GEOPHYSICS, VOL. 78, NO. 6 (NOVEMBER-DECEMBER 2013); P. E271–E280, 8 FIGS.
10.1190/GEO2013-0128.1



# Robust and accelerated Bayesian inversion of marine controlled-source electromagnetic data using parallel tempering

Anandaroop Ray[1], David L. Alumbaugh[2], G. Michael Hoversten[3], and Kerry Key[1]

## ABSTRACT

Bayesian methods can quantify the model uncertainty that is inherent in inversion of highly nonlinear geophysical problems. In this approach, a model likelihood function based on knowledge of the data noise statistics is used to sample the posterior model distribution, which conveys information on the resolvability of the model parameters. Because these distributions are multidimensional and nonlinear, we used Markov chain Monte Carlo methods for highly efficient sampling. Because a single Markov chain can become stuck in a local probability mode, we run various randomized Markov chains independently. To some extent, this problem can be mitigated by running independent Markov chains, but unless a very large number of chains are run, biased results may be obtained. We got around these limitations by running parallel, interacting Markov chains with "annealed" or "tempered" likelihoods, which enable the whole system of chains to effectively escape local probability maxima. We tested this approach using a transdimensional algorithm, where the number of model parameters as well as the parameters themselves were treated as unknowns during the inversion. This gave us a measure of uncertainty that was independent of any particular parameterization. We then subset the ensemble of inversion models to either reduce uncertainty based on a priori constraints or to examine the probability of various geologic scenarios. We demonstrated our algorithms' fast convergence to the posterior model distribution with a synthetic 1D marine controlled-source electromagnetic data example. The speed up gained from this new approach will facilitate the practical implementation of future 2D and 3D Bayesian inversions, where the cost of each forward evaluation is significantly more expensive than for the 1D case.

## INTRODUCTION

Marine controlled-source electromagnetic (CSEM) methods have been used to image geology with highly resistive contrasts for more than three decades (Young and Cox, 1981). Extensive research and commercialization of this technology over the last 10 years (Ellingsrud et al., 2002; Constable, 2010) has led to its being added to the staple suite of seismic methods as an exploration tool. Owing to the fact that electromagnetic skin depths are smaller in conductive media, marine geophysical EM methods almost always operate in the lower frequency quasistatic regime. This allows for deeper penetration of the CSEM fields into the earth, but as a consequence, it is more a diffusive process than wave-like (Loseth et al., 2006). Thus, the resolution of CSEM is never quite as good at depth as that of the seismic method, but the value of CSEM lies in its sensitivity to resistivity (which may be indicative of hydrocarbon saturation), and not acoustic impedance (which may be more indicative of geologic structure). Owing to this diffusive characteristic of marine CSEM, robust inferences made from a survey are necessarily from inversion of the data, and not merely from examination of the data itself (Weiss, 2007). Typically, regularized and linearized gradient based inversion methods have been used to arrive at models that in addition to minimizing data misfit are "optimal" in some user-defined sense. By means of regularization, highly oscillatory features in the model that are thought to be outside the resolution of CSEM are eliminated (e.g., Constable et al., 1987; Newman and Alumbaugh, 2000; Abubakar et al., 2008; Key, 2009; Sasaki, 2013). Though these methods are highly efficient and well understood, they provide only a single smooth model as a result, or a suite of smooth models. These models provide a limited insight into the

Manuscript received by the Editor 26 March 2013; revised manuscript received 13 June 2013; published online 4 October 2013.
[1]University of California, San Diego, Scripps Institution of Oceanography, La Jolla, California, USA. E-mail: anray@ucsd.edu; kkey@ucsd.edu.
[2]Formerly Chevron Energy Technology Company, San Ramon, California, USA; presently NEOS GeoSolutions, Pleasanton, California, USA.
[3]Chevron Energy Technology Company, San Ramon, California, USA. E-mail: dalumbaugh@chevron.com; hovg@chevron.com.

various classes of models that are compatible with the observed data given the noise. A clear understanding of the resolvability of various parts of the model space does not emerge from a linearized treatment of the nonlinear CSEM problem.

To quantify the uncertainty inherent in the inversion of CSEM data, one can use a Bayesian framework where all information is stored in probability density functions (PDFs). Because Bayesian probability (Bayes and Price, 1763) is a measure of information (Scales and Sneider, 1997) and because it is the aim of geophysical inversion to provide information about the earth's subsurface, it is natural to postulate geophysical inverse problems in a Bayesian framework (Tarantola and Valette, 1982). In such a framework, model parameters are treated as random variables, and their fit to the observed data given the observed statistical noise allow one to formulate a model likelihood. To make the connection with deterministic inversion methods, to the first order, models with low misfit possess a higher likelihood. After incorporating prior knowledge of the models independent of the data, the product of the prior model probability and the likelihood is known as the posterior model probability. Those parts of the model space that are more frequently required by the data than other parts manifest with greater posterior probability, and hence they are better resolved (Backus, 1988). However, fixing a particular model parameterization (e.g., fixing the number of layers, their thickness, and locations) for the inversion is known to produce posterior distributions, *only* for the given parametrization (Dettmer et al., 2010). This is where the "transdimensional" (Bodin and Sambridge, 2009) or "reversible jump" (Green, 1995) Markov chain Monte Carlo (RJ-MCMC) differs from traditional MCMC methods, in sampling from a posterior distribution where the number of unknowns and their positions are also treated as part of the inverse problem (Agostinetti and Malinverno, 2010). Such algorithms have a "parsimony" property (Malinverno, 2002), which refers to the fact that Bayes's theorem deems models that explain the data with simpler parameterizations more probable. MacKay (2003) discusses this aspect of Bayes's theorem in some detail. A good introduction to geophysical transdimensional Bayesian inversion can be found in Sambridge et al. (2013).

Parallel tempering is an accelerated MCMC technique sometimes known as "replica exchange" (Swendsen and Wang, 1987; Geyer, 1991; Earl and Deem, 2005). Using a sequence of parallel, interacting MCMC chains with "annealed" or "tempered" likelihoods allows the entire system of chains to effectively sample the model space without getting trapped in local modes of posterior probability. Recent examples of parallel tempering as applied to geophysical inversion can be found in Dosso et al. (2012), who use the method to discover multiple modes in the posterior model distribution while inverting underwater acoustic reverberation data and in Dettmer and Dosso (2012), who use underwater acoustic data to invert for seafloor sediment properties.

Bayesian uncertainty estimation for the CSEM problem has indeed been carried out in the past (e.g., Chen et al., 2007; Trainor-Guitton and Hoversten, 2011; Buland and Kolbjornsen, 2012). However, these methods considered the model parameterization to be fixed and did not address the model space at different scales. Tompkins et al. (2011) consider a scheme that does indeed address the issue of model parameterization at different scales, but their method is not Bayesian and requires the use of a starting model that fits the data well, and the resulting uncertainty retains some of the characteristics of this model. Gunning et al. (2010) use a "Bayesianized" hierarchical bootstrapping method to address the issue of model resolvability and escape from local probability maxima.

The advantage in using a transdimensional approach (Bodin and Sambridge, 2009) is that the theoretical framework and the practical implementation of it are straightforward — as explicitly shown in Ray and Key (2012), nothing more than a literal interpretation of Bayes's theorem (Bayes and Price, 1763), and the generation of random numbers is required to explore a realistic posterior model distribution. Recent applications of the transdimensional method to solve geophysical EM methods can be found in Minsley (2011) and Brodie and Sambridge (2012), who applied it the airborne EM problem, and in Ray and Key (2012), who tackle the marine CSEM problem. We extend the application of RJ-MCMC to CSEM by introducing parallel tempering, to hasten convergence of the RJ-MCMC chains and quickly escape local probability maxima. By using parallel tempering, we show that previously undetected modes in the posterior model distributions for CSEM data have come to light and that the total number of forward solves is reduced to less than half the number required without using it. This is significant because a realistic uncertainty appraisal for 2D CSEM data will require fast, efficient, and accurate sampling because 2D or 3D EM problems (Key and Ovall, 2011) are computationally far more expensive with complicated interactions between model parameters. We demonstrate our concept using synthetic 1D data, and we add RJ-MCMC and parallel tempering to an arsenal of tools for tackling 2D or 3D problems in the future. Last, we use the reversible jump or transdimensional method to provide a workflow for testing geologic hypotheses by subsetting the posterior distribution of inverted models by interrogating it with intelligent queries, without requiring further inversion. We see this as being particularly useful in exploration scenarios where new information keeps coming in over time, which can be used to hone the posterior model distribution without reinverting the acquired data.

## THEORY

### Bayesian inversion, Markov chains and the reversible jump

Bayesian information is contained in PDFs represented by $p(\cdot)$. Using Bayes's theorem (Bayes and Price, 1763), we write

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{m}|\mathbf{d}) \times p(\mathbf{m})}{p(\mathbf{d})}, \tag{1}$$

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior assumptions}}{\text{evidence}}. \tag{2}$$

For Bayesian geophysical inversion, the data vector $\mathbf{d}$ is a constant. All PDFs with a model dependence are functions of the random variable $\mathbf{m}$. The term $p(\mathbf{d}|\mathbf{m})$ can then be interpreted as the model likelihood, the functional form of which depends on the statistics of the noise distribution, and the value of which depends on the model $\mathbf{m}$ being sampled and its misfit. For Gaussian noise, the model likelihood is given as

$$p(\mathbf{d}|\mathbf{m}) \propto \exp\left(-\frac{[\mathbf{d}-f(\mathbf{m})]^T\mathbf{C_d}^{-1}[\mathbf{d}-f(\mathbf{m})]}{2}\right). \quad (3)$$

Here, $f(\mathbf{m})$ corresponds to the modeled data and $\mathbf{C}_d^{-1}$ is the data covariance matrix, and $[\mathbf{d}-f(\mathbf{m})]^T\mathbf{C_d}^{-1}[\mathbf{d}-f(\mathbf{m})]$ is the $\chi^2$ misfit for the evaluated model $\mathbf{m}$. The prior model distribution $p(\mathbf{m})$ represents our state of knowledge *independent* of the survey data. The evidence term $p(\mathbf{d})$ corresponds to a constant PDF normalizing factor equal to the integral over all models of the numerator in equation 1. Though the evidence can help in "model selection," i.e., decide which model parameterization is more probable than the other, it is very challenging to compute because it requires evaluation of a multidimensional integral over different models, evaluated for different model parameterizations. Another means of performing model selection is to calculate the full posterior model probability distribution $p(\mathbf{m}|\mathbf{d})$ by allowing the problem to be transdimensional (Dettmer et al., 2010), i.e., have a varying number of model parameters. This is the RJ-MCMC approach that we have used in this paper, which is different from the usual MCMC approach in the following manner. Treating the evidence as a proportionality constant, it follows from equation 1 that

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m}) \times p(\mathbf{m}), \quad (4)$$

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m}) \times p(\mathbf{m_k}|k) \times p(k). \quad (5)$$

What we have effectively done is not claimed to have known the optimal model parameterization a priori in equation 4. We have not fixed the prior model probability to be a constant over all models, which fixed-dimensional MCMC samplers do. We do not regularize our inversion because we do not know what an optimal regularization that preserves resolution yet removes spurious oscillatory behavior is. Similarly, we simply do not know how many layers to represent the earth with, or what their locations are (Bodin and Sambridge, 2009). For a given model $\mathbf{m}$, we split $p(\mathbf{m})$ into two parts. One part contains information about the number of interfaces $k$ in the model, $p(k)$. The other part $p(\mathbf{m_k}|\mathbf{k})$ contains information about where these interfaces are in depth, and what the resistivities in between these layers the top and bottom half-space, are given the number of interfaces $k$. Our task is to evaluate uncertainty in the models inverted from the observed data. To this end, we must arrive at the posterior distribution of models, most of which fit the data well, by evaluating their misfit and sampling models according to equation 5. However, it is nearly impossible to exhaustively sample the model space for more than a few parameters owing to the "curse of dimensionality"; hence, we resort to probing this highly nonlinear distribution using various MCMC methods (e.g., Liang et al., 2010) and focus on the RJ-MCMC or transdimensional method (Sambridge et al., 2013) in this work.

## Metropolis-Hastings MCMC and the acceptance probability

The RJ-MCMC sampler is a particular type of Metropolis-Hastings (MH) sampler (e.g., Hastings, 1970; Liang et al., 2010). A MCMC sampler such as the MH algorithm converges upon the posterior distribution using an acceptance probability $\alpha$. At every step of the Markov chain, a candidate model is sampled by perturbing the current model using a known probability distribution (the proposal distribution $q$) and the acceptance $\alpha$ is calculated. The proposal distribution $q$ is usually a simple distribution which should be easy to draw samples from (such as a Gaussian), and also be scaled somewhat like the posterior distribution we are sampling (Ray and Key, 2012). A random number $r$ is then drawn uniformly from the interval $[0, 1]$. If $r < \alpha$, then the proposed perturbation is accepted, else the old model is retained. The rationale behind this algorithm can be explained by examining in more detail the expression for $\alpha$ (Bodin and Sambridge, 2009), where

$$\alpha(\mathbf{m}'|\mathbf{m}) = \min\left[1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \times \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \times \frac{q(\mathbf{m}|\mathbf{m}')}{q(\mathbf{m}'|\mathbf{m})} \times |\mathbf{J}|\right]. \quad (6)$$

Here, $\mathbf{m}'$ is the new proposed model and $\mathbf{m}$ is the old model. Specifically, $\frac{p(\mathbf{m}')}{p(\mathbf{m})}$ is the prior ratio, $\frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})}$ is the likelihood ratio, and $\frac{q(\mathbf{m}|\mathbf{m}')}{q(\mathbf{m}'|\mathbf{m})}$ is the proposal ratio. The Jacobian term $|\mathbf{J}|$ is not to be confused with the model Jacobian needed for gradient-based inversions (e.g., Constable et al., 1987), but it is a matrix that incorporates changes in model dimension when moving from $\mathbf{m}$ to $\mathbf{m}'$. For a fixed number of dimensions in a classic MH algorithm, the prior ratio (for uniform priors), proposal ratio (for symmetric proposals), and Jacobian term are all one (Dettmer et al., 2010). Hence, the algorithm always moves toward areas of higher posterior probability if the data misfit improves (likelihood ratio greater than one). However, it can also move to areas of lower posterior probability with a probability $\alpha$ if the misfit does not improve (likelihood ratio less than one). This is essentially how MCMC samplers give us "regional" as opposed to "point" information about models that fit the data well, unlike gradient-based optimizers. Extensive details on the transdimensional RJ-MCMC prior and proposal ratios can be found in Bodin and Sambridge (2009), Dettmer et al. (2010), or Ray and Key (2012).

To be able to compare likelihoods between models with different numbers of parameters (i.e., with different dimensions), the Jacobian in the acceptance term in equation 6 needs to be evaluated. There are various implementations of RJ-MCMC, and in all the examples cited so far, a "birth-death" scheme has been used. As shown in Bodin and Sambridge (2009) for the birth-death RJ-MCMC scheme, this Jacobian term is unity. We have adopted the birth-death algorithm in this paper and shall not concern ourselves with this Jacobian term any further.

As to why the algorithm should not always look to improve the data fit by simply increasing the number of parameters (interfaces in the seabed), if we examine equation 6, we find that even if the likelihood ratio times the proposal ratio is greater than one for a proposed move that inserts a new interface into the model, the prior ratio will be less than one owing to the fact that the new prior PDF $p(\mathbf{m}')$ needs to integrate over a larger number of parameters to equal one. Hence, there is an opposition to the "birth" of a new layer (which may lead to improvement of data fit) by the prior ratio.

## Parallel tempering

If one were to examine the likelihood function for Gaussian noise in equation 3, it is apparent that the "peakiness" of the likelihood function can be manipulated if one were to introduce a term that plays the statistical mechanics analog of temperature in a partition function (Earl and Deem, 2005). Detailed balance is an aspect of

MCMC that enables unbiased sampling, allowing samples to be distributed in proportion to the target posterior PDF. If we can "anneal" or "temper" the likelihood, without violating the detailed balance (Earl and Deem, 2005), then local probability maxima in the posterior can be overcome. That is, if we were to say that our annealed likelihood should be

$$p(\mathbf{d}|\mathbf{m}) \propto \exp\left(-\frac{[\mathbf{d}-f(\mathbf{m})]^T\mathbf{C_d}^{-1}[\mathbf{d}-f(\mathbf{m})]}{2T_j}\right), \quad (7)$$

instead of the untempered version with $T_j = 1$, we would be sampling from a smoother likelihood with less extreme peaks and valleys than the target untempered likelihood. The first examples of accomplishing this without violating detailed balance can be found in Swendsen and Wang (1987) and in its more familiar form as a MCMC sampling algorithm in Geyer (1991). The basic idea is to run $N$ parallel Markov chains in concert, with the $j$th chain at temperature $T_j$, with $T_1 = 1$ and $T_j > 1$; $\forall j > 1$. Temperatures are usually arranged in ascending order, with Markov chains at adjacent temperatures being allowed to exchange their states (models) with a fixed probability or over a fixed number of steps. At the end of this joint simulation of $N$ Markov chains, the target chain at $T_1 = 1$ is used for posterior inference. An effective implementation is to have a temperature ladder of increasing temperatures, as is demonstrated in Figure 1. This figure describes how the likelihood function can be annealed as a consequence of using equation 7. Note how the likelihoods corresponding to higher temperature chains sample higher values of $\chi^2$ misfit over a broader range of probable likelihoods. Thus, these chains are never stuck in local probability maxima. Owing to their overlap with the narrower lower temperature chains, exchange of states (models) for adjacent chains is possible using the following MH acceptance criterion:

$$\alpha_{\text{swap}} = \frac{p_{\text{hot}}(\mathbf{d}|\mathbf{m}_{\text{cold}})}{p_{\text{hot}}(\mathbf{d}|\mathbf{m}_{\text{hot}})} \times \frac{p_{\text{cold}}(\mathbf{d}|\mathbf{m}_{\text{hot}})}{p_{\text{cold}}(\mathbf{d}|\mathbf{m}_{\text{cold}})}. \quad (8)$$

It is important to ensure that while using equation 8 the probability of selecting all pairs of temperatures is equal and that the pairs are chosen at random to maintain detailed balance (Dettmer and Dosso, 2012). Further, to ensure that one has configured a reasonable temperature ladder, accepted exchange rates between adjacent temperatures should be close to 25% (Dosso et al., 2012). The highest temperature should be high enough to allow the chain to escape local probability maxima, yet not so high for there to be no significant overlap between adjacent chains, which will make exchanges of information improbable. For a given model space, adding more chains to the configuration once an optimal exchange acceptance rate is found does not improve chain mixing. A detailed discussion on setting the temperature ladder can be found in Earl and Deem (2005).

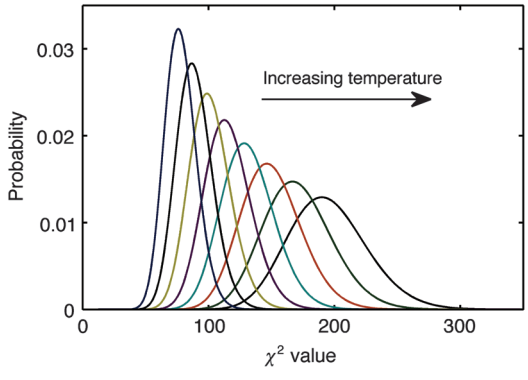


Figure 1. Annealed or tempered likelihood functions as a function of their $\chi^2$ misfit. Note how the leftmost likelihood function at $T = 1$ is narrow and peaked, which can manifest as a model space which is harder to sample.

The main motivation for using parallel tempering is that it is very efficient at escaping local probability modes (misfit minima) and the communication between adjacent chains significantly speeds up convergence to the posterior solution. This is an important factor in ultimately carrying out 2D inversion of CSEM data, where the computational complexity in forward model evaluation is far greater. One can escape local posterior probability modes by running lots of independent chains at the target temperature (Dettmer et al., 2010; Bodin et al., 2012; Ray and Key, 2012), but for a 2D problem where the evaluation of any one forward model itself can at present take up the resources of a cluster of computers, it is important to keep the number of parallel chains small. Parallel tempering achieves just this, bringing down the total number of evaluations as the number of communicating parallel chains required is small, a fact we demonstrate in this paper.

## SYNTHETIC INVERSION TESTS

As a test model for the algorithm, we use a 1D model already studied by Trainor and Hoversten (2009) and Tompkins et al. (2011). The model, as well as its phase and amplitude response are shown in Figure 2. This model is chosen for study by Trainor and Hoversten (2009) owing to a frustrating lack of well-defined convergence to a posterior distribution. Gaussian noise of 2% was added to the modeled data, and a standard source normalized amplitude of $10^{-15}$ V/Am$^2$ was used as the noise floor.

### Interface depths exactly fixed

In the first suite of tests, using the guidelines mentioned for temperature selection, we applied parallel tempering using four chains at temperatures $T = [1.00, 1.35, 1.84, 2.50]$. Note that the temperatures are equally spaced in the log domain (Dettmer and Dosso, 2012). We used an ordinary MH algorithm without the reversible jump for our initial studies. Using 30,000 samples per chain, with swaps attempted at every step, dismissing the first 3000 as the "burn-in" samples (low-probability, high-misfit region of model space), the results from the target chain at $T = 1$ are shown in Figure 3a, with the truth being shown in each layer by a red vertical line at the correct resistivity value. The vertical axis for all plots corresponds to the probability of resistivity in the given layer. The displayed marginal probabilities are simply obtained by binning the sampled posterior model resistivities at each depth. The same histograms for each layer, translated into depth and the $\log_{10}$-resistivity domain are displayed as an image in Figure 3b. Hotter colors represent higher probability, and cooler colors correspond to lower probability. The plots look far more natural in the log domain, which is more representative of subsurface resistivity than its linear counterpart. The true model is shown with a dotted yellow-black 1D model. Though we have exactly fixed the model interfaces to be at the true layer depths, we get a somewhat surprising result. In

addition to the probable layer resistivities being clustered near the truth (as we expected), in layers 3 and 4, we get $\log_{10}$ resistivities clustered around 0.05 and 0.32 that are less probable than the maximum probable in the layer, but are not negligible (which we did not expect). They have a slightly higher misfit and as expected from the likelihood formulation 3, a lower probability. However, they represent models that are within the data error. Root-mean-square (rms) misfit values are obtained by dividing the $\chi^2$ misfit by the number of data points and taking the square root. Models that belong to the "true" family of models have an rms misfit close to 0.9, whereas those models belonging to the "shadow" family of models have an rms closer to 1.1. Further, in carrying out untempered fixed-dimensional inversions, we always chanced upon one or the other family of models. In 10 independent fixed-dimensional trials, only two converged upon the true family. However, from Figure 3a and 3b it is apparent that the true family of models is more probable. Thus, we would probably have needed many more Markov chains running independently to obtain the true posterior probabilities of the two classes of models compatible with the data. In fact, though the results of Trainor and Hoversten (2009) and Tompkins et al. (2011) are generally compatible, albeit with certain distinct differences, neither of them show any hint of the shadow family of models in their posterior distributions. The fact that separate modes in a difficult probability landscape exist is not new, and it is the reason why parallel tempering is used in various sampling applications in the first place. We verified the existence of these distinct modes by computing a 5D grid searched posterior over the probable search ranges indicated by Figure 3b. Though the resulting grid was coarse (not shown here), it required $15 \times 15 \times 30 \times 30 \times 20 = 4{,}050{,}000$ evaluations to prove the existence of separate modes. In contrast, with parallel tempering and MCMC, we needed only $4 \times 30{,}000 = 120{,}000$ samples in total. This is a major reason why high-dimensional integrals (such as those required for marginal probability distributions) are computed using stochastic methods instead of a brute-force approach. Geophysical evidence of multiple modes in the model space has been amply demonstrated by Dosso et al. (2012) for geoacoustic inversion, and by Gunning et al. (2010) for the case of Bayesianized CSEM inversion. What is perhaps a little surprising is that as simple a model as the one we have studied, demonstrates such nonuniqueness with exact layer parameterization.

## Transdimensional inversion with parallel tempering

Given that we don't always know the exact layer parameterization required to perform an inversion, we elected to perform a transdimensional inversion with the noisy synthetic data. The number of interfaces is allowed to vary from one to 15, and the interfaces can be placed anywhere between 1002 and 3500 m. Models are allowed to have any $\log_{10}$ resistivities between $-1$ and 2.3 (0.1 to 200 linear ohm-m). In this section, we also applied parallel tempering to the RJ-MCMC framework. A remarkable aspect of most MCMC algorithms is their flexibility, which allowed us to run a transdimensional RJ-MCMC algorithm within a parallel tempering framework with very little modification to either set of codes.

### *Effect of different temperatures*

For the RJ-MCMC application, we used eight different temperatures at $T = [1.00, 1.14, 1.30, 1.48, 1.69, 1.92, 2.19, 2.50]$ and swapped randomly between two chains at every single step. For flexible model parameterization, eight chains were used as opposed to the four used earlier for fixed-dimensional MCMC. This is because we now have a different and larger posterior model probability space to sample. The results are shown in Figure 4a, where each row corresponds to the indicated temperature. The left panels show

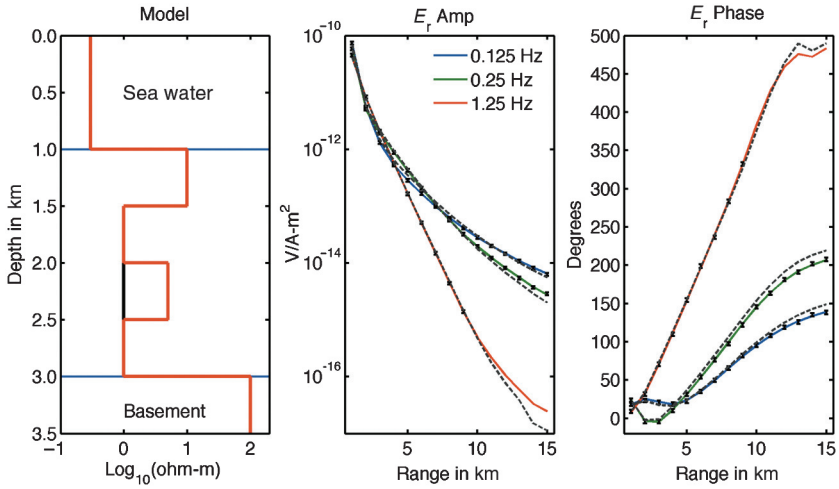


Figure 2. Synthetic 1D model (red) used for this study. Alternating, moderately resistive, then conductive sediments are terminated by a highly resistive basement at depth. The CSEM amplitude and phase responses at the seafloor receivers are plotted with range at three frequencies. The background responses in the absence of the middle layer (black model) are shown with dashed gray lines.

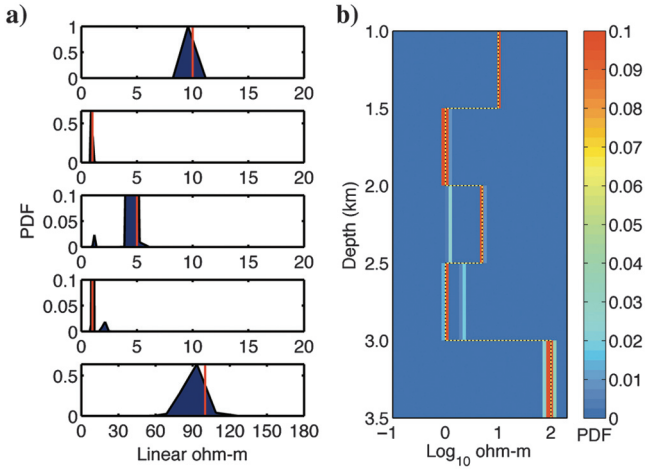


Figure 3. (a) Marginal resistivity distributions in each layer for a fixed-dimensional inversion. The truth is shown with a red vertical line. (b) Same as (a) except now higher probabilities are shown in hotter colors, and the resistivity scale is in $\log_{10}$ ohm-m. The true model is shown with a dotted yellow-black line.

probability of resistivity at depth, and the right panels show the probability of the presence of interfaces at depth. At each depth in the left panels, the 5% and 95% quantile lines of resistivity have been indicated with a dashed black line, and the truth is indicated with a solid black line. The dotted vertical line in the right panel corresponds to a uniform probability of interfaces between 1002 and 3500 m. As shown in Figure 1, the hotter temperature chains sample higher $\chi^2$ misfit values. Returning to Figure 4a, clearly, the hotter chains are not able to resolve well the middle $\log_{10}$ resistivity of 0.69 (5 ohm-m linear), preferring to be closer to the shadow value of $\log_{10}$ resistivity of 0.05 (1.12 ohm-m linear) with smoother posterior model ensembles. As we progress downward toward the cooler chains, we see sensitivity to the true value of the middle layer, though it is not uniquely resolved. This is in line with the observations of both Gunning et al. (2010) and Bodin et al. (2012) who explicitly state that larger data errors than actually observed in the data will lead to less structure (smoother model ensembles) in the inferred posterior. Because increasing the data errors for inversion is in some sense, increasing the temperature

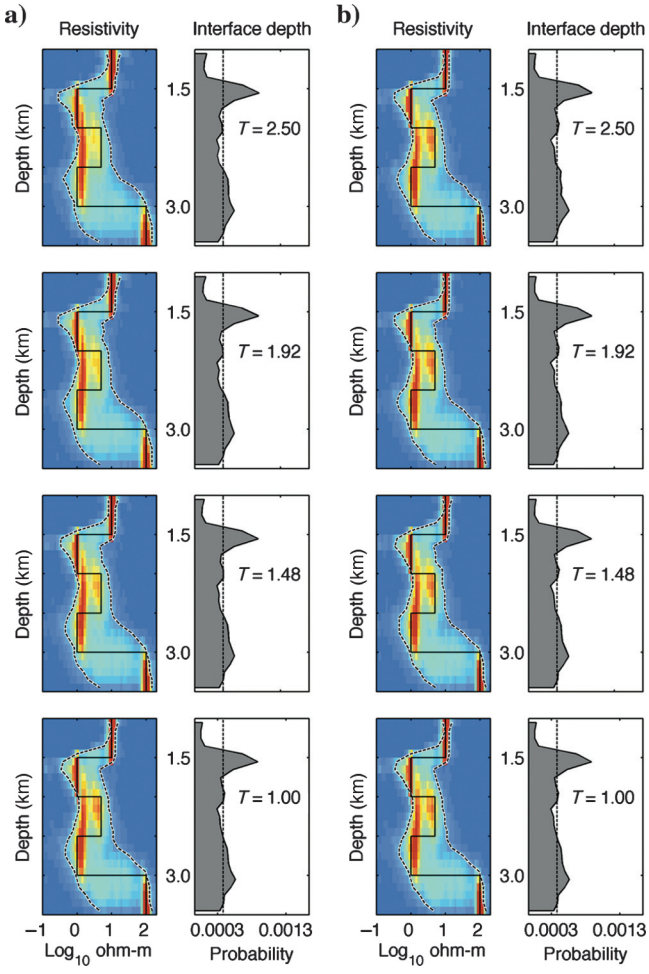


Figure 4. (a) Unweighted transdimensional posterior probability distributions for resistivity at depth (left panel) and interface probability with depth (right panel). Each row corresponds to the indicated temperature. The 5% and 95% quantile lines of resistivity at depth are shown with dashed black lines, and the truth is shown with a solid black line. (b) Samples at each temperature reweighted to the target temperature $T = 1$ to remove bias.

in the annealed likelihood, the two statements are equivalent. Thus, the higher temperature chains should be less sensitive to the middle layer, which is indeed what we observe in this study.

Looking at the target chain at $T = 1$, at some depths, the truth is in fact less probable than the shadow value. Given the results of our fixed-dimensional modeling, this should not come as a surprise. In this light, the result is also fairly intuitive: When we allow the layer parameterization to be variable, we are looking at a more realistic uncertainty estimate which should clearly be less certain than if we had fixed the interfaces to be at their true positions. Without more information, this is all the subsurface information content that we can glean from the observed data, within the relatively large prior bounds. In the following section, we deal with trying to reduce this uncertainty, a posteriori.

### *Sample reweighting*

We can also reweight the chains from different temperatures to obtain samples at the target temperature of $T = 1$ using a weighting factor (Brooks and Neil Frazer, 2005; Dosso et al., 2012). For $\mathbf{m}_i^j$, which is the $i$th model in a chain at temperature $T_j$ with a sampled misfit $\chi^2(\mathbf{m}_i^j)$, the weight is given as

$$w(\mathbf{m}_i^j) = \frac{\exp\left[-\frac{\chi^2(\mathbf{m}_i^j)}{2}\left(1-\frac{1}{T_j}\right)\right]}{\sum_i \exp\left[-\frac{\chi^2(\mathbf{m}_i^j)}{2}\left(1-\frac{1}{T_j}\right)\right]}. \quad (9)$$

Notice that at a higher temperature, a higher misfit implies a smaller weight. Further, the weight for all models in a chain are the same if $T_j = 1$, implying no reweighting. For all models in a chain at a particular temperature, the samples can be reweighted to the target temperature $T = 1$ using equation 9. For instance, marginal probabilities of resistivity at depth can be found by binning the resistivity values as before, but each histogram count needs to be multiplied by the weight corresponding to the model being binned. The results of this operation are shown in Figure 4b, where each row corresponds to the reweighted samples at a given temperature. The color scales are the same for all images in Figure 4a and 4b. All reweighted distributions look fairly similar, but reweighted samples from $T = 2.5$ have a slightly rougher posterior distribution — borne out by a minute observation of the interface probability curve at this temperature in Figure 4b and comparison with the interface probability curve for the $T = 1$ case. The $T = 2.5$ chain has undersampled the model space to a very small extent, a similar observation being borne out by Dosso et al. (2012). The extent of undersampling is problem dependent, and it should not be a cause for alarm if it is greater than shown here. This undersampling merely illustrates that there is a trade-off between traversing greater distances in the model space and fine sampling of the target areas. Because parallel tempering allows for the trickling down of information at higher temperatures to lower temperatures, the lower temperature chains will have sampled the model space adequately.

### *Posterior on the number of interfaces*

The marginal posterior distribution on the number of interfaces for the target temperature $T = 1$ is shown in Figure 5. The true number of interfaces is shown with a red vertical line. The dashed black horizontal line corresponds to a uniform probability on the number

of interfaces. The number of interfaces is a variable that changes from model to model and models with different numbers of interfaces can be swapped between different chains. Because the true number of interfaces is not known by the inversion a priori, the most probable number of interfaces is not equal to the true value of four, but turns out to be eight. This is not unusual, and a similar phenomenon has also been observed by Minsley (2011).

*Proposal variances at different temperatures*

If the posterior probability space has various different length scales associated with multiple modes, as long as equation 8 has been obeyed, different proposal distribution step sizes can be assigned to each chain. For instance, we can assign larger proposal variances to the higher temperature chains to ensure that they explore the model space with large steps. The step size for a chain at a particular temperature should be made smaller if the acceptance rates for that chain become very low (e.g., Bodin et al., 2012; Ray and Key, 2012) or if the accepted swap rate between adjacent chains falls far below 25%.

## Convergence to the posterior

A definitive statement about convergence to the true posterior is difficult to make, and it is still an area of active research, especially with RJ-MCMC, in which the number of parameters may change in the next step (Bodin and Sambridge, 2009). However, a method that works well is to keep sampling until it is apparent that the posterior distribution does not change appreciably by adding more samples (Dettmer and Dosso, 2012). For the transdimensional case, with a burn-in length of 5000 samples, we oversampled by 500,000 samples in each chain to test for convergence and stopped sampling at 1.75 million samples per chain.

## Comparison between ordinary and parallel-tempered MCMC

As is evident from Figure 3a and 3b, parallel tempering reveals hidden modes in the fixed-dimensional case by transitioning between them with relative ease. Further, because this problem can be mitigated by running numerous independent, noncommunicating chains, it is only fair to compare parallel tempering with the case when we run various independent MCMC chains. For the transdimensional case, to obtain the level of detail and convergence found in the last panel of Figure 4a, with parallel tempering we ran eight chains with 1.75 million samples in each chain. In total, we used 14 million samples with a slightly lower number of total forward evaluations (as proposed samples falling outside the prior bounds are not evaluated). For an equivalent level of detail without using parallel tempering, we needed at least 56 independent RJ-MCMC chains with 500,000 samples each, a total of 28 million samples, with a similar number of forward evaluations. Thus, we see that parallel tempering requires less chains and a lower total number of forward evaluations by a factor of two. To illustrate how much better chain mixing is using parallel tempering, a comparison has been made between sampled posteriors from parallel tempering and independent chains, for eight chains and 500,000 samples in each chain. Figure 6a shows the result from all eight independent chains, and Figure 6b shows the results using parallel tempering, inference made from only the target chain at $T = 1$. Though neither ensemble has achieved stationarity yet, an examination of the last panel of Figure 4a (which uses the same color scale) shows that the result with parallel tempering is closer to the final posterior distribution, as is evident in the shape of the interface probability curves and the smoothness of the 5% and 95% quantile lines of resistivity at depth. In fact, it may even appear that the interface probability bumps in the independent chain results are closer to the truth. This is not because they find the truth better than parallel tempering does, but merely because they have not sampled enough of the model space in an equivalent number of samples. Very similar behavior is reported by Dosso et al. (2012).

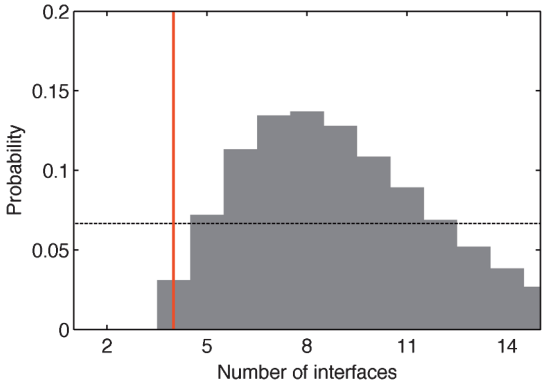


Figure 5. Marginal probabilities on number of interfaces required by the observed data at the target sampling temperature. The true number four is marked with a vertical red line. The dashed black horizontal line corresponds to a uniform probability on the number of interfaces.

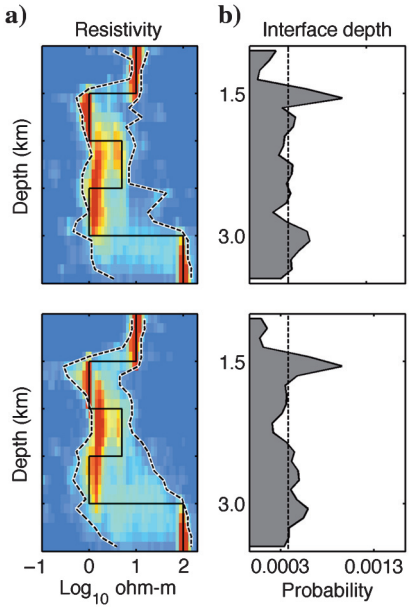


Figure 6. (a) Posterior from eight independent, noninteracting chains, 500,000 samples in each chain. (b) Posterior from parallel tempering with the same number of chains and samples, inference using only the target chain. As evidenced by the smoother 5% and 95% quantile lines of resistivity at depth, parallel tempering estimates are closer to the final sampled posterior.

## REDUCING UNCERTAINTY POSTINVERSION AND SCENARIO EVALUATION

If one has used large enough prior bounds, a lot of information is contained in PDFs pertaining to subsurface resistivity. In a typical exploration scenario, as a prospect is being evaluated, more information such as seismic data, well information, or even geologic models becomes available over the course of time. One can then incorporate this information when analyzing the posterior model distribution without performing further inversion. When using information from geologic models, one is effectively examining different geologic scenarios embedded in the posterior. An example application is shown in Figures 7 and 8. In step 1, we set up reasonable uniform bounded prior distributions (Ray and Key, 2012) for resistivity at depth (Figure 7, left panel), interfaces at depth (Figure 7, right panel), and the number of layers required by a model (Figure 8). In step 2, we introduce the data, and through the data misfit translated into a likelihood, we sample the product of likelihood and prior to obtain a posterior model distribution. Posterior inference is made only from the chain at the target temperature $T = 1$. Step 2 produces the marginal posterior distribution for the given prior bounds and data errors as shown in Figures 7 and 8. Step 2 in these figures represent the same case as Figures 4a and 5 for the target temperature $T = 1$. We could stop here, if this was all we knew about the subsurface, and indeed we should if that is the case. However, if new information from well logs or seismic imaging confirms the presence of certain horizons in the subsurface, we could mine the posterior model distribution using this information and go to step 3. In this step, we have selected all models in the posterior ensemble that have interfaces within $\pm 75$ m of the true location of interfaces, and the resulting posteriors from this subset are shown in Figures 7 and 8. Note how the presence of a five-layered structure is becoming quite apparent in Figure 7 step 3, and the middle layer is more likely to be resistive. All rows of Figure 7 are normalized to the same color scale. The last row shows the step 3 posterior overlain with examples of sampled models similar to both the true (black) as well as shadow (purple) family of models, given that sampled models can have different numbers of interfaces at different depths. The 5% and 95% quantile lines have been omitted for clarity. Note the similarity of this last row to Figure 3b for the fixed-dimensional case. The fixed-dimensional case shown in Figure 3b is a highly specific case of Figure 7 step 3, and MCMC samplers are not grid searches. However, that the fixed-dimensional case, which is a subset of the full posterior, should be so similar to the case shown in step 3 of Figure 7 is indeed reassuring.

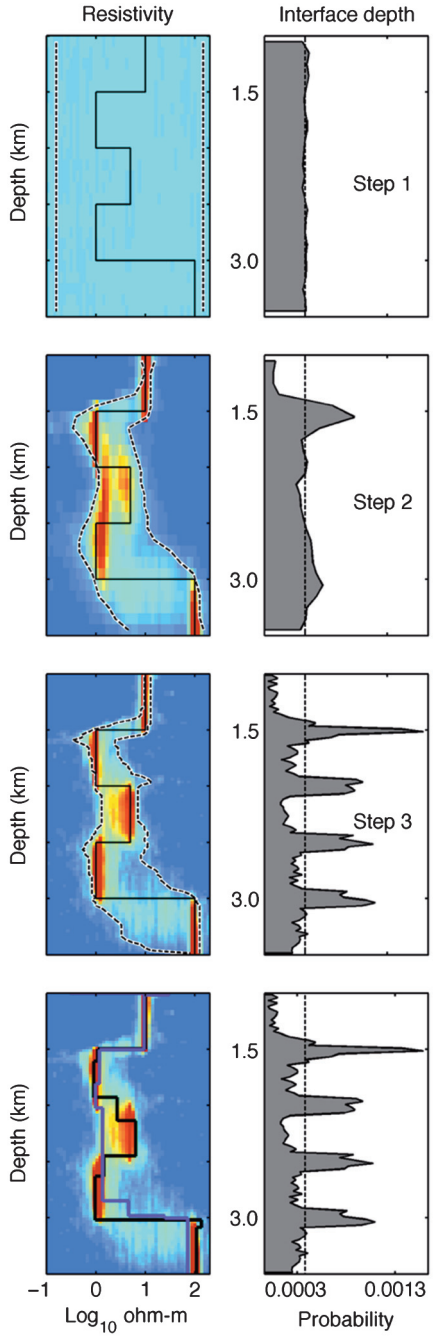


Figure 7. Workflow for inversion: (1) Start with broad uniform priors, (2) introduce the data and obtain full posterior model distribution, and (3) subset posterior using available knowledge of interfaces. The last row shows the step 3 posterior overlain with sampled models that are similar to the true model (in black) and the shadow model (in purple).

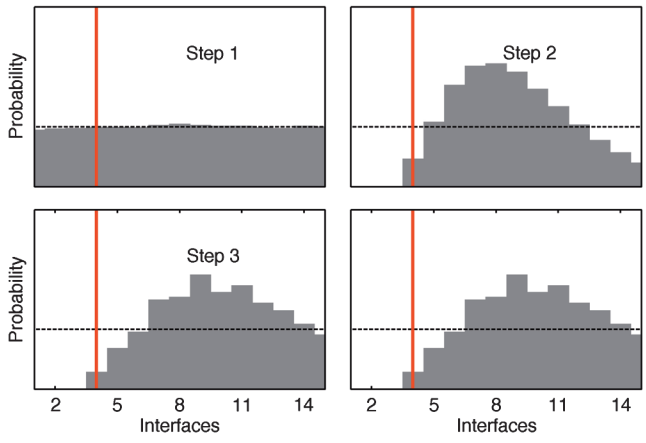


Figure 8. Marginal posterior probability distributions on the number of interfaces for the cases described in Figure 7.

This methodology can be encapsulated in the following manner: (1) Start by setting reasonably wide bounded prior distributions, (2) introduce the data to obtain the full posterior, and (3) subset the data based on available information. Geologic models can also be tested in this fashion. By subsetting parts of the model, we introduce a conditioning on the posterior. This conditioning affects other parts of the posterior model distribution. Depending on the available knowledge about the subsurface, this method can be used to examine different geologic hypotheses. At this juncture, all sampled likely models obey the model physics. Further, this step requires no further inversion, but simply mining of the posterior distribution (Ray and Key, 2012). We must caution though, that MCMC samplers are not exhaustive grid searches, so sampling for too specific a query may turn up no results, which does not imply that these specific cases do not exist. A corollary to postinversion conditioning on the posterior is that we could use a more restrictive prior distribution to reduce the model space being searched. However, this approach has a serious shortcoming in that one can "tune" the inversion drastically by overspecifying the prior distribution, when in reality our prior knowledge is limited. This can lead to serious bias in the resulting posterior distribution, as is demonstrated by Minsley (2011).

## CONCLUSIONS

Uncertainty is an inescapable aspect of geophysical inversion. Noisy observations, incomplete data sampling, insufficient knowledge of a suitable subsurface parameterization as well as model physics contribute to uncertainty and nonuniqueness in the inverted models. In this work, we have attempted to fully address these issues using a transdimensional Bayesian technique. To overcome strong nonlinearities in the posterior model probability distributions with multiple probability maxima, we have used parallel tempering to reliably transition between different modes. We have also demonstrated that parallel tempering speeds up convergence to the posterior distribution by reducing the total number of required samples to less than half of what would be required otherwise. We have attempted to use as few parallel chains as possible with as few forward calls per chain, to keep the total number of forward evaluations low. With the advent of cluster computing, computation times for accurate 2D and 3D model responses are rapidly getting smaller. However, the process of evaluating these model responses in themselves use up significant cluster resources, and it is not yet feasible, for tens of independent (noninteracting) Markov chains, to assign each Markov chain its own cluster for forward computations. This is where parallel tempering drastically reduces the number of required chains (eight for the RJ-MCMC), yet it keeps the number of forward evaluations per chain down to about 1.25 million for convergence. Given a hypothetical 2D forward evaluation time of 1 s, 1.25 million computations will take 14.5 days. Although this may seem like a significant amount of time, investments for drill decisions are significantly more expensive than this amount of computer time. With advances in GPU computing and the availability of fast parallel 2.5D forward solvers, we have demonstrated that when the jump to quick, higher dimensional forward modeling is made, parallel tempering and transdimensional RJ-MCMC will be valuable tools with which to evaluate the full uncertainty associated with observed data.

## ACKNOWLEDGMENTS

A. Ray would like to thank the modeling and inversion team at Chevron Energy Technology Company (San Ramon) for their willingness to discuss inverse problems in delightful detail. In particular, K. Nihei, J. Washbourne, and R. Modrak are thanked for their comments during the course of the project that produced these results. J. Dettmer is thanked for providing valuable clues in the hunt for a working parallel tempering algorithm, and T. Bodin is thanked for providing insight into the workings of the transdimensional method. A. Ray and K. Key would collectively like to thank the Scripps Seafloor Electromagnetic Methods Consortium for funding support. The authors collectively thank Chevron ETC for granting permission to publish this work and two anonymous reviewers for providing valuable comments that helped us express our ideas far more clearly.

## REFERENCES

Abubakar, A., T. M. Habashy, V. L. Druskin, L. Knizhnerman, and D. Alumbaugh, 2008, 2.5D forward and inverse modeling for interpreting low-frequency electromagnetic measurements: Geophysics, **73**, no. 4, F165–F177, doi: 10.1190/1.2937466.

Agostinetti, N. P., and A. Malinverno, 2010, Receiver function inversion by transdimensional Monte Carlo sampling: Geophysical Journal International, **181**, 858–872, doi: 10.1111/j.1365-246X.2010.04530.x.

Backus, G. E., 1988, Bayesian inference in geomagnetism: Geophysical Journal International, **92**, 125–142, doi: 10.1111/j.1365-246X.1988.tb01127.x.

Bayes, T., and R. Price, 1763, An essay towards solving a problem in the doctrine of chances: By the late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S.: Philosophical Transactions, **53**, 370–418, doi: 10.1098/rstl.1763.0053.

Bodin, T., and M. Sambridge, 2009, Seismic tomography with the reversible jump algorithm: Geophysical Journal International, **178**, 1411–1436, doi: 10.1111/j.1365-246X.2009.04226.x.

Bodin, T., M. Sambridge, H. Tkalčić, P. Arroucau, K. Gallagher, and N. Rawlinson, 2012, Transdimensional inversion of receiver functions and surface wave dispersion: Journal of Geophysical Research, **117**, B02301, doi: 10.1029/2011JB008560.

Brodie, R. C., and M. Sambridge, 2012, Transdimensional Monte Carlo inversion of AEM data: Presented at 22nd International Geophysical Conference and Exhibition.

Brooks, B. A., and L. Neil Frazer, 2005, Importance reweighting reduces dependence on temperature in Gibbs samplers: An application to the coseismic geodetic inverse problem: Geophysical Journal International, **161**, 12–20, doi: 10.1111/j.1365-246X.2005.02573.x.

Buland, A., and O. Kolbjornsen, 2012, Bayesian inversion of CSEM and magnetotelluric data: Geophysics, **77**, no. 1, E33–E42, doi: 10.1190/geo2010-0298.1.

Chen, J., G. 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**, no. 2, WA85–WA95, doi: 10.1190/1.2435082.

Constable, S., 2010, Ten years of marine CSEM for hydrocarbon exploration: Geophysics, **75**, no. 5, 75A67–75A81, doi: 10.1190/1.3483451.

Constable, S. C., R. L. Parker, and C. G. Constable, 1987, Occam's inversion — A practical algorithm for generating smooth models from electromagnetic sounding data: Geophysics, **52**, 289–300, doi: 10.1190/1.1442303.

Dettmer, J., and S. E. Dosso, 2012, Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains: Journal of the Acoustical Society of America, **132**, 2239–2250, doi: 10.1121/1.4746016.

Dettmer, J., S. E. Dosso, and C. W. Holland, 2010, Trans-dimensional geoacoustic inversion: Journal of the Acoustical Society of America, **128**, 3393–3405, doi: 10.1121/1.3500674.

Dosso, S. E., C. W. Holland, and M. Sambridge, 2012, Parallel tempering for strongly nonlinear geoacoustic inversion: Journal of the Acoustical Society of America, **132**, 3030–3040, doi: 10.1121/1.4757639.

Earl, D. J., and M. W. Deem, 2005, Parallel tempering: Theory, applications, and new perspectives: Physical Chemistry Chemical Physics, **7**, 3910–3916, doi: 10.1039/b509983h.

Ellingsrud, S., T. Eidesmo, S. Johansen, M. C. Sinha, L. M. MacGregor, and S. Constable, 2002, Remote sensing of hydrocarbon layers by seabed

logging (SBL): Results from a cruise offshore Angola: The Leading Edge, **21**, 972–982, doi: 10.1190/1.1518433.
Geyer, C. J., 1991, Markov chain Monte Carlo maximum likelihood: *in* E. M. Keramidas, ed., Computing science and statistics: Proceedings of the 23rd Symposium on the Interface: Interface Foundation of North America, 156–163.
Green, P. J., 1995, Reversible jump Markov chain Monte Carlo computation and Bayesian model determination: Biometrika, **82**, 711–732, doi: 10.1093/biomet/82.4.711.
Gunning, J., M. E. Glinsky, and J. Hedditch, 2010, Resolution and uncertainty in 1D CSEM inversion: A Bayesian approach and open-source implementation: Geophysics, **75**, no. 6, F151–F171, doi: 10.1190/1.3496902.
Hastings, W. K., 1970, Monte Carlo sampling methods using Markov chains and their applications: Biometrika, **57**, 97–109, doi: 10.1093/biomet/57.1.97.
Key, K., 2009, 1D inversion of multicomponent, multifrequency marine CSEM data: Methodology and synthetic studies for resolving thin resistive layers: Geophysics, **74**, no. 2, F9–F20, doi: 10.1190/1.3058434.
Key, K., and J. Ovall, 2011, A parallel goal-oriented adaptive finite element method for 2.5-D electromagnetic modelling: Geophysical Journal International, **186**, 137–154, doi: 10.1111/j.1365-246X.2011.05025.x.
Liang, F., C. Liu, and R. J. Carroll, 2010, Bayesian inference and Markov chain Monte Carlo: John Wiley & Sons, Ltd.
Loseth, L. O., H. M. Pedersen, B. Ursin, L. Amundsen, and S. Ellingsrud, 2006, Low-frequency electromagnetic fields in applied geophysics: Waves or diffusion?: Geophysics, **71**, no. 4, W29–W40, doi: 10.1190/1.2208275.
MacKay, D. J. C., 2003, Information theory, inference and learning algorithms: Cambridge University Press.
Malinverno, A., 2002, Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem: Geophysical Journal International, **151**, 675–688, doi: 10.1046/j.1365-246X.2002.01847.x.
Minsley, B. J., 2011, A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data: Geophysical Journal International, **187**, 252–272, doi: 10.1111/j.1365-246X.2011.05165.x.
Newman, G. A., and D. L. Alumbaugh, 2000, Three-dimensional magnetotelluric inversion using non-linear conjugate gradients: Geophysical Journal International, **140**, 410–424, doi: 10.1046/j.1365-246x.2000.00007.x.
Ray, A., and K. Key, 2012, Bayesian inversion of marine CSEM data with a transdimensional self parametrizing algorithm: Geophysical Journal International, **191**, 1135–1151, doi: 10.1111/j.1365-246X.2012.05677.x.
Sambridge, M., T. Bodin, K. Gallagher, and H. Tkalcic, 2013, Transdimensional inference in the geosciences: Philosophical Transactions of the Royal Society A, **371**, doi: 10.1098/rsta.2011.0547.
Sasaki, Y., 2013, 3D inversion of marine CSEM and MT data: An approach to shallow-water problem: Geophysics, **78**, no. 1, E59–E65, doi: 10.1190/geo2012-0094.1.
Scales, J. A., and R. Sneider, 1997, To Bayes or not to Bayes?: Geophysics, **62**, 1045–1046.
Swendsen, R. H., and J. S. Wang, 1987, Nonuniversal critical dynamics in Monte Carlo simulations: Physical Review Letters, **58**, 86–88, doi: 10.1103/PhysRevLett.58.86.
Tarantola, A., and B. Valette, 1982, Inverse problems = quest for information: Journal of Geophysics, **50**, 159–170.
Tompkins, M. J., J. L. Fernandez Martinez, D. L. Alumbaugh, and T. Mukerji, 2011, Scalable uncertainty estimation for nonlinear inverse problems using parameter reduction, constraint mapping, and geometric sampling: Marine controlled-source electromagnetic examples: Geophysics, **76**, no. 4, F263–F281, doi: 10.1190/1.3581355.
Trainor, W., and G. M. Hoversten, 2009, Practical challenges of stochastic inversion implementation for geophysical problems: 79th Annual International Meeting, SEG, Expanded Abstracts, 734–738.
Trainor-Guitton, W., and G. M. Hoversten, 2011, Stochastic inversion for electromagnetic geophysics: Practical challenges and improving convergence efficiency: Geophysics, **76**, no. 6, F373–F386, doi: 10.1190/geo2010-0223.1.
Weiss, C., 2007, The fallacy of the "shallow-water problem" in marine CSEM exploration: Geophysics, **72**, no. 6, A93–A97, doi: 10.1190/1.2786868.
Young, P. D., and C. S. Cox, 1981, Electromagnetic active source sounding near the East Pacific Rise: Geophysical Research Letters, **8**, 1043–1046, doi: 10.1029/GL008i010p01043.