Geophys. J. Int. (2002) **151,** 675–688

Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem

Alberto Malinverno

Schlumberger-Doll Research, Ridgefield, CT 06877, *USA. E-mail: malinverno@slb.com*

Accepted 2002 April 24. Received 2002 April 24; in original form 2001 April 5

SUMMARY

A key element in the solution of a geophysical inverse problem is the quantification of non-uniqueness, that is, how much parameters of an inferred earth model can vary while fitting a set of measurements. A widely used approach is that of Bayesian inference, where Bayes' rule is used to determine the uncertainty of the earth model parameters *a posteriori* given the data. I describe here, a natural extension of Bayesian parameter estimation that accounts for the posterior probability of how complex an earth model is (specifically, how many layers it contains). This approach has a built-in parsimony criterion: among all earth models that fit the data, those with fewer parameters (fewer layers) have higher posterior probabilities.

To implement this approach in practice, I use a Markov chain Monte Carlo (MCMC) algorithm applied to the nonlinear problem of inverting DC resistivity sounding data to infer characteristics of a 1-D earth model. The earth model is parametrized as a layered medium, where the number of layers and their resistivities and thicknesses are poorly known *a priori*. The algorithm obtains a sample of layered media from the posterior distribution; this sample measures non-uniqueness in terms of how many layers are effectively resolved by the data and of the range of layer thicknesses and resistivities consistent with the data.

Because the complexity of the model is effectively determined by the data, the solution does not need to be regularized. This is a desirable feature, because requiring the solution to be smooth beyond what is implied by prior information can lead to underestimating posterior uncertainty. Letting the number of layers be a free parameter, as done here, broadens the space of earth models possible *a priori* and makes the determination of posterior uncertainty less dependent on the parametrization.

Key words: Bayesian inversion, geophysical inversion, Markov chain Monte Carlo, MCMC, resistivity.

1 INTRODUCTION

In geophysical inverse problems we aim to reconstruct the distribution of subsurface properties (an 'earth model') given measurements that are usually acquired at the surface. The solution of these inverse problems is typically not unique. That is, significantly different earth models give rise to predicted data that fit equally well actual measurements. As clearly stated by Backus (1988), a geophysical inverse problem has an existence half, where we wish to obtain an earth model that fits the data, and a uniqueness half, where we want to measure how much the earth model may vary while fitting the data. A common way to solve the existence half of the inverse problem is to regularize, or damp the solution and obtain the smoothest earth model that fits the data within a given error. This model should only contain features that are required by the data and should be common to all the possible solutions (e.g. Constable *et al.* 1987).

It is often necessary, however, to address the uniqueness half of the problem, and a way to quantify non-uniqueness is to use a probabilistic Bayesian inference approach. The starting point of Bayesian inference is to specify the initial uncertainty of the earth model parameters in a prior probability distribution. Bayes' rule then combines this prior distribution with a likelihood function (which measures how probable an earth model is in light of the measurements) to give a posterior distribution, which is the solution of the problem and describes the final uncertainty of the earth model (e.g. Tarantola & Valette 1982; Jackson & Matsu'ura 1985; Duijndam 1988).

In Bayesian inference, the role of regularization is played by the prior distribution, which limits the space of plausible earth models by giving higher probability to those that agree more closely with prior knowledge. Still, it seems that sometimes smoothing is needed if the prior information does not constrain the earth model closely enough. For example, Grandis *et al.* (1999) and Schott *et al.* (1999) show that the posterior uncertainties in the resistivity of a finely layered medium obtained by inverting surface electromagnetic measurements are huge unless the solution is smoothed.

The smoothing factor was chosen to best trade off unrealistic oscillations in the solution and features that were well constrained by the data (Schott *et al.* 1999, p. 777). In Bayesian inference, however, the degree of smoothness of the solution ought to be set on the basis of *a priori* knowledge (Backus 1988; Scales & Tenorio 2001). If one initially knows little about the distribution of resistivity in depth, it is unlikely that one knows anything about the expected smoothness of the solution. Yet, if smoother earth models effectively have higher probabilities, the posterior distribution is not really based on what is known *a priori*, and thus it does not faithfully represent the posterior uncertainty.

The purpose of this paper is to describe an extension of the commonly used Bayesian parameter estimation approach to account for the posterior probabilities of different parametrizations of the earth model. Specifically, I will use a generic layered medium, where the number of layers, the depths to the interfaces between layers (or equivalently, layer thicknesses), and layer properties are free parameters. If the posterior probabilities of different parametrizations (that is, different numbers of layers) is considered, there is no need for regularization beyond what is dictated by scant prior information. This is because the posterior probability of model parametrizations obeys a principle of parsimony, or simplicity: among earth models that fit equally well the data, the models that have fewer degrees of freedom (fewer layers) have higher posterior probabilities. The net effect is that the data determine how complex the model parametrization ought to be. Also, by broadening the space of parametrizations possible *a priori*, the determination of posterior uncertainty does not depend on a particular choice of parametrization (e.g. a fixed number of layers) and gives a more comprehensive quantification of the non-uniqueness of the solution.

In previous work, I demonstrated the parsimony feature of Bayesian inference for linear inverse problems (Malinverno 2000), and described an efficient Markov chain Monte Carlo (MCMC) method applicable when the inverse problem can be partly linearized (Malinverno & Leaney 2000). In this paper I extend the previous work by describing a MCMC method that is applicable to a fully nonlinear problem. I use the same problem as that treated by Constable *et al.* (1987), Sen & Stoffa (1995), and Schott *et al.* (1999), inverting DC resistivity soundings obtained at the earth's surface for a 1-D layered model of resistivity. The MCMC algorithm samples the posterior distribution of layered models in a computationally efficient way and obtains a meaningful measure of the posterior uncertainty of the solution.

1.1 Notation

Vectors will be denoted by bold lowercase letters (e.g. $\mathbf{x}$) and matrices by bold uppercase letters (e.g. $\mathbf{A}$). Vectors are column vectors, so that the inner product of $\mathbf{x}$ is $\mathbf{x}^T\mathbf{x}$, where the symbol ^{T} denotes the transpose. In $\mathbf{y} = \mathbf{f}(\mathbf{x})$, the function $\mathbf{f}$ is a vector-valued function that returns a vector $\mathbf{y}$ for a given value of the vector $\mathbf{x}$. Values assumed *a priori* will be denoted with a bar; e.g. $\overline{x}$ is the prior value of x. Values estimated *a posteriori* will be denoted by a hat (e.g. $\widehat{x}$). The notation $p(\mathbf{x} \mid \mathbf{y})$ indicates the probability density function (*pdf*) of $\mathbf{x}$ conditional on $\mathbf{y}$.

2 DC RESISTIVITY SOUNDINGS IN A 1-D GENERIC LAYERED MEDIUM

The generic layered medium used here as an earth model is illustrated in Fig. 1. It is defined by a number of layers k, a vector of log-depths to the interfaces between the layers $\mathbf{z} = (z_1, z_2, \ldots, z_{k-1})$, and a vector of log-resistivities in each layer $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_k)$. Using logarithms ensures that depths and resistivities are positive, accounts for the decrease in resolving power of the resistivity sounding data with increasing depth, and covers the broad range of resistivities encountered in nature while avoiding numerical difficulties.

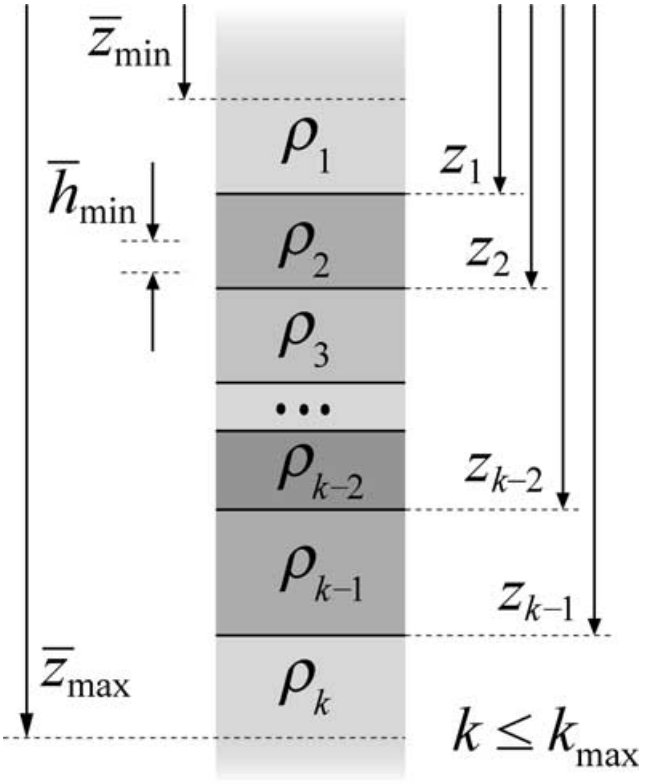

Figure 1. A generic layered medium with k layers. The layer interfaces are at depths between a minimum $\overline{z}_{\min}$ and a maximum $\overline{z}_{\max}$, and no layer can be thinner than $\overline{h}_{\min}$.

I use the logarithm of depths because the resolving power of the resistivity sounding data decreases with increasing depth; the logarithm of resistivity covers the broad range of resistivities encountered in nature while avoiding numerical difficulties. In short, the earth model can be written as a vector

$$\mathbf{m} = (k, \mathbf{z}, \boldsymbol{\rho}). \tag{1}$$

This is a 'generic' layered medium because it has enough degrees of freedom to describe media with different numbers of layers and layer interfaces that are not constrained to be at fixed depths.

The illustrative data I use are log-apparent resistivities measured at the surface using a Schlumberger array. For different spacings between the electrodes, the current lines sample different depth ranges in the subsurface, and the measured apparent resistivities (the resistivities that would be measured if the medium were homogeneous) vary accordingly (see e.g. Keller & Frischknecht 1966). Solving the forward problem requires specifying a forward modelling function $\mathbf{g}(\mathbf{m})$ that returns a vector of apparent resistivity data predicted by the generic layered medium in $\mathbf{m}$. I use here for $\mathbf{g}(\mathbf{m})$ the filter method described in the Appendix of Constable *et al.* (1987), with the 11-point filter of Guptasarma (1982).

3 BAYESIAN INFERENCE FOR A GENERIC LAYERED MEDIUM AND PARSIMONY

In Bayesian inference, the posterior *pdf* of $\mathbf{m}$ measures how well a generic layered medium agrees with prior information and data. It is helpful to write this posterior using the definition of a conditional *pdf* as

$$p(\mathbf{m} \,|\, \mathbf{d}, \mathcal{I}) = p(k \,|\, \mathbf{d}, \mathcal{I}) p(\mathbf{z}, \boldsymbol{\rho} \,|\, k, \mathbf{d}, \mathcal{I}), \tag{2}$$

where $\mathbf{d} = (d_1, d_2, \ldots, d_N)$ are the measured log-apparent resistivities and $\mathcal{I}$ denotes prior information. In Bayesian inference, all probabilities are conditional at least on $\mathcal{I}$, which represents prior knowledge about the parametrization of the earth model, realistic values of the parameters, the geometry of the subsurface, the forward model, etc. Let us consider the two terms in Eq. (2) separately.

3.1 Parameter estimation: $p(\mathbf{z}, \boldsymbol{\rho} \,|\, k, \mathbf{d}, \mathcal{I})$

This is the problem of determining the posterior *pdf* of the layered medium parameters for a given number of layers k. In statistical terminology, this is a 'parameter estimation' problem, and it has been widely treated in the geophysical literature (e.g. Tarantola & Valette 1982; Jaynes 1984; Jackson & Matsu'ura 1985; Backus 1988; Cary & Chapman 1988; Duijndam 1988; Mosegaard & Tarantola 1995; Sen & Stoffa 1995; Gouveia & Scales 1998). The posterior *pdf* of $\mathbf{z}$ and $\boldsymbol{\rho}$ can be written using Bayes's rule as

$$p(\mathbf{z}, \boldsymbol{\rho} \,|\, k, \mathbf{d}, \mathcal{I}) = \frac{p(\mathbf{z}, \boldsymbol{\rho} \,|\, k, \mathcal{I}) p(\mathbf{d} \,|\, \mathbf{z}, \boldsymbol{\rho}, k, \mathcal{I})}{p(\mathbf{d} \,|\, k, \mathcal{I})}, \tag{3}$$

where $p(\mathbf{z}, \boldsymbol{\rho} \,|\, k, \mathcal{I})$ is the prior *pdf* of $\mathbf{z}$ and $\boldsymbol{\rho}$ and $p(\mathbf{d} \,|\, \mathbf{z}, \boldsymbol{\rho}, k, \mathcal{I})$ is the likelihood function, which measures the probability of observing the data $\mathbf{d}$ when the parameters of a medium of k layers equal $\mathbf{z}$ and $\boldsymbol{\rho}$. The denominator of eq. (3) is commonly called the 'marginal likelihood' or 'evidence' and can be shown to be a normalizing factor for the posterior *pdf*:

$$p(\mathbf{d} \,|\, k, \mathcal{I}) = \int p(\mathbf{z}, \boldsymbol{\rho} \,|\, k, \mathcal{I}) p(\mathbf{d} \,|\, \mathbf{z}, \boldsymbol{\rho}, k, \mathcal{I}) \, d\mathbf{z} \, d\boldsymbol{\rho}. \tag{4}$$

As the marginal likelihood is not a function of $\mathbf{z}$ and $\boldsymbol{\rho}$, it is typically ignored in parameter estimation, so that for parameter estimation we need only to define the prior *pdf* and the likelihood.

3.1.1 *The prior pdf of* $\mathbf{z}$ *and* $\boldsymbol{\rho}$ *for a given number of layers* k

The natural choice for a prior *pdf* is the distribution that allows for the greatest uncertainty while obeying the constraints imposed by prior knowledge. I treat here the case where very little is known about $\mathbf{z}$ and $\boldsymbol{\rho}$ *a priori*, and it is natural to assume that prior knowledge of the depths of the layer interfaces is independent of prior knowledge of the layer resistivities. The prior *pdf* of $\mathbf{z}$ and $\boldsymbol{\rho}$ is then the product of the prior *pdf* s of each:

$$p(\mathbf{z}, \boldsymbol{\rho} \,|\, k, \mathcal{I}) = p(\mathbf{z} \,|\, k, \mathcal{I}) p(\boldsymbol{\rho} \,|\, k, \mathcal{I}). \tag{5}$$

Starting from the log-depths of the interfaces, I suppose that any configuration of k layers is as likely as any other *a priori*. In practice, the values of $\mathbf{z}$ will have to be between a minimum $\overline{z}_{\min}$ and a maximum $\overline{z}_{\max}$, and any layer will have a minimum log-thickness $\overline{h}_{\min}$. These values are set depending on the range of electrode spacings and on a maximum number of layers allowed (Appendix A). The appropriate prior *pdf* of $\mathbf{z}$ is then a modified version of the *pdf* of order-statistics:

$$p(\mathbf{z} \,|\, k, \mathcal{I}) = \frac{(k-1)!}{\prod_{i=1}^{k-1} \Delta_z(i-1)}, \tag{6}$$

where $\Delta_z(i) = (\overline{z}_{\max} - \overline{z}_{\min}) - (i+1)\overline{h}_{\min}$ is the log-depth interval available to place a layer interface when there are already i layer interfaces in the model, and $(k-1)!$ is the number of ways in which $(k-1)$ layer interfaces may be ordered (see 11.4 in Kendall & Stuart 1977).

For the log-resistivities in each layer, I assume that prior knowledge only gives a most probable value and a multiplicative factor within which the resistivity is expected to be; for example, one may expect *a priori* the resistivity to be within a factor of ten about a value of 100 Ω m. This information translates into a multivariate normal prior *pdf* for the log-resistivities $\boldsymbol{\rho}$:

$$p(\boldsymbol{\rho} \,|\, k, \mathcal{I}) = \frac{1}{\left[(2\pi)^k \det \overline{\mathbf{C}}_\rho\right]^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{\rho} - \overline{\boldsymbol{\rho}})^{\mathsf{T}} \overline{\mathbf{C}}_\rho^{-1} (\boldsymbol{\rho} - \overline{\boldsymbol{\rho}})\right], \tag{7}$$

where the prior mean vector $\overline{\boldsymbol{\rho}}$ and the prior covariance matrix $\overline{\mathbf{C}}_\rho$ are set on the basis of a most probable value and of the factor within which resistivity is expected to vary (Appendix A). Keeping with the principle of having as much uncertainty in the prior *pdf* as allowed by prior knowledge, $\overline{\mathbf{C}}_\rho$ is a diagonal matrix, so that there is no correlation *a priori* between the log-resistivities in different layers and no smoothing will be imposed. A diagonal prior covariance is the least informative state *a priori* because a prior covariance matrix with non-zero off-diagonal terms implies that knowledge of the resistivity of any single layer (say, from a hypothetical point measurement) has consequences on the resistivities of other layers.

3.1.2 *The likelihood function*

The likelihood function depends on the magnitude of the measurement error vector $\mathbf{e}$, defined as the difference between the N observed data $\mathbf{d}$ and the data predicted for a given value of the parameter vector:

$$\mathbf{e} = \mathbf{d} - \mathbf{g}(\mathbf{z}, \boldsymbol{\rho}, k) = \mathbf{d} - \mathbf{g}(\mathbf{m}). \tag{8}$$

I follow here the common assumption that the measurement errors have a normal distribution with zero mean and are uncorrelated. The likelihood function is then

$$p(\mathbf{d} \,|\, \mathbf{z}, \boldsymbol{\rho}, k, \mathcal{I}) = \frac{1}{\left[(2\pi)^N \det \overline{\mathbf{C}}_\mathbf{e}\right]^{1/2}} \exp\left(-\frac{1}{2}\mathbf{e}^{\mathsf{T}} \overline{\mathbf{C}}_\mathbf{e}^{-1} \mathbf{e}\right), \tag{9}$$

where the diagonal prior covariance matrix of the measurement error vector $\overline{\mathbf{C}}_\mathbf{e}$ contains the variances of the errors expected for each of the measured log-apparent resistivities in $\mathbf{d}$ (Appendix A). In practice, the likelihood decreases as the quadratic form $\mathbf{e}^{\mathsf{T}} \overline{\mathbf{C}}_\mathbf{e}^{-1} \mathbf{e}$ becomes larger, and thus it quantifies how likely the layered medium parameters are in light of the data.

3.2 Model selection: $p(k \,|\, \mathbf{d}, \mathcal{I})$

This is the problem of determining the posterior probability of having k layers in the earth model. In statistical terminology, it is a 'model selection' problem, in the sense that one wishes to extend parameter estimation and compare the relative merits of different

model parametrizations. Bayesian model selection has been described in the statistics and signal processing literature (e.g. Jeffreys 1939; Gull 1988; Bretthorst 1988; MacKay 1992; Jaynes 1995; Ó Ruanaidh & Fitzgerald 1996; Sivia 1996). An important feature of Bayesian model selection is that earth models with fewer layers will have greater posterior probabilities, as long as k is sufficiently large to fit the data. This is the key characteristic that makes Bayesian inference parsimonious without requiring an artificial smoothing of the solution.

Compared to parameter estimation, model selection is relatively new to geophysical inverse problems. For example, inversions of surface electromagnetic measurements are typically done using either a fixed, relatively small number of layers (Sternberg 1979; Sen & Stoffa 1995) or many thin layers where the values of resistivities are smoothed (Grandis *et al.* 1999; Schott *et al.* 1999). When solutions with different numbers of layers are compared (Jones & Hutton 1979; Tarits *et al.* 1994), the merits of each parametrization are not measured by their posterior probabilities.

In Bayesian inference, one can define a posterior probability distribution for the number of layers in the solution in the same way as for the layer parameters $\mathbf{z}$ and $\boldsymbol{\rho}$. Using Bayes's rule, the posterior *pdf* of k can be written as

$$p(k \mid \mathbf{d}, \mathcal{I}) = \frac{p(k \mid \mathcal{I}) p(\mathbf{d} \mid k, \mathcal{I})}{p(\mathbf{d} \mid \mathcal{I})}, \tag{10}$$

where $p(k \mid \mathcal{I})$ is the prior *pdf* of the number of layers and $p(\mathbf{d} \mid k, \mathcal{I})$ the marginal likelihood of eq. (4). Assuming that there is no definite *a priori* knowledge about the Earth's layering, I set the prior *pdf* of k to be uniform between one and $\overline{k}_{\max}$ layers. In other words, any number of layers is just as probable *a priori* as long as it is no more than $\overline{k}_{\max}$. The denominator of Bayes's rule in eq. (10) is again a normalizing factor; as k is a discrete variable, it is

$$p(\mathbf{d} \mid \mathcal{I}) = \sum_{i=1}^{\overline{k}_{\max}} p(k \mid \mathcal{I}) p(\mathbf{d} \mid k, \mathcal{I}) = \frac{1}{\overline{k}_{\max}} \sum_{i=1}^{\overline{k}_{\max}} p(\mathbf{d} \mid k, \mathcal{I}). \tag{11}$$

Except for a normalizing constant, if $p(k \mid \mathcal{I})$ is uniform the posterior probability of having k layers equals the marginal likelihood of eq. (4).

To see why the marginal likelihood is higher for parametrizations that fit the data and have relatively few layers, recall that it is the integral of the product prior-likelihood in parameter estimation (eq. 4). Consider a simplified case, where the prior *pdf* is uniform for both log-depths of layer interfaces and log-resistivities over a range $\overline{\Delta z}$ and $\overline{\Delta \rho}$, respectively, so that we can write the prior *pdf* of $\mathbf{z}$ and $\boldsymbol{\rho}$ as

$$p(\mathbf{z}, \boldsymbol{\rho} \mid k, \mathcal{I}) = \frac{1}{\overline{\Delta z}^{k-1} \, \overline{\Delta \rho}^{k}}. \tag{12}$$

Suppose also that the likelihood constrains the log-depths of layer interfaces and the log-resistivities to be in a posterior range $\widehat{\Delta z}$ and $\widehat{\Delta \rho}$, centered on maximum *a posteriori* values $\widehat{\mathbf{z}}$ and $\widehat{\boldsymbol{\rho}}$, respectively. Assuming that the likelihood is approximately constant over these posterior ranges, we can approximate the marginal likelihood integral of eq. (4) as

$$\begin{aligned} p(k \mid \mathbf{d}, \mathcal{I}) &\approx \left[\frac{\widehat{\Delta z}^{k-1} \, \widehat{\Delta \rho}^{k}}{\overline{\Delta z}^{k-1} \, \overline{\Delta \rho}^{k}} \right] \cdot [p(\mathbf{d} \mid \widehat{\mathbf{z}}, \widehat{\boldsymbol{\rho}}, k, \mathcal{I})] \\ &= [\text{Ockham factor}] \cdot [\text{Max. likelihood factor}] \end{aligned} \tag{13}$$

As the number of layers in the earth model increases, the 'Max. likelihood factor' in eq. (13) will typically increase, because with more layers the data will be fitted more closely. The 'Ockham factor' in eq. (13) (Gull 1988; MacKay 1992; Jaynes 1995; Ó Ruanaidh & Fitzgerald 1996; Sivia 1996) will instead typically decrease with increasing k, because the ratio of posterior and prior ranges for $\mathbf{z}$ and $\boldsymbol{\rho}$ will be less than one.

In other words, the Ockham factor measures how much prior probability is contained in the region of parameter space occupied by the posterior *pdf*, which is the region where the earth model parameters give a good fit to the data. As one allows for more parameters, the amount of prior probability in the region of parameter space where the data are fitted well (and hence the Ockham factor) will typically decrease, while the best fit to the data will generally improve (and the max. likelihood factor will increase).

A similar criterion to determine an optimal number of parameters in an inverse problem has been proposed by Ulrych *et al.* (2001), and named the 'Akaike Bayesian Information Criterion,' or ABIC. From eq. (22) of Ulrych *et al.* (2001), the minimum of

$$\text{ABIC} = -2 \log[p(\mathbf{d} \mid k, \mathcal{I})] + 2k \tag{14}$$

should be where the number of layers k is optimal. In terms of posterior probabilities, the ABIC equals two times the negative logarithm of the posterior *pdf* of the number of layers $p(k \mid \mathbf{d}, \mathcal{I})$ for a prior distribution $p(k \mid \mathcal{I}) \propto k^{-1}$. However, while Ulrych *et al.* (2001) suggest fixing the number of layers to the value that minimizes the ABIC, the full posterior *pdf* of k will be considered here. As it will become clear later, using the whole range that k can take a *a posteriori* accounts for a broader range of possibilities and thus gives a better measure of the final uncertainty.

In conclusion, Bayesian inference can evaluate the posterior probabilities of layered media with different numbers of layers, given a set of measurements. Models with fewer layers are more probable *a posteriori*, as long as there are enough layers to fit the data. This approach is a natural extension of the well-known Bayesian parameter estimation procedure; in fact, if a Bayesian approach is used to determine the posterior *pdf* of the parameters in the problem, there is no reason not to use the same approach for the posterior *pdf* of the number of layers. If Bayesian inference is properly applied to determine the posterior distribution of the number of layers, there should be no need to artificially smooth the solution. The number of layers needed will be effectively determined by the data.

4 MARKOV CHAIN MONTE CARLO (MCMC) ALGORITHM

In the previous section, I defined a posterior *pdf* for an earth model vector $\mathbf{m}$ that includes as a variable the number of layers in the earth model. Formally, the value of this posterior *pdf* can be computed (except for a constant) as the product of the marginal likelihood (eq. 4), the prior *pdf*s of $\mathbf{z}$ and $\boldsymbol{\rho}$ (eqs 6 and 7), and the likelihood function (eq. 9). An obvious practical difficulty, however, is to evaluate the marginal likelihood integral of eq. (4). If the inverse problem is linear and the prior *pdf*s of the earth model parameters are Gaussian, this integration is straightforward (Malinverno 2000). In the general nonlinear case examined here, however, the marginal likelihood does not have an easily integrable form, and evaluating its integral in a high-dimensional space is a difficult computational problem. This is a common problem in Bayesian inference applications, where the main difficulties are not in the formulation but in the calculations needed to characterize the solution, which typically require evaluating integrals of the posterior *pdf* in model space (Smith 1991).

Monte Carlo methods can efficiently evaluate integrals in high-dimensional space (e.g. Press *et al.* 1992, p. 304). In particular, Markov chain Monte Carlo (MCMC) algorithms have been found to be well suited for problems of Bayesian inference. References in the statistical literature are Gelman & Rubin (1992); Geyer (1992); Neal (1993); Tierney (1994); Besag *et al.* (1995); Gilks *et al.* (1996); applications to geophysical inverse problems are Mosegaard & Tarantola (1995); Sen & Stoffa (1995); Mosegaard *et al.* (1997); Grandis *et al.* (1999); Schott *et al.* (1999). MCMC algorithms consist of a random walk where different states (i.e. different values of a parameter vector) are visited and where the choice of the next state depends only on the value of the current state. After an initial 'burn-in' period in which the random walker moves toward the high posterior probability region, the chain samples a desired posterior *pdf*, that is, returns a number of parameter vectors that are distributed as in the posterior *pdf* (for a proof, see the references cited above). It is then easy to compute any desired characteristic of the posterior *pdf* (e.g. posterior means and standard deviations of the parameters) from the sample obtained by the chain, effectively carrying out an integration. It should be emphasized that MCMC algorithms are not the same as Monte Carlo methods where models are chosen independently and tested against the data (e.g. Press 1968; Sternberg 1979). A MCMC algorithm effectively has a memory mechanism that makes the chain stay in the high posterior probability region of model space, and thus it is much more efficient than a sequence of independently picked models, most of which will have a very low posterior probability.

A widely used MCMC algorithm is the Metropolis–Hastings algorithm (Metropolis *et al.* 1953; Hastings 1970; Chib & Greenberg 1995), which is a two-step procedure. In the first step, the current model parameter vector $\mathbf{m}$ in the Markov chain is modified at random to obtain a candidate vector $\mathbf{m}'$. This candidate is drawn from a proposal distribution $q(\mathbf{m}' \mid \mathbf{m})$, where the choice of $\mathbf{m}'$ depends on the current vector $\mathbf{m}$. In the second step, the candidate is accepted with an acceptance probability

$$\begin{aligned}\alpha &= \min\left[1, \frac{p(\mathbf{m}' \mid \mathbf{d}, \mathcal{I})}{p(\mathbf{m} \mid \mathbf{d}, \mathcal{I})} \cdot \frac{q(\mathbf{m} \mid \mathbf{m}')}{q(\mathbf{m}' \mid \mathbf{m}}\right] \\ &= \min\left[1, \frac{p(\mathbf{m}' \mid \mathcal{I})}{p(\mathbf{m}\mathcal{I})} \cdot \frac{p(\mathbf{d} \mid \mathbf{m}', \mathcal{I})}{p(\mathbf{d} \mid \mathbf{m}, \mathcal{I})} \cdot \frac{q(\mathbf{m} \mid \mathbf{m}')}{q(\mathbf{m}' \mid \mathbf{m})}\right] \\ &= \min[1, (\text{Prior ratio}) \cdot (\text{Likelihood ratio}) \cdot (\text{Proposal ratio})].\end{aligned} \tag{15}$$

If the candidate is accepted, the state of the chain is changed to $\mathbf{m}'$; otherwise, the chain stays at $\mathbf{m}$. If one were to always accept the candidate states drawn from the proposal *pdf*, the chain would take a random walk in the space of $\mathbf{m}$. The acceptance probability in eq. (15) effectively biases the random walk toward regions of model space that have higher posterior probabilities. The 'Proposal ratio' term in eq. (15) applies a correction if the move from $\mathbf{m}$ to $\mathbf{m}'$ has a different probability of being proposed than the opposite move from $\mathbf{m}'$ to $\mathbf{m}$.

The Metropolis–Hastings algorithm is typically applied to cases where the dimension of the parameter vector $\mathbf{m}$ is fixed at the outset. Recently, Green (1995) showed that a Metropolis–Hastings algorithm with the acceptance probability of eq. (15) can be applied to inference problems where the number of parameters in $\mathbf{m}$ is not fixed *a priori*, i.e. to problems where 'the number of things you don't know is one of the things you don't know'. Green's expression for the acceptance probability, on his p. 720, also contains a 'Jacobian ratio' term, which is always 1 for the candidates proposed here. The method only requires the ability to evaluate the prior *pdf* and the likelihood function for any value of $\mathbf{m}$, and there is no need to compute the marginal likelihood integral of eq. (4). Green (1995) calls this a 'reversible jump' MCMC algorithm, because the chain can jump back and forth between different spaces corresponding to different numbers of parameters. These jumps are done by occasionally proposing to add a layer interface and split a layer in two, or to delete a layer interface and merge two layers.

In the following, I give a schematic outline of how the method of Green (1995) can be implemented for a nonlinear geophysical inverse problem. The details are given in the appendixes: Appendix A describes the parameters that need to be specified *a priori* and the criteria used to set them; Appendix B describes the method used to choose candidate models, paying particular attention to making the MCMC sampling efficient; and Appendix C gives an explicit expression for the acceptance probability of eq. (15).

To start, the Markov chain needs an arbitrary initial earth model $\mathbf{m}$ (e.g. two layers with an interface placed in the middle of the depth interval of interest and an initial log-resistivity equal to the prior mean). The sampling then follows this loop:

(i) Start the sampling of a candidate model $\mathbf{m}'$ based on $\mathbf{m}$ by choosing a candidate number of layers k' and candidate log-depths of layer interfaces $\mathbf{z}'$; this is done by taking k and $\mathbf{z}$ of $\mathbf{m}$ and choosing at random between one of the four following moves (Appendix B):

(a) Add a layer interface at a random location
(b) Delete a layer interface chosen at random
(c) Perturb the depth of a layer interface chosen at random
(d) Leave the layer interfaces unchanged

(ii) Complete constructing $\mathbf{m}'$ by sampling a candidate value for the log-resistivities $\boldsymbol{\rho}'$ in each layer from the proposal *pdf* of eq. (B.2) (Appendix B)

(iii) Compute probability of acceptance α as in eq. (C5) (Appendix C)

(iv) Sample a random number r between 0 and 1; if $r < \alpha$, accept the candidate and set $\mathbf{m} = \mathbf{m}'$, otherwise leave $\mathbf{m}$ unchanged

(v) If the burn-in period is over (see below), output $\mathbf{m} = (k, \mathbf{z}, \boldsymbol{\rho})$

(vi) Go to step (i)

As noted earlier, the MCMC algorithm will not be immediately sampling the posterior *pdf*, but it will need a 'burn-in' period of some number of iterations to converge to the high-probability region. A simple way to diagnose the end of the burn-in period in the problem discussed here is to monitor how well the data predicted by the current model $\mathbf{g}(\mathbf{m})$ match the measured data $\mathbf{d}$. As the chain is started from an arbitrary point, at the beginning the fit to the data will be very poor, and it will improve as the iterations progress. The burn-in period is estimated to end the first time that the standard deviation of the data misfit vector $\mathbf{e} = \mathbf{d} - \mathbf{g}(\mathbf{m})$ is less than the standard deviation of the expected measurement errors.

Once the burn-in period is over, the parameter vectors output by the chain should be distributed as in the posterior *pdf*. The next question is obviously when to stop, i.e. how to decide when the sample obtained by the MCMC algorithm is large enough to characterize the posterior *pdf*. This problem is the subject of active MCMC research (Gelman & Rubin 1992; Cowles & Carlin 1996; Brooks & Gelman 1998; Brooks & Giudici 1999), and will not be treated here. I will use a simple practical criterion, and continue the MCMC iterations until the characteristics of the posterior *pdf* stop changing significantly.

Similar MCMC algorithms have been described and applied to geophysical inverse problems by other authors. Mosegaard & Tarantola (1995) and Mosegaard *et al.* (1997) described a method

based on the Metropolis–Hastings algorithm where the number of layers in the earth model is allowed to vary. The two main differences with the approach described here are that these authors did not consider the issue of model simplicity by evaluating the posterior probability of having different numbers of layers, and that in their implementation the candidate models in the chain are chosen from the prior *pdf*. As noted in Appendix B, different ways to choose a candidate can significantly affect the efficiency of Metropolis–Hastings MCMC sampling, and setting the proposal *pdf* on the basis of an approximation to the posterior *pdf* as done here is more efficient than picking candidates from the prior. Other MCMC approaches allow only for a fixed number of layers (Sen & Stoffa 1995), or introduce a smoothing factor that is not determined by *a priori* information and thus casts doubt on the final quantification of uncertainty (Grandis *et al.* 1999; Schott *et al.* 1999).

5 RESULTS (SYNTHETIC DATA)

In this section I apply the MCMC algorithm described above to invert a synthetic data set of DC resistivity soundings. The purpose of this exercise is to check that the inferences one makes *a posteriori* from the Markov chain Monte Carlo sampling results are consistent with a true, known earth model. The simple three-layer earth model I use is the same as that employed by Sen & Stoffa (1995) and Schott *et al.* (1999), and is shown in Fig. 2. The synthetic measurements were computed with the method described earlier in the forward model section, adding to the log-apparent resistivities random Gaussian noise with zero mean and a standard deviation equal to 10 per cent of the apparent resistivity value.

The prior *pdf* used in the inversion was set assuming that little was known *a priori*. The prior probability of the number of layers is uniform from one up to $\overline{k}_{\max} = 30$ layers, and any configuration of layer interfaces is equally probable *a priori* in a depth interval of 0.1–1000 m determined by the span of electrode spacings in the data (for details, see Appendix A). The resistivity in each layer was assumed to be within a factor of five about a most probable value of 50 Ω m, corresponding to a broad prior 95 per cent confidence interval of 2–1250 Ω m.

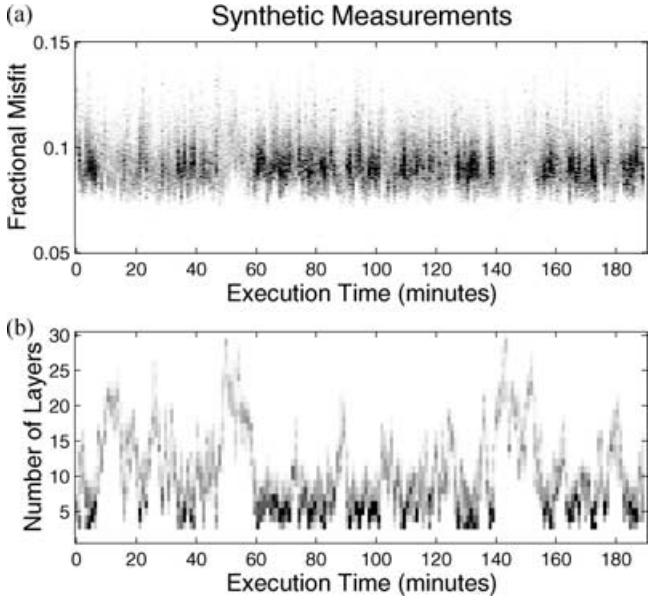

Figure 3. Progress of the algorithm in terms of how well each sampled layered model fits the measured data (a) and of how many layers each model contains (b).

The MCMC sampling algorithm was started from a two-layer model with an interface in the middle and layer resistivities equal to the prior mean. The sampling was then run for 500 000 iterations, which took about 3 h on a Sun Ultra 60 workstation. After these many iterations, the results of the sampling did not appear to change appreciably. The progress of the algorithm is illustrated in Fig. 3, which plots the average fractional misfit (the root-mean-square difference between the measured data and the data predicted by each sampled model) and the number of layers in each sampled layered model. The starting model fit the data very poorly (the fractional misfit being about 2.5), but the expected misfit level of 0.1 was reached quickly after 657 iterations (17 s of execution time). The layered media sampled during this short 'burn-in' period were

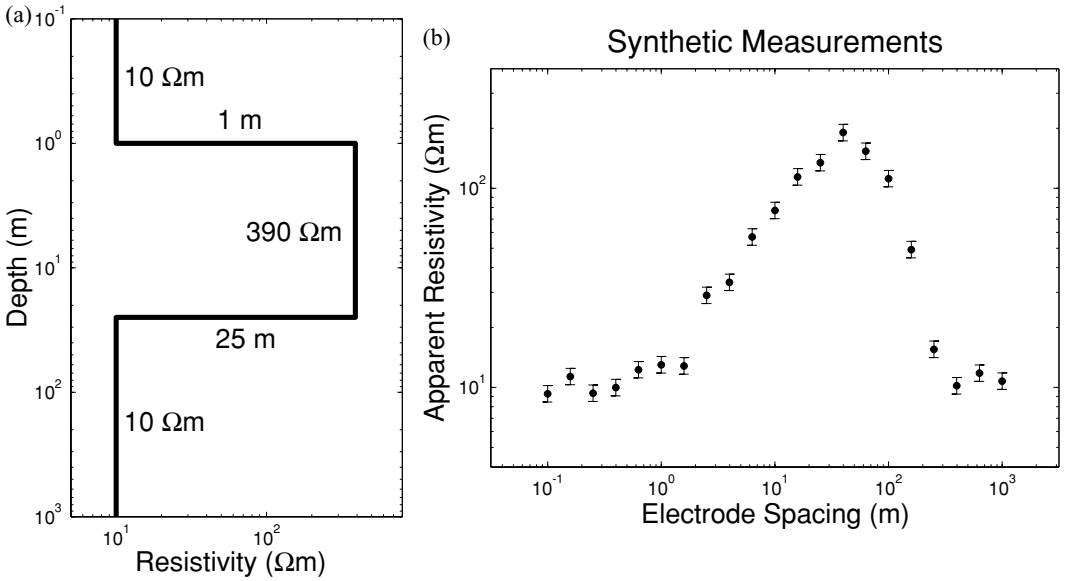

Figure 2. Three-layer earth model (a) and data (b) used in the synthetic example.

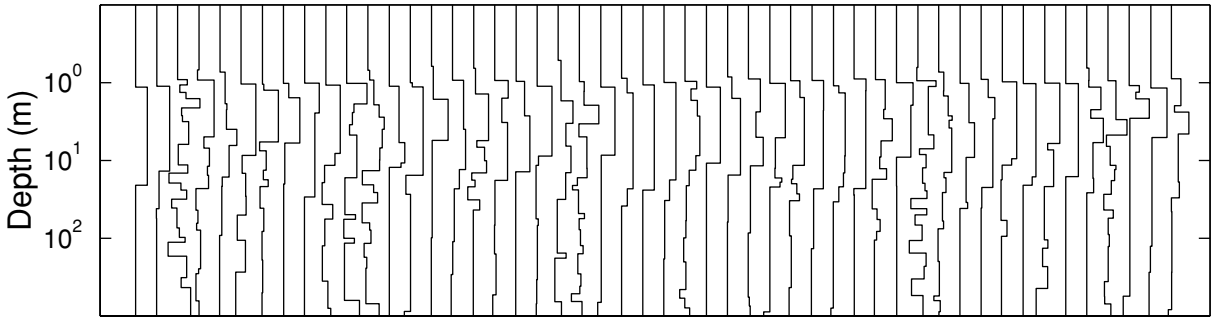

Figure 4. Fifty layered media sampled by the MCMC algorithm for the synthetic data in Fig. 2. The horizontal separation between layered media corresponds to a factor of 1000 in resistivity.

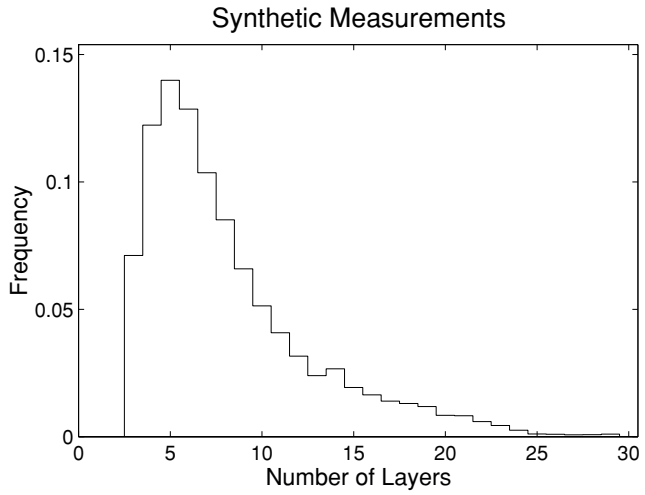

Figure 5. Histogram of the number of layers sampled by the MCMC algorithm for the synthetic data in Fig. 2. This histogram approximates the posterior *pdf* of the number of layers; note that the prior *pdf* is uniform.

discarded; after this initial interval, the chain sampled the posterior *pdf* of layered media consistent with the data, and all the sampled layered media were retained. Fig. 4 shows fifty layered media sampled by the MCMC algorithm at intervals of 10 000 iterations over the whole run.

Figs 3 and 4 also show that the chain mostly samples models that contain relatively few layers, and only occasionally returns models with more than 10 layers. A histogram of the number of layers sampled by the chain is in Fig. 5; as the chain is sampling the posterior *pdf*, this histogram is an estimate of the posterior *pdf* of the number of layers. No models with fewer than three layers are ever sampled, because three layers are the absolute minimum needed to adequately fit the data; 91 per cent of the models sampled have 15 or fewer layers.

While the most probable number of layers in Fig. 5 is five, a posterior image obtained by superimposing all values of resistivity in the sampled layered models defines a clear three-layer structure with a resistive layer sandwiched between two conductive layers (Fig. 6). A three-layer structure is also evident in a posterior histogram of the depths of the layer interfaces (Fig. 7), which shows two well defined peaks, a sharp peak centered at a depth of 1 m and a broader peak between 5 and 40 m (the layer interfaces in the true model are at depths of 1 m and 25 m). The many sampled layered media that have more than three layers contain minor departures from this basic three-layer structure.

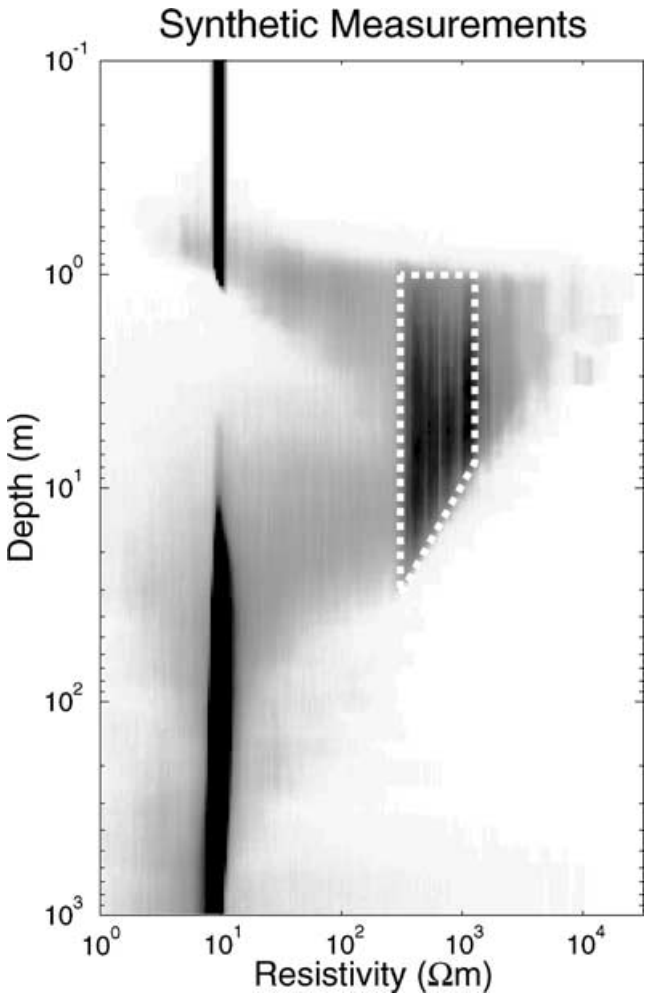

Figure 6. Image obtained by superimposing the values of resistivity in the layered media sampled by the MCMC algorithm for the synthetic data in Fig. 2. This image is an estimated display of the posterior marginal *pdf* of resistivity at different depths. The dotted white line shows the range of resistivity and thickness of the middle layer that are consistent with the data of Fig. 2 (see the text and Fig. 8).

The results of Fig. 6 are consistent with a simple non-uniqueness analysis of the inverse problem. It is well known that there are many equivalent layered media that fit equally well DC resisitivity sounding data, and that the data only constrain some combination of the layer resistivities and thicknesses (Keller & Frischknecht 1966, Chapter 3). The flat parts of the measured apparent resistivity curve

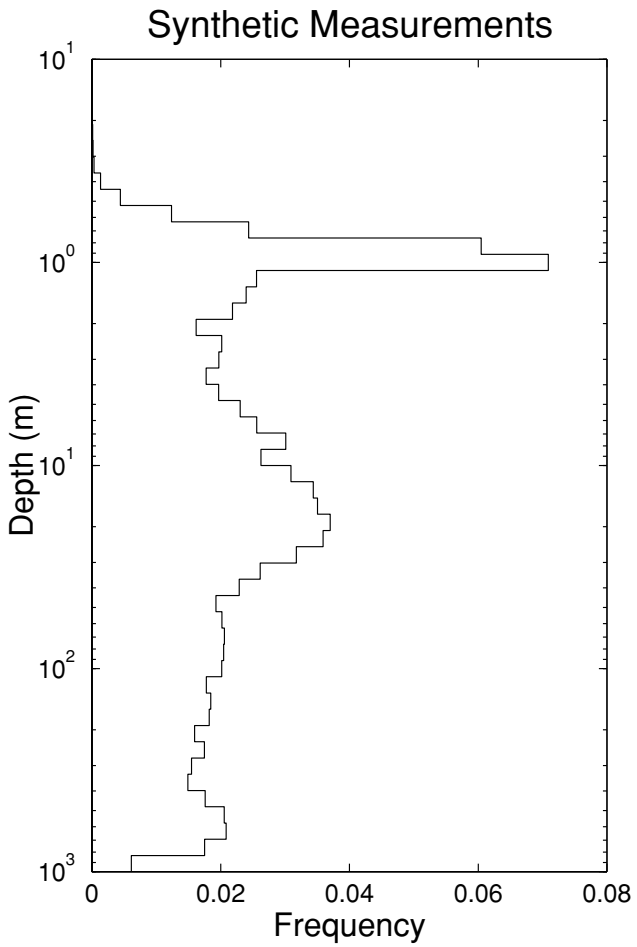

Figure 7. Histogram of the depths to interfaces in the layered media sampled by the MCMC algorithm for the synthetic data in Fig. 2. This histogram approximates the posterior *pdf* of the depths to layer interfaces.

in Fig. 2 at short (<1 m) and long spacings (>400 m) constrain the value of the resistivity of the top and bottom layers. The linear increase in apparent resistivity for electrode spacings between 1 and 40 m is a function of the ratio resistivity/thickness of the top layer; as the resistivity of the top layer is constrained to be around 10 Ω m, its thickness is well determined by the data to be about 1 m. Fig. 6 indeed shows that the resistivities of the top and bottom layer and the thickness of the top layer have small posterior uncertainties.

The main non-uniqueness in the results of Fig. 6 is in the resistivity and thickness of the middle layer. In the case of three layers with a resistive middle layer, the shape of the hump in the measured apparent resistivity curve depends primarily on the product resistivity-thickness of the middle layer. To determine the range of resistivity and thickness consistent with the measurements of Fig. 2, I plot in Fig. 8 how well the data are fitted as resistivity and thickness of the middle layer vary while keeping their product constant to 390 Ω m × 24 m. The residual misfit is below the expected value of 0.1 for resistivities of the middle layer between about 300 and 1240 Ω m, corresponding to thicknesses between about 31 and 7.5 m. This interval of resistivities and thicknesses is marked by a dotted line in Fig. 6, and it encompasses the area of highest posterior probabilities for the middle layer.

The results presented here are similar to those obtained by Sen & Stoffa (1995). The main difference is that these authors consid-

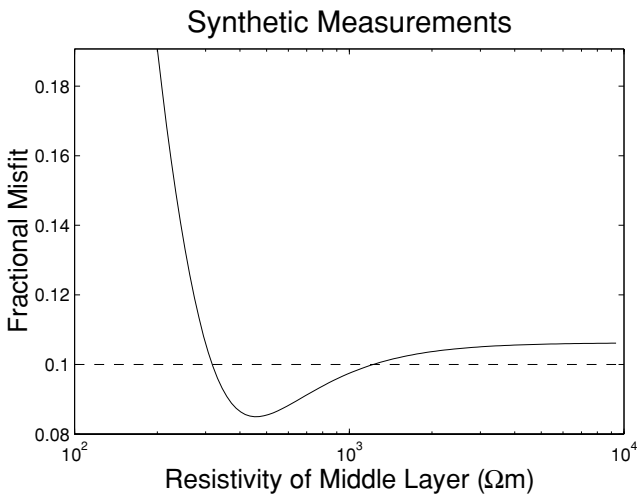

Figure 8. Fractional misfit to the data of Fig. 2 for three-layer layered models whose top and bottom layers have a resistivity of 10 Ω m, while the resistivity and thickness of the middle layer are varied keeping constant their product to 390 Ω m × 24 m (solid line). The data are fitted below the expected misfit level of 0.1 (dashed line) for resistivities of the middle layer between about 300 and 1240 Ω m (thicknesses between about 31 and 7.5 m). This interval of resistivities and thicknesses is marked by a dotted white line in the posterior image of Fig. 6.

ered only models with three layers and carried out a search on a finite interval for each model parameter (three layer resistivities and two layer thicknesses). Letting the number of layers vary and not restricting the resistivity value to a finite interval results in a greater range of plausible models. For example, while Sen & Stoffa (1995) allowed for possible values of resistivity between 100 and 450 Ω m in the middle resistive layer, the results of Figs 6 and 8 show that resistivities greater than a thousand Ω m are consistent with the data in the middle layer.

Schott *et al.* (1999) used a regularization factor in their MCMC and thus obtained a sample of solutions that were constrained to be smooth (compare their Fig. 1 with Fig. 6 in this paper). All these smooth solutions show a broad transition both at the top and at the bottom of the middle resistive layer. The results of Fig. 6, however, suggest that the depth to the top of the middle layer is much better resolved than the depth to the bottom of the middle layer. Regularization smears these transitions uniformly throughout the depth interval being investigated, and thus does not allow for distinguishing layer interfaces that are well resolved from those that are not. In addition, the posterior uncertainties of resistivity in Fig. 1 of Schott *et al.* (1999) are relatively small (about a factor of two) and uniform over the whole depth interval considered. In contrast, the results of Fig. 6 indicate that posterior uncertainties in resistivities are much greater at the center than at the top and bottom.

6 RESULTS (FIELD DATA)

In this section I apply the MCMC algorithm to the inversion of field DC resistivity soundings using the data set from the central Australia Wauchope station of Constable *et al.* (1984) (Fig. 9). As in the synthetic data example, I assume that very little is known *a priori*: the prior probability of the number of layers is uniform from one up to $\overline{k}_{\max} = 30$ layers, and any configuration of layer interfaces is equally probable *a priori* in a depth interval of 0.1–100 000 m. The resistivity in each layer was assumed to be *a priori* within a factor

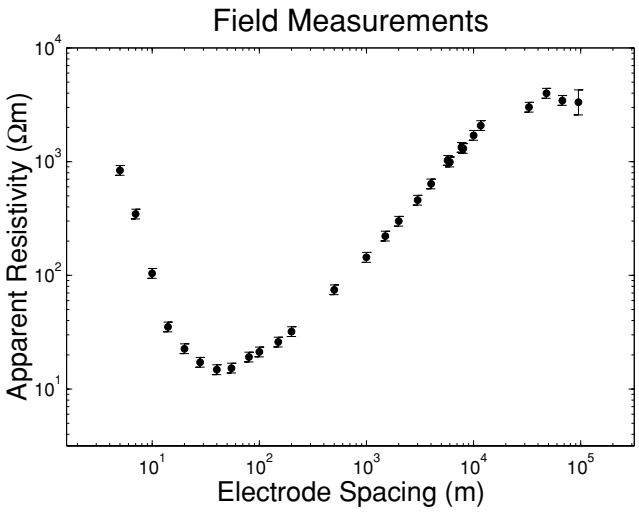

Figure 9. Field measurements from the Wauchope DC resistivity sounding, central Australia, after Constable *et al.* (1984).

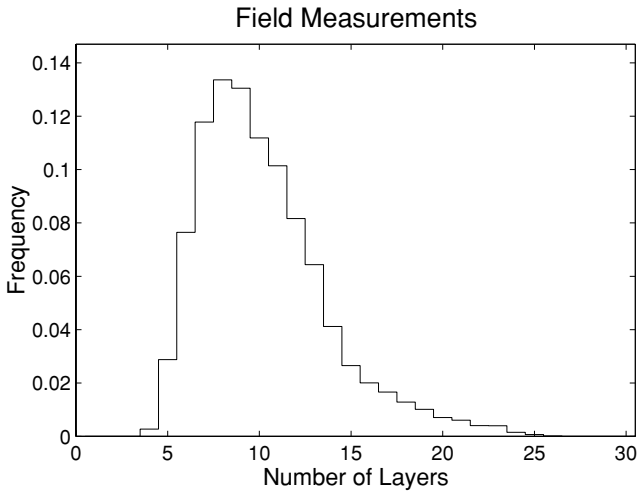

Figure 11. Histogram of the number of layers sampled by the MCMC algorithm for the field data in Fig. 9. This histogram approximates the posterior *pdf* of the number of layers; note that the prior *pdf* is uniform.

of ten about a most probable value of 1000 Ω m, corresponding to a very broad prior 95 per cent confidence interval of 10–100 000 Ω m. The measurement errors were set to the values given by Constable *et al.* (1984). The MCMC sampling started from an initial featureless two-layer model and continued for 500 000 iterations. As in the synthetic data example, after these many iterations the results of the sampling did not appear to change significantly. Fig. 10 shows fifty layered media sampled by the MCMC algorithm at intervals of 10 000 iterations over the whole run.

Fig. 11 is a histogram of the number of layers sampled by the chain after the initial 'burn-in' period; this histogram estimates the posterior *pdf* of the number of layers. At least four layers are required by the data, and most layered models sampled by the chain have less than 15 layers. The posterior image of resistivity (Fig. 12) and the posterior histogram of the depths to the layer interfaces (Fig. 13) obtained from all the sampled layered models define four to five major layers. While the resistivities of the layers and the depth to the interfaces are relatively well constrained *a posteriori* down to about 100 m depth, there are large posterior uncertainties in the resistivities and the interface depths of the deeper layers.

The inversion results presented here are generally in agreement with those of Constable *et al.* (1984) and Constable *et al.* (1987). Constable *et al.* (1984) obtain a six-layer model whose resistivities and layer thicknesses are close to those illustrated in Fig. 12. These authors analyse the correlation of the inverted resistivities and thicknesses and note that only the product resistivity-thickness is well determined for the resistive fifth layer. The MCMC sampling carried out here shows this fundamental non-uniqueness for the resistive crustal layer with a top at about 100 m and a bottom between 500 and 10 000 m (Fig. 12; see also the earlier discussion of non-uniqueness in the inversion of synthetic data). While Constable *et al.* (1984) note that the data seem to indicate a low-resistivity bottom layer, they express some doubts about the significance of this result, because their analysis shows that the resistivity in the deep layer is poorly resolved. The MCMC sampling results agree with their conclusion, as Fig. 12 shows that the posterior distribution of resistivity in the deepest layer is very close to the prior distribution (which has a mean of 1000 Ω m and a 95 per cent confidence interval of 10–100 000 Ω m), meaning that the data do not provide significant information on resistivity at this depth. Finally, the median posterior resistivity in Fig. 12 (thick solid line) is close to the maximally smooth solution of Constable *et al.* (1987) (see their Fig. 2). The main difference is that the change in resistivity around 100 m depth is more abrupt in the results obtained here, because there is no smoothing enforced *a priori*.

7 DISCUSSION AND CONCLUSIONS

In this paper, I presented an extension of the often used Bayesian parameter estimation to include the number of free parameters in the earth model being inverted (specifically, the number of layers). As the posterior probability of earth models that fit the data becomes

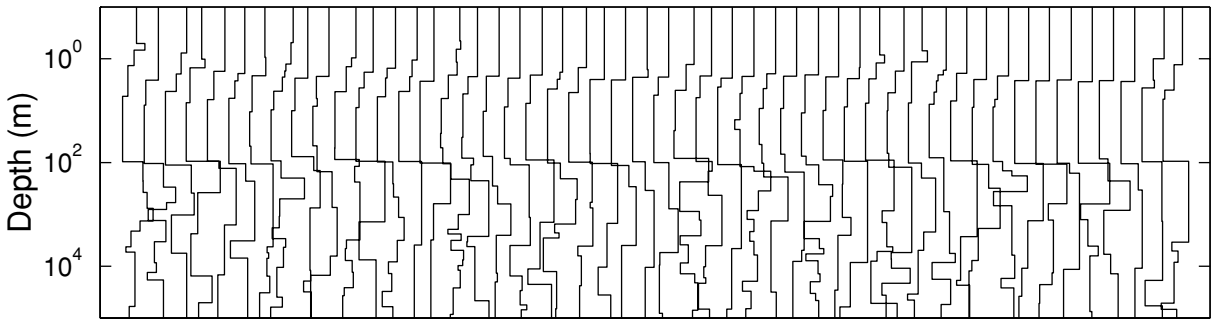

Figure 10. Fifty layered media sampled by the MCMC algorithm for the field data in Fig. 9. The horizontal separation between layered media corresponds to a factor of 1000 in resistivity.

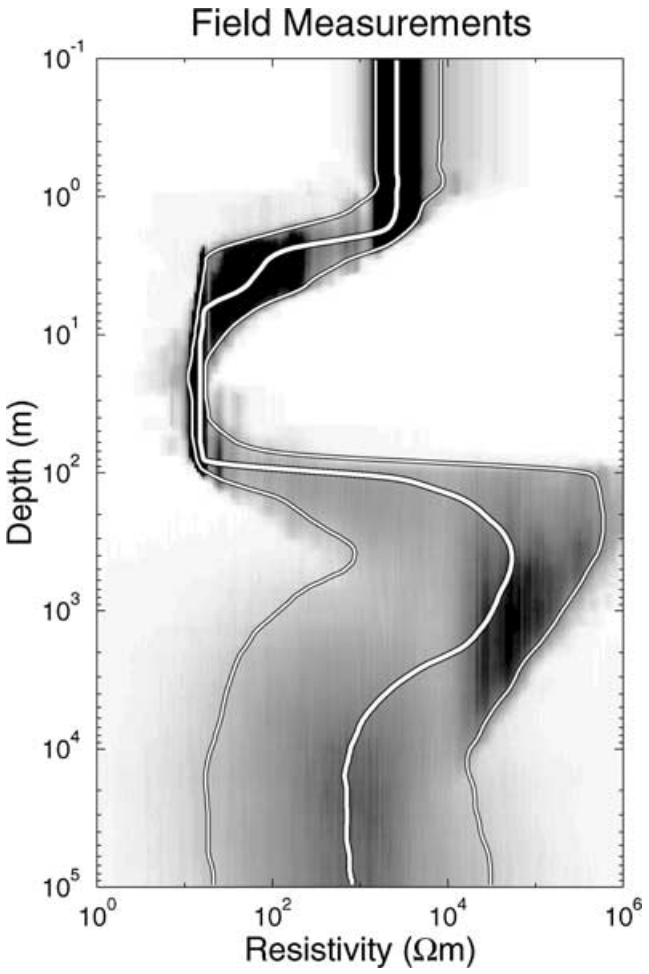

Figure 12. Image obtained by superimposing the values of resistivity in the layered media sampled by the MCMC algorithm for the field data in Fig. 9. This image is an estimated display of the posterior marginal *pdf* of resistivity at different depths. The continuous curves are the median value of resistivity (thick line) and the 5 per cent and 95 per cent bounds (thin lines).

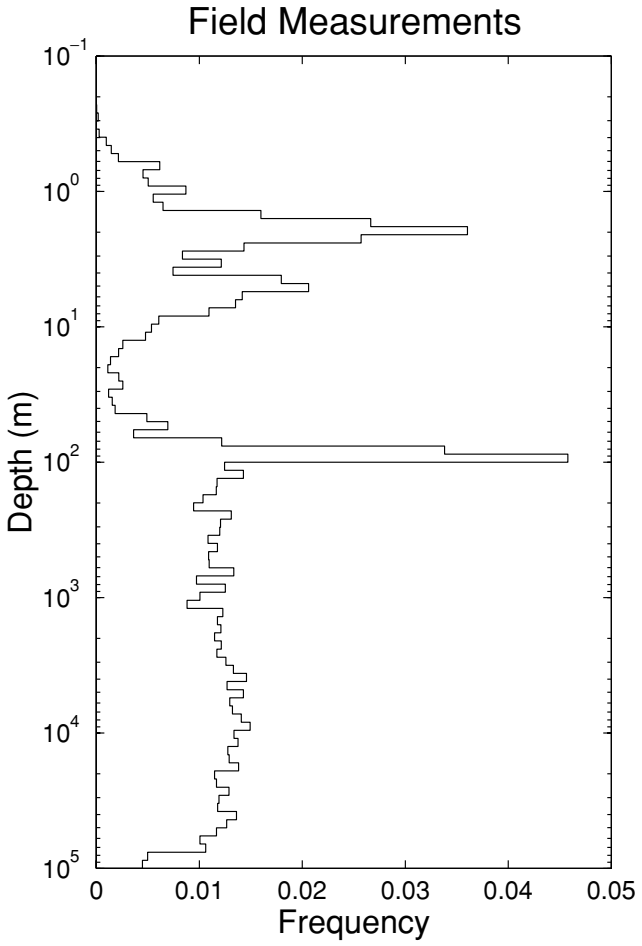

Figure 13. Histogram of the depths to interfaces in the layered media sampled by the MCMC algorithm for the field data in Fig. 9. This histogram approximates the posterior *pdf* of the depths to layer interfaces.

lower as the number of layers increases, this procedure ensures that parsimonious model descriptions are preferred without imposing an additional simplicity requirement. This formulation has been implemented using a Markov chain Monte Carlo algorithm that returns a sample of layered models distributed as in the posterior *pdf*. This sample spans the space of earth models that fit the data within a specified measurement error (Fig. 14).

The formulation presented here addresses two basic problems in solving the non-uniqueness half of inverse problems: the difficulty of setting a prior distribution when little is known *a priori* and the dependence of the posterior uncertainty on a particular earth model parametrization. While Bayesian inference has many desirable qualities for geophysical data inversion, setting the prior *pdf* on the basis of available knowledge is often problematic (e.g. Backus 1988; Ulrych *et al.* 2001; Scales & Tenorio 2001). Setting the prior distribution is equivalent to formulating an initial hypothesis that should then be modified by the data as needed (Ellis & Oldenburg 1994). If little is known *a priori*, the prior hypothesis should allow for a variety of possible earth models and for a broad range for the parameters in these models. But this does not seem generally feasible: to solve the problem in practice, it seems that one must choose a particular parametrization (e.g. fix the number of layers) and impose a regularization factor that goes well beyond prior knowledge.

This paper shows how it is possible to set a simple prior *pdf* with earth model parameters that are effectively unknown *a priori* (such as the number of layers) being left as such. The *a posteriori* span of all the unknown parameters is then effectively determined by the data. Because earth models that have fewer layers have higher posterior probabilities, there is no need to impose smoothness on the solution beyond what is implied by prior information. In the examples presented here, any layer configuration up to a maximum number of layers was equally probable *a priori*, while the prior *pdf* of log-resistivity in each layer was set to a normal *pdf* with a large prior standard deviation and no correlations between different layers (that is, no regularization).

Broadening the space of possible earth model parametrizations also addresses a fundamental problem noted by Constable *et al.* (1987), namely, that the results of inversion are subject to the choice of parametrization. The answer offered to this problem by Constable *et al.* (1987) is to drastically restrict the space of possible solutions and choose the smoothest earth model that fits the data. While this approach obviously provides a single, sensible earth model that fits the data, the non-uniqueness or uncertainty of the inversion cannot be properly assessed because of regularization: smoothed solutions can only span a narrow range of possibilities.

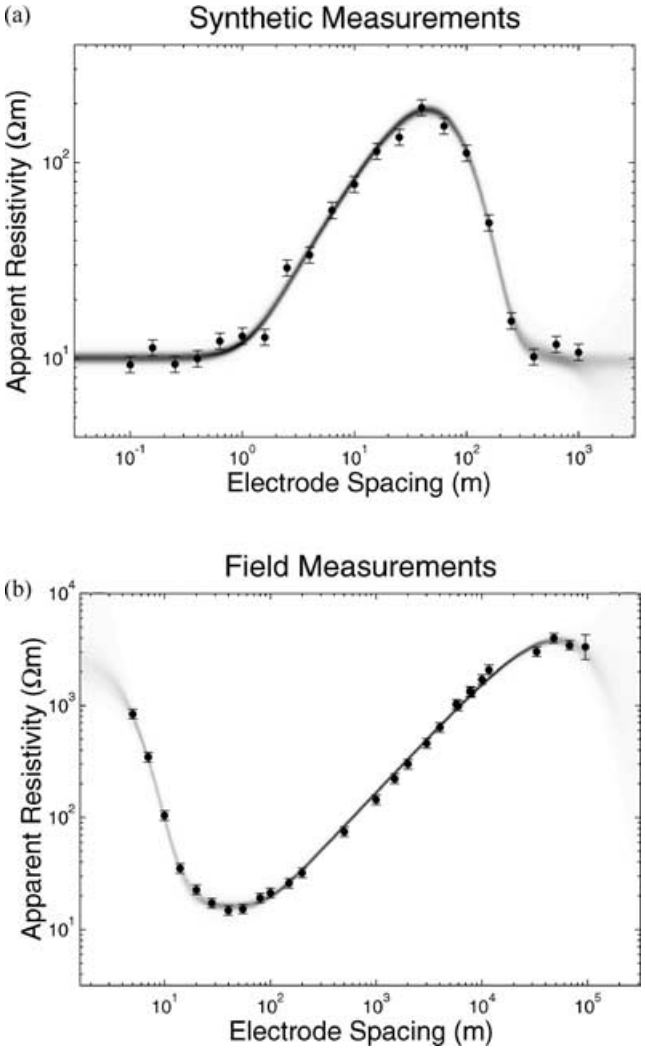

Figure 14. The images in the background are histograms of the apparent resistivity curves predicted by the layered media sampled by the MCMC algorithm in the synthetic (a) and field data example (b).

The alternative offered here to make inference less sensitive to the parametrization is to first allow for a range as broad as possible of different and potentially complex earth model parametrizations, and then to assess their relative merits (i.e. posterior probabilities) in light of the data. If Bayesian inference is properly applied to quantify the posterior probability of the various candidate parametrizations, the data effectively choose the level of complexity needed in the earth model. This parsimony characteristic ensures that the *pdf* of parametrizations and associated parameters gives a realistic measure of the posterior uncertainty of the earth model.

It should be emphasized that while the approach described here accounts properly for the uncertainty due to not knowing the number of layers in the earth model, the posterior uncertainty will still be subject to the *a priori* choice of a class of parametrizations. If the earth models considered are layered media with no smoothness imposed *a priori*, then solutions with a few layers and large resistivity jumps will be the most probable *a posteriori*. If the earth model were a continuous function, the results might be different. Constable *et al.* (1987) correctly point out that resistivity sounding data do not necessarily require sudden jumps in resistivity; their inverted earth model for the Wauchope data with a gradual transition from low to high resistivity between depths of 50–1000 m fits the data as well as the sampled layered media in Fig. 12, most of which contain a jump in resistivity at 100 m depth.

The approach presented here, however, can be immediately extended to other classes of parametrizations where the earth model consists of many thin layers. For example, just as done in this paper for the number of layers, one may determine the posterior distribution of the number of nodes in a cubic spline interpolation (e.g. Malinverno 2000) or of the value of a correlation length in the prior distribution of resistivity. In principle, the results obtained with different classes of parametrizations could be ranked on the basis of their respective posterior probabilities. Again, the point is that the more parametrizations one examines, the more extensively one can address the uniqueness half of the inverse problem. The intent of this paper is to show how to decrease the dependence on the parametrization of the inverse problem and thus obtain a more general quantification of the uncertainty in the inferred earth model.

ACKNOWLEDGMENTS

Thanks to Michael Prauge, Albert Tarautole and an anonymous reviewer for discussions and valuable suggestions.

REFERENCES

Backus, G.E., 1988. Bayesian inference in geomagnetism, *Geophys. J.,* **92,** 125–142.

Bard, Y., 1974. *Nonlinear Parameter Estimation,* Academic Press, New York.

Besag, J., Green, P., Higdon, D. & Mengersen, K., 1995. Bayesian computation and stochastic systems (with discussion), *Stat. Sci.,* **10,** 3–66.

Bretthorst, G.L., 1988. *Bayesian Spectrum Analysis and Parameter Estimation,* Vol. 48, Lecture Notes in Statistics, Springer Verlag, New York.

Brooks, S.P. & Gelman, A., 1998. General methods for monitoring convergence of iterative simulations, *J. Comput. Graph. Stat.,* **7,** 434–455.

Brooks, S.P. & Giudici, P., 1999. Convergence assessment for Reversible Jump MCMC simulations, in *Bayesian Statistics, Vol. 6,* pp. 733–742, eds Bernardo, J.M., Berger, J.O., Dawid, A.P. & Smith, A.F.M., Oxford University Press, Oxford.

Cary, P.W. & Chapman, C.H., 1988. Automatic 1-D waveform inversion of marine seismic refraction data, *Geophys. J.,* **93,** 527–546.

Chib, S. & Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm, *Amer. Stat.,* **49,** 327–335.

Constable, S.C., McElhinny, M.W. & McFadden, P.L., 1984. Deep Schlumberger sounding and the crustal resistivity structure of central Australia, *Geophys. J. R. astr. Soc.,* **79,** 893–910.

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

Cowles, M.K. & Carlin, B.P., 1996. Markov chain Monte Carlo convergence diagnostics: A comparative review, *J. Amer. Stat. Ass.,* **91,** 883–904.

Denison, D.G.T., Mallick, B.K. & Smith, A.F.M., 1998. Automatic Bayesian curve fitting, *J. R. Statist. Soc. B,* **60,** 333–350.

Duijndam, A.J.W., 1988. Bayesian estimation in seismic inversion, part I: Principles, *Geophys. Prospect.,* **36,** 878–898.

Ellis, R.G. & Oldenburg, D.W., 1994. Applied geophysical inversion, *Geophys. J. Int.,* **116,** 5–11.

Gelman, A. & Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences (with discussion), *Stat. Sci.,* **7,** 457–511.

Gelman, A., Roberts, G.O. & Gilks, W.R., 1996. Efficient Metropolis jumping rules, in *Bayesian Statistics, Vol. 5,* pp. 599–607, eds Bernardo, J.M., Berger, J.O., Dawid, A.P. & Smith, A.F.M., Oxford University Press, Oxford.

Geyer, C.J., 1992. Practical Markov chain Monte Carlo (with discussion), *Stat. Sci.,* **7,** 473–511.

Gilks, W.R., Richardson, S. & Spiegelhalter, D.J., (eds), 1996. *Markov Chain Monte Carlo in Practice,* Chapman and Hall, London.

Gouveia, W.P. & Scales, J.A., 1998. Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis, *J. geophys. Res.,* **103,** 2759–2779.

Grandis, H., Menvielle, M. & Roussignol, M., 1999. Bayesian inversion with Markov chains–I. The magnetotelluric one-dimensional case, *Geophys. J. Int.,* **138,** 757–768.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika,* **82,** 711–732.

Gull, S.F., 1988. Bayesian inductive inference and maximum entropy, in *Maximum Entropy and Bayesian Methods in Science and Engineering, Vol. 1,* pp. 53–74, eds Erickson, G.J. & Smith, C.R., Kluwer Academic, Dordrecht.

Guptasarma, D., 1982. Optimization of short digital linear filters for increased accuracy, *Geophys. Prospect.,* **30,** 501–514.

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

Jackson, D.D. & Matsu'ura, M., 1985. A Bayesian approach to nonlinear inversion, *J. geophys. Res.,* **90,** 581–591.

Jaynes, E.T., 1984. Prior information and ambiguity in inverse problems, in *Inverse Problems, Vol. 14,* pp. 151–166, ed., McLaughlin, D., of SIAM-AMS Proceedings, American Mathematical Society, Providence.

Jaynes, E.T., 1995. *Probability Theory: The Logic of Science,* Fragmentary edition available on the World Wide Web at http://bayes.wustl.edu.

Jeffreys, H., 1939. *Theory of Probability,* 1st edn, Clarendon Press, Oxford.

Jones, A.G. & Hutton, R., 1979. A multi-station magnetotelluric study in southern Scotland–II. Monte-Carlo inversion of the data and its geophysical and tectonic implications, *Geophys. J. R. astr. Soc.,* **56,** 351–368.

Keller, G.V. & Frischknecht, F.C., 1966. *Electrical Methods in Geophysical Prospecting,* Pergamon Press, Oxford.

Kendall, M. & Stuart, A., 1977. *The Advanced Theory of Statistics, Vol. 1,* Macmillan Publishing Company, New York.

MacKay, D.J.C., 1992. Bayesian interpolation, *Neural Comp.,* **4,** 415–447.

Malinverno, A., 2000. A Bayesian criterion for simplicity in inverse problem parametrization, *Geophys. J. Int.,* **140,** 267–285.

Malinverno, A. & Leaney, S., 2000. A Monte Carlo method to quantify uncertainty in the inversion of zero-offset VSP data, in *SEG 70th Annual Meeting, Calgary, Alberta,* The Society of Exploration Geophysicists, Tulsa, Oklahoma.

Malinverno, A. & Torres-Verdín, C., 2000. Bayesian inversion of DC electrical measurements with uncertainties for reservoir monitoring, *Inverse Problems,* **16,** 1343–1356.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.,* **21,** 1087–1092.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100,** 12 431–12 447.

Mosegaard, K., Singh, S., Snyder, D. & Wagner, H., 1997. Monte Carlo analysis of seismic reflections from Moho and the W reflector, *J. geophys. Res.,* **102,** 2969–2981.

Neal, R.N., 1993. *Probabilistic Inference Using Markov Chain Methods, Technical Report CRG-TR-93-1, Department of Computer Science,* University of Toronto.

Ó Ruanaidh, J.J.K. & Fitzgerald, W.J., 1996. *Numerical Bayesian Methods Applied to Signal Processing,* Springer-Verlag, New York.

Press, F., 1968. Earth models obtained by Monte Carlo inversion, *J. geophys. Res.,* **73,** 5223–5234.

Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T., 1992. *Numerical Recipes in C: The Art of Scientific Computing,* Cambridge University Press, New York.

Scales, J.A. & Tenorio, L., 2001. Prior information and uncertainty in inverse problems, *Geophysics,* **66,** 389–397.

Schott, J.J., Roussignol, M., Menvielle, M. & Nomenjahanary, F.R., 1999. Bayesian inversion with Markov chains–II. The one-dimensional DC multilayer case, *Geophys. J. Int.,* **138,** 769–783.

Sen, M.K. & Stoffa, P.L., 1995. *Global Optimization Methods in Geophysical Inversion, Advances in Exploration Geophysics, Vol. 4,* Elsevier, Amsterdam.

Sivia, D.S., 1996. *Data Analysis: A Bayesian Tutorial,* Clarendon Press, Oxford.

Smith, A.F.M., 1991. Bayesian computational methods, *Phil. Trans. R. Soc. Lond.* A., **337,** 369–386.

Sternberg, B.K., 1979. Electrical resistivity structure of the crust in the southern extension of the Canadian Shield—Layered earth models, *J. geophys. Res.,* **84,** 212–228.

Tarantola, A. & Valette, B., 1982. Inverse problems = quest for information, *J. Geophys.,* **50,** 159–170.

Tarits, V., Jouanne, V., Menvielle, M. & Roussignol, M., 1994. Bayesian statistics of non-linear inverse problems: example of the magnetotelluric 1-D inverse problem, *Geophys. J. Int.,* **119,** 353–368.

Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion), *Ann. Stat.,* **22,** 1701–1762.

Ulrych, T.J., Sacchi, M.D. & Woodbury, A., 2001. A Bayes tour of inversion: A tutorial, *Geophysics,* **66,** 55–69.

APPENDIX A: PARAMETERS TO BE SET *A PRIORI*

A1 Layering parameters

The parameters that need to be defined are a maximum number of layers $\overline{k}_{\max}$, and a minimum $\overline{z}_{\min}$ and maximum log-depth $\overline{z}_{\max}$. The maximum number of layers can be set to a value that is large compared to the number of layers needed to explain the data (in the examples shown here, $\overline{k}_{\max} = 30$). The minimum log-depth $\overline{z}_{\min}$ and the maximum $\overline{z}_{\max}$ can be set to a value near the minimum and the maximum electrode spacings in the measured data, respectively. This depth range covers the interval probed by the measurements and is the range typically used when inverting these kinds of measurements (e.g. Constable *et al.* 1987; Schott *et al.* 1999). The minimum log-layer thickness is computed as $\overline{h}_{\min} = (\overline{z}_{\max} - \overline{z}_{\min})/(2\overline{k}_{\max})$; this choice of $\overline{h}_{\min}$ ensures that even when k approaches $\overline{k}_{\max}$ at least half of the interval $\overline{z}_{\max} - \overline{z}_{\min}$ is available to place a new interface.

A2 Prior values of resistivity

The prior mean log-resistivity $\overline{\rho}$ is set to the logarithm of the most likely value of resistivity, and the prior standard deviation $\overline{\sigma}_\rho$ to the logarithm of one plus the factor within which the resistivity is expected to vary. The prior mean vector and the prior covariance matrix in the prior *pdf* of log-resistivity (eq. 7) are then set to $\overline{\boldsymbol{\rho}} = \overline{\rho}\mathbf{i}$ and $\overline{\mathbf{C}}_\rho = \overline{\sigma}_\rho^2\mathbf{I}$, respectively (where $\mathbf{i}$ is the unit vector of length k and $\mathbf{I}$ the $k \times k$ identity matrix).

A3 Measurement errors

The errors of the measured log-apparent resistivities are assumed to be independent and normally distributed with zero mean. Their prior covariance matrix $\overline{\mathbf{C}}_\mathbf{e}$ (used in the likelihood function of eq. 9) is diagonal, and contains the variances of the errors expected for the logarithm of each measured log-apparent resistivity in $\mathbf{d}$.

APPENDIX B: THE CHOICE OF A CANDIDATE MODEL $\mathbf{m}'$

This appendix described how a candidate model $\mathbf{m}'$ is selected on the basis of the current model $\mathbf{m}$ from a proposal *pdf* $q(\mathbf{m}' \mid \mathbf{m})$. It is helpful to break the selection of a candidate in two steps: first pick a candidate number of layers k' and a candidate log-depth of the layer interfaces $\mathbf{z}'$ based on k and $\mathbf{z}$ of the current earth model; then pick

a candidate log-resistivity $\boldsymbol{\rho}'$ based on k', $\mathbf{z}'$, and $\boldsymbol{\rho}$ in the current model. The proposal *pdf* can then be written as

$$q(\mathbf{m}' \mid \mathbf{m}) = q(k', \mathbf{z}' \mid \mathbf{m})\, q(\boldsymbol{\rho}' \mid k', \mathbf{z}', \mathbf{m}) \tag{B1}$$

and the two steps in selecting a candidate models will be illustrated separately.

B1 Picking candidates k', $\mathbf{z}'$

Similarly to the implementation of Green (1995) and Denison *et al.* (1998), the choice of candidates k' and $\mathbf{z}'$ on the basis of k and $\mathbf{z}$ in the current model is done by selecting at random one of four possible moves:

(i) **Birth:** The candidate number of layers is $k' = k + 1$, and the candidate log-depth of layer interfaces $\mathbf{z}'$ is the same as $\mathbf{z}$ with an interface added at a random log-depth between $\overline{z}_{\min}$ and $\overline{z}_{\max}$ (making sure that no layers thinner than $\overline{h}_{\min}$ are created);

(ii) **Death:** $k' = k - 1$, and $\mathbf{z}'$ is the same as $\mathbf{z}$ with an interface chosen at random and deleted;

(iii) **Perturbation:** $k' = k$, and $\mathbf{z}'$ is the same as $\mathbf{z}$, but for a randomly chosen interface whose log-depth is changed at random within an interval $\pm\overline{h}_{\min}$ about its current value (making sure that no layers thinner than $\overline{h}_{\min}$ are created);

(iv) **No Perturbation:** $k' = k$, and $\mathbf{z}' = \mathbf{z}$.

Each of these moves will have some probability of being selected; in the examples shown in this paper, the probabilities of birth, death, and perturbation are $1/6$, and the probability of no perturbation $1/2$.

B2 Picking a candidate $\boldsymbol{\rho}'$

The next step in choosing a candidate earth model is to pick values of the candidate log-resistivities $\boldsymbol{\rho}'$ given k', $\mathbf{z}'$, and the log-resistivities $\boldsymbol{\rho}$ in the current model. To decide what to do at this stage, it is important to realize that the choice of the proposal *pdf* will have a major effect on the efficiency of the algorithm. For example, suppose that the current parameter vector $\mathbf{m}$ is in a high-posterior probability region and that a 'no perturbation' move was chosen in the previous step. If the proposed $\boldsymbol{\rho}'$ typically differs from $\boldsymbol{\rho}$ by an amount that is large compared to the spread of the posterior *pdf*, $\mathbf{m}'$ will have a low likelihood, α in eq. (15) will be small, and it will take many iterations before a proposed move is accepted. At the other extreme, if the changes proposed are small, α will be large, but many iterations will be needed to cover the high-posterior probability region of the parameter space (Chib & Greenberg 1995; Gelman *et al.* 1996). The best choice for $q(\boldsymbol{\rho}' \mid k', \mathbf{z}', \mathbf{m})$ would be a *pdf* that approximates the posterior *pdf* of $\boldsymbol{\rho}$; in this case, the acceptance probability α in eq. (15) would generally be close to one.

The proposal *pdf* for $\boldsymbol{\rho}'$ used here is a multivariate normal *pdf* with a mean vector $\tilde{\boldsymbol{\rho}}'$ and a covariance matrix $\tilde{\mathbf{C}}_{\rho'}$:

$$q(\boldsymbol{\rho}' \mid k', \mathbf{z}', \mathbf{m}) = \frac{1}{\left[(2\pi)^{k'} \det \tilde{\mathbf{C}}_{\rho'}\right]^{1/2}} \cdot \exp\left[-\frac{1}{2}(\boldsymbol{\rho}' - \tilde{\boldsymbol{\rho}}')^{\mathsf{T}} \tilde{\mathbf{C}}_{\rho'}^{-1} (\boldsymbol{\rho}' - \tilde{\boldsymbol{\rho}}')\right]. \tag{B2}$$

The mean vector $\tilde{\boldsymbol{\rho}}'$ is obtained from the log-resistivities $\boldsymbol{\rho}$ in the current model. If the first step was a move where $k' = k$, $\tilde{\boldsymbol{\rho}}'$ is simply set to equal $\boldsymbol{\rho}$. If the first step was a birth (or death), the log-resistivities in $\tilde{\boldsymbol{\rho}}'$ are the same as those in $\boldsymbol{\rho}$ for all layers that were not split by the birth (or combined by the death). In a birth, the log-resistivities in $\tilde{\boldsymbol{\rho}}'$ for the two layers that were generated by splitting a single layer in $\boldsymbol{\rho}$ are the same as that of the original layer. In a death, the log-resistivity in $\tilde{\boldsymbol{\rho}}'$ for the layer that resulted from the combination of two layers in $\boldsymbol{\rho}$ is the average of the log-resistivities in the two original layers.

The reasoning above suggests that the proposal *pdf* should approximate the posterior *pdf* of $\boldsymbol{\rho}$. An approximate posterior covariance matrix for $\boldsymbol{\rho}$ can be obtained from a linearization about the point at $\tilde{\boldsymbol{\rho}}'$ (Bard 1974) as

$$\widehat{\mathbf{C}}_{\rho} \approx \left[\mathbf{J}^T \overline{\mathbf{C}}_{\mathbf{e}} \mathbf{J} + \overline{\mathbf{C}}_{\rho}\right]^{-1}, \tag{B3}$$

where $\mathbf{J}$ is the Jacobian matrix (the matrix of derivatives of the predicted data with respect to each element of $\boldsymbol{\rho}$) evaluated at $\tilde{\boldsymbol{\rho}}'$. Constable *et al.* (1987) show in their Appendix how to compute the Jacobian matrix in the DC resistivity problem (the only difference here is that the elements of $\mathbf{J}$ are derivatives of predicted log-apparent resistivities with respect to model log-resistivities). Using the recipe suggested by Gelman *et al.* (1996) for optimally sampling spherical distributions, I set $\tilde{\mathbf{C}}'_{\rho'}$ to be a diagonal matrix with entries equal to the diagonal elements of the approximate $\widehat{\mathbf{C}}_{\rho}$ in eq. (B 3) multiplied by a scaling factor chosen so that the acceptance rate of the Markov chain is around 25 per cent. Gelman *et al.* (1996) show that this acceptance rate is optimal in the sense that it maximizes the diffusion rate of the random walk taken by the Markov chain. A scaling factor that worked well was $1.5/\sqrt{k'}$ (in terms of standard deviations), which is slightly less than the value of $2.4/\sqrt{k'}$ suggested by Gelman *et al.* (1996).

Following the principle that the proposal *pdf* should approximate the posterior, it may seem that a better strategy would be to simply set $\tilde{\mathbf{C}}'_{\rho}$ to the approximate value of $\widehat{\mathbf{C}}_{\rho}$ estimated by eq. (B3). This strategy gives good results when the inverse problem is linear for the material properties in the layers (Malinverno & Leaney 2000) or where the Jacobian matrix is evaluated at the mode of the posterior *pdf* (Malinverno & Torres-Verdín 2000). In the DC resistivity sounding problem, however, this strategy does not work well because the approximation of eq. (B3) is often not very good in the off-diagonal elements and the choice of a candidate model is correspondingly poor.

APPENDIX C: THE ACCEPTANCE PROBABILITY α

Once the proposal *pdf* is defined (Appendix B), the acceptance probability in eq. (15) can be written out explicitly. To start, note that there is an obvious symmetry between the terms in the prior and the proposal ratio of eq. (15):

$$\text{Prior ratio} = \frac{p(\mathbf{m}' \mid \mathcal{I})}{p(\mathbf{m} \mid \mathcal{I})} = \frac{p(k', \mathbf{z}' \mid \mathcal{I})}{p(k, \mathbf{z} \mid \mathcal{I})} \frac{p(\boldsymbol{\rho}' \mid k', \mathbf{z}', \mathcal{I})}{p(\boldsymbol{\rho} \mid k, \mathbf{z}, \mathcal{I})} \tag{C1}$$

$$\text{Proposal ratio} = \frac{q(\mathbf{m} \mid \mathbf{m}')}{q(\mathbf{m}' \mid \mathbf{m})} = \frac{q(k, \mathbf{z} \mid \mathbf{m}')}{q(k', \mathbf{z}' \mid \mathbf{m})} \frac{q(\boldsymbol{\rho} \mid k, \mathbf{z}, \mathbf{m}')}{q(\boldsymbol{\rho}' \mid k', \mathbf{z}', \mathbf{m})} \tag{C2}$$

As noted by Mosegaard & Tarantola (1995), if the candidate models are sampled from the prior *pdf*, the proposal *pdf* equals the prior *pdf* and the corresponding terms in the acceptance probability cancel out. This is in fact the case here for the number of layers k and the log-depths of the layer interfaces $\mathbf{z}$. If the move proposed is a perturbation of $\mathbf{z}$ or no perturbation, so that $k' = k$, then clearly

$$\frac{p(k', \mathbf{z}' \mid \mathcal{I})}{p(k, \mathbf{z} \mid \mathcal{I})} = \frac{q(k, \mathbf{z} \mid \mathbf{m}')}{q(k', \mathbf{z}' \mid \mathbf{m})} = 1. \tag{C3}$$

If instead a birth move is proposed, the proposal ratio is

$$\frac{q(k, \mathbf{z} \mid \mathbf{m}')}{q(k', \mathbf{z}' \mid \mathbf{m})} = \frac{\Delta z(k)}{k}, \tag{C4}$$

as the probability of a birth move equals the probability of a death move, the probability of choosing any layer interface for death equals $1/k$, and the probability of placing a new layer at a given travel-time is $1/\Delta z(k)$ (where $\Delta z(k)$ is the log-depth interval available to place a layer interface when there are already k interfaces in the model; see eq. 6). This is the inverse of the ratio of the prior *pdf*'s from eq. (6), so that the proposal ratio and prior ratio terms for k and $\mathbf{z}$ cancel each other. The same reasoning holds for a death move.

The acceptance probability for the proposal *pdf* used here is then

$$\begin{aligned}\alpha &= \min\left[1, \frac{p(\boldsymbol{\rho}' \mid k', \mathbf{z}', \mathcal{I})}{p(\boldsymbol{\rho} \mid k, \mathbf{z}, \mathcal{I})} \cdot \frac{p(\mathbf{d} \mid \mathbf{m}', \mathcal{I})}{p(\mathbf{d} \mid \mathbf{m}, \mathcal{I})} \cdot \frac{q(\boldsymbol{\rho} \mid k, \mathbf{z}, \mathbf{m}')}{q(\boldsymbol{\rho}' \mid k', \mathbf{z}', \mathbf{m})}\right] \\ &= \min\left[1, (\text{Prior ratio}) \cdot (\text{Likelihood ratio}) \cdot (\text{Proposal ratio})\right],\end{aligned} \tag{C5}$$

where the prior ratio can be obtained from eq. (7), the likelihood ratio from eq. (9), and the proposal ratio from eq. (B2).