

# Practical Bayesian Inversions for Rock Composition and Petrophysical Endpoints in Multimineral Analysis

Liwei Cheng<sup>1\*</sup>, Ge Jin<sup>1</sup>, Reinaldo Michelena<sup>2</sup>, and Ali Tura<sup>1</sup>

<sup>1</sup>Colorado School of Mines

<sup>2</sup>SeisPetro Geosoftware, LLC.

## Summary

Rock composition can be related to conventional well logs through theoretical equations and petrophysical endpoints. Multimineral analysis is a formation evaluation tool that uses inversions to quantify rock composition from well logs. However, because of data errors and the multivariate selection of petrophysical endpoints, solutions from the multimineral analysis are nonunique. Many plausible realizations exhibit comparable data misfits. Therefore, the uncertainties in rock composition and petrophysical endpoints must be quantified but cannot be fulfilled by deterministic solvers. Stochastic Bayesian methods have been applied to assess the uncertainties, but the high run time, tedious parameter tuning, and need for specific prior information hinder their practical use. We implement Markov chain Monte Carlo with ensemble samplers (MCMCES) to assess the uncertainties of rock composition or petrophysical endpoints in the Bayesian framework. The resultant posterior probability density functions (PDFs) quantify the uncertainties. Our method has fewer tuning parameters and is more efficient in convergence than the conventional random walk Markov chain Monte Carlo (MCMC) methods in high-dimensional problems. We present two independent applications of MCMCES in multimineral analysis. We first apply MCMCES to assess the uncertainties in volume fractions with a suite of well logs and petrophysical endpoints. However, defining the petrophysical endpoints can be challenging in complex geological settings because the values of standard endpoints may not be optimal. Next, we use MCMCES to estimate petrophysical endpoints' posterior PDFs when the endpoints are uncertain. Our methods provide posterior volume-fraction or petrophysical-endpoint realizations for interpreters to evaluate multimineral solutions. We demonstrate our approach with synthetic and field examples. Reproducible results are supplemented with the paper.

## Introduction

Estimating the rock composition, such as mineral volume fractions, porosity, and fluid saturations, is essential to hydrocarbon reservoir characterization and development. The emerging need for CO<sub>2</sub> sequestration also calls for a better understanding of rock composition for site assessment as the injected CO<sub>2</sub> interacts with host rock for mineral trapping (Benson and Cole 2008; Ajayi et al. 2019). Multimineral analysis is a formation evaluation tool using conventional well logs to quantify rock composition (or constituent volume fractions). Mayer and Sibbit (1980) optimize rock composition by minimizing the misfits between measured logs and theoretical values. Quirein et al. (1986) propose linear mixing approximations between well logs and rock composition through a set of petrophysical endpoints (endpoints hereafter). The linear mixing assumption significantly reduces the computational time in optimization. Early multimineral developments mostly rely on deterministic optimizations that obtain solutions efficiently. However, because of data errors and uncertainties of petrophysical models, multimineral results are nonunique. If available, multimineral solutions are recommended to reconcile with core data, mud logs, and local geological knowledge. Assessing the uncertainties is critical but cannot be easily achieved by deterministic methods. Therefore, using stochastic methods helps to evaluate the solutions associated with the uncertainties.

There have been stochastic approaches applied to multimineral analysis to assess the uncertainties in rock composition. For instance, Aldred (2018) uses the Monte Carlo processing that randomly draws samples from a probability distribution function to approximate the desired function. The uncertainty assessment can be achieved by finding many realizations that fit the data. However, the exhaustive Monte Carlo sampling may not explore the model space efficiently, especially for high-dimensional problems, thereby costing significant computational time. In addition, the assumption of parameter distributions, such as normal or triangle, is generally required. The MCMC method in the Bayesian framework provides an alternate approach to random sampling high-dimensional probability distributions to estimate posterior PDFs. The resultant posterior PDFs are proportional to the known function, thereby quantifying the uncertainties. A random walker MCMC approach is formulated by Spalburg (2004), but the application is a low-dimension implementation that only focuses on porosity, net-to-gross ratio, and saturation. Deng et al. (2019, 2020) implement a comprehensive MCMC to estimate rock composition with associated uncertainties for thinly bedded formations. However, their approach generally requires assuming prior normal distributions to assist convergence. In addition, to account for thinly bedded formations, the multimineral model implemented in Deng et al. (2019) incorporates nonlinear nuclear log responses, which costs considerable computing time (approximately 1 minute per depth sample). Therefore, a more practical and robust MCMC method for estimating uncertainties in rock composition is needed.

In addition to rock composition, endpoint selection is another challenge in multimineral analysis. Faithful endpoint selection leads to reasonable volume-fraction inversion. Even though endpoints are commonly treated as known parameters, the standard values of endpoints may not be optimal, especially in complex geological settings where the formation has more constituents than the number of well logs. In practice, the solutions are reconciled with other available information by altering tool responses or endpoints iteratively. However, the tedious heuristic iteration to the best results only means the best that one can achieve in the allotted time. In addition, there can be variations in endpoints from place to place, which may contain valuable geological information. Michelena et al. (2020) employ a genetic algorithm to solve the rock composition and endpoints from well-log data simultaneously. Their field example shows that the change in

\*Corresponding author; email: lcheng@mines.edu

Copyright © 2022 Society of Petroleum Engineers

Original SPE manuscript received for review 1 February 2022. Revised manuscript received for review 26 April 2022. Paper (SPE 210576) peer approved 17 May 2022.

kerogen grain density from different wells may indicate the variation in the thermal maturity of kerogen. However, the uncertainties associated with endpoints have not been explored.

We use MCMCES in the Bayesian framework in two independent applications, rock composition or endpoint inversion. MCMCES provides efficient convergence in high-dimensional problems with fewer tuning parameters than the conventional MCMC method. Furthermore, the linear mixing model simulated by MCMCES allows for efficient estimation of posterior probability distributions without detailed prior information. Given the advantage of assessing uncertainties, optimizing Bayesian inversion can be challenging. Therefore, this paper aims to illustrate comprehensive implementations of Bayesian inversions for multiminer analysis. The paper is organized as follows. We first review multiminer analysis and the sources of uncertainties. Then, we explain how to use MCMCES to address the uncertainties. Finally, we demonstrate our method with synthetic and field examples for volume fraction and endpoint applications and provide reproducible results.

## Review of Multiminer Analysis and Its Uncertainty

In multiminer analysis, rock composition is inverted jointly from the conventional well logs through a set of theoretical equations and endpoints. The typical input data are a triple- or quad-combo logging set that may include gamma ray (GR), resistivity (RT), neutron porosity (NPHI), bulk density (RHOB), photoelectric absorption factor (PEF), and sonic slowness (DT) logs. To demonstrate our method, we use the linear mixing model (Quirein et al. 1986; Mitchell and Nelson 1988; Doveton 1994) that has already been implemented in commercial software for theoretical forward modeling and inversion. In the absence of gas in the formation, conventional well logs ( $\vec{d}$ ) at every depth can be linearly related through an endpoint matrix ( $G$ ) and rock composition ( $\vec{V}$ ), given by

where  $\vec{V}$  is an  $N$  by one vector of constituent volume fractions, given by

$$\vec{V} = [V_1 \ V_2 \ \dots \ V_i]^T, \quad (2)$$

where  $V_i$  is the volume fraction of the  $i$ th constituent and  $T$  denotes transpose.  $N$  represents the number of constituents considered in the analysis. In practice, resistivity is linearized as near-wellbore square root conductivity (CX) through the Archie equation, where the porosity and saturation exponents are assumed to be two (Doveton 1994). A disperse shale model for resistivity is assumed if clay minerals are present. Also, the linear superposition is used for the volumetric cross section (U) derived from the multiplication of PEF and RHOB.

The inverse problem is solved in a depth-by-depth fashion and is subject to the unity equation, which is expressed as

$$\sum_{i=1}^N V_i = 1. \quad (3)$$

In addition, each constituent volume fraction must be greater than zero and less than unity, as follows

$$0 \leq V_i \leq 1. \quad (4)$$

As expressed in Eq. 1, preconditioning well log data ( $d$ ) and selecting proper endpoints ( $G$ ) for each constituent are critical steps to estimate the rock composition ( $\vec{V}$ ) correctly. However, both have associated uncertainties. For instance, well log data may have measurement errors because of borehole conditions, calibration, operational issues, etc. Therefore, one must account for the data errors in the inversion to avoid obtaining erroneous solutions by overfitting the data.

In practice, using well-log crossplots can help to approximate endpoints from the data. However, the endpoints can be uncertain in some situations. First, even though most common minerals have suggested endpoint values, there are variations. For instance, the density of quartz may range from 2.64 to 2.68 g/cm<sup>3</sup> as the mineral may not be in its pure compositional form in the study interval. The change in endpoints from place to place may provide important insights into geology (Michelena et al. 2020). Second, some endpoints do not have exact values but ranges. For instance, clay GR can range from tens of API to a few hundreds of API, depending on the radioactive component in the clay composite. In such a case, without local knowledge, estimating appropriate endpoints can be a time-consuming process, typically done by a trial-and-error approach.

Furthermore, the number of constituents in the formation is generally greater than the number of available well logs in a complex geological setting. Therefore, grouping constituents with trace volume fractions or similar properties is common to make the inverse problem well determined (Quirein et al. 1986). For example, one may group quartz, K-feldspar, and plagioclase into a category of clastics. However, after grouping, the collective endpoints of the mixture are different from their standard values. In addition, the optimal endpoints may be from the averaging of several consecutive layers for thinly bedded formations because of the limited vertical resolutions of well-log data. Therefore, the endpoints must be somewhat adjusted for optimal inversion in most field cases, but their uncertainties cannot be easily evaluated in current commercial implementations of multiminer inversion.

In addition to the uncertainties in data and endpoints, the inverse problem can be ill-conditioned if there is a high similarity between constituent endpoints—the high similarity results in a high condition number of the matrix ( $G$ ). Thus, the uncertainties are higher for the constituents with similar endpoints as many realizations result in comparable data misfits. For instance, resolving quartz and plagioclase can be difficult because they have similar densities, low gamma ray reading, low neutron porosity, etc., even if the inverse problem is well determined. Cheng et al. (2022) show uncertainties in mineral volume fractions, resulting in violations in rock-physics models. Furthermore, the uncertainties are not constant along depths but a function of data errors, rock composition, and the weighting of well logs/equations. Therefore, understanding and accounting for this mathematical limitation embedded in the multiminer analysis are critical. However, the imposed constraints (Eqs. 3 and 4) increase the difficulty of estimating the uncertainties analytically. Hence, we propose nonlinear stochastic methods in endpoints to assist in assessing the uncertainties of endpoints from well-log data. The next section demonstrates how to evaluate the uncertainties numerically using the Bayesian inversion.

## Bayesian Inversion

Bayesian statistics or inversion can translate the uncertainties in the physical model and data into the computed functions. Bayes' theorem can be formulated as

$$P(\theta|d) = \frac{P(d|\theta) P(\theta)}{P(d)}, \quad (5)$$

where  $d$  indicates the data,  $\theta$  is the model parameter,  $P(\theta|d)$  represents the posterior PDF given the data ( $d$ ),  $P(d|\theta)$  is the likelihood function given the model parameter ( $\theta$ ),  $P(\theta)$  is the prior information, and  $P(d)$  represents the evidence. Prior information describes our belief and uncertainty on the model parameter before seeing the data. Bayes' theorem updates the prior information through the likelihood function based on the observed data. MCMC is a computational tool to perform Bayesian statistics and approximates the posterior PDF that expresses the uncertainty after seeing the data. The resultant posterior PDF is commonly displayed as a histogram and described by the mean, median, and standard deviation (STD in figures) to quantify the uncertainty. The evidence term ( $P(d)$ ) represents the probability of all possible models.  $P(d)$  usually can be eliminated if using MCMC sampling. Therefore, the prior information and likelihood function are the two primary functions performing MCMC.

Unlike Monte Carlo sampling methods that draw independent samples from the model distribution, the MCMC method draws samples where the next sample depends on the existing sample, called a Markov chain (a chain hereafter). According to the neighborhood transition probabilities, the chain moves from state to state on each iteration (Shonkwiler and Mendivil 2009). Each chain is independent and can be seen as a series of walker steps exploring the model space. There are different approaches in MCMC depending on how to propose the next step in a chain. For instance, random walk Metropolis-Hastings MCMC (Metropolis et al. 1953; Hastings 1970; Tierney 1994) that draws samples from symmetric proposal function is a common approach for performing Bayesian statistics. On the other hand, Hamiltonian or hybrid Monte Carlo (Duane et al. 1987) uses gradient information for the continuous space to avoid the basic random walk behavior. Moreover, Hamiltonian or hybrid Monte Carlo reduces the correlation between successive sampled states by proposing moves to distant states, requiring fewer chain samples to approximate Bayesian integrals. However, both methods require tedious tuning to reach optimal results, especially in high-dimensional problems.

This paper employs the MCMCES (or affine-invariant ensemble samplers) (Goodman and Weare 2010; Forman-Mackey et al. 2013; Grinsted 2020) to estimate the posterior PDFs. The main difference from other MCMC methods is the step proposal. Unlike the random walk method with independent walkers, the ensemble method uses ensemble walkers to automatically generate a properly tuned proposal distribution from the ensemble (Hogg and Foreman-Mackey 2018). The walkers are first randomly distributed within the preset ranges in the model space. To update the  $j$ th walker ( $X_j$ ), MCMCES selects a complementary walker  $Y$  and proposes a new step ( $\tilde{X}$ ), given by

$$\tilde{X} = ZX_j + (1 - Z)Y, \quad (6)$$

where  $Z$  is a real-valued stretching variable that defines how far the walker  $X_j$  to a new position is along the line connecting  $X_j$  to  $Y$  (Huijser et al. 2022). The proposed  $\tilde{X}$  is accepted to replace  $X_j$  with the probability ( $p$ ), given by

$$p(X, Y, Z) = \min \left( 1, Z^{n-1} \frac{\pi(\tilde{X})}{\pi(X)} \right), \quad (7)$$

where  $n$  represents the number of model parameters and  $\pi(X)$  indicates the target density function. Eq. 7 demonstrates a Metropolis-Hastings type of acceptance probability that depends on the ratio of the target densities at the current and proposed steps. Therefore, MCMCES makes a multivariate proposal for each walker at each iteration, given a scaled difference in position by parameters between the current walker and another randomly selected walker. The ensemble-tuned proposal distribution includes gauging the step sizes and directions for the new proposal. For more details on MCMCES, see Goodman and Weare (2010).

Compared with conventional MCMC methods, random walk Metropolis-Hastings, for instance, requires tuning the proposal covariance matrix to optimize the posterior results. In addition, random walk Metropolis-Hastings may have  $\sim n^2$  tuning parameters for a problem with  $n$  model parameters. We find that fewer tuning parameters and fast convergence in high-dimensional problems are the main advantages of MCMCES over other methods. However, MCMCES may not be applicable when the number of model parameters is greater than 50 (Huijser et al. 2022).

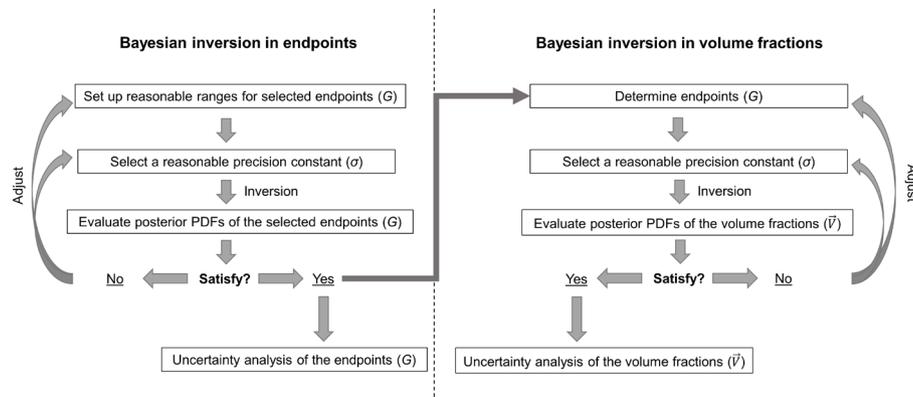
Our MCMCES application has six tuning parameters, the precision constant, the number of walkers (chains), steps per walker (chain), burn-in, step size, and thinning regardless of the model dimensions. The precision constant is the most important parameter among the tuning parameters, determining how precise the target distribution is. The precision constant is the denominator of a likelihood function—a smaller constant aims for distributions with smaller data misfits. Hence, knowing the data's noise level is key to determining the value of the precision constant. The number of walkers defines the number of independent Markov chains used to sample the model space. MCMCES requires a complementary ensemble of walkers to tune the proposal distribution. Only a few walkers are not sufficient to generate a proper proposal distribution. However, employing many ensemble walkers may sample the model space thoroughly but costs high computational time. One hundred ensemble walkers are typically appropriate and are used for the examples in this paper. The steps per walker define how long a chain is. A chain includes the initial random state evolving to convergence or stationary distributions. A longer chain records more statistics but also costs more computing time. Burn-in represents a process to discard the steps of a chain from the initial state before reaching the stationary distribution. For example, a burn-in of 0.5 means the first 50% of steps are discarded, and only the last 50% of steps are recorded for the posterior PDF. The step size (as  $Z$  in Eq. 6) determines the distance of random proposals for the next chain step. Because the ensembles automatically tune the proposal distribution, the step size in MCMCES is a low-maintenance parameter. A step size of 5 is used for our examples. Finally, we have a common option for keeping every  $k$ th sample to reduce the autocorrelation between steps as "thinning," where  $k$  is a positive integer. However, Link and Eaton (2012) find that the thinning process may not efficiently help the convergence or autocorrelation. Therefore, thinning is set to one (no thinning) for the examples in the paper. Thinning can be used for applications that require many steps (>5,000) to converge, saving the computer memory but costing  $k$  times of computing.

Diagnosing and optimizing Bayesian inversion can be challenging. Therefore, we present comprehensive diagnostic panels for every example shown in this paper to ensure the chains function properly. First, trace plots of combined chains illustrate how chains evolve along with steps. We use trace plots to determine how many steps per walker are needed and the burn-in fraction and identify convergence. After convergence, further steps are required to obtain samples for posterior inference. More steps gather more accurate

posterior PDF estimates and need more time to compute. The final posterior PDF is the collective posterior PDFs from all the chains after burn-in. Next, autocorrelation is an important measurement to gauge the efficiency of a chain. If an MCMC chain is strongly autocorrelated, the chain produces clumpy samples unrepresentative of the true underlying PDF. A good chain should have rapid mixing as the stationary distribution is reached quickly from an arbitrary position, thereby having a shorter autocorrelation time. In addition, the acceptance rate is another heuristic proxy statistic to diagnose the efficiency of chains. The acceptance rate is computed from the ratio of accepted steps over the total proposed steps. If the acceptance rate is too high, the chains accept most proposed steps, indicating the step sizes are too small and vice versa. For best performance in high-dimensional problems with Metropolis algorithms, the optimal acceptance rate is 0.234 (Gelman et al. 1997). However, the range from 0.25 to 0.5 is generally acceptable (Hogg and Foreman-Mackey 2018).

Note that a chain only accepts proposed steps inside the prior limits—inaccurate prior information results in erroneous posterior solutions. In addition, accurate prior information, such as narrow search limits with a normal distribution that centers at the most probable solutions, helps the chains converge quickly to reliable posterior PDFs. For instance, a normal prior distribution initiates more chains around the targeted distribution, thereby costing fewer steps to reach convergence. However, such accurate assumptions may not be available in most field applications. Our MCMCES applications only require reasonable upper and lower search limits with equally distributed probabilities (random distributions) for prior information. After evaluating the likelihood function, the posterior PDFs estimated in the following examples are shown in probability histograms. The most probable solutions are expressed by means and STDs. This paper shows the posterior PDFs as color histograms in probability for demonstration purposes. However, the uncertainty statistics can also be presented in percentiles or error bars.

**Fig. 1** depicts the workflow of our method. We develop two separate MCMCES applications for multiminer analysis. First, we recommend assessing the posterior PDFs in endpoints to evaluate the uncertainties from well-log data. Then, the validated endpoints compute the posterior PDFs in volume fractions. Depending on the goal of the Bayesian inversion, the model parameter ( $\theta$ ) in Eq. 5 can be the volume fraction ( $V$ ), which is a linear simulation, or the endpoints ( $G$ ), which is a nonlinear simulation. Iteration of our method can be done by changing the search limits or precision constant to optimize the results. The following sections demonstrate the two different applications with synthetic and field examples. The model space of volume fractions is linear and thereby easier to illustrate than the model space of endpoints. Thus, we first present our method for estimating the posterior PDFs in volume fractions.



**Fig. 1—The suggested workflow for MCMCES applications in multiminer analysis. The workflow starts with estimating the uncertainties in endpoints to identify the most probable endpoint matrices. After determining the endpoints, we assess the uncertainties in volume fractions.  $\sigma$  represents the precision constant to determine the precision of the targeted distributions. The two applications are independent and can be applied separately.**

Evaluating computational time is a critical factor for Monte Carlo methods to be practical. The time estimated for the examples is based on a Windows 10 desktop with an i7 CPU at 3.47 GHz and 24 GB memory. MCMCES can be programmed with parallel computation by separating the ensembles into different subsets. However, parallel computation is not implemented in our codes for demonstration purposes.

### Estimation of Uncertainty in Volume Fractions

We present a synthetic case with three constituents and a field example of a carbonate reservoir using a quad-combo logging set. MCMCES is applied to the volume fractions ( $\vec{V}$ ), while the endpoints ( $G$ ) are constants in these applications. We use the linear mixing law (Eq. 1) for simplicity and practical purpose, which is also a well-accepted approach and demands less computational time. However, the linear approximation is not a requirement for employing MCMCES. Our method can potentially be applied to nonlinear models (Heidari et al. 2012) as well.

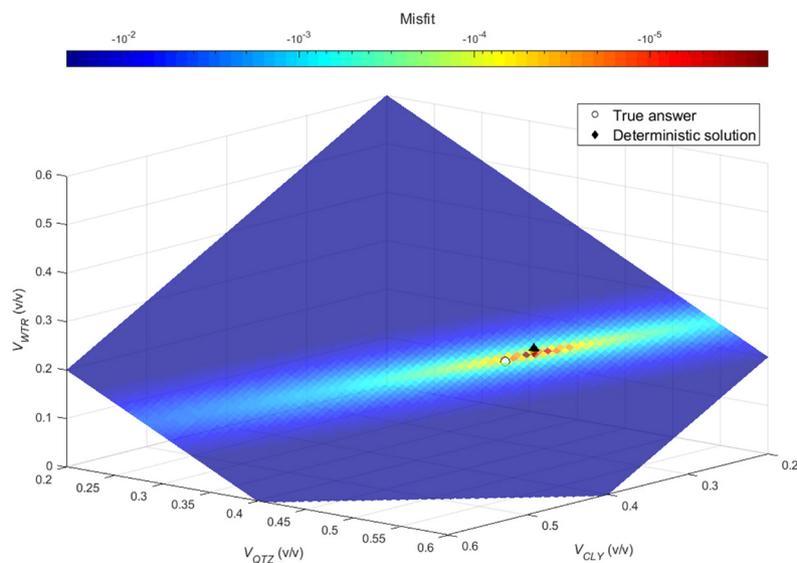
**Synthetic Data Example.** We illustrate our method with a simple synthetic example with a one-depth sample of three constituents (clay, quartz, and water). This model helps to illustrate how MCMCES samples the model space to assess the uncertainties in volume fractions. The simulated well log data are NPHI and RHOB using Eq. 1 with 2% random noises. The endpoints used in this example are NPHI [0.35, -0.02, 1.0] (v/v) and RHOB [2.79, 2.65, 1.0] (g/cm<sup>3</sup>) for clay, quartz, and water, respectively. The unity is an additional equation (Eq. 3) to ensure the inverse problem is well-determined and the solutions are realistic.

The likelihood function ( $p(\vec{d} | \vec{V})$ ) for estimating the posterior PDF in volume fractions at every depth is expressed by

$$p(\vec{d}|\vec{V}) \propto \exp \left[ -\frac{1}{2} \frac{(\vec{G}\vec{V} - \vec{d})^T (\vec{G}\vec{V} - \vec{d})}{\sigma} \right], \quad (8)$$

where  $\sigma$  is the precision constant determining the acceptable misfit. The term in the exponential square bracket represents the evaluation of the data misfit from a given set of volume fractions. The proposed step is accepted in each chain if the resultant misfit is lower than the current step. If the misfit is higher, the step can only be acceptable by a probability determined from the Metropolis-Hastings algorithm (Eq. 7). When the misfit is close to zero, the likelihood function is close to one. Thus, the targeted distribution is stationary steps around one to determine the convergence in a trace plot.

First, a grid search approach is performed to illustrate the model space of the volume fractions, as shown in Fig. 2. The grid search returns the misfits of quartz and clay combinations ( $V_{QTZ}$  and  $V_{CLY}$ ) from 0.2 to 0.6 in a 0.005 spacing where the water volume fractions are computed from  $1 - (V_{QTZ} + V_{CLY})$ . The three axes represent quartz, clay, and water volume fractions. The colored surface in Fig. 2 illustrates the misfits of combinations with the sums of three-volume fractions equaling one. The oval shape of the misfits depicts the different uncertainties in volume fractions. The uncertainty in the vertical axis representing the water content is smaller than the other two horizontal axes (quartz and clay). The uneven uncertainties result from the values in endpoints as quartz and clay have greater similarity in RHOB and NPFI than water. Because of the data errors, the correct solution is away from the minimum misfit that the deterministic inversion finds. Therefore, finding plausible realizations within an acceptable misfit is more reasonable than finding a single solution with the lowest misfit. The Bayesian inversion draws samples from the model space and maps the posterior distributions with the occurrence/probability of these realizations proportional to misfits' absolute values.



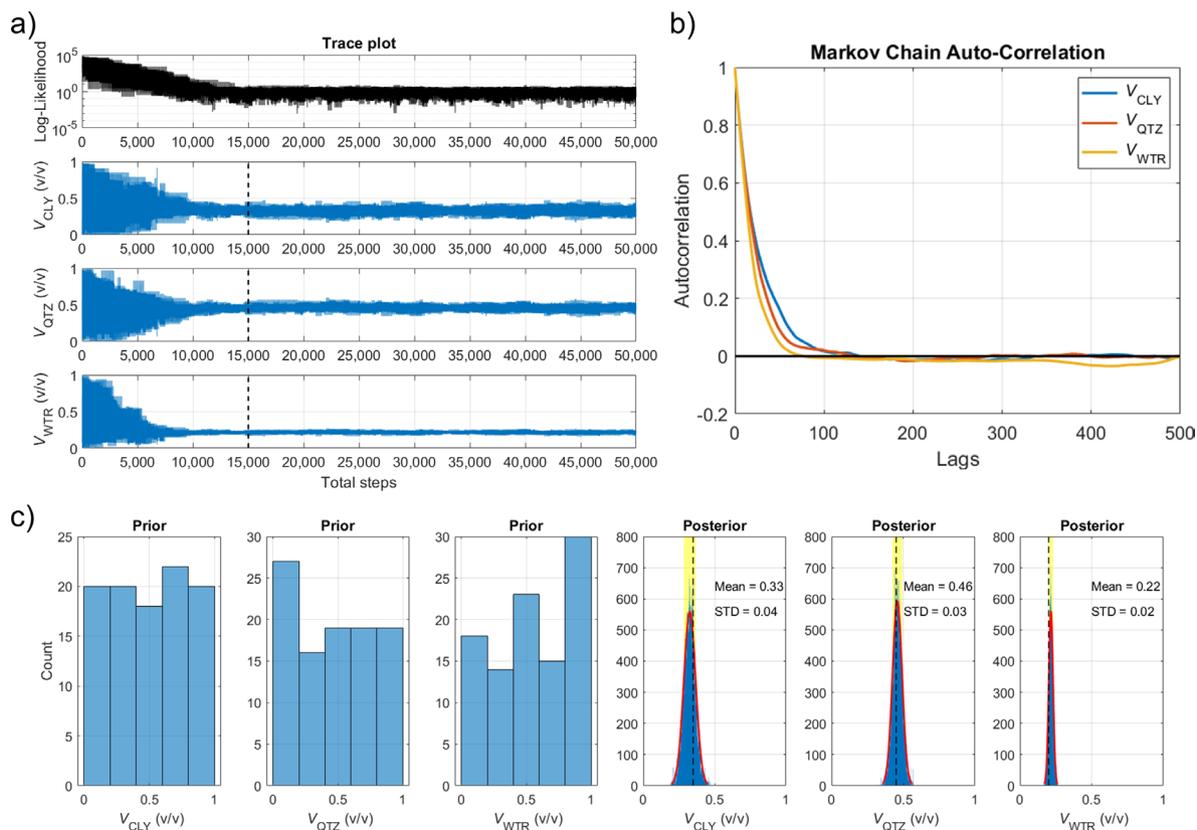
**Fig. 2—The model space of the synthetic composite with three constituents (quartz, clay, and water). The surface represents the combinations tested by a grid search approach. The quartz and clay volume fractions ( $V_{QTZ}$  and  $V_{CLY}$ ) are evaluated from 0.2 to 0.6 with a spacing of 0.005, and the water volume fraction is estimated from  $(1 - V_{QTZ} - V_{CLY})$ . Only the sums of three constituent volume fractions equaling one are shown. The colors on the surface illustrate the misfits of the evaluated combinations.**

Our method assumes random distributions within searching ranges for the prior information. The broadest range for the volume fractions is  $[0, 1]$ . This example employs 100 ensemble walkers randomly deployed in the model space. Each walker takes 500 steps. Trace plots of the combined steps are shown in Fig. 3a. The first row of Fig. 3a illustrates the likelihood function decreasing from high initial values to around one as the steps increase. Rows 2–4 of Fig. 3a show the corresponding volume fraction trace plots. The burn-in of 0.3 is determined where the likelihood function has reached the target distribution and the volume fractions converge. The posterior PDFs are the solutions after burn-in. Fig. 3b shows the autocorrelations of each model parameter. The rapid decreases in autocorrelations along lags indicate fast convergence. Fig. 3c shows random prior and posterior PDFs for clay, quartz, and water volume fractions. The posterior PDFs estimate the means and standard deviations. The correct solutions conform to the ranges defined by the standard deviations. As shown in Fig. 2, clay and quartz volume fractions exhibit higher standard deviations (0.04 and 0.03) than the standard deviation of water volume fraction (0.01).

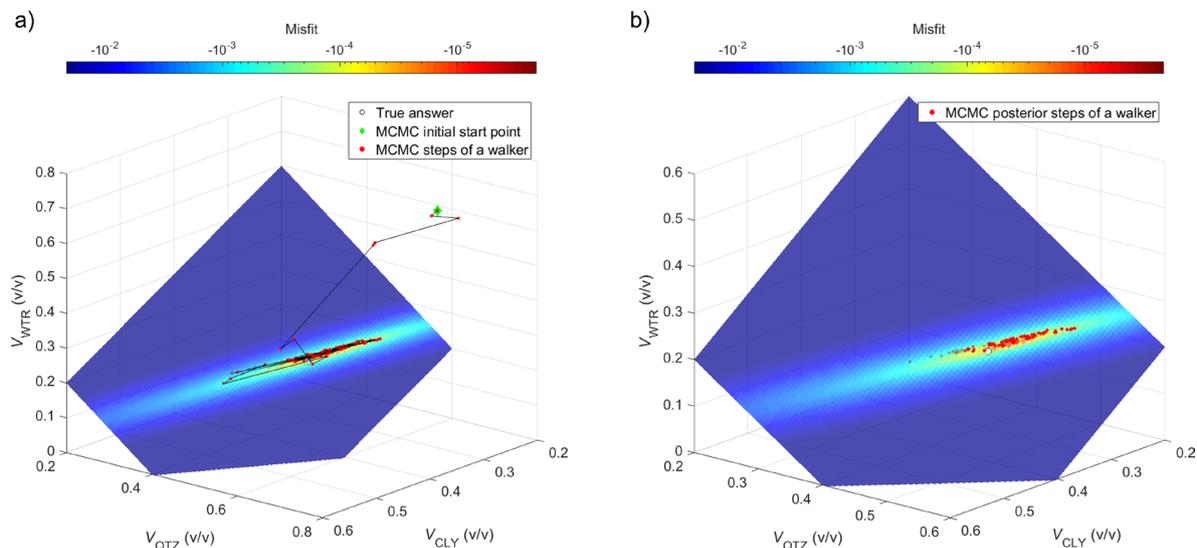
Fig. 4a shows an example of the path one of the ensemble walkers has traveled. Even though the initial random state is unrealistic, the walker finds the surface (as unity constraint) in five steps, locates the low misfit area in ten steps, and samples the area ever since. In MCMCES, even though each walker travels simultaneously and trades the proposal distribution, each walker explores the model space independently. Fig. 4b shows the posterior steps of the walker that only sample the area whose misfit is lower than the targeted value. The final posterior PDFs represent the unbiased contributions from all the ensemble walkers and sample the model space in volume fractions given a preset precision constant.

The synthetic example demonstrates that overfitting the data leads to erroneous inverted solutions because of data errors. Also, determining a proper precision constant is key to estimating uncertainties from noisy data. Posterior PDFs represent plausible solutions, thereby quantifying the uncertainties. Above are the general rules applied to both Bayesian inversions in volume fractions and endpoints, explained in the next section.

**Field Data Example.** Next, we demonstrate our method using a set of quad-combo logs from a carbonate reservoir. The zone of interest is from the measured depth of 11,450 to 11,700 ft with a sample interval of 0.5 ft. The host rock consists of anhydrite, dolomite, calcite,



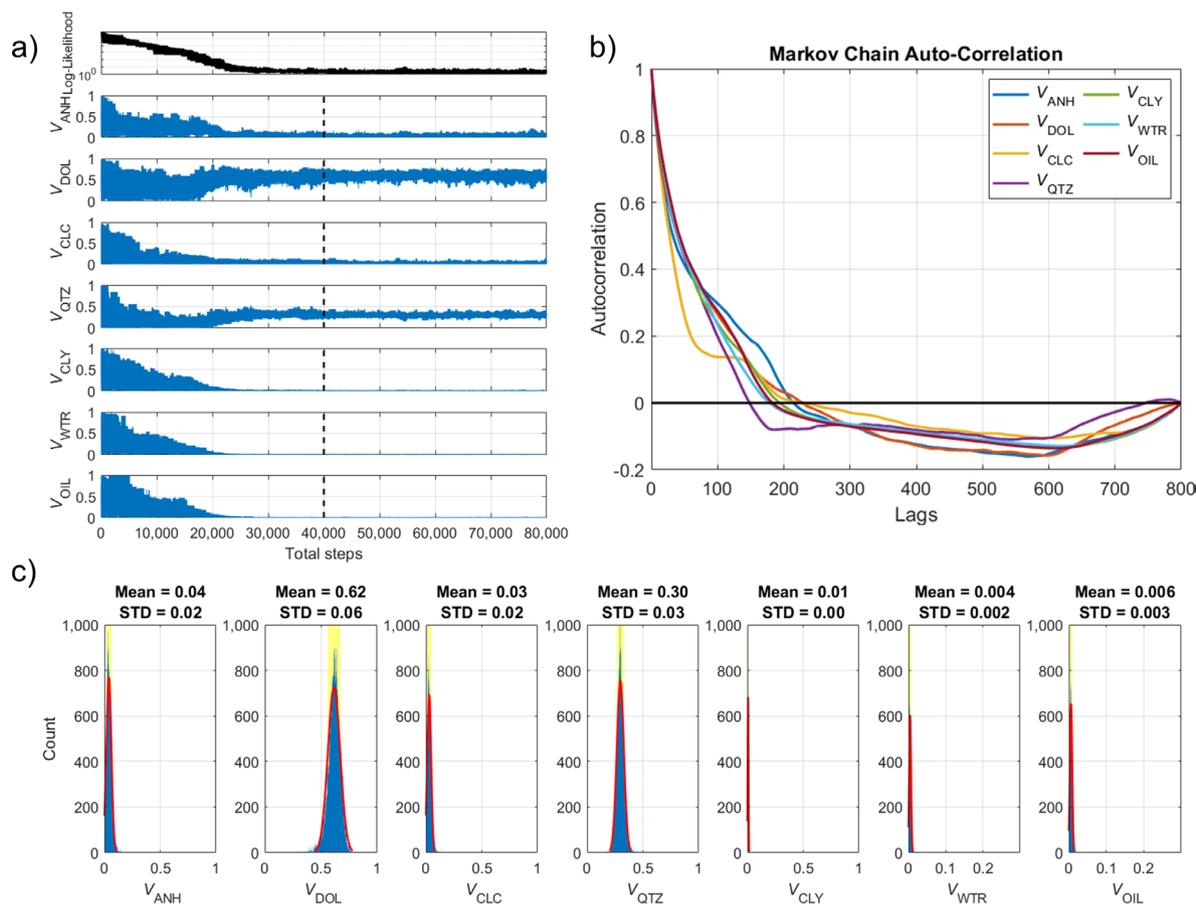
**Fig. 3—MCMCES diagnostic panels for the Bayesian inversion of the three-constituent composite. (a)** Trace plots of the misfits, volumes of clay, volumes of quartz, and volumes of water. The burn-in is determined where the likelihood function and traces converge to stationary iterations, marked by the black dashed line. **(b)** Autocorrelations of model parameters. The autocorrelations reduce rapidly along with the lag, indicating fast convergences. **(c)** Prior and posterior distributions. MCMCES does not require specifying a distribution for prior information. Random distributions are used for the cases in the paper. Posterior PDFs are characterized by histograms, means, and standard deviations. The correct solutions (black dashed lines) fall between the ranges defined by means and standard deviations (yellow areas).



**Fig. 4—Steps of one ensemble walker for the inversion of the three-constituent composite. This example employs 100 ensemble walkers. However, only one of the ensemble walkers is shown for demonstration purposes. (a)** Total 500 steps of the ensemble walker. The green diamond indicates the random initial state of the walker. **(b)** The posterior steps after burn-in.

quartz, and clay. Water and oil are the fluids in the pore space. Available well logs are GR, RHOB, NPHI, CX, DT, and U. The inversion is well-determined. The input well logs are weighted by the inverse of their standard deviations.

The endpoints used in this case are shown in **Table 1**. First, we arbitrarily select one of the depth samples to parameterize MCMCES. **Fig. 5a** shows the trace plots of MCMCES at measured depth of 11,570 feet after employing 100 ensemble walkers with 800 steps per



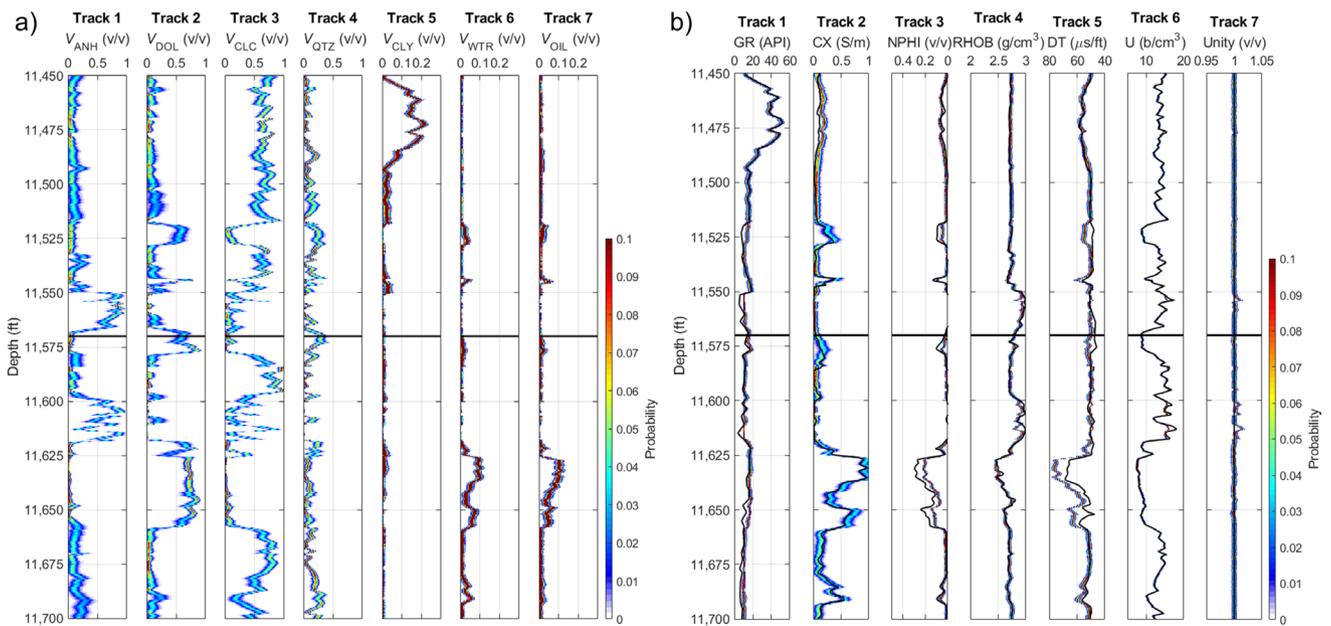
**Fig. 5—MCMCES diagnostic panels for the Bayesian inversion of one depth sample (measured depth of 11,570 ft). (a) Trace plots of the likelihood function and model parameters. The burn-in is determined at the total steps of 40,000 because of the convergences in misfits and volume fractions. (b) Autocorrelations of model parameters show reasonable decays in lags. (c) Posterior PDFs of volume fractions. The dolomite volume fraction exhibits the highest uncertainty (standard deviation = 0.08).**

walker. After 40,000 total steps, the likelihood function reaches the target distribution, and the iterations in volume fraction converge. Therefore, the burn-in section is selected at 50% of the total steps. The acceptance rate is ideal at the value of 0.23. **Fig. 5b** illustrates the autocorrelations of model parameters. The autocorrelations decrease rapidly along with lags, indicating efficient convergence. The posterior PDFs of volume fractions are shown in **Fig. 5c** as histograms. Dolomite exhibits a mean of 0.62 in volume fraction with the highest uncertainty of 0.06 in standard deviation, whereas clay, water, and oil have less than 0.01.

Constituent	Gamma Ray (API)	Conductivity (S/m)	Density (g/cm <sup>3</sup> )	Neutron Porosity (v/v)	Compressional Slowness (μs/ft)	Volumetric Cross Section (b/cm <sup>3</sup> )
Anhydrite	10	0	2.98	0.01	50.0	14.9
Dolomite	20	0	2.88	0.05	42.0	10.0
Calcite	10	0	2.71	0.0	47.0	16.0
Quartz	5	0	2.68	-0.02	53.0	4.8
Clay	200	0.8	2.79	0.3	85.0	8.0
Water	0	10	1.0	1.0	189.0	0.4
Oil	0	0	0.8	0.95	190.0	0.1

Table 1—Endpoints used for the carbonate field example.

We repeat the same parameters for the entire interval of interest. **Fig. 6a** shows the results of the depth-by-depth MCMCES. The posterior PDFs in volume fractions are displayed in color schemes. The logs reconstructed from posterior PDFs compared with the input logs are shown in **Fig. 6b**. The faithful agreement between the input and reconstructed well logs indicates plausible realizations, conforming to the unity constraint (Track 7 of **Fig. 6b**). The solutions from deterministic and Bayesian inversions are consistent, but the latter explores the model spaces and returns realizations that honor the injection well logs within the acceptable misfits. As shown in the previous synthetic example, the posterior PDFs depict the model space of each constituent volume fraction. The minerals generally have higher uncertainties than the fluids because of their similarity in endpoints compared with the endpoints of fluids. However, the clay volume fraction is less uncertain than other minerals because clay with a relatively high GR reading can be well-constrained by the gamma ray log. The

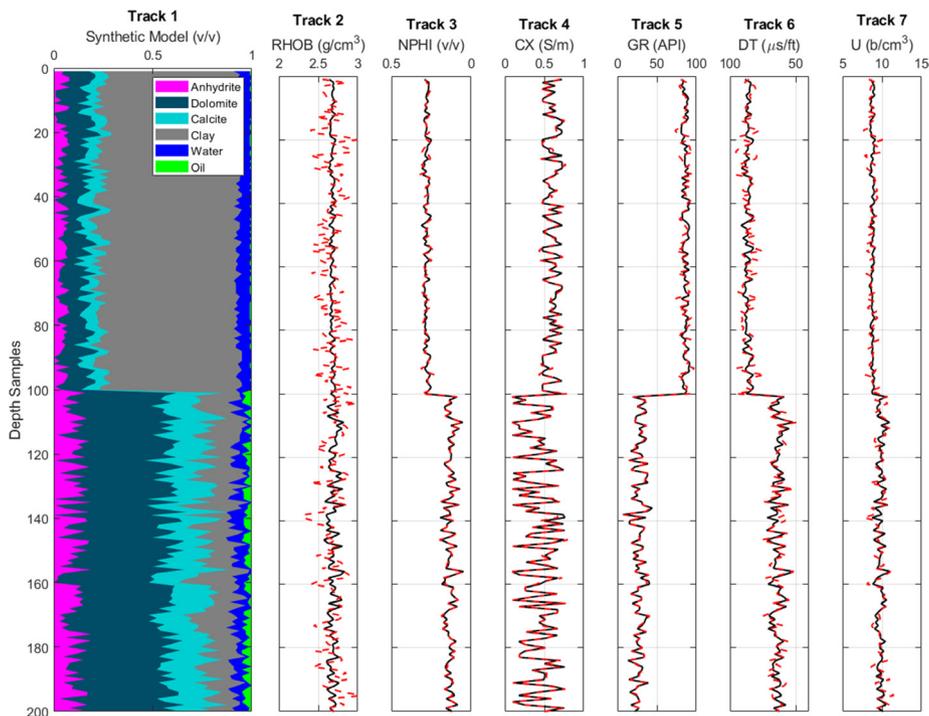


**Fig. 6—Depth-by-depth posterior PDFs and reconstructed logs for the field example. (a) Posterior PDFs in volume fractions are shown in the color schemes. For example, the dolomite volume fraction (Track 2) exhibits the highest uncertainties, whereas clay, water, and oil volume fractions show the lowest. (b) The faithfully reconstructed logs from the posterior solutions indicate the plausibility of the solutions. The solutions also conform to the unity constraint (Track 7).**

uncertainties in volume fractions may change with (1) the noise level in the data, (2) similarity between constituent endpoints, and (3) the weighting of each well log in the inversion. Finally, the computation for this field example with 501 depth samples takes 87 seconds, which is practical for day-to-day use.

### Estimation of Uncertainty in Petrophysical Endpoints (G)

In practice, optimizing endpoints in the multimaterial analysis is time-consuming and commonly adjusted with a trial-and-error approach. Nevertheless, the variation in endpoints may provide local geological information. Moreover, the uncertainties of endpoints have not been



**Fig. 7—A synthetic model and simulated well logs. Track 1 shows that the synthetic model consists of anhydrite, dolomite, calcite, clay, water, and oil with 200 depth samples. Tracks 2–7 are simulated well logs from the model and endpoints. The black curves are logs without any noise. 5% Gaussian noises are added to the simulated well logs as the red dashed curves. The noise levels appear differently because of the different display scales of the logs.**

statistically assessed by other studies so far. Bayesian inversion of this section aims to automate the endpoint selection with associated uncertainties, which is a nonlinear multivariate task and more computationally intense than estimating uncertainties in volume fractions. The model parameter ( $\theta$ ) in Eq. 5 is now the endpoint matrix ( $G$ ). The misfits from depth-by-depth deterministic inversions are summed along the depths for every step in MCMCES to evaluate a given endpoint matrix. We assume that the endpoints do not change along the depths of interest, which is also the prerequisite for conventional multiminer analysis. Also, the well logs are recorded in a reasonable wellbore condition. The uncertainty assessment provides interpreters with additional statistical information to assist the endpoint selection and analyze local geology by assessing the posterior PDFs in endpoints. We present three synthetic examples to validate our approach before applying MCMCES to a field example.

**Synthetic Data Examples.** We design a synthetic rock model as shown in Track 1 of Fig. 7. The formation consists of anhydrite, dolomite, calcite, clay, water, and oil with 200 depth samples. Depending on rock composition, the formation can be divided into two intervals. The top 100 depth samples represent an interval with a clay-dominated and low-porosity matrix. The underlying interval represents a carbonate-rich and medium-porosity oil-saturated matrix. The well logs, including RHOB, NPHI, CX, GR, DT, and U, are modeled from the volume fractions and endpoints using Eq. 1. The endpoints used in this example are shown in Table 2. The well logs are simulated without and with 5% Gaussian noises (Tracks 2–7 of Fig. 7). The objectives of the synthetic examples are to (1) provide a practical application to assist the endpoint selection, (2) assess the uncertainties associated with endpoints, and (3) understand the limitations when the data are noisy.

Constituent	Gamma Ray (API)	Conductivity (S/m)	Density (g/cm <sup>3</sup> )	Neutron Porosity (v/v)	Compressional Slowness (μs/ft)	Volumetric Cross Section (b/cm <sup>3</sup> )
Anhydrite	10	0	2.98	−0.03	49.0	14.9
Dolomite	10	0	2.87	0.03	43.5	9.0
Calcite	15	0	2.71	0.0	47.5	13.7
Clay	120	0.1	2.79	0.3	90.0	8.7
Water	0	8.2	1.0	1.0	189.0	0.4
Oil	0	0	0.7	0.95	210.0	0.1

Table 2—Endpoints used to simulate the synthetic example.

**Estimating Gamma Ray Endpoints Using Perfect Data.** Among commonly used endpoints, gamma ray endpoints may be inconsistent and vary from place to place. This example validates our method by estimating gamma ray endpoints using perfect data. Unlike the conventional approach, which requires endpoints to be determined before inverting for volume fractions, our approach requires only reasonable upper and lower bounds for uncertain endpoints as prior information for Bayesian inversion. Furthermore, random distributions are assumed. In this case, MCMCES estimates the posterior PDFs of gamma ray endpoints with the given precision constant while other endpoints are correct and constant. The upper and lower limits for anhydrite, dolomite, and calcite gamma ray endpoints are set between [5, 20] API. Whereas clay is commonly more radioactive, the initial steps of ensemble walkers are selected randomly from 80 to 140 API. Note that the fluids generally are not radioactive, and their gamma ray endpoints are zero. For forward modeling, the explicit form of Eq. 1 is given by

$$\begin{bmatrix} RHOB \\ NPHI \\ CX \\ GR \\ DT \\ U \\ 1 \end{bmatrix} = \begin{bmatrix} 2.98 & 2.87 & 2.71 & 2.79 & 1.0 & 0.7 \\ -0.03 & 0.03 & 0.0 & 0.3 & 1.0 & 0.9 \\ 0.0 & 0.0 & 0.0 & 0.1 & 8.2 & 0.0 \\ 5-20 & 5-20 & 5-20 & 80-140 & 0 & 0 \\ 49 & 43.5 & 47.5 & 90 & 189 & 210 \\ 14.9 & 9.0 & 13.7 & 8.7 & 0.4 & 0.1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} V_{ANH} \\ V_{DOL} \\ V_{CLC} \\ V_{CLY} \\ V_{WTR} \\ V_{OIL} \end{bmatrix} \quad (9)$$

For the likelihood function in endpoints, Cheng et al. (2021) use a two-step approach in which the first step is the linear least-square optimization given by

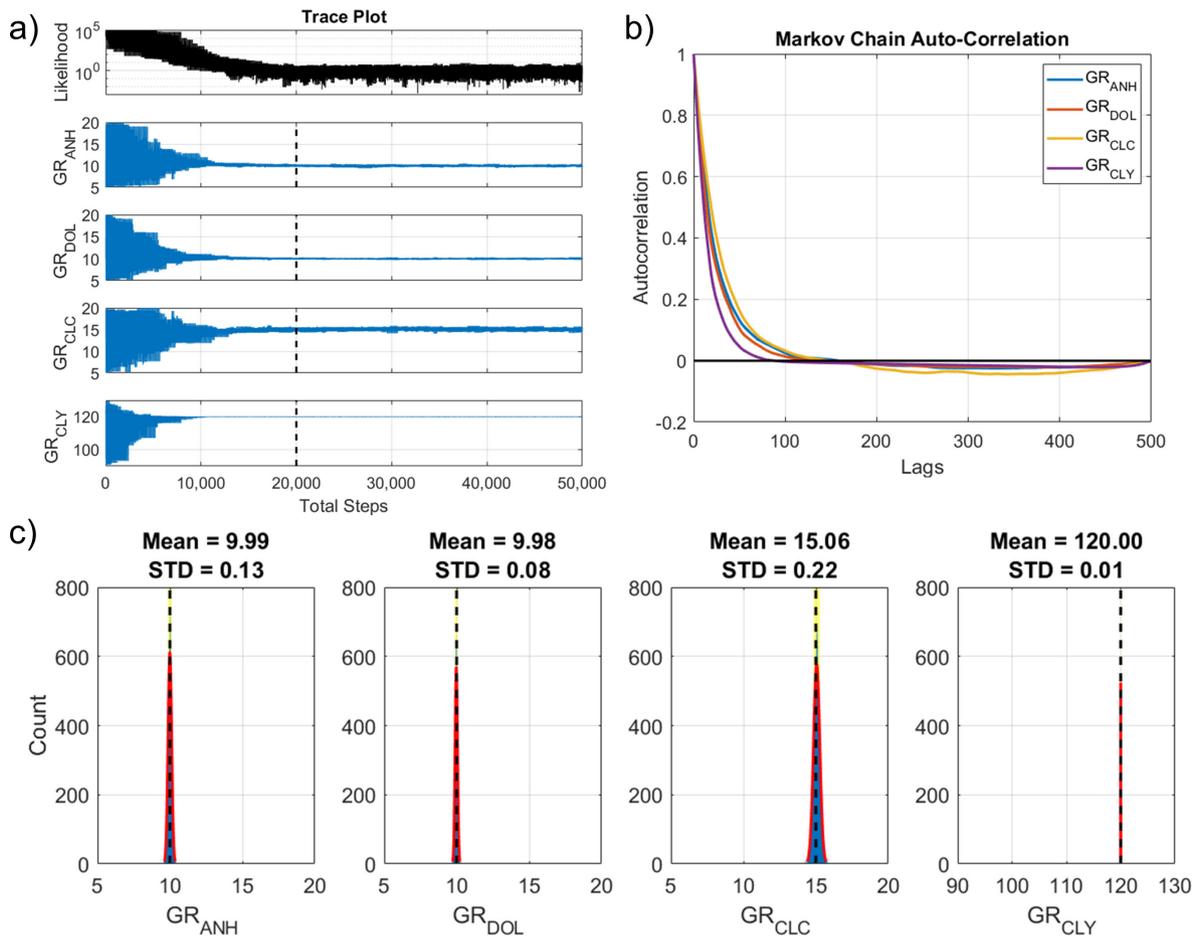
$$\min_{\vec{V}} \left\| G\vec{V} - \vec{d} \right\|_2^2, \quad (10)$$

with the constraint (Eq. 4,  $0 \leq V_i \leq 1$ ) at every depth. The depth-by-depth misfits are then summed along the depths for the likelihood function of each endpoint matrix. A matrix of endpoints that exhibits a low summed misfit is more likely to be accepted than an endpoint matrix with a high misfit. The method proposed by Cheng et al. (2021) shows promising results in estimating the posterior PDFs of endpoints but requires significant computational time for solving constrained linear inversions. For an example of 200 depth samples, if we employ 100 walkers with 500 steps per walker, a total of 10 million ( $200 \times 100 \times 500$ ) depth-by-depth inversions are performed. In this paper, we improve the efficiency of estimating endpoint posterior PDFs by replacing the depth-by-depth constrained inversion with an unconstrained one. The fitness of an endpoint matrix is evaluated by how the resultant volume fractions meet the material balance (Eq. 3). Accordingly, the proposed likelihood function for endpoints is expressed as

$$p(d|G) \propto \exp \left[ -\frac{1}{2} \sum_{j=1}^K \frac{\left(1 - \sum_{i=1}^N V_{ij}\right)^2}{\sigma} \right], \quad (11)$$

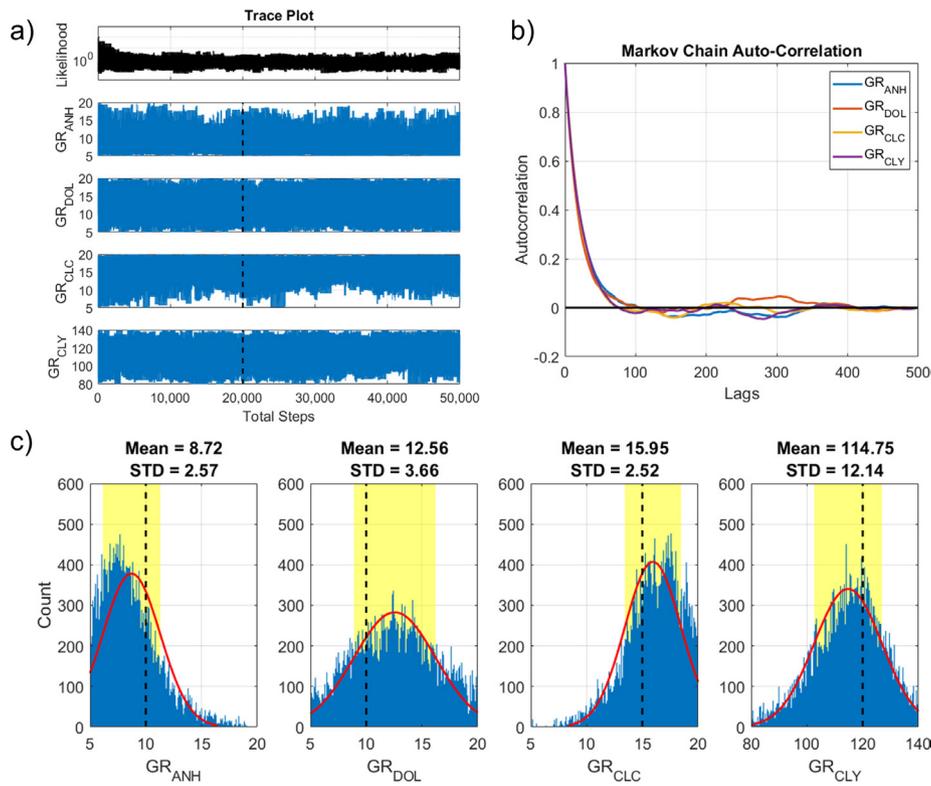
where  $j$  represents the  $j$ th depth sample,  $K$  is the total number of depth samples, and  $N$  is the number of constituents. The prerequisite for MCMCES in endpoints using our proposed likelihood function is that the number of well logs must be equal to or greater than the number of constituents ( $M \geq N$ ). Even though the resultant volume fractions may be unrealistic ( $>1$  or  $<0$ ), the sum of solutions meeting the material balance determines the fitness of the given endpoint matrix. The unconstrained linear inversion reduces the computational time for a more practical application. At each iteration, the ensemble walkers propose a new value in selected endpoints for each chain. If the proposed step is outside the given range, the current value is not evaluated and stays unchanged until the next proposal is inside the expected range. The likelihood function only evaluates the endpoint matrix when the proposal is inside the preset range.

In this synthetic case where input well-log data are noise-free, a small precision constant of  $1E-5$  is used to approximate the correct gamma ray endpoints. The number of constituents (6) equals the number of input logs (6). We employ 100 ensemble walkers, and each walker takes 500 steps. **Fig. 8a** shows the trace plots of the likelihood function and gamma ray endpoints for anhydrite, dolomite, calcite, and clay. While the iterations converge before the combined steps of 20,000, a burn-in of 0.4 is selected. **Fig. 8b** exhibits the autocorrelations of gamma ray endpoints whose values decrease quickly along with the lags, indicating optimal convergence time. The acceptance rate of 0.28 is in a reasonable range. **Fig. 8c** shows the posterior PDFs of gamma ray endpoints, which correctly predict true endpoint values with standard deviations. The gamma ray endpoint of clay is the most sensitive parameter with the smallest standard deviation (0.01), whereas calcite has the least sensitivity. The precise results from the synthetic example with noise-free data validate the feasibility of using the proposed likelihood function (Eq. 11) for MCMCES in endpoints. The computation for this example costs 50 seconds. Using the approach described in Cheng et al. (2021) with the same parameters requires more than one hour of computational time to produce similar results.

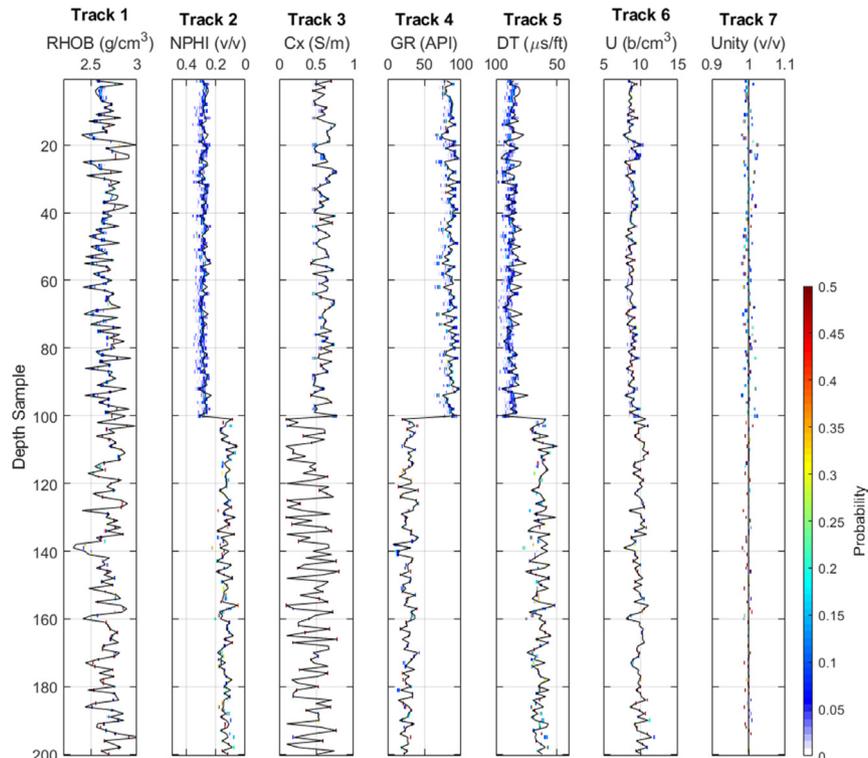


**Fig. 8—MCMCES diagnostic panels for the Bayesian inversion in endpoints. The figures show a synthetic case without noises in the data. (a) Trace plots of the likelihood function and model parameters (GRs for anhydrite, dolomite, calcite, and clay). (b) Autocorrelations of model parameters. (c) Posterior PDFs of model parameters, which correctly predict the solutions with uncertainties.**

**Estimating Gamma Ray Endpoints Using Noisy Data.** The next synthetic case assesses the posterior PDFs of the same gamma ray endpoints, but 5% Gaussian noises are added to all the well logs, as shown in red dashed curves in **Fig. 7**. Even though the prior information keeps the same as the previous noise-free case, the precision constant cannot be as small to account for the noises in the data. Otherwise, overfitting the data leads to erroneous estimations. Hence, a larger precision constant of  $5E-4$  is used for MCMCES in endpoints. In this case, 100 ensemble walkers with 500 steps per walker are used. **Fig. 9a** shows the trace plots of all the steps, and a burn-in parameter of 0.4 is selected. The acceptance rate is 0.33, and the computational time is 24 seconds. **Fig. 9b** illustrates the reasonable



**Fig. 9—MCMCES diagnostic panels for the Bayesian inversion in endpoints. The figures show a synthetic case with noisy data. (a) Trace plots of the likelihood function and model parameters (GRs for anhydrite, dolomite, calcite, and clay). (b) Autocorrelations of model parameters. (c) Posterior PDFs of model parameters. Black dashed lines represent the correct solutions. The yellow areas represent the ranges defined by the means and standard deviations of the posterior PDFs.**



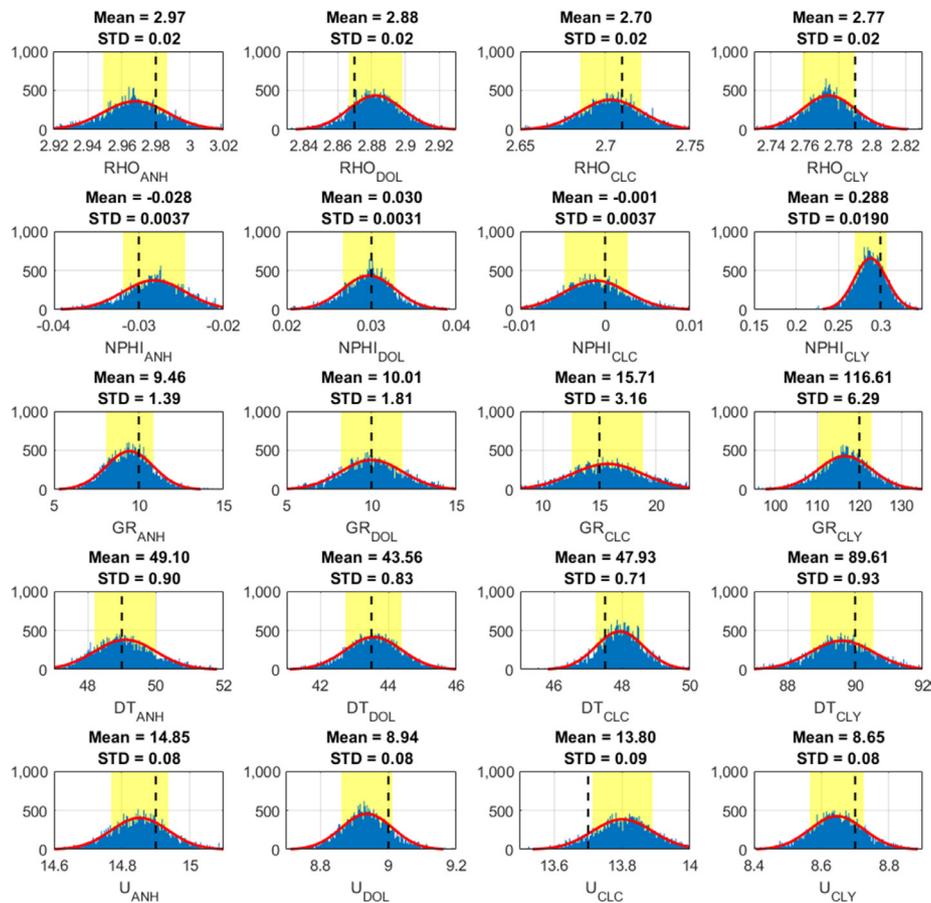
**Fig. 10—Input and reconstructed logs. The input logs are shown in black curves. The reconstructed logs are simulated from constrained inversions with posterior endpoint matrices. The unity constraint is plotted in Track 7.**

convergence time from autocorrelations of model parameters. **Fig. 9c** shows the posterior PDFs of gamma ray endpoints with means and standard deviations under the given precision constant. The predicted ranges defined by means and standard deviations, marked by yellow colors, are the most probable combinations estimated from MCMCES. Note that the uncertainties of endpoints are higher than in the previous case because of the higher value of the precision constant to account for noises. In addition, the correct solutions are all within the predicted ranges but are not the modes of posterior PDFs. The combinations in modes represent the solutions with the lowest misfits.

**Fig. 10** depicts the reconstructed logs in probability compared with the noisy input logs. Note that we use Eqs. 10 and 11 to efficiently evaluate the fitness of individual endpoint matrices. Reconstructing logs to validate the posterior endpoint matrices is required. The reconstructed logs are computed from the posterior endpoint PDFs and the volume fractions, inverted from constrained linear least-square inversions. The faithful matches between input and reconstructed well logs indicate that the computed posterior PDFs in endpoints are plausible. The reconstructed well logs in the top interval show higher uncertainties than the bottom interval because of the high gamma ray reading from the high clay content of the top interval. DT and NPFI exhibit higher uncertainties among logs because their lower weightings are used in the inversion. This example shows that estimating the endpoint uncertainties is feasible even with noisy well-log data.

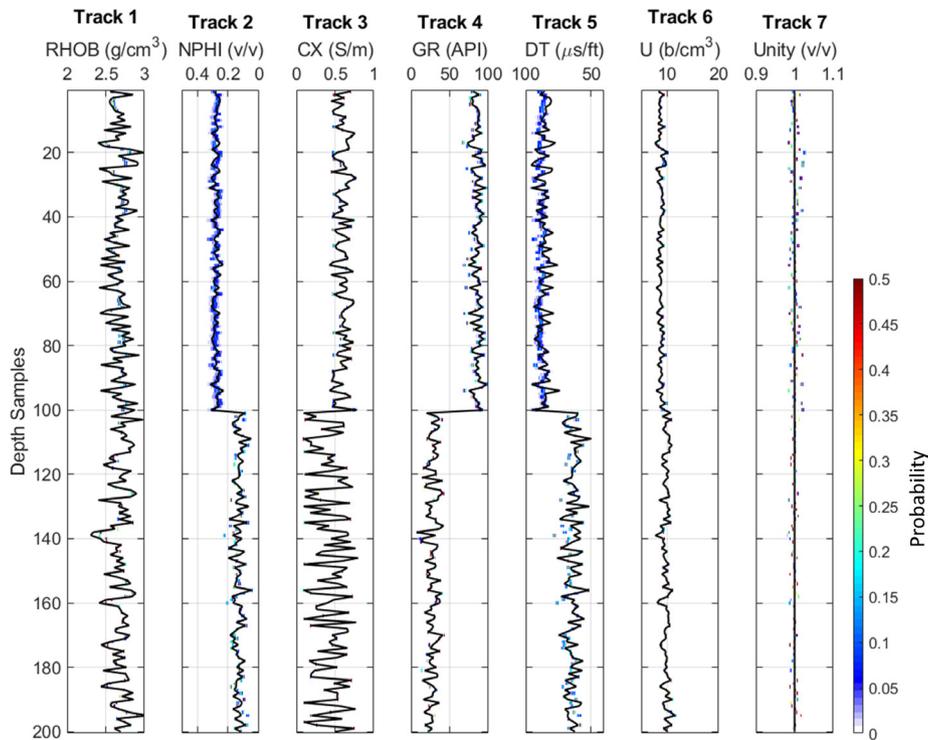
**Estimating 20 Endpoints Using Noisy Data.** The previous examples illustrate how to estimate uncertainties in selected endpoints through the posterior PDFs, while the rest of the endpoints remain fixed. However, adjusting endpoints is a multivariate problem. If any endpoints are inappropriately determined, the resultant posterior PDFs of selected endpoints are likely affected. Therefore, this example assesses up-to-20-endpoint posterior PDFs, similar to the field application where interpreters are uncertain about selecting endpoints. Our method helps estimate posterior endpoint combinations from the given well-log data. We continue using the noisy well logs (**Fig. 7**) and compute the posterior PDFs for RHOB, NPFI, GR, DT, and U endpoints of all the solid constituents (anhydrite, dolomite, calcite, and clay). Note that the endpoints for CX are not included because the conductivities of nonmetal minerals are commonly assumed to be zero, except for minor conductivity from clay. The explicit ranges of endpoints are given by

$$\begin{bmatrix} RHOB \\ NPFI \\ CX \\ GR \\ DT \\ U \\ 1 \end{bmatrix} = \begin{bmatrix} 2.92 - 3.02 & 2.83 - 2.93 & 2.65 - 2.75 & 2.73 - 2.83 & 1.0 & 0.7 \\ -0.04 - -0.02 & 0.02 - 0.04 & -0.01 - 0.01 & 0.15 - 0.35 & 1.0 & 0.9 \\ 0.0 & 0.0 & 0.0 & 0.1 & 8.2 & 0.0 \\ 5 - 15 & 5 - 15 & 8 - 23 & 95 - 135 & 0 & 0 \\ 47 - 52 & 41 - 46 & 45 - 50 & 87 - 92 & 189 & 210 \\ 14.6 - 15.1 & 8.7 - 9.2 & 13.5 - 14.0 & 8.4 - 8.9 & 0.4 & 0.1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} V_{ANH} \\ V_{DOL} \\ V_{CLC} \\ V_{CLY} \\ V_{WTR} \\ V_{OIL} \end{bmatrix}. \quad (12)$$



**Fig. 11—Endpoint posterior PDFs with means and standard deviations. The black dashed lines mark the correct solutions. The most probable ranges defined by the means and standard deviations are colored in yellow.**

In Eq. 12, endpoints without ranges remain constants during iterations. The precision constant is set to 0.002, while 100 ensemble walkers with 500 steps per walker are used for MCMCES. The computation takes 22 seconds, while the acceptance rate is 0.28. **Fig. 11** demonstrates the posterior PDFs of selected endpoints with marked correct solutions. The most probable endpoints with uncertainty statistics are displayed with means and standard deviations. The true answers generally fall within the predicted standard deviations. For instance, even though the NPFI endpoint for clay starts with a wide range in [0.15, 0.35], MCMCES identifies the faithful posterior PDF that is not at the center of the prior range. **Fig. 12** shows the comparison of the reconstructed and injection well logs, including the unity constraint. The good matches validate the plausibility of the estimated endpoint matrices. Finally, the 20-endpoint example shows that our MCMCES in endpoints can assess the uncertainties and provide useful parameter information to assist the selection of endpoints.

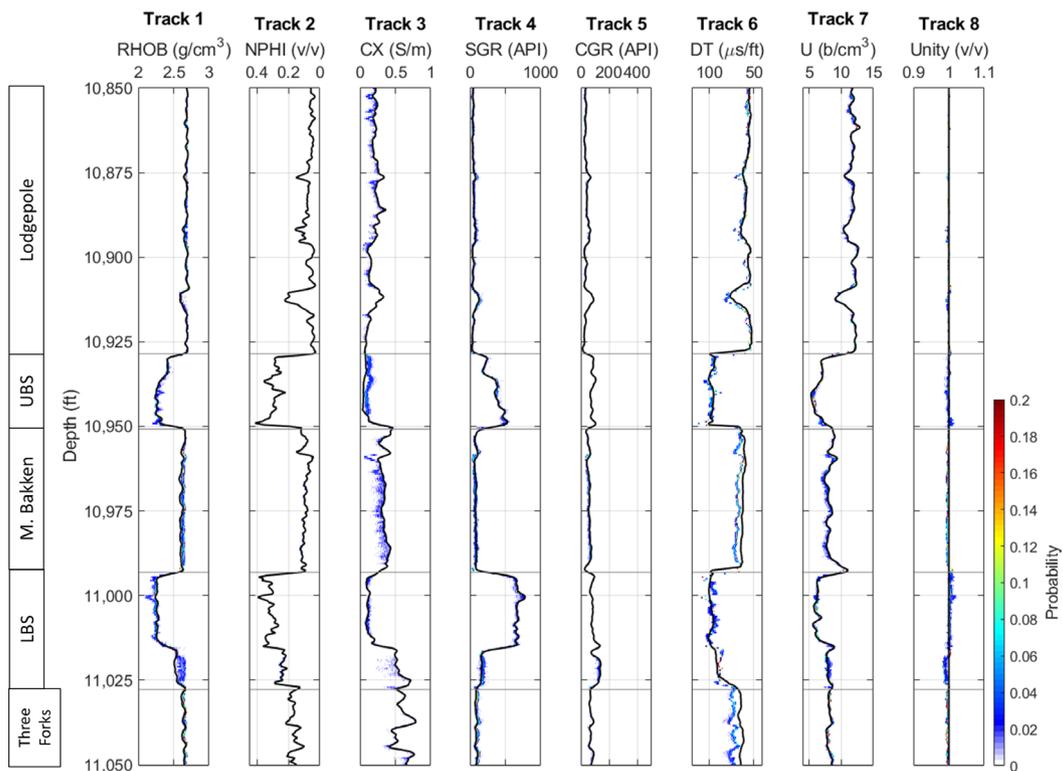


**Fig. 12**—Input and reconstructed logs for the synthetic example of 20 endpoints with noisy data. The reconstructed logs are simulated from constrained deterministic inversions with endpoint posterior PDFs and are shown in color schemes. The unity constraint is plotted in Track 7.

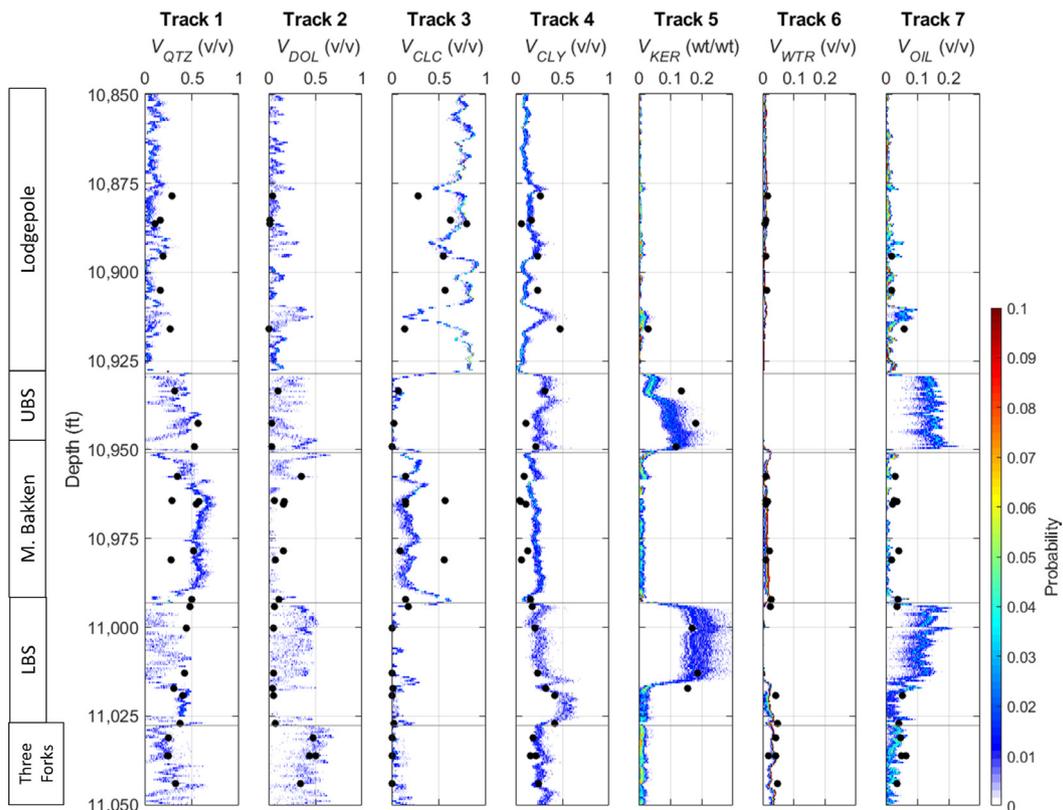
**Field Data Example.** We apply our method to a suite of well logs from the unconventional Bakken Shale Formation (North Dakota). The available well logs are RHOB, NPFI, CX, standard gamma ray (SGR), computed gamma ray (CGR) from thorium and potassium, DT, and U, as shown in **Fig. 13**. The interval of interest includes the Lodgepole, Upper Bakken Shale, Middle Bakken, Lower Bakken Shale (including a section of Pronghorn), and Three Forks Formations. Core X-ray diffraction (XRD) data identify quartz, dolomite, calcite, clay, and kerogen composing the host rock with quantitative measurements. Clay mostly consists of illite. Water and oil are the fluids in the pore space. The SGR in Upper Bakken Shale and Lower Bakken Shale exhibits extremely high readings (~800 API) because of the high uranium content of kerogen in place. Approximating the kerogen gamma ray endpoint can be challenging because there is a wide range of plausibility. In addition, other gamma ray endpoints may affect the approximation in the trial-and-error process. Therefore, applying MCMCES in SGR and CGR endpoints helps identify posterior combinations from the given well logs by only setting reasonable ranges of endpoints as prior information. The ranges of endpoints and explicit form of Eq.1, in this case, can be expressed as

$$\begin{bmatrix} RHOB \\ NPFI \\ CX \\ SGR \\ CGR \\ DT \\ U \\ 1 \end{bmatrix} = \begin{bmatrix} 2.65 & 2.85 & 2.71 & 2.79 & 1.35 & 1.2 & 0.8 \\ -0.03 & 0.08 & 0.0 & 0.25 & 0.03 & 1.0 & 0.95 \\ 0.0 & 0.0 & 0.0 & 0.1 & 0.0 & 12.0 & 0.0 \\ 0 - 20 & 0 - 20 & 0 - 20 & 150 - 400 & 1,500 - 4,500 & 0 & 0 \\ 0 - 15 & 0 - 15 & 0 - 15 & 100 - 300 & 20 - 170 & 0 & 0 \\ 55 & 45 & 47 & 110 & 120 & 189 & 210 \\ 6.0 & 9.4 & 13.0 & 10.0 & 0.6 & 0.6 & 0.1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} V_{QTZ} \\ V_{DOL} \\ V_{CLC} \\ V_{CLY} \\ V_{KER} \\ V_{WTR} \\ V_{OIL} \end{bmatrix} \quad (13)$$

In this field case, we use 100 ensemble walkers with 600 steps per walker and a precision factor of 0.1. The burn-in is 0.4, with an acceptance rate of 0.32. **Fig. 13** shows the reconstructed logs in probability compared with the input logs. The inverted volume fractions are shown in **Fig. 14** and are consistent with XRD and core measurements. **Fig. 15** illustrates the crossplots of the core measurements and inverted results with error bars. **Fig. 16** demonstrates the posterior PDFs in gamma ray endpoints of solid constituents. Note that the volume fractions in **Fig. 14** are realizations that are inverted linearly from the well logs and endpoint posterior PDFs with the constraint of Eq. 4. The faithfully reconstructed logs and good matches between inverted volume fractions and XRD data indicate plausible endpoint

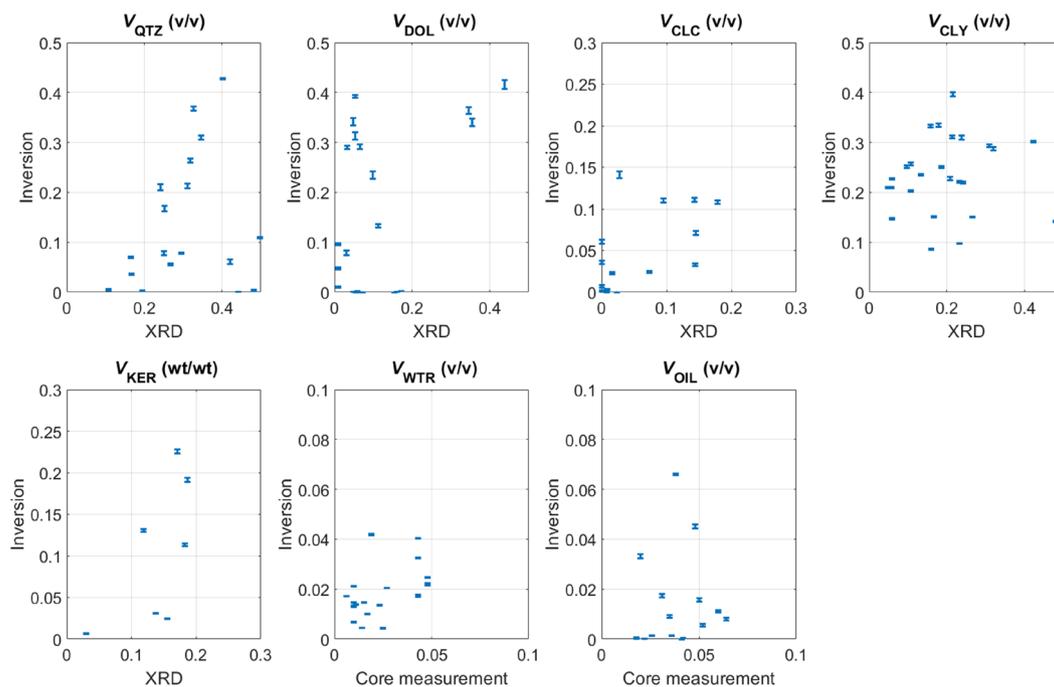


**Fig. 13—Input (black) and reconstructed (color scheme) logs for the Bakken field example.**

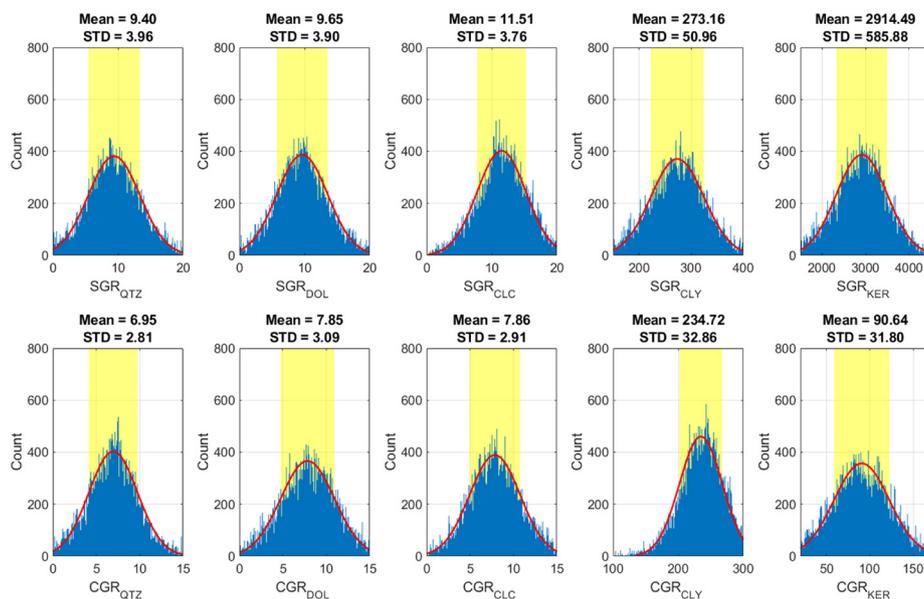


**Fig. 14—The resultant volume fractions from linear constrained inversions with posterior endpoint PDFs. XRD and core data are plotted as black dots to cross-check the inversion results.**

combinations. The posterior PDFs in gamma ray endpoints assist interpreters in determining the most probable combinations from the given well-log data. For instance, the prior SGR limits for kerogen start with a wide range of [1,500, 4,500], but the posterior PDF estimates a mean of 2,914 API with a standard deviation of 585 API. The high uncertainties in kerogen volume fraction result from the wide



**Fig. 15—Crossplots of the core measurements and inversion results. The error bars represent 95% confidence intervals.**



**Fig. 16—Posterior PDFs in SGRs and CGRs. The most probable range for kerogen in SGR is a mean of 2,914 API with a standard deviation of 585 API.**

range of kerogen gamma ray searching limits. Additional iterations of MCMCES can be done by narrowing the prior limits of kerogen gamma ray to approximate its posterior distribution further. The computation of this example requires 53 seconds.

## Discussion

Bayesian inversion is a statistical tool to sample the targeted distributions of the model parameters. The posterior PDFs represent the most probable realizations suggested by the well log data, targeted misfit, selected constituents, and petrophysical models (equations). The posterior PDFs can be affected by data errors, rock composition, similarity in endpoints, and weightings of equations. Our implementation shows posterior PDFs in histograms quantified by the mean and standard deviation. The mode of a posterior PDF may not be a good measure to represent the Bayesian statistics as the synthetic example shows that adapting only the modes is likely overfitting the data. Posterior solutions within the ranges defined by the mean and standard deviation are equally plausible. Determining the value of the precision constant is key to estimating meaningful posterior PDFs and depends on the noise level in the data. If the noise level is unknown, we suggest starting with a larger value and reducing it iteratively.

Our methods automate the selection of model parameters and only require the upper and lower limits as prior information with random distributions. However, setting reasonable limits helps estimate reliable posterior PDFs and reduces convergence time. Defining the

proper bounds and diagnosing posterior PDFs still rely on the interpreters' knowledge. For instance, the range of 0–1 is the largest possible range for volume fractions. However, the fluid volume fractions should be between 0 and 0.5 if the formation is known to be consolidated. The ranges of endpoints can be from interpreters' local experience or 15% above and below the standard values. In addition, some endpoints should be determined as prior information to reduce the model dimensions. For instance, the gamma ray of fluid or conductivity of most minerals (except clay or metals) should be zero or close to zero. We recommend using the endpoint values directly measured in the research area and using our method for endpoints with higher uncertainty. For instance, our approach may approximate water conductivity, but we suggest obtaining conductivities from available water sample measurements or Pickett plots.

## Conclusions

We implement Bayesian inversions to quantify the uncertainties of multiminer analysis. MCMCES in the Bayesian framework shows efficient convergence with fewer tuning parameters, most suitable for multiminer analysis. We evaluate the uncertainties in volume fractions or endpoints with different likelihood functions. The suggested workflow starts with endpoint estimation and then assesses the uncertainties in volume fractions with the given endpoint matrix. However, the two Bayesian inversions are independent and can be applied separately for different purposes. For instance, the well-by-well variation in endpoints may provide important insights into the geology of the research area. In addition, the endpoint estimation may be used to identify normalization issues of well logs if the posterior endpoint PDFs of a certain log drift from expected values. Moreover, if one is confident in their endpoint matrix, one may directly estimate the posterior PDFs in volume fractions. One limitation of our method is that estimating endpoint uncertainties may not be used on the well-log data recorded in a bad-hole condition. The resultant endpoint estimation may be erroneous.

We show comprehensive diagrams of synthetic and field examples to diagnose the performance of the Bayesian inversions. Our method only requires defining reasonable upper and lower limits for model parameters with random distributions as prior information. The synthetic examples validate our method of estimating the posterior PDFs that conform to the target distributions. The number of uncertain endpoints can be up to 20. The iterative finetuning can be done by adjusting the search ranges of the model parameters to better approximate the targeted distributions. The field examples demonstrate efficient computations in posterior PDFs of volume fractions or endpoints to assess uncertainties and assist interpretation.

## Nomenclature

$d$	= well log data
$G$	= constituent petrophysical endpoints
$k$	= thinning parameter
$K$	= number of depth samples
$M$	= number of well logs used in the multiminer analysis
$n$	= number of model parameters
$N$	= number of constituents selected in multiminer analysis
$V$	= vector of volume fraction, $v/v$
$X$	= walker in a Markov chain
$Y$	= randomly selected ensemble walker to propose a step from the current state
$Z$	= ensemble-tuned proposal distribution
$\theta$	= model parameter
$\pi$	= probability density function
$\sigma$	= precision constant

## Acknowledgments

The authors would like to express gratitude to the reviewers and editors of *SPE Reservoir Evaluation & Engineering* for their technical and editorial feedback. We thank Chris Skelt for the technical discussion and Patricia Rodrigues for her help finding the Bakken data set. The authors also thank the members of the Reservoir Characterization Project (RCP) at Colorado School of Mines for their financial support. The MATLAB code for the synthetic example in estimating volume fractions can be found at [https://github.com/lchengCSM/SPE\\_Bayesian\\_Multimin](https://github.com/lchengCSM/SPE_Bayesian_Multimin).

## References

- Ajayi, T., Gomes, J. S., and Bera, A. 2019. A Review of CO<sub>2</sub> Storage in Geological Formations Emphasizing Modeling, Monitoring and Capacity Estimation Approaches. *Pet Sci* **16** (5): 1028–1063. <https://doi.org/10.1007/s12182-019-0340-8>.
- Aldred, R. 2018. Monte Carlo Processing of Petrophysical Uncertainty. Paper presented at the SPWLA 59th Annual Logging Symposium, London, UK, 2–6 June. SPWLA-2018-TTT.
- Benson, S. M. and Cole, D. R. 2008. CO<sub>2</sub> Sequestration in Deep Sedimentary Formations. *Elements* **4** (5): 325–331. <https://doi.org/10.2113/gselements.4.5.325>.
- Cheng, L., Jin, G., and Michelena, R. J. 2021. Bayesian Inversion for Rock Composition and Petrophysical Endpoints in Multiminer Analysis. In *SEG Technical Program Expanded Abstracts*. Houston, Texas: Society of Exploration Geophysicists. <https://doi.org/10.1190/segam2021-3582154.1>.
- Cheng, L., Prasad, M., Michelena, R. J. et al. 2022. Using Rock-Physics Models to Validate Rock Composition from Multiminer Log Analysis. *Geophysics* **87** (2): MR49–MR62. <https://doi.org/10.1190/geo2020-0918.1>.
- Deng, T., Ambia, J., and Torres-Verdín, C. 2019. Fast Bayesian Inversion Method for the Generalized Petrophysical and Compositional Interpretation of Multiple Well Logs with Uncertainty Quantification. Paper presented at the SPWLA 60th Annual Symposium, The Woodlands, Texas, USA, 15–19 June. SPWLA-2019-FFFF. [https://doi.org/10.30632/T60ALS-2019\\_FFFF](https://doi.org/10.30632/T60ALS-2019_FFFF).
- Deng, T., Ambia, J., Torres-Verdín, C. et al. 2020. Bayesian Method for Rapid Multi-Well Interpretation of Well Logs and Core Data in Unconventional Formations. Paper presented at the SPWLA 61st Annual Online Symposium, Virtual, 24 June–29 July. SPWLA-5040. <https://doi.org/10.30632/SPWLA-5040>.
- Doveton, J. H. 1994. *Geologic Log Analysis Using Computer Methods*. Vol. 2. Tulsa, Oklahoma, USA: American Association of Petroleum Geologists. <https://doi.org/10.1306/CA2580>.
- Duane, S., Kennedy, A. D., Pendleton, B. J. et al. 1987. Hybrid Monte Carlo. *Phys Lett* **195** (2): 216–222. [https://doi.org/10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X).
- Forman-Mackey, D., Hogg, D. W., Lang, D. et al. 2013. Emcee: The MCMC Hammer. *PASP* **125** (306). <https://doi.org/10.1086/670067>.

- Gelman, A., Gilks, W. R., and Roberts, G. O. 1997. Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *Ann Appl Probab* 7 (1): 110–120. <https://doi.org/10.1214/aoap/1034625254>.
- Goodman, J. and Weare, J. 2010. Ensemble Samplers with Affine Invariance. *Commun Appl Math Comput Sci* 5 (1): 65–80. <https://doi.org/10.2140/camcos.2010.5.65>.
- Grinsted, A. 2020. Ensemble MCMC Sampler. GitHub. <https://github.com/grinsted/gwmcnc> (accessed 3 December 2020).
- Hastings, W. K. 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* 57 (1): 97–109. <https://doi.org/10.1093/biomet/57.1.97>.
- Heidari, Z., Torres-Verdín, C., and Preeg, W. E. 2012. Improved Estimation of Mineral and Fluid Volumetric Concentrations from Well Logs in Thinly Bedded and Invaded Formations. *Geophysics* 77 (3): WA79–WA98. <https://doi.org/10.1190/geo2011-0454.1>.
- Hogg, D. W. and Foreman-Mackey, D. 2018. Data Analysis Recipes: Using Markov Chain Monte Carlo. *ApJS* 236 (1): 11. <https://doi.org/10.3847/1538-4365/aab76e>.
- Huijser, D., Goodman, J., and Brewer, B. J. 2022. Properties of the Affine-invariant Ensemble Sampler’s ‘Stretch Move’ in High Dimensions. *Aus NZ J of Statistics* 64 (1): 1–26. <https://doi.org/10.1111/anzs.12358>.
- Link, W. A. and Eaton, M. J. 2012. On Thinning of Chains in MCMC. *Methods Ecol Evol* 3 (1): 112–115. <https://doi.org/10.1111/j.2041-210X.2011.00131.x>.
- Mayer, C. and Sibbit, A. 1980. Global, A New Approach To Computer-Processed Log Interpretation. Paper presented at the SPE Annual Technical Conference and Exhibition, Dallas, Texas, USA, 21–24 September. SPE-9341-MS. <https://doi.org/10.2118/9341-MS>.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. et al. 1953. Equation of State Calculations by Fast Computing Machines. *J Chem Phys* 21 (6): 1087–1092. ; <https://doi.org/10.1063/1.1699114>.
- Michelena, R. J., Godbey, K. S., Uland, M. J. et al. 2020. Petrophysical Multiminerals Analysis Using Genetic Optimization to Solve Complex Mineral Composition in Unconventional Reservoirs. In *90th Annual International Meeting, SEG*, 2455–2463. Houston, Texas, USA: Society of Exploration Geophysicists. <https://doi.org/10.1190/segam2020-3425780.1>.
- Mitchell, W. K. and Nelson, R. J. 1988. A Practical Approach to Statistical Log Analysis. Paper presented at the SPWLA 29th Annual Logging Symposium, San Antonio, Texas, USA, 5–8 June. SPWLA-1988-S.
- Quirein, J., Kimminau, S., LaVigne, J. et al. 1986. A Coherent Framework for Developing and Applying Multiple Formation Evaluation Models. Paper presented at the SPWLA 27th Annual Logging Symposium, Houston, Texas, USA, 9–13 June. SPWLA-1986-DD.
- Shonkwiler, R. W. and Mendivil, F. 2009. *Explorations in Monte Carlo Methods*. New York, New York, USA: Springer. <https://doi.org/10.1007/978-0-387-87837-9>.
- Spalburg, M. R. 2004. Bayesian Uncertainty Reduction for Log Evaluation. Paper presented at the Abu Dhabi International Conference and Exhibition, Abu Dhabi, United Arab Emirates, 10–13 October. SPE-88685-MS. <https://doi.org/10.2118/88685-MS>.
- Tierney, L. 1994. Markov Chains for Exploring Posterior Distributions. *Ann Statist* 22 (4): 1701–1728. <https://doi.org/10.1214/aos/1176325750>.