, $\mathbf{r}$, $\boldsymbol{\alpha}_0$, $\mathbf{Z}_0$, $\boldsymbol{\rho}_0$, and $\mathbf{r}_0$), reservoir parameters ($\boldsymbol{\varphi}$ and $\mathbf{S}$), lithotypes ($\mathbf{L}$), and other parameters ($\boldsymbol{\theta}_s$ and $\boldsymbol{\theta}_e$) to achieve good convergence (Tierney, 1994).

Sampling of geophysical attributes and error-related parameters

We first need to derive the conditional distributions of the unknown geophysical attributes before sampling. Since both seismic AVA and EM data are nonlinear functions of the geophysical parameters, we cannot obtain analytical forms of their full conditional distributions. We primarily use block-sampling strategies. For completeness, we list them as follows:

$$
\begin{aligned}
&f(\boldsymbol{\alpha}_0, \mathbf{Z}_0, \boldsymbol{\rho}_0 \mid \cdot) \propto f(\mathbf{A} \mid \boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \boldsymbol{\alpha}_0, \mathbf{Z}_0, \boldsymbol{\rho}_0, \boldsymbol{\theta}_s) \times f(\boldsymbol{\alpha}_0, \mathbf{Z}_0, \boldsymbol{\rho}_0),\\
&f(\boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho} \mid \cdot) \propto f(\mathbf{A} \mid \boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \boldsymbol{\alpha}_0, \mathbf{Z}_0, \boldsymbol{\rho}_0, \boldsymbol{\theta}_s) \times f(\boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \mathbf{r}, \boldsymbol{\varphi}, \mathbf{S} \mid \mathbf{L}),\\
&f(\boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \boldsymbol{\alpha}_0, \mathbf{Z}_0, \boldsymbol{\rho}_0 \mid \cdot) \propto f(\mathbf{A} \mid \boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \boldsymbol{\alpha}_0, \mathbf{Z}_0, \boldsymbol{\rho}_0, \boldsymbol{\theta}_s)\\
&\qquad\qquad \times f(\boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \mathbf{r}, \boldsymbol{\varphi}, \mathbf{S} \mid \mathbf{L}) \times f(\boldsymbol{\alpha}_0, \mathbf{Z}_0, \boldsymbol{\rho}_0),\\
&f(\mathbf{r}_0 \mid \cdot) \propto f(\mathbf{E} \mid \mathbf{r}, \mathbf{r}_0, \boldsymbol{\theta}_e) \times f(\mathbf{r}_0),\\
&f(\mathbf{r} \mid \cdot) \propto f(\mathbf{E} \mid \mathbf{r}, \mathbf{r}_0, \boldsymbol{\theta}_e) \times f(\boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \mathbf{r}, \boldsymbol{\varphi}, \mathbf{S} \mid \mathbf{L}),\\
&f(\mathbf{r}, \mathbf{r}_0 \mid \cdot) \propto f(\mathbf{E} \mid \mathbf{r}, \mathbf{r}_0, \boldsymbol{\theta}_e) \times f(\boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \mathbf{r}, \boldsymbol{\varphi}, \mathbf{S} \mid \mathbf{L}) \times f(\mathbf{r}_0).
\end{aligned}
\tag{8}
$$

We can similarly obtain the conditional probability distributions of the error-related parameters below:

$$
\begin{aligned}
&f(\boldsymbol{\theta}_s \mid \cdot) \propto f(\mathbf{A} \mid \boldsymbol{\alpha}, \mathbf{Z}, \boldsymbol{\rho}, \boldsymbol{\alpha}_0, \mathbf{Z}_0, \boldsymbol{\rho}_0, \boldsymbol{\theta}_s) \times f(\boldsymbol{\theta}_s),\\
&f(\boldsymbol{\theta}_e \mid \cdot) \propto f(\mathbf{E} \mid \mathbf{r}, \mathbf{r}_0, \boldsymbol{\theta}_e) \times f(\boldsymbol{\theta}_e).
\end{aligned}
\tag{9}
$$

We use a hybrid sampling strategy as used by Chen et al. (2007). The sampling methods include: (1) single variable Metropolis-Hastings methods (SMH) (Metropolis et al., 1954; Hastings et al., 1970), (2) multivariate Metropolis-Hastings methods (MMH), (3) single variable slice sampling methods (SSS) (Neal, 2003), (4) multivariate slice sampling methods (MSS), and (5) adaptive Metropolis sampling methods (AMH) (Haario et al., 2001).

Sampling of lithotypes

Since lithotypes play a key role in the Bayesian model, we need to develop an efficient method to sample them. From Equations 1, 5, and 7, we can obtain the conditional probability distribution of lithotypes given all other information as follows:

$$f(\mathbf{L}\,|\,\cdot) \propto \exp\left\{\sum_{i \sim j} \beta\ \mathrm{Ind}[L_i = L_j]\right\} \prod_{i=1}^{n} f(\alpha_i, Z_i, \rho_i, r_i, \varphi_i, S_i\,|\,L_i). \tag{10}$$

Drawing samples from the joint distribution given in Equation 10 typically suffers from low convergence using Gibbs or traditional Metropolis-Hasting methods only (Swendsen and Wang, 1987). To treat such slow convergence, in this study, we also use a hybrid sampling strategy, which includes the Swendsen-Wang algorithm (Swendsen and Wang, 1987) and the Wolf algorithm (Wolf, 1989). Both Swendsen-Wang and Wolf algorithms are cluster based methods for the sampling, which have been generalized to become an efficient sampling strategy for complex posterior distributions (Smith and Smith, 2006; Barbu and Zhu, 2007).

Sampling of reservoir parameters

We use a Gibbs sampler (Geman and Geman, 1984) to draw many samples of porosity and water saturation from the joint posterior distribution. This is because we can derive analytically the full conditional distributions of porosity and water saturation at each layer given

other parameters; the full conditionals are Gaussian distributions and can be directly sampled without any rejection. We assume that porosity (φ_i) and water saturation (S_i) are conditionally independent given lithotype (L_i), i.e., $f(\varphi_i, S_i \mid L_i) = f(\varphi_i \mid L_i) \times f(S_i \mid L_i)$. Thus, we have

$$f(\alpha_i, Z_i, \rho_i, r_i, \varphi_i, S_i \mid L_i) = f(\alpha_i, Z_i, \rho_i, r_i \mid \varphi_i, S_i, L_i) \times f(\varphi_i \mid L_i) \times f(S_i \mid L_i). \quad (11)$$

Since the conditional distribution $f(\alpha_i, Z_i, \rho_i, r_i, \varphi_i, S_i \mid L_i)$ has a multivariate Gaussian distribution, the conditional distribution $f(\alpha_i, Z_i, \rho_i, r_i \mid \varphi_i, S_i, L_i)$ also has a multivariate Gaussian distribution (Stone, 1996). Let $\boldsymbol{\mu}_c(L_i)$ and $\boldsymbol{\Sigma}_c(L_i)$ be the conditional mean vector and covariance matrix of $f(\alpha_i, Z_i, \rho_i, r_i \mid \varphi_i, S_i, L_i)$, both of which can be derived directly from the mean vector $\boldsymbol{\mu}(L_i)$ and $\boldsymbol{\Sigma}(L_i)$. Let $\mu_k(L_i)$, where $k = 1, 2, \cdots, 6$, be the mean of P- and S-wave velocity ratio, acoustic impedance, density, resistivity, porosity, and water saturation of the *ith* lithotype, respectively. Let $\mathbf{g}(L_i) = (\mu_1(L_i), \mu_2(L_i), \mu_3(L_i), \mu_4(L_i))^T$ and let $\mathbf{x}_i = (\alpha_i, Z_i, \rho_i, r_i)^T$. Let vectors $\mathbf{b}_1(L_i)$ and $\mathbf{b}_2(L_i)$ represent two coefficient vectors with the length of four (see Appendix A). Thus, we have:

$$\begin{aligned} &f(\alpha_i, Z_i, \rho_i, r_i \mid \varphi_i, S_i, L_i) \propto \left|\boldsymbol{\Sigma}_c(L_i)\right|^{-1/2} \exp\left\{-0.5(\mathbf{x}_i - \boldsymbol{\mu}_c(L_i))^T \boldsymbol{\Sigma}_c(L_i)^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_c(L_i))\right\}, \\ &\boldsymbol{\mu}_c(L_i) = \mathbf{g}(L_i) + \mathbf{b}_1(L_i)(\varphi_i - \mu_5(L_i)) + \mathbf{b}_2(L_i)(S_i - \mu_6(L_i)). \end{aligned} \quad (12)$$

To obtain the full conditional distribution of porosity (φ_i) and water saturation (S_i), we need their conditional prior distribution given the lithotype (L_i). Based on borehole logs, we assume that $f(\varphi_i \mid L_i) \sim N(\mu_\varphi^0(L_i), \tau_\varphi^0(L_i))$, where $\mu_\varphi^0(L_i)$ is the mean porosity and $\tau_\varphi^0(L_i)$ is the corresponding inverse variance. Similarly, if the lithotype is not shale, we have $f(S_i \mid L_i) \sim N(\mu_s^0(L_i), \tau_s^0(L_i))$, where $\mu_s^0(L_i)$ denotes the mean water saturation and $\tau_s^0(L_i)$ denotes the corresponding inverse variance. However, for shale, since water saturation typically

is very close to one, i.e., the logarithmic water saturation (S_i) is close zero. Therefore, we assume that the negation of S_i has the exponential distribution, i.e., $f(S_i \mid L_i) = e^{S_i/\beta_s} / \beta_s$ for $S_i \le 0$ and zero, otherwise.

From Equation 10, we have (see Appendix B)

$$f(\varphi_i \mid \cdot) \propto f(\alpha_i, Z_i, \rho_i, r_i \mid \varphi_i, S_i, L_i) \times f(\varphi_i \mid L_i) \sim N(\mu_{\varphi_i}, \tau_{\varphi_i}), \tag{13}$$

where

$$\begin{cases} \tau_{\varphi_i} = \tau_\varphi^0(L_i) + \tau_\varphi(L_i), \\ \mu_{\varphi_i} = (\mu_\varphi^0(L_i)\tau_\varphi^0(L_i) + \mu_\varphi(L_i)\tau_\varphi(L_i)) / \tau_{\varphi_i}, \end{cases} \tag{14}$$

and

$$\begin{cases} \tau_\varphi = \mathbf{b}_1^T(L_i)\boldsymbol{\Sigma}_c(L_i)^{-1}\mathbf{b}_1(L_i) \\ \mu_\varphi = \mathbf{d}_1^T(L_i)\boldsymbol{\Sigma}_c(L_i)^{-1}\mathbf{b}_1(L_i) \\ \mathbf{d}_1(L_i) = \mathbf{x}_i - \mathbf{g}(L_i) + \mathbf{b}_1(L_i)\mu_5(L_i) - \mathbf{b}_2(L_i)(S_i - \mu_5(L_i)) \end{cases} \tag{15}$$

We can similarly obtain the full conditional distribution of water saturation as follows:

$$f(S_i \mid \cdot) \propto f(\alpha_i, Z_i, \rho_i, r_i \mid \varphi_i, S_i, L_i) \times f(S_i \mid L_i) \sim N(\mu_{S_i}, \tau_{S_i}). \tag{16}$$

If the lithotype is not shale, we have

$$\begin{cases} \tau_{S_i} = \tau_S^0(L_i) + \tau_S(L_i), \\ \mu_{S_i} = (\mu_S^0(L_i)\tau_S^0(L_i) + \mu_S(L_i)\tau_S(L_i)) / \tau_{S_i}. \end{cases} \tag{17}$$

If the lithotype is shale, we have

$$\begin{cases} \tau_{S_i} = \tau_S(L_i), \\ \mu_{S_i} = (\mu_S(L_i)\tau_S(L_i) - 0.5/\beta_s) / \tau_{S_i}. \end{cases} \tag{18}$$

where

$$\begin{cases} \tau_{\varphi} = \mathbf{b}_2^T(L_i)\boldsymbol{\Sigma}_c(L_i)^{-1}\mathbf{b}_2(L_i), \\ \mu_{\varphi} = \mathbf{d}_2^T(L_i)\boldsymbol{\Sigma}_c(L_i)^{-1}\mathbf{b}_2(L_i), \\ \mathbf{d}_2(L_i) = \mathbf{x}_i - \mathbf{g}(L_i) - \mathbf{b}_1(L_i)(\varphi_i - \mu_5(L_i)) + \mathbf{b}_2(L_i)\mu_6(L_i). \end{cases} \tag{19}$$

MCMC sampling procedures

We use a hybrid strategy to draw many samples from the conditional probability distributions given in Equations 1-19. The main procedures are summarized as follows:

(1) Assign initial values to all the unknowns and referred them to as $\boldsymbol{\alpha}^{(0)}$, $\mathbf{Z}^{(0)}$, $\boldsymbol{\rho}^{(0)}$, $\boldsymbol{\alpha}_0^{(0)}$, $\mathbf{Z}_0^{(0)}$, $\boldsymbol{\rho}_0^{(0)}$, $\mathbf{r}^{(0)}$, $\mathbf{r}_0^{(0)}$, $\boldsymbol{\varphi}^{(0)}$, $\mathbf{S}^{(0)}$, $\boldsymbol{\theta}_s^{(0)}$, $\boldsymbol{\theta}_e^{(0)}$, and $\mathbf{L}^{(0)}$, respectively. Set $k = 0$.

(2) Update geophysical attributes by first randomly picking one of the five sampling methods (i.e., SMH, MMH, SSS, MSS, and AMH) with a pre-set probability and then randomly picking a subset of unknowns given in Equation 8 for updating. Refer those samples to as $\boldsymbol{\alpha}^{(k)}$, $\mathbf{Z}^{(k)}$, $\boldsymbol{\rho}^{(k)}$, $\boldsymbol{\alpha}_0^{(k)}$, $\mathbf{Z}_0^{(k)}$, $\boldsymbol{\rho}_0^{(k)}$, $\mathbf{r}^{(k)}$, and $\mathbf{r}_0^{(k)}$.

(3) Update error-related parameters by first picking one of the five sampling methods and then drawing samples from the conditional probability distributions from Equation 9. We refer those samples to as $\boldsymbol{\theta}_s^{(k)}$, and $\boldsymbol{\theta}_e^{(k)}$.

(4) Update reservoir parameters using Gibbs samplers and full conditional distributions given in Equations 13 and 16. We refer those samples to as $\boldsymbol{\varphi}^{(k)}$, and $\mathbf{S}^{(k)}$.

(5) Update lithotypes by randomly picking the Swendsen-Wang or the Wolf algorithms with a pre-set probability and refer those samples to as $\mathbf{L}^{(k)}$.

(6) If the pre-set total number of iterations has been reached, stop; otherwise, let $k = k + 1$ and go to Step 2.

The pre-set probabilities for choosing sampling methods and subsets of unknowns are tuning parameters, and they affect the speed of convergence but not the results. We set them being uniform in the preliminary stage and change them after gaining some experience through the preliminary runs. For example, we can update seismic attributes more frequently then electrical properties because the convolution based seismic forward calculation is much faster than the EM forward simulation.

CASE STUDIES

We demonstrate the effectiveness of the developed Bayesian model for parameter estimation by applying it to a synthetic dataset extracted from an elaborately designed 2D model by Chevron Energy Technology Company for inversion algorithm benchmarking. The generated model mimics real scenarios in the Kuito area of Angola with built-in non-stationarity. The model includes several reservoirs embedded in a background shale/sand mixture model. The reservoir properties were taken mostly from particular reservoir distributions of porosity, water saturation, etc. The geophysical attributes (such as seismic and electric properties) were calculated using suitable rock-physics relationships by taking account of deterministic porosity compaction trend. The final model has properties whose horizontal and vertical trends agree with wellbore data in the Kuito area and whose specific fluctuations away from the mean trend approximately agree with the variance observed in the well logs. Figure 1 shows the 2D cross-section of logarithmic resistivity with borehole locations, and the thickness of seawater is 500 m and that of the overburden is approximately 1650 m. A target zone with injected CO2 is located at x = 5500 m.

For the purpose of this study, we assume that there is a borehole at x = 5770 m that passes through the target zone. We use borehole logs collected at this location to derive statistical

rock-physics relationships, which are used later for estimating reservoir parameters at other locations. Specifically, we will estimates reservoir parameters at three locations: (1) x = 5784 m, which is very close to the training borehole; (2) x = 6684 m, which is away from the target zone and near the right edge of the target zone, and (3) x = 4584 m, which is at the left edge of the target zone.

Borehole logs and statistical rock-physics models

Figure 2 shows numerous borehole logs from the training well, including porosity, water saturation, seismic P- and S-wave velocity ratio, seismic P-wave impedance, density, and logarithmic electrical resistivity, between depths 2100 m and 2800 m. There is a reservoir with relatively high porosity, low water saturation and high resistivity at depths between 2300 m and 2600 m. To have a rock-physics relationship for estimating reservoir parameters, we derive site-specific statistical relationships among those geophysical attributes and reservoir parameters. We classify the borehole data into three lithotypes based on shale content and water saturation. We refer to a lithotype as 'shale' if shale content is greater than 0.5, 'brine sand' if shale content is less than or equal to 0.5 and water saturation is greater than 0.7, and 'CO2 sand' if shale content is less than or equal to 0.5 and water saturation is less than or equal to 0.7. We also performed cluster analysis on the dataset using the k-mean algorithm (Venables and Ripley, 2002) and found that the results are almost identical to those obtained from our empirical classification.

Figure 3 shows pair-wise cross-plots among various geophysical attributes with the green crosses showing CO2 sand, the red crosses showing brine sand, and the black triangles showing shale. We can see that CO2 sand is well separate from shale and brine sand with low P- and S-wave velocity ratio, low density and high resistivity. However, the separation between shale and brine sand is less significant. We can only distinguish them from the cross-plots of seismic P-

wave and S-wave velocity ratio versus impedance and versus density. Since shale and brine sand are close to each other, we may alternatively group them together and have only two lithotypes, i.e., CO2 sand and non-CO2 sand-shale. The effects of using two versus three lithotypes on inversion will be investigated in the next section. As given in Section 2, we assume that the geophysical attributes have a multivariate Gaussian distribution and therefore they can be described by the mean and covariance matrix among those quantities.

Synthetic seismic and CSEM data

We first generated synthetic seismic data using the convolution models based on Zoeppritz equations (Aki and Richards, 1980) and seismic P- and S-wave velocity and density along the profiles at x = 5784 m, 4584 m, and 6684 m between depths 2100 m and 2800 m. We use a time interval of 4 ms and Ricker waveforms with two different central frequencies, representing low and high frequency seismic data. The seismic profile consists of a thirty-layer target zone with a layer thickness of 20 m and one layer above and one layer below the target zone with a thickness of 50 m.

We added temporally correlated Gaussian random noise to the synthetic data. We use the S/N of 5 to 1 from the smallest (i.e., 0 degree) to the largest (i.e., 55 degrees) incident angles with the signal amplitude defined by the maximum amplitude in the chosen time windows. We use the exponential variogram and a temporal correlation length of three times of the time intervals (i.e., 12 ms) to calculate the temporal covariance. Figure 4 shows the synthetic seismic data with noise included using 5 Hz and 20 Hz central frequencies, respectively. We can see that the seismic data generated using the 20 Hz Ricker waveform have more details compared to those generated using the 5 Hz Ricker waveform. For the 5 Hz seismic data, the signals are weak at the near offsets. In the synthetic study, we first use the seismic data at all the 56 incident

angles, which represents an ideal situation; we then use the data at only five widely separate angles (i.e., 0, 10, 20, 30, and 40 degrees), which represent a common situation in practice.

We generated 1D EM data according to the resistivity logs along the same location using the layered EM model. Although it is more suitable to use a 2D forward model for generating EM data based on the profile given in Figure 1, we choose to use a 1D model because it is fast and our primary focus is on the use of statistical rock-physics models. The profile for EM forward simulation includes the seawater with a thickness of 500 m, the overburden with a thickness of 1650 m, the target zone with thirty 20m-thick layers, and infinitely thick bedrock. We use a typical CSEM survey configuration and assume that a receiver is deployed on the seafloor and a transmitter is towed 50 m above the seafloor. We collect data using ten offsets (i.e., 1, 2, 3, 4, 5, 6, 7, 8, 10, 12 km) and five frequencies (i.e., 0.125, 0.25, 0.5, 0.875, and 1 Hz). We also added 3% to 5% independent random noise to the synthetic data relative to the amplitudes of the data from the near to the far offsets. Figure 5 shows the amplitudes and phases of the noised synthetic EM data.

INVERSION RESULTS AND DISCUSSION

We apply the developed Bayesian model to estimate geophysical attributes and reservoir parameters under various situations. Since the forward models for calculating seismic and EM data and for inversion are the same, our focus is on the role of statistical rock-physics models. First, we estimate seismic attributes without using any rock-physics model. The results serve as benchmark for other analysis. We will also investigate some features of the current Bayesian model, some may be general and others may be specific to the current algorithms.

Inversion without using statistical rock-physics models

We first invert the 5 Hz synthetic seismic data along the profile at x = 5784 m without using the rock-physics models. This represents a naive condition at the early stage of exploration when we do not have any borehole logs. Because we do not have a relationship to link seismic attributes to reservoir properties such as porosity and water saturation, we only estimate seismic attributes, including P-wave and S-wave velocity ratio, impedance, and density, at each target layer and the layers outside the target zone. The priors for the parameters in the target zone are assumed uniform within the bounds determined by our prior knowledge. We choose the bounds as the extreme values derived at the well location by fluid substitution of 100% brine and 100% CO2 into the reservoir sand. We use the obtained high and low values of seismic P- and S-wave velocity, density, and electrical resistivity to calculate the extreme values of seismic velocity ratios, impedance, density, and resistivity. The lower prior bounds are then set as the 80% of the extreme low values and the prior upper bounds are set as the 120% of the extreme high values.

We ran four Markov chains using the same initial values with different random seeds for 240,000 iterations but only keep samples at every 10 iterations (i.e., a thinning of 10) to save memory. The use of different random seeds makes chains move in different paths, which is equivalent to start from different initial values. Figure 6 shows the root-mean-squares (RMS) of the differences between the data and simulated results for the four chains; three of them (excluding the black curve) are visually converged after a few hundred iterations. We use the potential scale reduction factor (PSRF) to monitor the convergence of the four chains, which is a measure of the between-chain variability relative to the within-chain variability (Gelman and Rubin, 1992; Brooks and Gelman, 1998). A large value of the PSRF suggests that either the between-chain variability can be further decreased by more simulations or further simulations

will increase the within-variability since the simulated sequences have not yet made a full tour of the target distribution (Brooks and Gelman, 1998). With that approach, if the scale reduction score is less than 1.2, the Markov chain is considered converged; otherwise, more runs are needed. We calculate the PSRF for all the parameters based on the three Markov chains, and most of them are less than the recommended threshold value of 1.2.

Figure 7 shows the estimated medians (black curves), the 95% lower and upper bounds (red curves), and the corresponding true values (blue curves with triangles). The black dashed lines are the prior bounds, which are very wide and provide limited information on the parameters. Overall, the estimated medians follow the true values well and most of the measurements are within the 95% predictive intervals. We can also see that the uncertainty associated with the estimation is large, especially for the P-wave and S-wave velocity ratio and density.

Inversion using seismic data and statistical rock-physics models

We invert the same 5 Hz seismic data along the profile at x = 5784 m with the use of the statistical rock-physics models. Because the rock-physics models provide correlation between reservoir porosity and seismic attributes as well as information about clustering (see Figure 3), we can jointly estimate seismic attributes, porosity, and lithotypes. We started from the same initial values, prior bounds and random seeds as used for the inversion without using rock-physics models. We similarly ran four chains and pooled the samples in the second half of the two chains that are deemed as converged for the results shown in Figure 8.

Compared to Figure 7, we can see that the use of rock-physics models improves the estimates significantly: (1) reducing the differences between the true and estimated values, (2) reducing the uncertainty in the estimation, and (3) smoothing the estimates spatially. The

improvement is particularly remarkable for P-wave and S-wave velocity ratio and density. To quantify the benefits of using rock-physics models, we calculate the RMS of the differences between the true values and the estimated medians as a measure of prediction errors and the averaged width of the 95% predictive intervals as a measure of uncertainty. As shown in the columns 2-3 and 5-6 of Table 1, the use of statistical rock-physics models helps to reduce the RMS of the differences and the associated uncertainty for all the unknown parameters.

Another significant benefit is that the incorporation of rock-physics models allows us to estimate reservoir porosity directly. As shown in Figure 8(a), inversion of the seismic data provides very accurate estimates of porosity; all the estimated medians follow the true values closely with tight bounds. When not using the statistical rock-physics models and jointly estimating porosity, porosity has to be estimated using relationships between seismic attributes and porosity by deterministic or stochastic methods. In either case, uncertainty in the alternative porosity estimates will be large because the two-step approach sums up the uncertainty in the estimated seismic attributes as shown in Figure 7 and the uncertainty in rock-physics models as demonstrated by Chen et al. (2005).

An additional byproduct of using the statistical rock-physics models is the estimates of lithotypes. Figure 9 shows the estimated lithotypes as a function of iterations when we use the rock-physics models with two lithotypes, where we use a thin of 10 for iterations. Compared to the true lithotypes shown on Figure 9(a), we can see that after 10,000 iterations, the sampling algorithms find the two major reservoir layers but miss the thin layer because of the limited resolution of the lower frequency data.

Inversion using seismic and CSEM data and rock-physics models

To estimate water saturation, we need to jointly invert seismic and CSEM data because seismic data are insensitive to water saturation as shown by Hoversten et al. (2006) and Chen et al. (2007). To have a better resolution, we use the 20 Hz seismic data. We can follow two procedures: (1) jointly inverting the CSEM and 20 Hz seismic data directly; (2) first inverting the 5 Hz seismic data to provide narrow bounds for porosity and seismic attributes and then inverting the CSEM data and 20 Hz seismic data jointly. The main difference between the two procedures is that the second approach uses information from low frequency seismic data. The first approach produces high fluctuation in parameters and converges slowly because seismic attributes get no information from the EM data and porosity is primarily constrained by the seismic attributes. The second procedure, however, converges much faster as the porosity is better constrained, and we only need to modify water saturation and resistivity to fit the EM data.

Figure 10 shows the comparison between the true values and the estimated results obtained using the second approach with two lithotypes for joint inversion. Due to the use of rock-physics models, we are able to estimate porosity and water saturation as well as seismic and electrical attributes. In the figure, the black dashed lines show the bounds obtained from the inversion of the 5 HZ seismic data and the blue lines with triangles are the true values along the profile at x = 5584 m. The true values closely follow the estimated medians and the uncertainty bounds are tight except for resistivity. This is reasonable because CSEM data have much low resolution than seismic data. Comparing the estimated 95% lower and upper bounds with the prior bounds, we found that all the parameters are updated because the 20 Hz seismic data provide higher resolution than the 5 Hz seismic data as shown by the columns 3-4 and 6-7 of Table 1.

Inversion using two versus three lithotypes

Physically we may need to classify the lithotypes into three rather than two lithotypes. To see its effects on the estimation, we invert the same sets of geophysical data using three lithotypes and compare their results with those obtained using two lithotypes. Similar to the inversion of the 5 Hz seismic data (see Figure 9), with the use of two lithotype rock-physics models, we miss the detection of the thin CO2 sand layer even if we use the high frequency (i.e., 20 Hz) seismic data together with the EM data. However, if we use three lithotype rock-physics models, as shown in Figure 11, we can correctly predict each lithotype. The use of three lithotype rock-physics models also improves the estimates of water saturation and electrical resistivity as shown in Figure 12(b)(c). Figure 13 compares the estimated posterior probabilities of being shale, brine sand, and CO2 sand (solid lines) with their corresponding true values (dashed lines with dots). We can see that the uncertainty in predicting each lithotype is very small. Similar results were also found by Ulvmoen and Omre (2010) when they use Markov random fields as prior models.

Estimation of total CO2 volumes with uncertainty

To show the advantage of using the stochastic method over deterministic approaches, we estimate the CO2 volume and associated uncertainty in each layer by the product of porosity and CO2 saturation (i.e., $1-S$) in the layer. Figure 14 shows the estimated CO2 volumes and their corresponding true values. We can see that the estimated medians (black solid lines) closely follow the true values (blue lines with triangles) except in the second reservoir layer. The inversion results of using three lithotypes are better for estimating the CO2 volume in the second CO2 layer; the RMS of the differences between the true and estimated values using two and

three lithotypes are 0.0430 and 0.0344, respectively. Most importantly, they provide the lower and upper bounds of 95% predictive intervals (red solid lines). Such uncertainty information about the values for the given estimation may be important for risk assessments and other purposes.

Application of the statistical rock-physics models to other locations

The borehole that we used for developing the statistical rock-physics models is located at x = 5570 m and it is 14 m away from x = 5884 m. To evaluate the applicability of the regression-based rock-physics models to other locations, we invert seismic and EM data at two other locations far away from x = 5570 m. The first one is located at x = 4584 m, 986 m left of the training borehole; the second one is located at x = 6684 m, 1114 m right of the training borehole. We follow the same two-step procedure to first invert lower frequency seismic data and then jointly invert CSEM and high frequency seismic data, and the results at x = 4584 m and 6684 m are given in Figures 15 and 16, respectively. Similar to the case at x = 5784 m, the joint inversion of CSEM and seismic data provide good estimates of reservoir properties and geophysical attributes. These results demonstrate that even if the training borehole is about 1 km away, the derived statistical rock-physical models are applicable to other locations for the synthetic case.

Inversion using seismic data at five widely separate incident angles

To show how the information content of seismic data affects the joint inversion results, we follow the same two-step procedure using the seismic data at only five incident angles (i.e., 0, 10, 20, 30 and 40 degrees) to estimate geophysical and reservoir parameters. Figures 17, 18, and 19 show the results without using the rock-physics models, using the rock-physics models, and the joint inversion; Table 2 summarizes the data misfits and averaged widths of predictive

intervals. Although the uncertainties in each of those inversions are significantly larger than those obtained using all the seismic data (i.e., Figures 7, 8, and 12 and Table 1), the main conclusion remains unchanged. The use of statistical rock-physics models and prior Markov random fields provides significant information for improved estimation of geophysical attributes and reservoir parameters. Figure 20 shows the estimated posterior probabilities of being shale, brine sand, and CO2 sand. Even if the accuracy of predicting shale and brine sand is much lower than that when all the seismic data are used (see Figure 13), the prediction of CO2 sand is still very accurate, only missing a thin layer around the depth of 2350 m.

CONCLUSIONS

We have developed a Bayesian model to invert CSEM and seismic data for reservoir parameters based on statistical rock-physics models and Markov random fields and have shown the advantages of incorporating statistical rock-physics models through synthetic studies. The use of statistical rock-physics models together with Markov random fields significantly improves the estimates of geophysical attributes and allows for directly estimation of reservoir parameters with associated uncertainty. Although the statistical rock-physics models are obtained just from one borehole, we have shown that they are applicable to other locations. This provides a convenient and cost-effective way to invert geophysical data for estimating reservoir parameters at the early stage of exploration, in which collecting borehole samples and performing laboratory or field rockphysics experiments to obtain detailed physical-based rock-physics models is difficult.

One main assumption of the current method is that the statistical-based rock-physics models derived from the borehole logs are representative of the sites of interest. This is a common limitation of learning based rock-physics models due to the lack of information on the

underlying physics, geology and mechanics. To have confidence on the developed rock-physics models, suitable analysis should be carried out before applying them in estimation, for example, whether the geology at this site is relatively similar. It is desirable that if possible we collect data from more than one borehole because it allows for cross-validation. However, since the main goal of geophysical exploration at the early stage is to determine whether it is worthy of further investigation, the use of statistical based rock-physics models is useful. The application of this approach in production settings with a high density of wells to derive and cross-validate the rock-physics models is obvious.

The current Bayesian model is developed for estimating geophysical attributes and reservoir parameters along a 1D profile. It can be extended to estimate parameters on 2D cross-sections or in 3D spaces. If the spatial variability of geophysical and reservoir properties along lateral directions is small, we can first estimate reservoir parameters on each profile and then spatially interpolate the estimates to 2D or 3D domains using geostatistics methods (e.g., kriging, or indicator kriging). However, if the spatial variability is large, we need to consider the unknown geophysical attributes and reservoir properties as 2D or 3D random fields and jointly estimate them by conditioning to multiple seismic CDP data and 2D or 3D CSEM data. This is an ongoing but challenging task because the total number of unknowns is much larger and the forward modeling of 2D or 3D problems is computationally more demanding. A major breakthrough in the direction needs effective parameterizations to reduce the total number of unknowns, fast forward modeling to reduce computational time, and efficient MCMC sampling algorithms to speed the mixing and convergence of multiple chains.

We applied a two-step (or sequential) procedure to combine seismic and CSEM data: first inverting lower frequency seismic data to narrow down the ranges of unknown parameters and

then jointly inverting high frequency seismic and CSEM data under the constraints of the previous inversion results. We demonstrated that the two-step approach is more efficient than the one-step approach for joint stochastic inversion of seismic and EM data in terms of the mixing and convergence of MCMC sampling methods. This is because lower frequency seismic data are the responses of large-scale structures and their likelihood surface is relatively smooth. The smoothing likelihood surface allows MCMC chains move freely and thus speeds up the convergence of samples to the stationary distribution. The two-step approach is practical because we can always obtain low frequency seismic data by using low-pass filter. We can also incorporate other types of low frequency information, such as the depth dependent trends of geophysical and reservoir parameters, into the current Bayesian model through prior distributions. This certainly will provide valuable information to reduce ambiguity in the joint estimation. The two-step strategy can be used in many ways for seismic and EM inversion. For example, since porosity is primarily determined by seismic data, we can first estimate it using multiple-frequency seismic data through a sequential approach, and then estimate water saturation using EM data by keeping it fixed.

ACKNOWLEDGMENTS

The work is funded by Chevron Energy Technology Company. We wish to thank Joe Stefani from Chevron for generating the 2D synthetic model and Chevron for permission to publish this work. This work was also partially supported by the U.S. Department of Energy and LBNL under Contract No. DE-AC02-05CH11231.

Appendix A: Conditional distribution of geophysical attributes given reservoir parameters and lithotypes

Since the conditional distribution of geophysical attributes given lithotypes, i.e., $f(\alpha_i, Z_i, \rho_i, r_i, \varphi_i, S_i \mid L_i)$, has a multivariate Gaussian distribution, the conditional distribution of geophysical attributes given reservoir parameters and lithotypes, $f(\alpha_i, Z_i, \rho_i, r_i \mid \varphi_i, S_i, L_i)$, also has a multivariate Gaussian distribution (Stone, 1996). Let $\mu_k(L_i)$, where $k = 1, 2, \cdots, 6$, be the mean of P- and S-wave velocity ratio, acoustic impedance, density, resistivity, porosity, and water saturation of the ith lithotype, respectively. Let $\mathbf{\mu}(L_i)$ and $\mathbf{\Sigma}(L_i)$ be the mean vector and covariance matrix of conditional distribution $f(\alpha_i, Z_i, \rho_i, r_i, \varphi_i, S_i \mid L_i)$. We partition them as follows:

$$\mathbf{\mu}(L_i) = \begin{pmatrix} g(L_i) \\ h(L_i) \end{pmatrix}, \text{ and } \mathbf{\Sigma}(L_i) = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \tag{A1}$$

where $g(L_i) = (\mu_1(L_i), \mu_2(L_i), \mu_3(L_i), \mu_4(L_i))^T$ and $h(L_i) = (\mu_5(L_i), \mu_6(L_i))^T$, representing the mean vectors of geophysical attributes and reservoir parameters. The matrices C_{11}, C_{12}, C_{21}, and C_{22} have dimensions of 4×4, 4×2, 2×4, and 2×2, respectively.

Let $\mathbf{\mu}_c(L_i)$ and $\mathbf{\Sigma}_c(L_i)$ be the conditional mean vector and covariance matrix of $f(\alpha_i, Z_i, \rho_i, r_i \mid \varphi_i, S_i, L_i)$. We thus have (Gelman et al, 2004):

$$\mathbf{\mu}_c(L_i) = g(L_i) + C_{12}C_{22}^{-1}((\varphi_i, S_i)^T - h(L_i)) \tag{A2}$$

and

$$\mathbf{\Sigma}_c(L_i) = C_{11} - C_{12}C_{22}^{-1}C_{21}. \tag{A3}$$

Let $\mathbf{b}_1(L_i)$ and $\mathbf{b}_2(L_i)$ be the first and second column of the product of C_{12} and C_{22}^{-1}. We thus have:

$$\boldsymbol{\mu}_c(L_i) = g(L_i) + \mathbf{b}_1(L_i)(\varphi_i - \mu_5(L_i)) + \mathbf{b}_1(L_i)(S_i - \mu_6(L_i)). \tag{A4}$$

Appendix B: Full conditional distributions of porosity and water saturation

Since the derivation of full conditional distributions of porosity and water saturation is similar, we only give details on the derivation of porosity. From Equation 13, we have

$$\begin{aligned} f(\varphi_i \mid \cdot) &\propto f(\alpha_i, Z_i, \rho_i, r_i \mid \varphi_i, S_i, L_i) \times f(\varphi_i \mid L_i) \\ &\propto \left|\boldsymbol{\Sigma}_c(L_i)\right|^{-1/2} \exp\left\{-0.5(\mathbf{x}_i - \boldsymbol{\mu}_c(L))^T \boldsymbol{\Sigma}_c^{-1}(L_i)(\mathbf{x}_i - \boldsymbol{\mu}_c(L))\right\} \\ &\quad \times \tau_\varphi^0(L_i)^{1/2} \exp\left\{-0.5\tau_\varphi^0(L_i)(\varphi_i - \mu_\varphi^0(L_i))^2\right\} \\ &\propto \exp\left\{-0.5(\varphi_i \mathbf{b}_1(L_i) - \mathbf{d}_1(L_i))^T \boldsymbol{\Sigma}_c^{-1}(L_i)(\varphi_i \mathbf{b}_1(L_i) - \mathbf{d}_1(L_i))\right\} \\ &\propto \exp\left\{-0.5\tau_{\varphi_i}(\varphi_i - \mu_{\varphi_i})^2\right\} \propto N(\mu_{\varphi_i}, \tau_{\varphi_i}), \end{aligned} \tag{B1}$$

where $\mathbf{d}_1(L_i)$ is given in Equation 15.

REFERENCES

Aki, K., and P. G. Richards, 1980, Quantitative seismology: Theory and methods: W. H. Freeman and Co.

Barbu, A., and S. C. Zhu, 2007, Generalizing Swendsen-Wang for imaging analysis: Journal of Computational and Graphical Statistics, **16**, 877-900.

Buland, A., and H. Omre, 2003, Bayesian linearized AVO inversion: Geophysics, **68**, 185-198.

Chen, J., M. Hoversten, D. Vasco, Y. Rubin, and Z. Hou, 2007, A Bayesian model for gas saturation estimation using marine seismic AVA and CSEM data: Geophysics, **72**, WA85.

Chen, J., and T. Dickens, 2009, Effects of uncertainty in rock-physics models on reservoir parameter estimation using seismic amplitude variation with angle and controlled-source electromagnetics data: Geophysical Prospecting, **57**, 61.

Eidsvik, J., P. Avseth, H. Omre, T. Mukerji, and G. Mavko, 2004, Stochastic reservoir characterization using pre-stack seismic data: Geophysics, **69**, 978.

Fregoso, E., and L. A. Gallardo, 2009, Cross-gradients joint 3D inversion with applications to gravity and magnetic data: Geophysics, **74**, No. 4, L31-L42

Gallardo, L. A., and M. A. Meju, 2004, Joint two-dimensional DC resistivity and seismic traveltime inversion with cross-gradients constraints: J. Geophys. Res., **109**, B03311.

Gassmann, F., 1951, Elastic waves through a packing of spheres: Geophysics, **16**, 673-685.

Gelman, A., and D. B. Rubin, 1992, Inference from iterative simulation using multiple sequences: Statistical Science, **7**, 457-472.

Geman, S. and D. Geman, 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images: IEEE Transactions on Pattern Analysis and Machine Intelligence, **6**, 721-741.

Gilks, W., S. Richardson, and D. Spiegelhalter, 1996, Markov Chain Monte Carlo in Practice: Chapman & Hall/CRC, New York.

Haario, H., E. Saksman, and J. Tamminen, 2001, An adaptive Metropolis algorithm: Bernoulli, **7**, 223-242.

Hastings, W. K., 1970, Monte Carlo sampling methods using Markov chains and their applications: Biometrika, **57**, 97-109.

Higdon, D. M., 1998, Auxiliary variable methods for Markov chain Monte Carlo with applications: Journal of the American Statistical Association, **93**, 585.

Hoversten, G., F. Cassassuce, E. Gasperikova, G. Newman, J. Chen,, D. Vasco, Y. Rubin, and Z. Hou,, 2006, Direct reservoir parameter estimation using joint inversion of marine seismic AVA and CSEM data: Geophysics, **71**, C1-C13.

Mavko, G., T. Mukerji, and J. Dvorkin, 1998, The rock physics handbook: Tools for seismic analysis in porous media: Cambridge University Press, New York, USA.

Metropolis, N., A. W. Rosenbluth, M.N. Rosenbluth, A. H. Teller, and E. Teller, 1953, Equations of state calculations by fast computing machines: Journal of Chemical Physics, **21**, 1087-1092.

Neal, R. M., 2003, Slice sampling: The Annuals Statistics, **31**, 705-767.

Smith, D., and M. Smith, 2006, Estimation of binary Markov random fields using Markov chain Monte Carlo: Journal of Computational and Graphical Statistics, **15**, 207-227.

Swendsen, R. H. and J. S. Wang, 1987, Non-universal critical dynamics in Monte Carlo simulations: Physical Review Letter, **58**, 86.

Venables, W. N., and B. D. Ripley, 2002, Modern applied statistics with S, the fourth edition: Springer, New York.

Wolf, U., 1989, Collective Monte Carlo updating for spin systems: Physics Review Letter, **62**, 4.

Yilmaz, O., 1987, Seismic data processing: Society of Exploration Geophysicists, Tulsa, Oklahoma.

FIGURE CAPTIONS

Figure 1: 2D cross-section of the synthetic model with boreholes and estimation locations. The borehole at x=5784 m is visually at the same location as the training borehole at x=5770 m. The color images are plotted based on resistivity with the red color being high values and the blue color being low values.

Figure 2: Various logs along the training borehole at x=5770 m.

Figure 3: Statistical rock-physics model with three lithotypes based on the borehole logs at x=5770 m, where the green crosses are CO2 sand, the red crosses are brine sand, and the black triangles are shale.

Figure 4: Synthetic seismic data at 56 angles with signal-to-noise ratios (S/N) from 5 to 1 and the spatial correlation length of 12 ms, where the time 0 corresponds to the depth of 2100 m.

Figure 5: Synthetic CSEM data with 3% to 5% relative errors from the near to the far offsets.

Figure 6: The root-mean-square (RMS) of the differences between the 5 Hz seismic data and the simulated results for four chains (e.g., the black, red, blue, and green curves) with the same initial values but different random seeds. The total number of actual iterations is 240,000 and the plot shows the results with a thinning of 10.

Figure 7: Inversion results of the 5 Hz seismic data at x=5784 m without using rock-physics models. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

Figure 8: Inversion results of the 5 Hz seismic data at x=5784 m using the statistical rock-physics models with two lithotypes. The black dashed lines are the prior ranges, the red lines

are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

Figure 9: Estimated lithotypes at x=5784 m as a function of iterations (using a thinning of 10), where the yellow colors are CO2 sand and the red colors are shale or brine sand.

Figure 10: Joint inversion results of the 20 Hz seismic and CSEM data at x=5784 m using the statistical rock-physics models with two lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

Figure 11: Estimated lithotypes at x=5784 m as a function of iterations (using a thinning of 4) based on three lithotype rock-physics models, where the red, orange, and yellow colors are shale, brine sand, and CO2 sand, respectively.

Figure 12: Joint inversion results of the 20 Hz seismic and CSEM data at x=5784 m using the statistical rock-physics models with three lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

Figure 13: Comparison between the estimated posterior probabilities being shale, brine sand, and CO2 sand (solid lines) and their corresponding true values (dashed lines with dots) at x=5784 m using the 20 Hz seismic and CSEM data.

Figure 14: Estimated CO2 volumes along the profile at x=5784 m for (a) two lithotypes and (b) three lithotypes. The red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values. The RMS of errors using two and three lithotypes rock-physics models are 0.0430 and 0.0344, respectively.

Figure 15: Joint inversion results at x=4584 m using 20 Hz seismic and CSEM data and the statistical rock-physics models with three lithotypes. The black dashed lines are the prior ranges, the red lines are the 95\% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

Figure 16: Joint inversion results at x=6684 m using 20 Hz seismic and CSEM data and the statistical rock-physics models with three lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

Figure 17: Inversion results of the 5 Hz seismic data at the incident angles of 0, 10, 20, 30, 40 degrees at x=5784 m without using rock-physics models. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

Figure 18: Inversion results of the 5 Hz seismic data at the incident angles of 0, 10, 20, 30, 40 degrees at x=5784 m using the statistical rock-physics models with two lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

Figure 19: Joint inversion results of the CSEM data and the 20 Hz seismic data at the incident angles of 0, 10, 20, 30, 40 degrees at x=5784 m using the statistical rock-physics models with three lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

Figure 20: Comparison between the estimated posterior probabilities being shale, brine sand, and CO2 sand (solid lines) and their corresponding true values (dashed lines with dots) at x=5784 m using the CSEM data and the 20 Hz seismic data at the incident angles of 0, 10, 20, 30, 40 degrees.

Table 1: The root-mean-square (RMS) of the differences between the measured and estimated parameter values and the averaged width of 95% predictive domains along the profile at x = 5784 m using the seismic data at 56 incident angles.

Parameters	Root-mean-square of differences			Averaged width of predictive intervals		
	Inversion of seismic data		Joint Inversion	Inversion of seismic data		Joint inversion
	Without rockphysics	With rockphysics		Without rockphysics	With rockphysics	
Vp/Vs	0.0980	0.0561	0.0294	0.2732	0.2017	0.0814
Zp (MPa.s/m)	0.1454	0.1344	0.1123	0.6380	0.3895	0.2001
Density (g/cc)	0.0594	0.0235	0.0161	0.3243	0.0774	0.0421
Porosity		0.0166	0.0109		0.0513	0.0247
Sw			0.1404			0.1614
Resistivity (ohm-m)			24.3506			27.9352

Table 2: The root-mean-square (RMS) of the differences between the measured and estimated parameter values and the averaged width of 95% predictive domains along the profile at x = 5784 m using the seismic data at five incident angles (i.e., 0, 10, 20, 30, and 40 degrees).

Parameters	Root-mean-square of differences			Averaged width of predictive intervals		
	Inversion of seismic data		Joint Inversion	Inversion of seismic data		Joint inversion
	Without rockphysics	With rockphysics		Without rockphysics	With rockphysics	
Vp/Vs	0.1379	0.1226	0.1417	0.6134	0.3000	0.1741
Zp (MPa.s/m)	0.4052	0.4200	0.2513	2.2347	1.0683	0.5959
Density (g/cc)	0.0715	0.0410	0.0468	0.5284	0.1228	0.0683
Porosity		0.0164	0.0202		0.0726	0.0341
Sw			0.1604			0.2261
Resistivity (ohm-m)			26.5325			27.3670

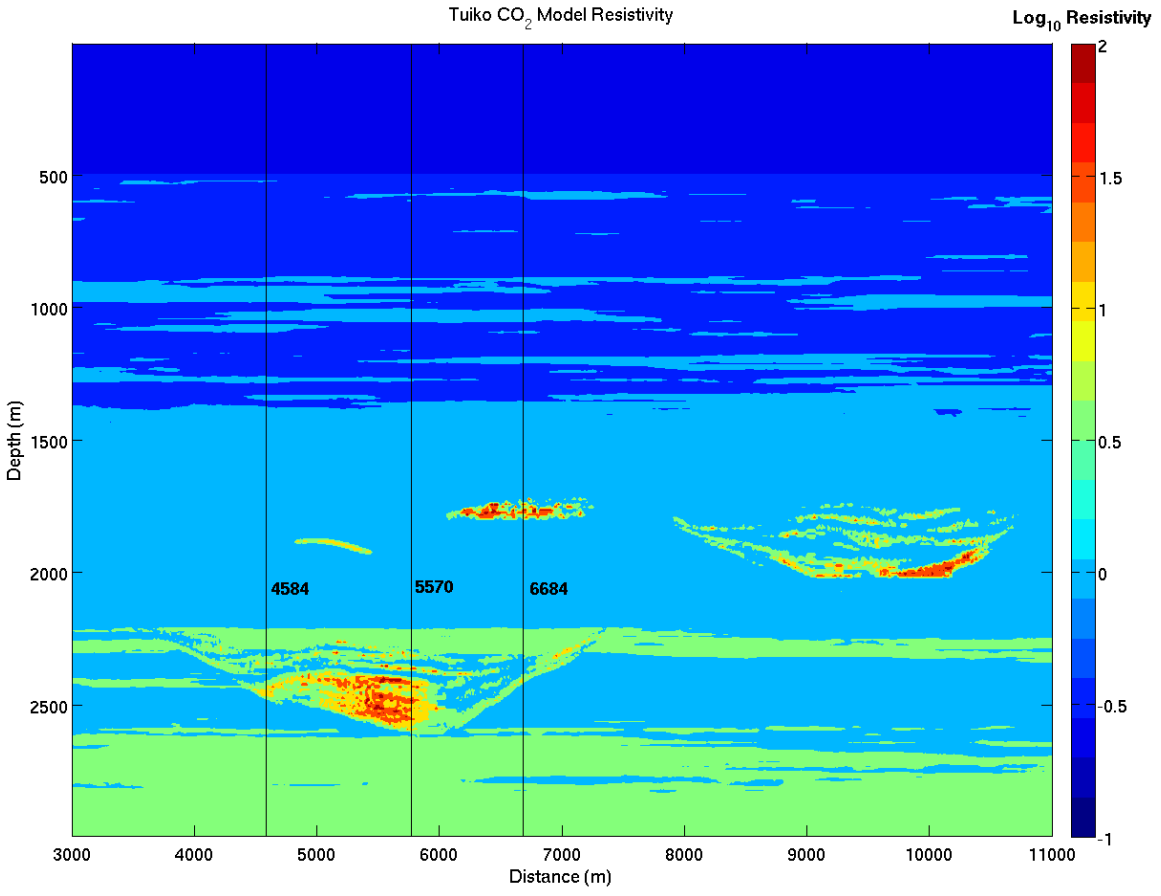

Figure 1. 2D cross-section of the synthetic model with boreholes and estimation locations. The borehole at x=5784 m is visually at the same location as the training borehole at x=5770 m. The color images are plotted based on resistivity with the red color being high values and the blue color being low values.

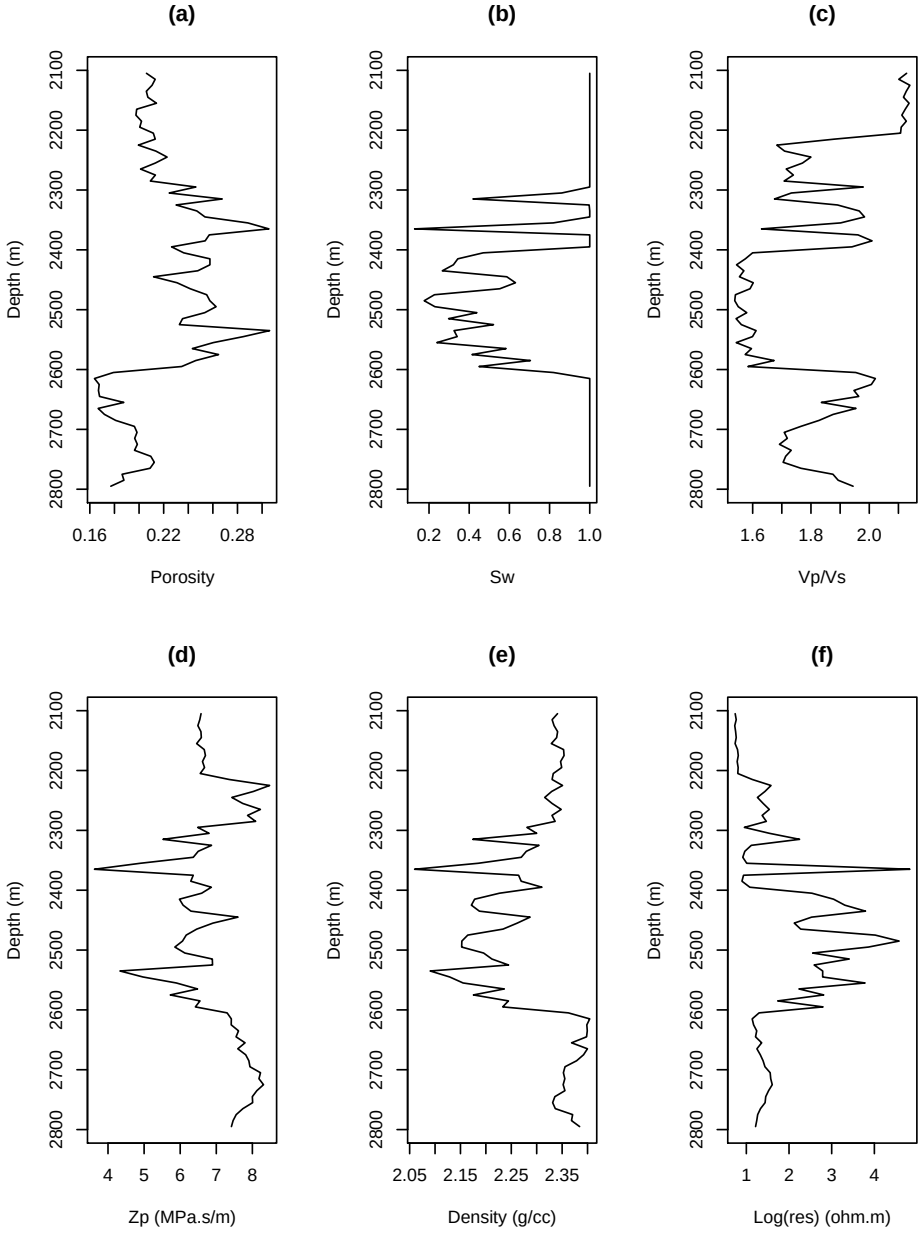

Figure 2. Various logs along the training borehole at x=5770 m.

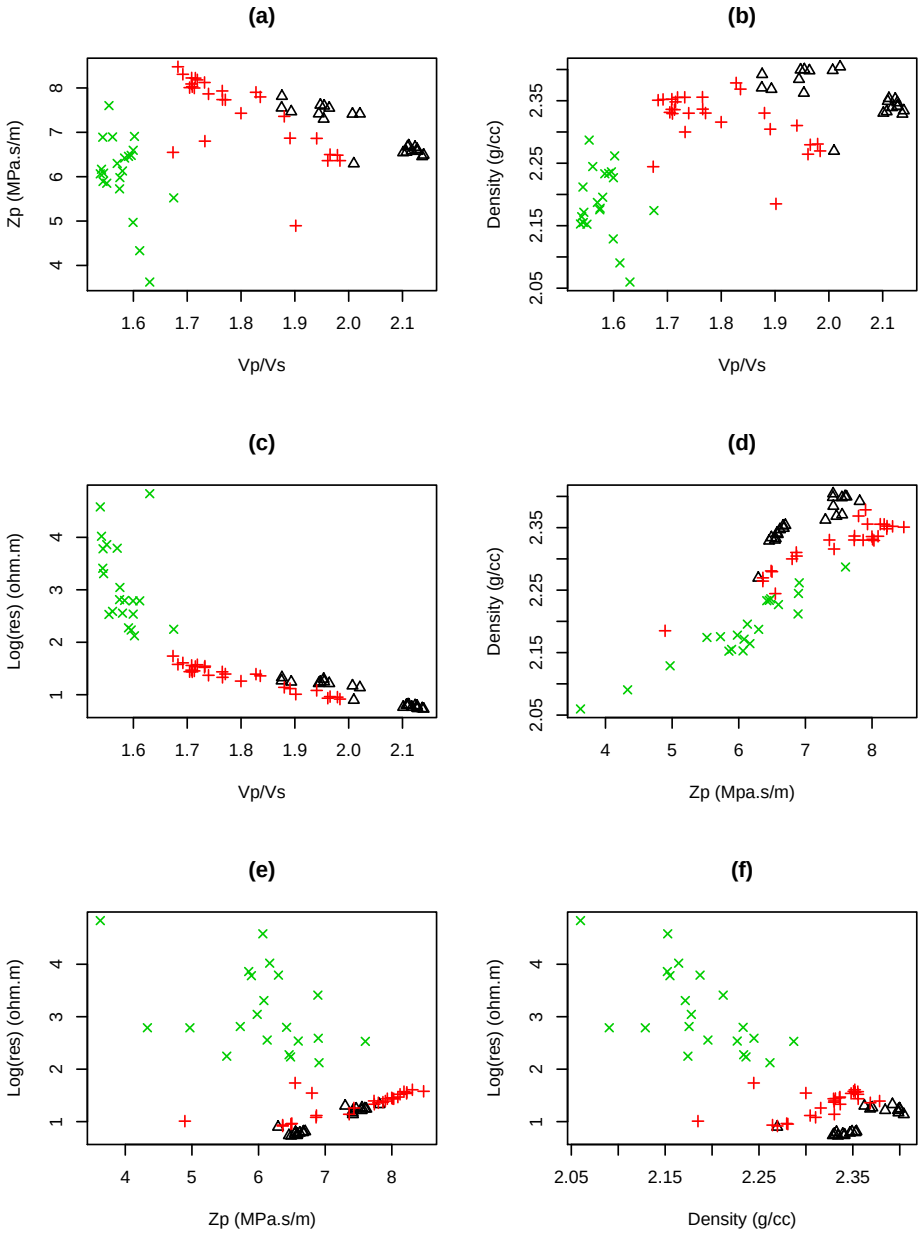

Figure 3. Statistical rock-physics model with three lithotypes based on the borehole logs at x=5770 m, where the green crosses are CO2 sand, the red crosses are brine sand, and the black triangles are shale.

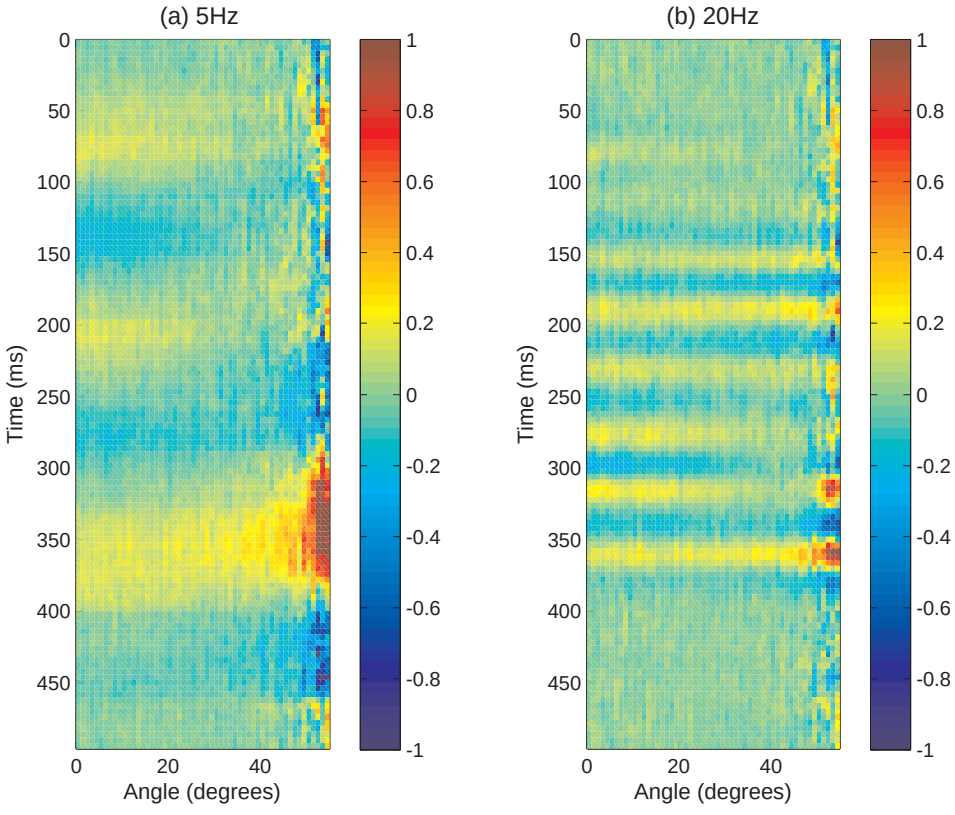

Figure 4. Synthetic seismic data at 56 angles with signal-to-noise ratios (S/N) from 5 to 1 and the spatial correlation length of 12 ms, where the time 0 corresponds to the depth of 2100 m.

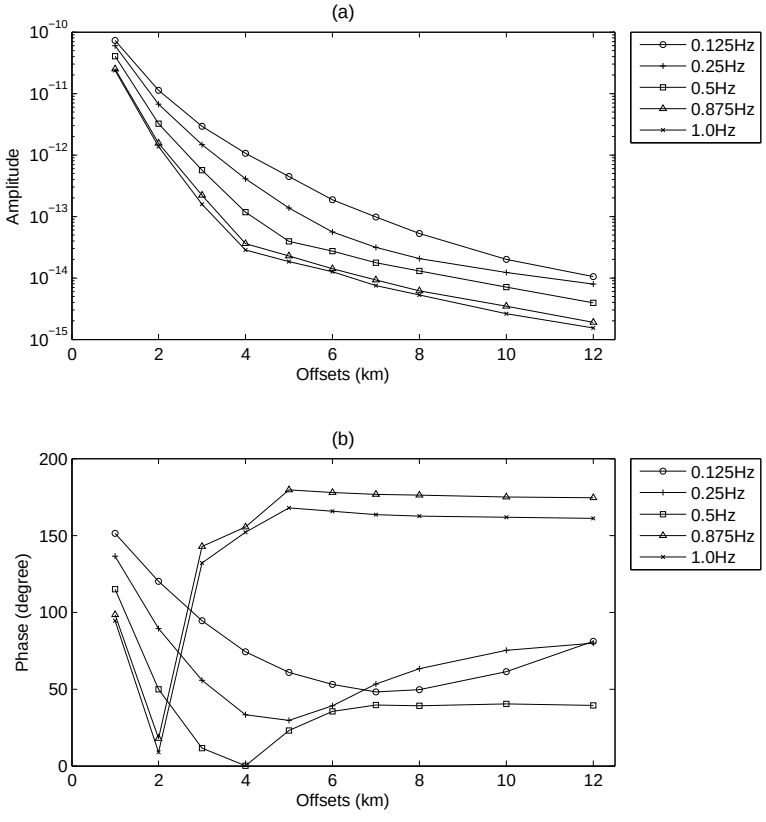

Figure 5. Synthetic CSEM data with 3% to 5% relative errors from the near to the far offsets.

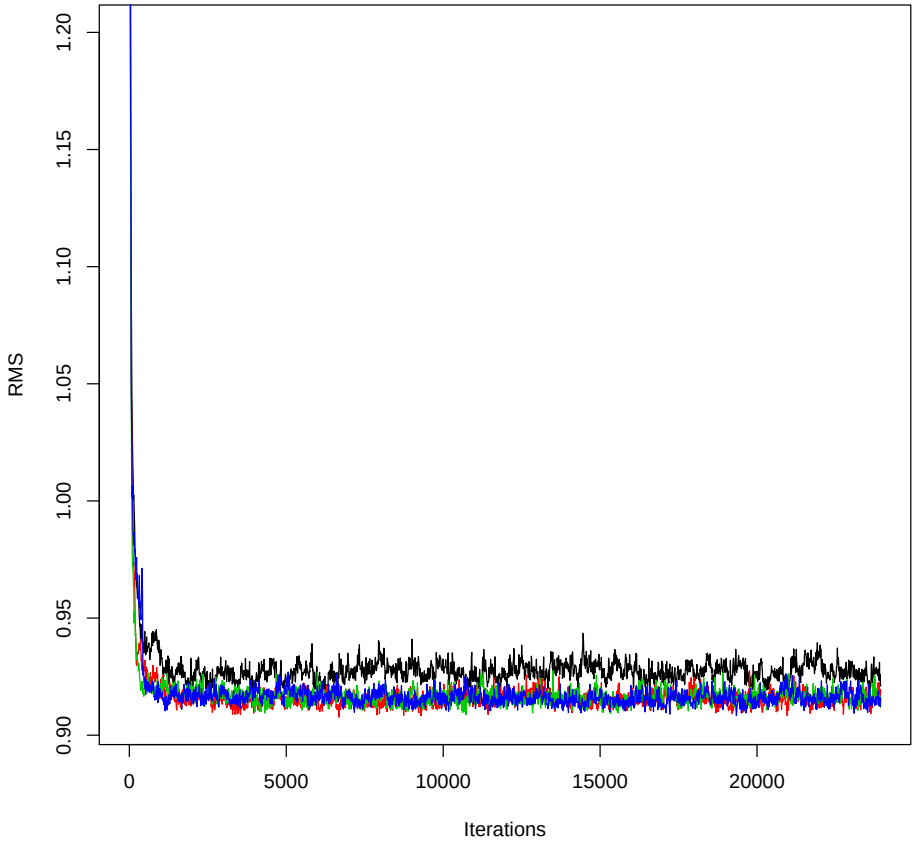

Figure 6. The root-mean-square (RMS) of the differences between the 5Hz seismic data and the simulated results for four chains (e.g., the black, red, blue, and green curves) with the same initial values but different random seeds. The total number of actual iterations is 240,000 and the plot shows the results with a thinning of 10.

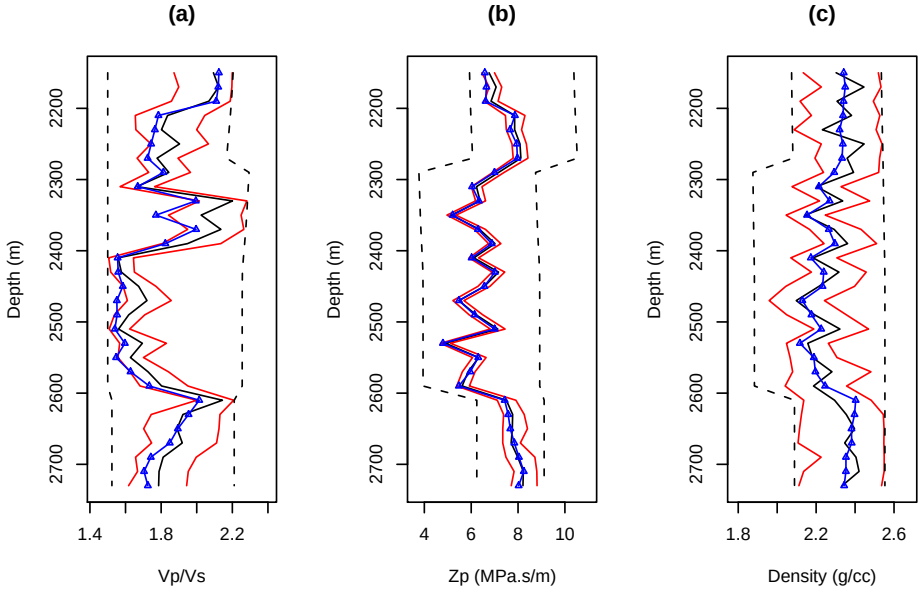

Figure 7. Inversion results of the 5 Hz seismic data at x=5784 m without using rock-physics models. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

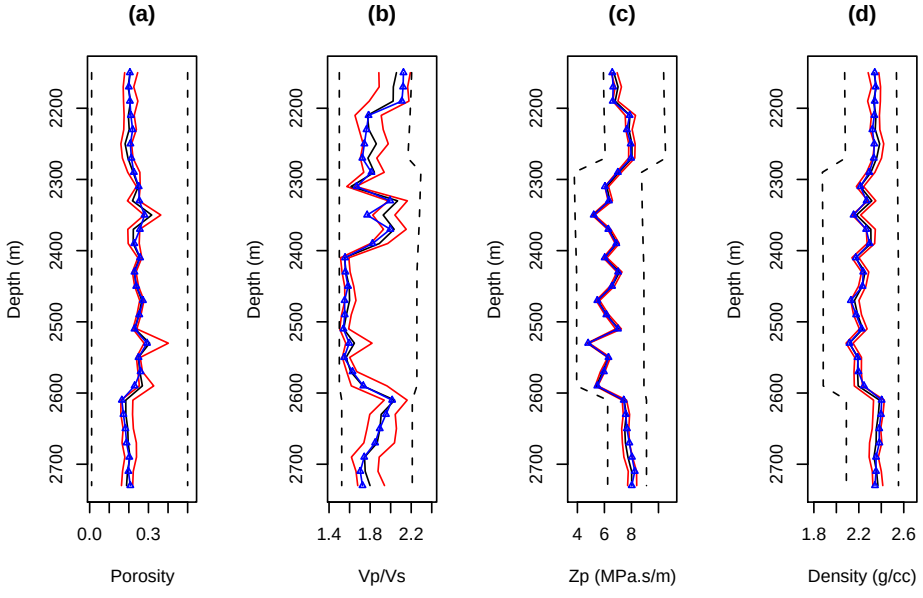

Figure 8. Inversion results of the 5 Hz seismic data at x=5784 m using the statistical rock-physics models with two lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

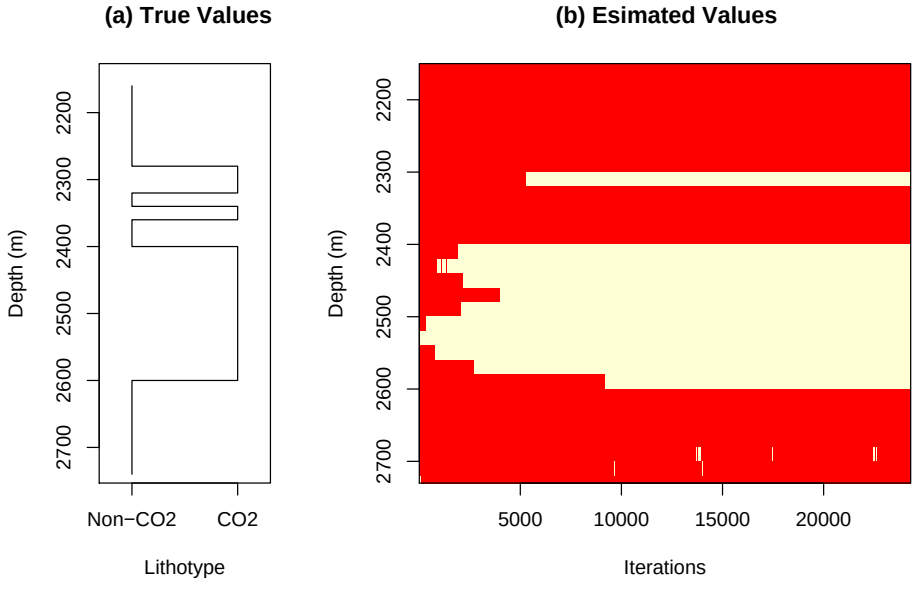

Figure 9. Estimated lithotypes at x=5784 m as a function of iterations (using a thinning of 10), where the yellow colors are CO2 sand and the red colors are shale or brine sand.

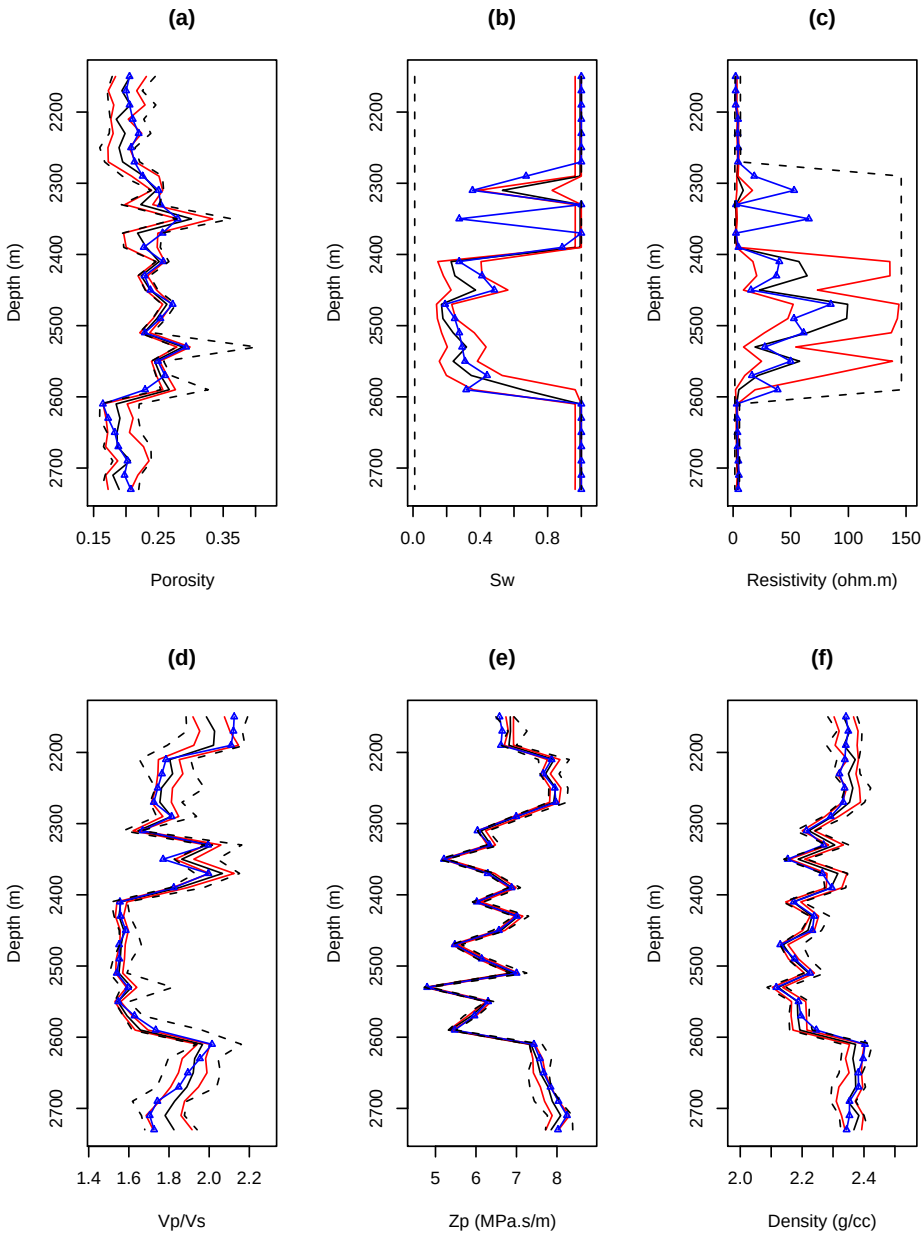

Figure 10. Joint inversion results of the 20 Hz seismic and CSEM data at x=5784 m using the statistical rock-physics models with two lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

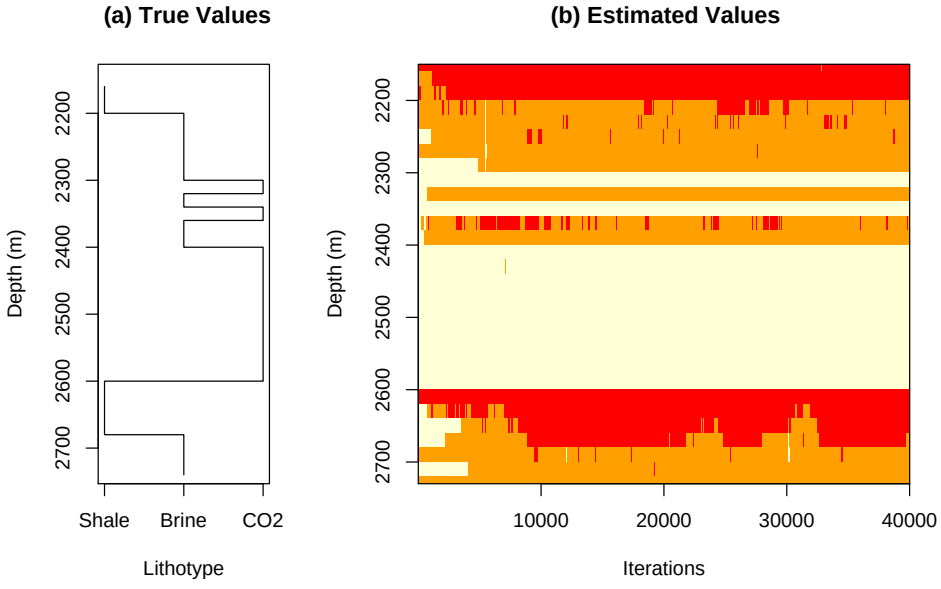

Figure 11. Estimated lithotypes at x=5784 m as a function of iterations (using a thinning of 4) based on three lithotype rock-physics models, where the red, orange, and yellow colors are shale, brine sand, and CO2 sand, respectively.

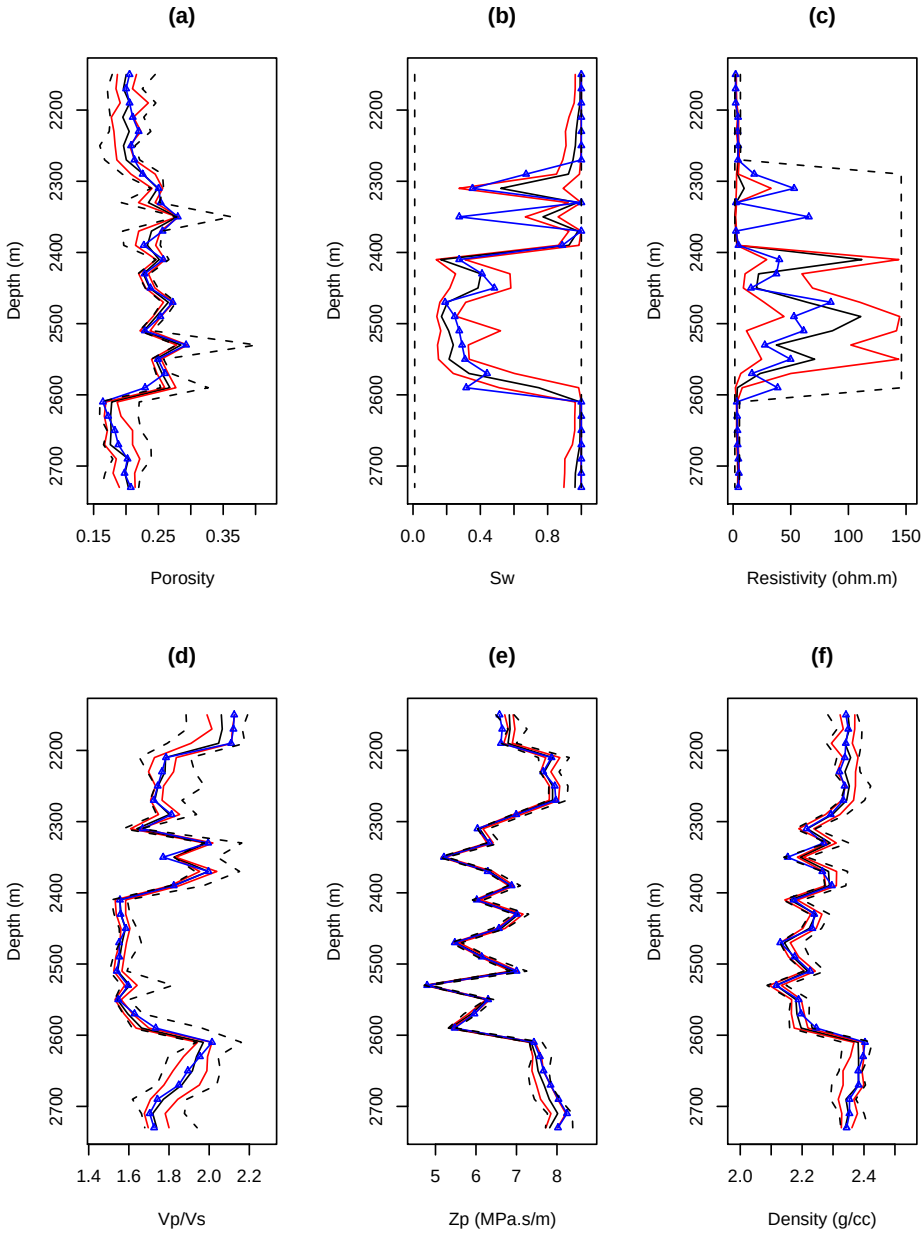

Figure 12. Joint inversion results of the 20 Hz seismic and CSEM data at x=5784 m using the statistical rock-physics models with three lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

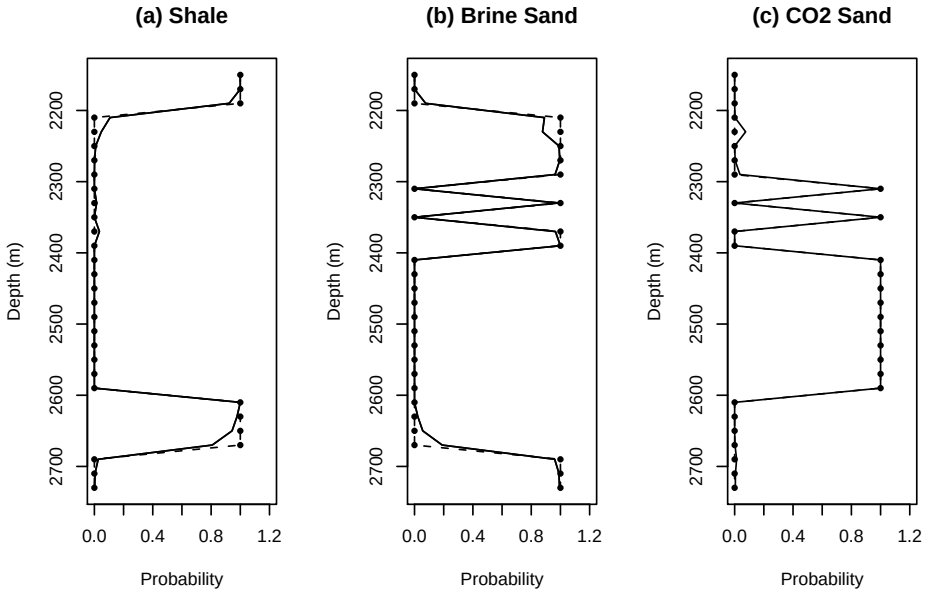

Figure 13. Comparison between the estimated posterior probabilities being shale, brine sand, and CO2 sand (solid lines) and their corresponding true values (dashed lines with dots) at x=5784 m using the 20 Hz seismic and CSEM data.

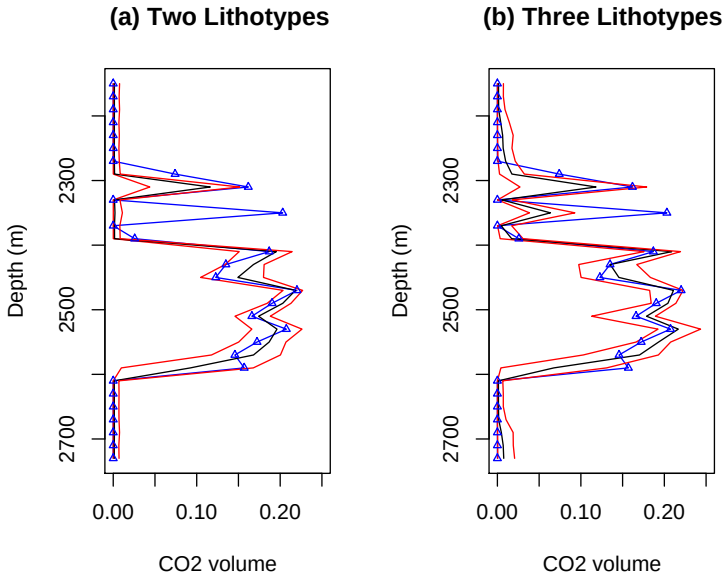

Figure 14. Estimated CO2 volumes along the profile at x=5784 m for (a) two lithotypes and (b) three lithotypes. The red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values. The RMS of errors using two and three lithotypes rock-physics models are 0.0430 and 0.0344, respectively.

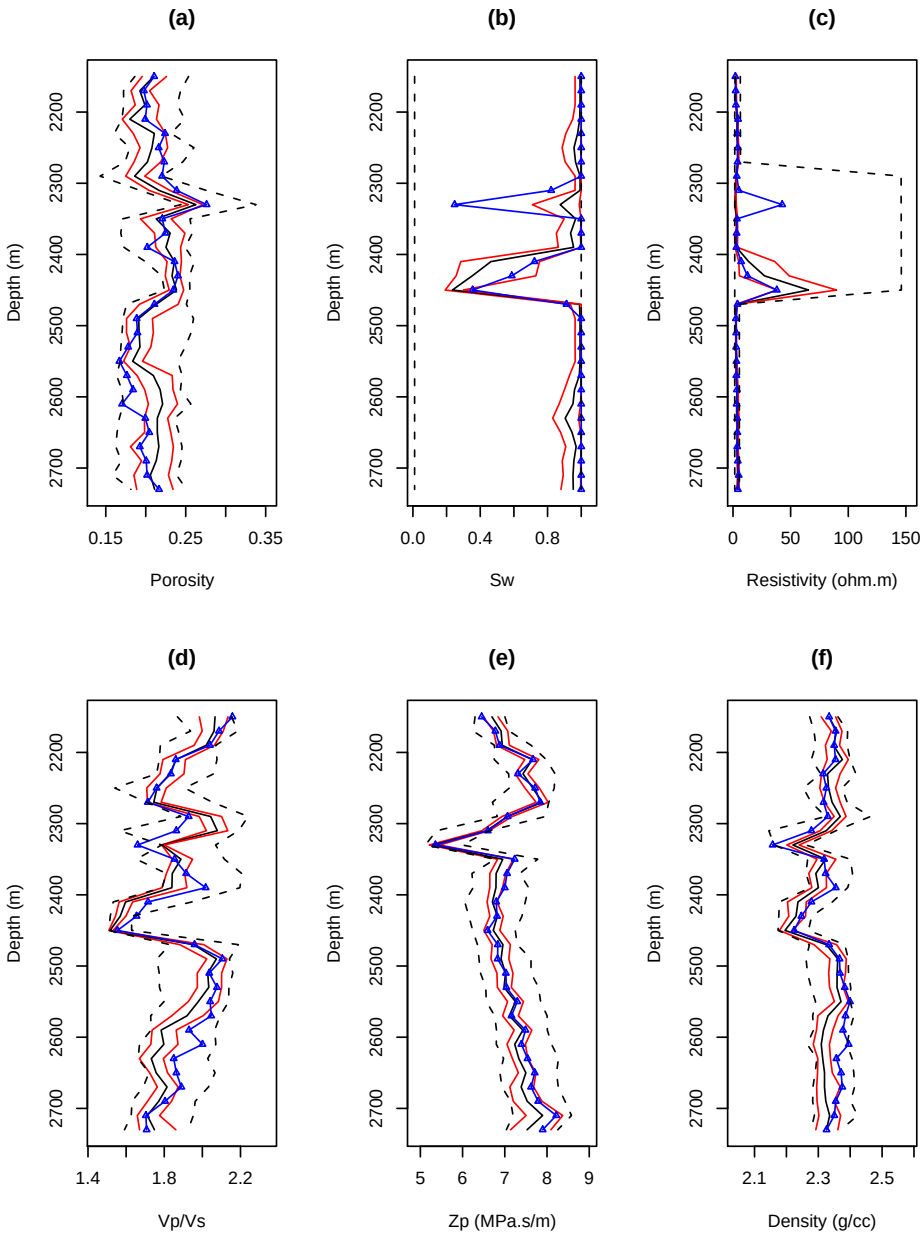

Figure 15. Joint inversion results at x=4584 m using 20 Hz seismic and CSEM data and the statistical rock-physics models with three lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

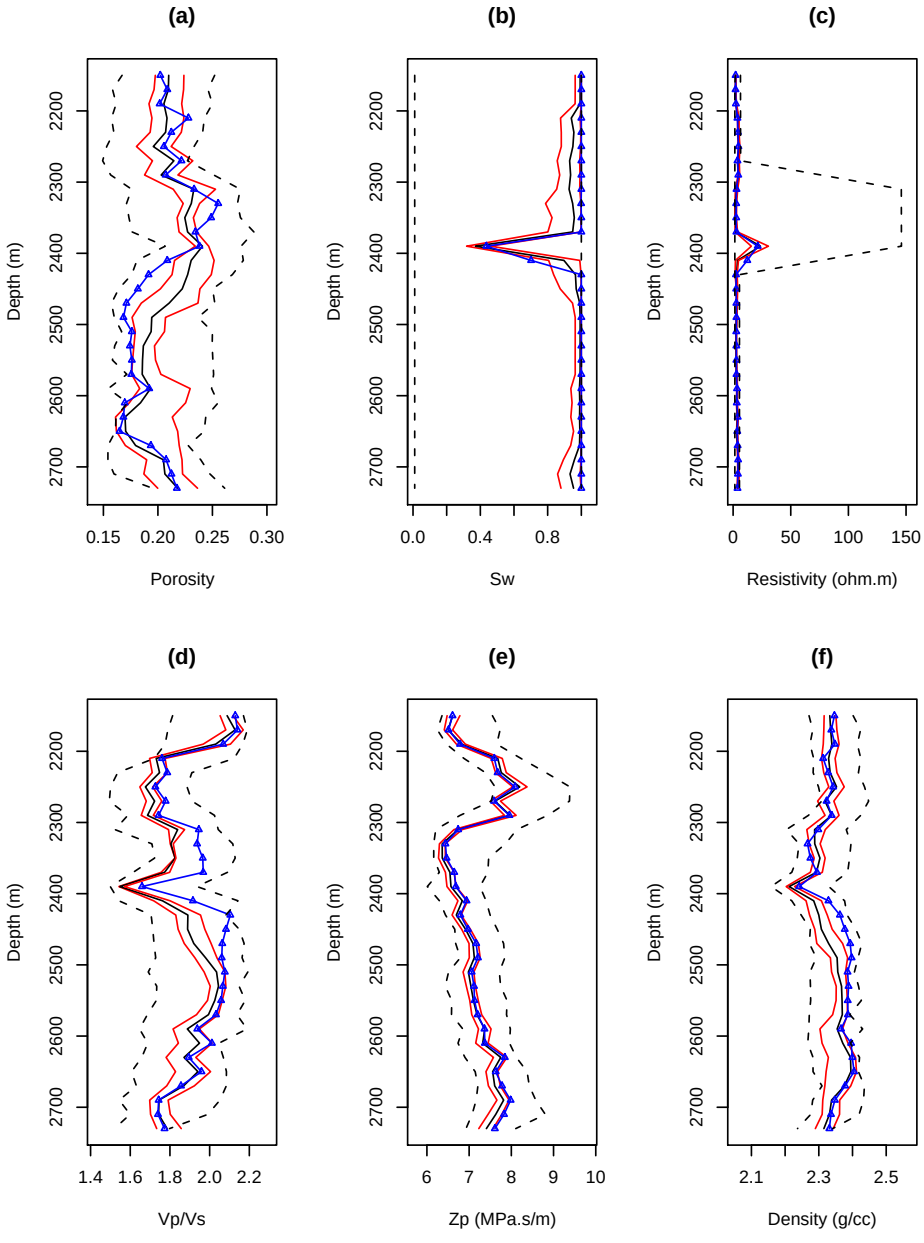

Figure 16. Joint inversion results at x=6684 m using 20 Hz seismic and CSEM data and the statistical rock-physics models with three lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

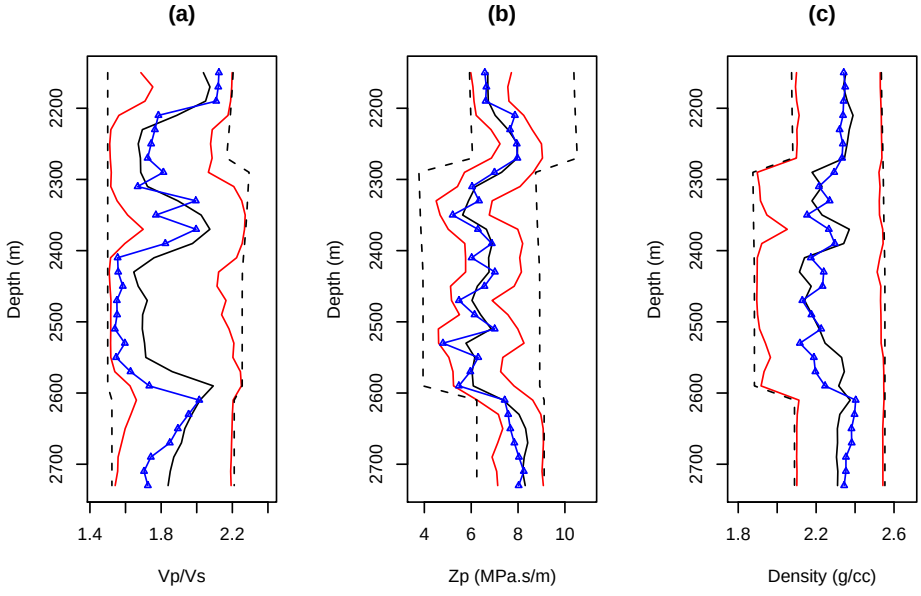

Figure 17. Inversion results of the 5 Hz seismic data at the incident angles of 0, 10, 20, 30, 40 degrees at x=5784 m without using rock-physics models. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

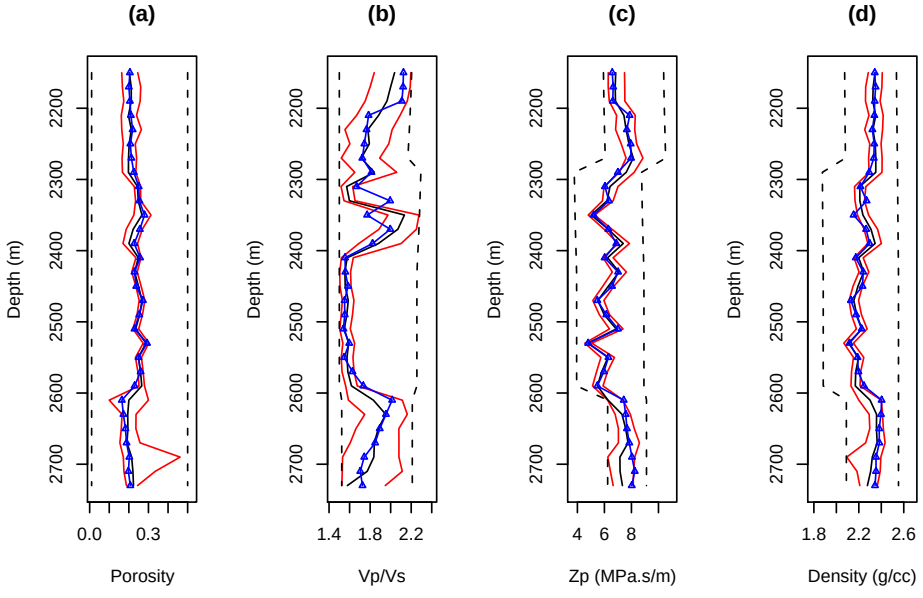

Figure 18. Inversion results of the 5 Hz seismic data at the incident angles of 0, 10, 20, 30, 40 degrees at x=5784 m using the statistical rock-physics models with two lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

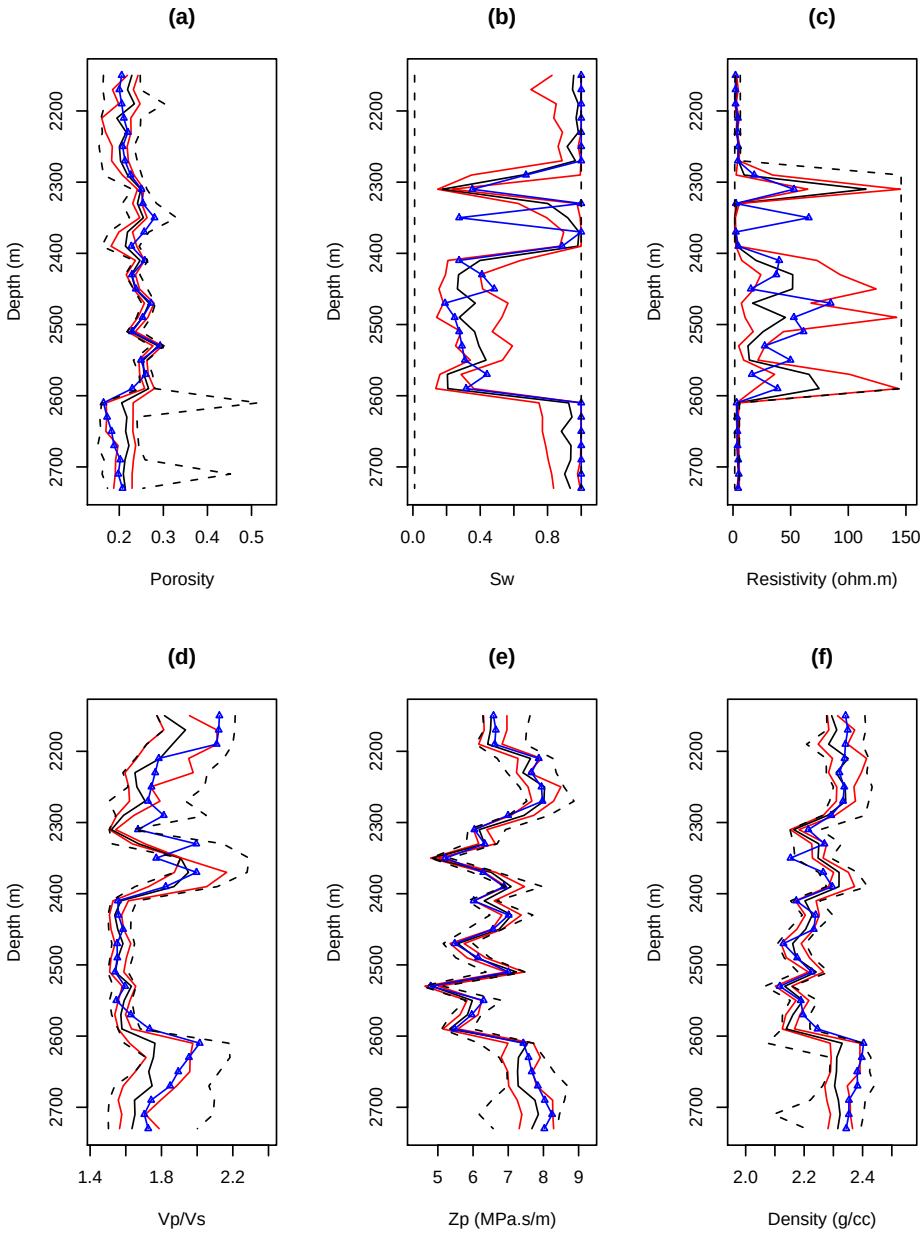

Figure 19. Joint inversion results of the CSEM data and the 20 Hz seismic data at the incident angles of 0, 10, 20, 30, 40 degrees at x=5784 m using the statistical rock-physics models with three lithotypes. The black dashed lines are the prior ranges, the red lines are the 95% lower and upper bounds, the black lines are the estimated medians, and the blue lines with triangles are the true values.

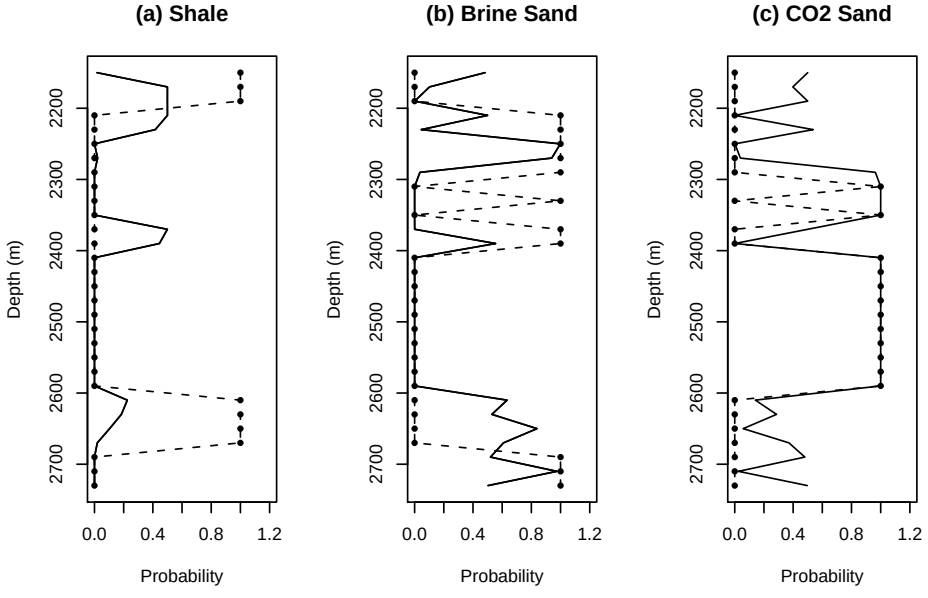

Figure 20. Comparison between the estimated posterior probabilities being shale, brine sand, and CO2 sand (solid lines) and their corresponding true values (dashed lines with dots) at x=5784 m using the CSEM data and the 20 Hz seismic data at the incident angles of 0, 10, 20, 30, 40 degrees.