Geophys. J. Int. (1999) **138,** 757–768

Bayesian inversion with Markov chains—I. The magnetotelluric one-dimensional case

H. Grandis,[1,*] M. Menvielle[1,†] and M. Roussignol[2]

[1] *Laboratoire de Géophysique de l'Environnement, UMR UPS et CNRS, Université Paris Sud-Bât.* 504, F-91405 *Orsay Cedex, France*
[2] *Equipe d'Analyse et de Mathématiques Appliquées, Université de Marne la Vallée,* 5 *Boulevard Descartes,* F-77454 *Marne La Vallee Cedex* 2, *France.*
E-mail: michel.menvielle@cetp.ipsl.fr

Accepted 1999 April 26. Received 1999 April 26; in original form 1998 March 3

SUMMARY

We use Monte Carlo Markov chains to solve the Bayesian MT inverse problem in layered situations. The domain under study is divided into homogeneous layers, and the model parameters are the conductivity of each layer. We use an *a priori* distribution of the parameters which favours smooth models. For each layer, the *a priori* and *a posteriori* distributions are digitized over a limited set of conductivity values.

The Markov chain relies on updating the model parameters during successive scanning of the domain under study. For each step of the scanning, the conductivity is updated in one layer given the actual value of the conductivity in the other layers. Thus we designed an ergodic Markov chain, the invariant distribution of which is the *a posteriori* distribution of the parameters, provided the forward problem is completely solved at each step.

We have estimated the *a posteriori* marginal probability distributions from the simulated successive values of the Markov chain. In addition, we give examples of complex magnetotelluric impedance inversion in tabular situations, for both synthetic models and field situations, and discuss the influence of the smoothing parameter.

Key words: Bayesian inversion, magnetotellurics, Markov chains, stochastic algorithms.

1 INTRODUCTION

It is well known that interpreting actual field data from electromagnetic soundings does not yield a unique solution even though theoretical considerations demonstrate that ideal observations can do so (e.g. Langer 1933; Bailey 1970; Weidelt 1972; Parker 1980, 1982; Backus 1988). This non-uniqueness of the solution obviously results from both the finite number of frequencies at which data are available and the existence of random noise on these data.

Basically, the aim of solving the inverse problem is finding a model whose theoretical response is as close as possible to the observations. The difference between the observed data and those computed with a given model is estimated by a function of the model parameters, and the selected model is the one which minimizes this function. In a standard approach to inversion, the difference between two models is estimated by means of a distance in the model space. In a probabilistic approach, it defines the probability of the model, and the smaller the distance the more likely the model.

In order to overcome the problem of the non-uniqueness of the solution, geophysicists put constraints on the models they seek, both to stabilize the solution and to get what is often called the best model accounting for the data. In the general case, this comes down to restricting the solution to a particular class of models. The solution obtained in this manner actually depends upon both the chosen distance function and the class of models. For a given data set, the more relevant the chosen selected distance function and class of models, the more representative the solution.

The dependence of the solution upon the distance function is obvious, and is clearly illustrated by the difference between the results of the least-square and L1-norm minimization (see e.g. Parker & McNutt 1980). It is also clear that selecting a class of model implies a hypothesis on the actual resistivity distribution in the Earth. For instance, selecting models made of homogeneous blocks implicitly assumes the existence of sharp discontinuities in the resistivity distribution. This may be quite realistic in some situations (e.g. the limit between the sedimentary cover and the crystalline basement), but is obviously not a general feature of the resistivity distribution inside the Earth. Selecting such models could be misleading in

* Permanent address: Jurusan Geofisika, Institut Teknologi Bandung, Jalan Ganesha 10, Bandung-40132, Indonesia.
† Now at: Centre d'Etudes des Environnements Terrestre et Planétaire, 4, Avenue de Neptune, F-94107 Saint Maur des Fossés Cedex, France.

some situations. For instance, in 1-D situations such models tempt one to believe that there are really large discontinuities between the layers at depths discovered by the computer program, whilst this may just be the result of the choice of class of model. Interpreting electromagnetic data in terms of resistivity distribution in the Earth can, therefore, be considered as the statistical problem of characterizing random variables (the parameters of the model) given observations and their statistical distributions on the one hand, and further constraints on the models' parameters on the other.

Bayesian statistics provide a theoretical framework well suited to set up such a problem. Solving the inverse problem then comes down to expressing the knowledge of the parameters of the model after having processed the data by updating the information available before the data processing with the information provided by the data (Box & Tiao 1973; Berger 1985; Press *et al.* 1989). The *a priori* and *a posteriori* knowledge is expressed in a statistical way as *a priori* and *a posteriori* probability distribution functions of the parameters. Solving the inverse problem is then determining the *a posteriori* probability distribution function of the parameters knowing their *a priori* probability distribution and the probability distribution of the data. The selection of a class of models is expressed in terms of *a priori* probability distribution of the parameters. The choice of the distance function is then straightforward because it directly corresponds to the statistical distribution of errors on the data.

In the general case, the direct computation of the *a posteriori* distribution is, in practice, impossible, because it involves sums over the very large number of possible models. Furthermore, its analytical expression is actually possible only if (i) the distributions on the data and the parameters are Gaussian, and (ii) the forward problem is linear. In the case of electromagnetism, (i) the forward problem is non-linear and (ii) the studied parameter, that is, the resistivity of geological materials, may vary by several orders of magnitude, from a few thousand Ω m in the case of dry, cold rocks to a few tenths of an Ω m in the case of porous rocks saturated with H_2O-rich fluid, or of partial melting. The solutions of the inverse problem involving analytical expressions of the forward problem have, therefore, a very limited interest because (i) linearization may induce instabilities, and (ii) the Gaussian hypothesis implies knowledge of the more likely values of the parameters (see e.g. Mackie *et al.* 1988). Solving the electromagnetic inverse problem, therefore, actually requires numerical estimation of the *a posteriori* distributions of the parameters.

Monte Carlo methods are in many cases fairly efficient for multiple integral computation, and one can, therefore, use such methods to achieve numerical estimates of the *a posteriori* distribution. Using a method already proposed by Wiggins (1969, 1972) to optimize multiple integral computation by means of Monte Carlo methods, Jouanne (1991) and Tarits *et al.* (1994) proposed a method allowing, in practice, the numerical computation of the *a posteriori* marginal distributions of the parameters in the case of non-linear inverse problems, provided the number of parameters remains small.

Tarits *et al.* (1994) considered models with a few homogeneous layers, the parameters of the model being the depth of the interfaces between two consecutive layers, and the layer resistivity. For such models, the number of layers is part of the *a priori* knowledge. Tarits *et al.* investigated the influence on the result of errors in the *a priori* determination of the number of layers. They showed that the analysis of the *a posteriori* marginal distributions easily allows determination of the actual number of layers in case of *a priori* overestimation of this number. On the other hand, *a priori* underestimation of the number of layers cannot be shown through the analysis of the *a posteriori* marginal distributions, and the results thus obtained may be significantly biased and, accordingly, misleading. A solution to overcome this problem is to digitize the variation of the resistivity with depth as a stack of L thin homogeneous layers with fixed thicknesses.

Because of their cost in terms of CPU time, the Monte Carlo methods actually become almost impossible to implement for models with a large number of parameters. We use, therefore, a simulation method to estimate *a posteriori* marginal distributions involving integrals impossible to compute directly. We present in this paper a method based on an ergodic Markov chain, the invariant probability of which is equal to the *a posteriori* distribution of the parameters (Robert 1996; Fishman 1996). It consists of simulating the Markov chain and estimating the *a posteriori* marginal probability distributions from the successive values of the Markov chain. Such methods, frequently referred to as MCMC methods, have already been developed by some authors, both in electromagnetics (Jouanne 1991; Roussignol *et al.* 1993; Grandis 1994) and in seismics (see e.g. Lavielle 1991).

The theoretical background is briefly recalled in Section 2, and the algorithm we developed for 1-D situations is described in Section 3. Examples of the inversion of both synthetic and field data are presented in Section 4 and are used to illustrate the main features of the proposed algorithm.

2 BAYESIAN ELECTROMAGNETIC INVERSION USING MARKOV CHAINS

Let the *a priori* knowledge be the information available in the model before processing the data, and the *a posteriori* knowledge the information available after processing the data. *A priori* and *a posteriori* distributions account in a probabilistic way for the *a priori* and *a posteriori* knowledge respectively. In the Bayesian context, solving the inverse problem thus comes down to determining the *a posteriori* knowledge by updating the *a priori* knowledge with the information gleaned from the data (Box & Tiao 1973; Berger 1985; Press *et al.* 1989).

Solving the inverse problem first requires the direct problem to be solved. Let $\mathbf{F}$ be the direct problem function which enables computation of the observations $\mathbf{y}$ for a model $\mathbf{x}$. Assuming that the error $\boldsymbol{\varepsilon}$ is only related to the data acquisition, $\mathbf{y}$ becomes

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \varepsilon . \tag{1}$$

We mention here that the *a posteriori* probability of the parameter vector $\mathbf{X}$ taking the value $\mathbf{x}$ given the observations $\mathbf{y}$ is given by Bayes' formula (Bayes 1763). Noting $P(\mathbf{X}=\mathbf{x}/\mathbf{Y}=\mathbf{y})$, the conditional probability of $\mathbf{X}$ given $\mathbf{Y}$, it becomes

$$P(\mathbf{X}=\mathbf{x}/\mathbf{Y}=\mathbf{y}) = \frac{P(\mathbf{Y}=\mathbf{y}/\mathbf{X}=\mathbf{x})P(\mathbf{X}=\mathbf{x})}{\sum_{\mathbf{x}} P(\mathbf{Y}=\mathbf{y}/\mathbf{X}=\mathbf{x})P(\mathbf{X}=\mathbf{x})}, \tag{2}$$

where $P(\mathbf{X}=\mathbf{x})$ is the *a priori* distribution of the parameters $\mathbf{x}$. Let us denote hereafter $P(\mathbf{X}=\mathbf{x})$ by $g(\mathbf{x})$ and $P(\mathbf{Y}=\mathbf{y}/\mathbf{X}=\mathbf{x})$ by $f(\mathbf{y}/\mathbf{x})$. The value $\mathbf{y}$ of the random vector $\mathbf{Y}$ corresponds to the observed data. Since $\mathbf{y}$ is not modified during the inversion process, we will hereafter omit $\mathbf{Y}$ in the expression of the probability distributions.

The data are a set of electromagnetic responses observed at a finite number of frequencies. We are only concerned with the case of complex impedances, but it is clear that any other electromagnetic response could be used as well. In most of the cases, the available data do not constrain the model parameters enough, and solving the inverse problem requires further *a priori* knowledge of the model, such as, for example, geometrical features of the studied medium, domains of variation of the parameters or smoothness in the variations of the parameters. In some cases, the algorithm used implies making choices which govern the result to a certain unknown extent. For instance, the result obtained with a deterministic steepest-gradient method may depend on the chosen starting model, and it may be very difficult to determine to what extent the result has been affected. In the Bayesian approach, the hypotheses made are expressed in a statistical way by the *a priori* distribution, and the comparison between *a posteriori* and *a priori* distributions gives a quantitative estimate of the information provided by the data.

Consider a system which can take a certain number of states and evolves at random with time. At a given time, the state of the system can be described as a random variable $\mathbf{X}$, the values of which, $\mathbf{x}$, belong to the set of possible states, $\boldsymbol{E}$. The system is a Markovian process if, at any time, its future evolution depends only on its present state. This means that a Markovian system depends only on its past through its present state. Markov chains are the particular class of Markovian processes such that (i) $\boldsymbol{E}$ is finite or countable, and (ii) the successive times for the evolution are denoted by integers. A series $[\mathbf{S}^{(n)}; n = 0, 1, 2, \ldots]$ of random variables with values in a finite or countable space is then a Markov chain if the state of the system at the time $n+1$, $\mathbf{S}^{(n+1)}$, depends only on its past through the state of the system at time n, $\mathbf{S}^{(n)}$. The behaviour of a Markov chain is defined by the set of its transition probabilities,

$$\mathrm{P}_n(\mathbf{x}, \mathbf{x}') = P(\mathbf{S}^{(n+1)} = \mathbf{x}'/\mathbf{S}^{(n)} = \mathbf{x}), \tag{3}$$

where $\mathbf{x}$ and $\mathbf{x}'$ belong to $\boldsymbol{E}$.

If (i) the transition probabilities do not depend on time n, (ii) $\boldsymbol{E}$ is finite, and (iii) each possible state can be reached from any other, the chain is a homogeneous ergodic aperiodic Markov chain, and there exists one and only one probability density function Π on $\boldsymbol{E}$ which is invariant for the Markov chain. When n increases towards infinity, the behaviour of ergodic Markov chains is such that the average fraction of time at which the chain is at a state $\mathbf{x}$ ($\mathbf{x} \in \boldsymbol{E}$) tends towards the invariant probability of $\mathbf{x}$, $\Pi(\mathbf{x})$.

Consider now situations for which the Bayesian inverse problem comes down to determining the L parameters of a model. Assume further that there are only M_l possible values, $\theta_{l,i}; i = 1, 2, \ldots, M_l$, associated with each parameter x_l of the model $\mathbf{x}$. The choice of the possible values $\theta_{l,i}$ expresses an *a priori* knowledge of the parameters. According to Bayes' rule (eq. 2), the *a posteriori* marginal probability of the element X_l taking the value $\theta_{l,i}$ is then

$$P(X_l = \theta_{l,i}/\mathbf{y}) = \frac{\sum\limits_{\mathbf{x} \in \boldsymbol{E}_{l,i}} f(\mathbf{y}/\mathbf{x}) g(\mathbf{x})}{\sum\limits_{\mathbf{x} \in \boldsymbol{E}} f(\mathbf{y}/\mathbf{x}) g(\mathbf{x})}, \tag{4}$$

where $\boldsymbol{E}$ is the set of possible images and $\boldsymbol{E}_{l,i}$ the set of images $\mathbf{x}$ for which $x_l = \theta_{l,i}$. X_l is a random scalar which is the component number l of the random vector $\mathbf{X}$.

The normalization constant, which appears in the denominator of eq. (4), is actually very difficult to estimate, because it is a sum over the set of possible images, the dimension of which is the product $M_1 M_2 M_3 \ldots M_L$. On the other hand, it is easy to compute the conditional probability of $X_l = \theta_{l,i}$ given the values of all the other parameters and the observations:

$$P(X_l = \theta_{l,i}/X_k = x_k, k \neq l; \mathbf{y}) = \frac{f(\mathbf{y}/\mathbf{x}^{(l,i)}) g(\mathbf{x}^{(l,i)})}{\sum\limits_{j=1}^{M_l} f(\mathbf{y}/\mathbf{x}^{(l,j)}) g(\mathbf{x}^{(l,j)})}, \tag{5}$$

where $\mathbf{x}$ and $\mathbf{x}^{\{l,j\}} \in \boldsymbol{E}$ and

$$x_k^{(l,j)} = x_k \forall\, k \neq l,$$

$$x_l^{(l,j)} = \theta_{l,j}.$$

The normalization constant in eq. (5) is a sum over the set of M_l possible values $\theta_{l,i}$ of the lth component x_l of the model $\mathbf{x}$. Its computation then becomes possible provided M_l remains reasonable.

Consider now a random algorithm which updates the value of the parameter for the component x_l by selecting at random its new value from the set of possible values $\theta_{l,i}$ according to the probability distribution

$$P(X_l = \theta_{l,i}/X_k = x_k, k \neq l; \mathbf{y})$$

given in eq. (5).

Considering each element successively and using this algorithm to update its parameter value defines an iterative random process in the set $\boldsymbol{E}$ of possible images. This process is a Markov chain on $\boldsymbol{E}$ called the Gibbs sampler (Robert 1996).

Consider first the case for which the component to be updated, x_l, is selected at random over the L components of $\mathbf{x}$ with a uniform probability distribution. The sequence of models thus obtained is a homogeneous Markov chain, because the probability of an image $\mathbf{S}^{(n+1)}$ depends only on the previous image $\mathbf{S}^{(n)}$, and does not change with step n. Its transition probability $\mathrm{P}_n(\mathbf{x}, \mathbf{z})$ is

$$\mathrm{P}_n(\mathbf{x}, \mathbf{z}) = \frac{1}{L} P(X_l = z_l/X_k = x_k, k \neq l; \mathbf{y}) \qquad \text{if } x_k = z_k, k \neq l, \tag{6}$$

$$\mathrm{P}_n(\mathbf{x}, \mathbf{z}) = 0 \text{ otherwise}.$$

This Markov chain is ergodic, and its invariant probability Π is the *a posteriori* probability given by eq. (2) (Robert 1996).

Let us illustrate these definitions, and consider the case of a two-layer model ($L = 2$) consisting of a homogeneous layer of resistivity x_1 above a homogeneous half-space of resistivity x_2. Both x_1 and x_2 have the same set of M possible values, $(\theta_1, \ldots, \theta_M)$. The parameter vector $\mathbf{x}$ is then the pair (x_1, x_2) and the space $\boldsymbol{E}$ of possible images has M^2 elements.

According to formula 4, the *a posteriori* probability density function of the Markov chain is

$$\Pi((x_1, x_2)) = \frac{f(\mathbf{y}/(x_1, x_2)) g((x_1, x_2))}{\sum\limits_{i=1}^{M} \sum\limits_{j=1}^{M} f(\mathbf{y}/(\theta_i, \theta_j)) g((\theta_i, \theta_j))}.$$

Let us define a Markov chain on $\boldsymbol{E}$ by its transition probability between any two elements of $\boldsymbol{E}$. From formula 6,

$$\mathrm{P}((x_1, x_2), (z_1, x_2)) = \frac{1}{2} \frac{f(\mathbf{y}/(z_1, x_2))g((z_1, x_2))}{\sum_{j=1}^{M} f(\mathbf{y}/(\theta_i, x_2))g((\theta_i, x_2))},$$

$$\mathrm{P}((x_1, x_2), (x_1, z_2)) = \frac{1}{2} \frac{f(\mathbf{y}/(x_1, z_2))g((x_1, z_2))}{\sum_{i=1}^{M} f(\mathbf{y}/(x_1, \theta_i))g((x_1, \theta_i))}$$

and

$$\mathrm{P}((x_1, x_2), (z_1, x_2)) = 0 \qquad \text{if } x_1 \neq z_1 \text{ and } x_2 \neq z_2.$$

The first of the above formulae describes transitions which update the value of the resistivity of layer 1, and the second describes those which update the resistivity of layer 2. The third expresses the fact that there is no transition which updates simultaneously the resistivity of the two layers.

It can be shown that, for any pair (x_1, x_2),

$$\Pi((x_1, x_2)) = \sum_{i=1}^{M} \sum_{j=1}^{M} \Pi((\theta_i, \theta_j))\mathrm{P}((\theta_i, \theta_j), (x_1, x_2)),$$

which implies that the probability density function Π is the invariant probability of this Markov chain.

If the element to be updated is selected according to a deterministic procedure, the sequence of images obtained is no longer a homogeneous Markov chain. However, the sequence of images $\mathbf{S}^{(\nu)}$ obtained after each cycle ν of scanning of all the elements remains an ergodic Markov chain. Its invariant probability is again the *a posteriori* probability given by eq. (4).

Because the Markov chain is irreducible and aperiodic in a finite space, the invariant probability distribution Π on $\boldsymbol{E}$ is the unique invariant probability distribution (Feller 1970). The ergodic theorem holds and

$$\frac{1}{N} \sum_{n=1}^{N} \Phi(\mathbf{S}^{(n)}) \xrightarrow[N \to \infty]{} \sum_{\mathbf{x} \in E} \Pi(\mathbf{x})\Phi(\mathbf{x}),$$

where Φ is any numerical function on $\boldsymbol{E}$. This implies that the empirical average of the transition probability distributions of $\mathbf{S}^{(n)}$ tends towards the invariant distribution of the chain, Π, whatever the initial state $\mathbf{S}^{(0)}$ of the system. This average then easily provides an estimate of the marginal distributions of Π. In fact, we estimate the marginal distributions of the invariant distribution as the empirical average of the transition probability distributions (see e.g. Robert 1996; Fishman 1996). Denoting by $\mathrm{P}(\bullet, \bullet)$ the transition probability of the chain, and by $\mathbf{S}^{(n)}$ and $[\mathbf{S}^{(n)}]^{(l,j)}$ the images defined by

$$[S^{(n)}]_k^{(l,j)} = S_k^{(n)} \qquad \text{for } k \neq l$$

and

$$[S^{(n)}]_l^{(l,j)} = \theta_{l,j},$$

the lth marginal *a posteriori* distribution $\Pi_l(\theta_{l,j})$ is estimated after N steps of the Markov chain by

$$\Pi_l^{(N)}(\theta_{l,j}) = \frac{1}{N} \sum_{n=1}^{N} \mathrm{P}[\mathbf{S}^{(n)}, [\mathbf{S}^{(n)}]^{(l,j)}]. \tag{7}$$

Thus, estimators of the parameter x_l are either the mean value of $\Pi^{(N)}$ or the value for which the probability $\Pi^{(N)}$ is maximum.

In the case of the two-layer model that we introduced above as example, formula (7) for the first marginal *a posteriori* distribution leads to

$$\Pi_1^{(N)}(\theta_j) = \frac{1}{N} \sum_{n=1}^{N} \mathrm{P}[(S_1^{(n)}, S_2^{(n)}), (\theta_j, S_2^{(n)})]$$

$$= \frac{1}{N} \sum_{n=1}^{N} \frac{1}{2} \frac{f[\mathbf{y}/(\theta_j, S_2^{(n)})]g[(\theta_j, S_2^{(n)})]}{\sum_{i=1}^{M} f[\mathbf{y}/(\theta_i, S_2^{(n)})]g[(\theta_i, S_2^{(n)})]}.$$

3 THE ALGORITHM

The MT 1-D inverse problem is a classical problem, and it has already been addressed by many authors. In fact, various algorithms based on deterministic or probabilistic approaches are at present available (see e.g. Tarlowski 1982; Parker 1983); Oldenburg 1990; Raiche 1994). To our knowledge, however, none of them involves the use of Markov chains.

MT data sets consist of complex coefficients (also called MT impedances) determined from the frequency analysis of simultaneous records of natural fluctuations of electric and magnetic fields. The frequency range of the fluctuations is quite broad (thousands of Hertz to less than one cycle per year). The frequency range, measured during MT sounding experiments, is fixed by the objectives. The bandwidth is approximately 1000–0.001 Hz for crustal investigations. It has to be extended towards the low-frequency range for deeper investigations, and shifted towards the high-frequency range for investigations focusing on the shallow cover. The MT data set is generally limited to 5–10 data points per decade for a single MT sounding. Each coefficient is an independent statistical estimate. In most cases, the probability distribution function of the MT coefficients is Gaussian, or may be approximated by a Gaussian distribution after robust processing (see e.g. Chave *et al.* 1987), and is characterized by an expected value and a variance.

Efficient algorithms which solve the forward MT 1-D problem at very low CPU time consumption do exist, even for a large number of homogeneous layers. The MT-1D problem is then well suited for a first application of the algorithm for Bayesian inversion with Markov chains which we described in Section 2.

3.1 The model

The model consists of a stack of L thin homogeneous layers of fixed thicknesses (Fig. 1). The parameters are then the resistivities of the layers,

$$\mathbf{x} = (\rho_1, \rho_2, \ldots, \rho_L). \tag{8}$$

The distribution of the depths of the interfaces between two consecutive layers, z_l, is a logarithmic one. This ensures a good balance between the data, in particular when the frequencies f_i at which the data are estimated are regularly distributed (Smith & Booker 1988).

The forward problem is solved at the surface of the tabular medium using recursive procedures. The impedance at a given interface z_{l-1} is deduced from that at the next deeper one, z_l, using a relation involving only field characteristics (frequency and source space wavelength) and the parameters of the lth layer (thickness and resistivity). The 1-D model of electric resistivity versus depth which we use is shown in Fig. 1. Starting at the top of the underlying homogeneous half-space and moving upwards step by step therefore allows easy computation of the impedance at the surface of the stratified medium.

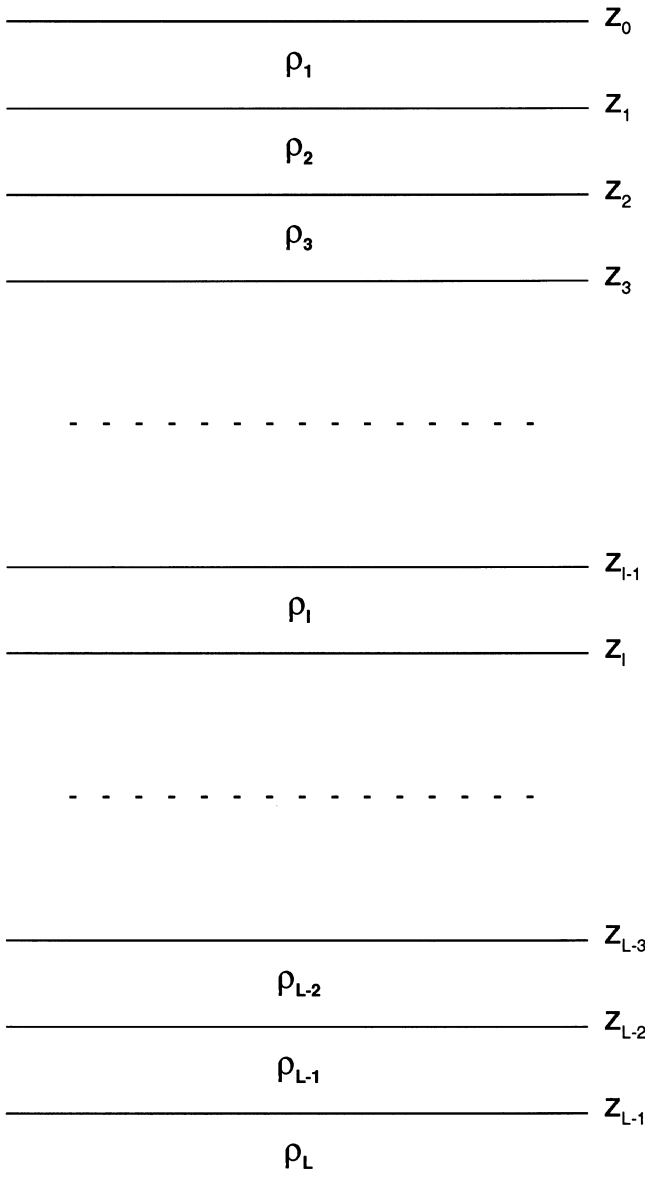

Figure 1. The 1-D model of electric resistivity versus depth. The depths of the interfaces (z_l) are *a priori* defined and remain fixed. The parameters of the model are the resistivities ρ_l of the layers.

Two recursive procedures are used in this paper. The synthetic models used to test our algorithm have a limited number of layers, and the corresponding impedances are computed using the recursive relation generally known as the Lipskaya formula:

$$Z_{l-1}(z_{l-1}) = -\frac{i\omega\mu}{k_l}\coth\left[-h_l k_l - \coth^{-1}\left(k_l\frac{Z_l(z_l)}{i\omega\mu}\right)\right]$$

for $l = 1, \ldots L-1$,

$$Z_{L-1}(z_{L-1}) = \frac{i\omega\mu}{k_L},$$

where $k_l^2 = i\omega\mu/\rho_l$, and $h_l = z_l - z_{l-1}$ and ρ_l are respectively the thickness and resistivity of layer l (see Fig. 1). $Z_l(z_l)$ is the impedance computed at the bottom of layer l.

The inversion scheme we propose involves numerous forward computations for models with a large number of layers. These forward computations are performed using a matrix formalism well suited for such situations (Madden & Nelson 1964; Ward & Hohmann 1988). The electromagnetic field at the bottom of layer l is linked to that at the bottom of layer $l-1$ through a transfer matrix which depends only on the thickness h_l and resistivity ρ_l of layer l

$$\begin{bmatrix} E_{x,l-1} \\ H_{y,l-1} \end{bmatrix} = \begin{bmatrix} \cosh(k_l h_l) & \zeta_l \sinh(k_l h_l) \\ \zeta_l^{-1}\sinh(k_l h_l) & \cosh(k_l h_l) \end{bmatrix} \begin{bmatrix} E_{x,l} \\ H_{y,l} \end{bmatrix},$$

where $\zeta_l^2 = i\omega\mu\rho_l$ and $k_l^2 = i\omega\mu/\rho_l$.

3.1.1 *The pdf of the observational errors*

As already mentioned, the data are classical complex scalar MT impedances estimated at a set of n_F frequencies f_i, $i = 1, \ldots, n_F$, which can be considered as a realization of a set of $2n_F$ independent random real scalar variables with centred Gaussian noise of variance σ_i^2. The conditional probability of the observations $\mathbf{y}$ given a value $\mathbf{x}$ of the parameter vector $\mathbf{X}$ is then

$$f(\mathbf{y}/\mathbf{x}) = C\exp\{-X^2(\mathbf{x},\mathbf{y})\} = C\exp\left\{-\sum_{i=1}^{n_F}\frac{|y_i - [\mathbf{F}(\mathbf{x})]_i|^2}{2\sigma_i^2}\right\}, \tag{9}$$

where $[\mathbf{F}(\mathbf{x})]_i$ is the ith component of the vector $\mathbf{F}(\mathbf{x})$ computed for the model $\mathbf{x}$ with the direct problem function F relating the model parameters and the data (eq. 1).

3.1.2 *The* a priori *distribution of the parameters*

Except for very particular situations which are very unlikely to be encountered in geophysics, the *a priori* distribution is not a Gaussian one. On the other hand, no evidence exists for a general analytical expression of the *a priori* distribution. We therefore digitized it over a set of resistivity values, hereafter called possible resistivity values, the choice of which depends on the *a priori* knowledge of the medium considered. There is no constraint on this choice, but it is clear that the larger this number, the better the determination of the *a posteriori* distribution, but the greater the computer time. The *a priori* distribution, as well as the possible resistivity values, may depend on the layer. The *a posteriori* distribution will accordingly be expressed as the *a posteriori* probability of these possible resistivity values.

In the following, we will only consider cases for which the *a priori* knowledge, and then the *a priori* distribution, are the same for all the layers. This comes down to selecting a set of possible resistivity values in the medium under study, and to digitizing the *a priori* distribution on these values. Thus we select M possible values ρ_m of the resistivity, such that

$$\rho_{\min} = \rho_1 < \rho_2 < \ldots < \rho_M = \rho_{\max}.$$

The *a priori* probability of an L-layer model $\mathbf{x}$ is then

$$g(\mathbf{x}) = P(X_1 = x_1, X_2 = x_2, \ldots, X_L = x_L),$$

where x_k $(k = 1, 2, \ldots, L)$ are the parameters of the model which belong to the set of possible resistivity values ρ_m $(m = 1, 2, \ldots, M)$.

In situations for which no particular information is available, we choose as the *a priori* distribution the product of a uniform pdf on each layer. In this case $g(\mathbf{x})$ is a uniform distribution over all the possible models and the different parameters (that is, the resistivity values in different layers) are independent. Because of the non-uniqueness of the MT 1D inverse problem with a finite number of noisy observations, this *a priori* distribution leads to marginal *a posteriori* distributions that are almost uniform over a large range of resistivity values, as illustrated in Fig. 2 for the five-layer synthetic model and set of apparent resistivities with 5 per cent added Gaussian noise, as described in Table 1(a).

Overcoming this problem requires a better specification of $g(\mathbf{x})$. In most cases, the available *a priori* information does not allow the distribution on each layer to be specified; one possible solution is then to introduce a dependence between the parameters which will select a particular class of models.

Theoretical considerations suggest using 'regular' models. Elementary differential equation theory restricts attention to the space of relatively smooth functions, and it is accordingly the case for most works on the geophysical inverse problem. For example, Weidelt (1972) only considered resistivity profiles that were continuous with respect to depth. Parker (1980) showed, however, that the natural setting for studies with finite data sets is a space permitting delta functions in the conductivity. He thus demonstrated that the best-fitting model (which he called D^+) consists of a finite number of delta functions in conductivity. He also showed that other types of models that approach the same level of misfit develop oscillations as they try to mimic the delta functions of D^+. When looking at models without delta functions, adding a further smoothing constraint in order to restrict the class of models provides an efficient solution to the problem of getting a stable and realistic solution (see e.g. Constable *et al.* 1987). In the same way, we introduce the following *a priori* distribution:

$$g(\mathbf{x}) = \gamma(x_1) \prod_{k=1}^{L-1} h(x_k, x_{k+1}), \tag{10}$$

where h is an $M \times M$ Markovian matrix (Robert 1996) which depends on the set of possible resistivities and $\boldsymbol{\gamma}$ is the normalized left eigenvector of the matrix $\mathbf{h}$ associated with the eigenvalue 1. The definition of γ ensures that the distribution g is homogeneous with respect to the layers. This implies in particular that all the marginal distributions of g are equal to $\boldsymbol{\gamma}$ (Robert 1996); the components of $\boldsymbol{\gamma}$ are therefore a solution of the set of equations

$$\sum_{i=1}^{M} \gamma(\rho_i) h(\rho_i, \rho_j) = \gamma(\rho_j), \qquad j = 1, \ldots, M. \tag{11}$$

We take as matrix $\mathbf{h}$ the following:

$$h(\rho_i, \rho_j) = C(\rho_i) \exp\{-\lambda[\log(\rho_i) - \log(\rho_j)]^2\}, \tag{12}$$

where $C(\rho_i)$ is a normalization constant such that, for each ρ_i,

$$\sum_{j=1}^{M} h(\rho_i, \rho_j) = 1.$$

The definition of $\mathbf{h}$ ensures that the smoother the model, the larger its *a priori* probability.

3.1.3 *The transition distribution of the Markov chain*

As shown is Section 2, the transition distribution of the Markov chain is the conditional probability of $X_l = \theta_i$ given the values of all the other parameters and the observations. Its expression is deduced from eq. (5) by replacing f and g by

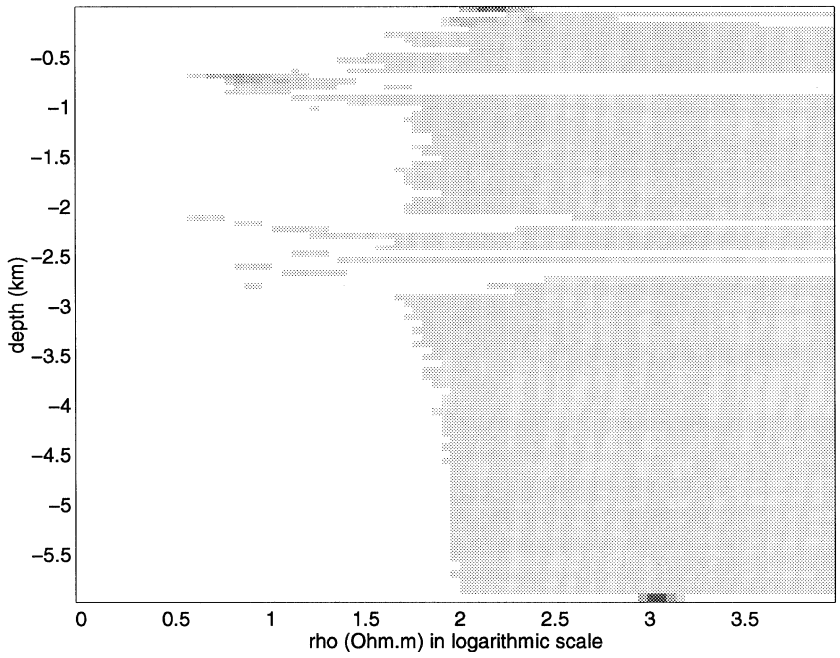

Figure 2. *A posteriori* distribution of the parameters for the five-layer model described in Table 1(a) for a smoothing factor $\alpha = 0$ (i.e. without smoothing). The results are presented as images, with the resistivities along the horizontal (x) axis and the depths along the vertical (y) axis. In the images, the level of grey of the (m, l) pixel corresponds to the value of the marginal *a posteriori* distribution of the parameter X_l for $X_l = \rho_m$: the darker the pixel, the higher the marginal *a posteriori* probability.

Table 1. Description of the models used to test the Markov chain Bayesian algorithm in magnetotelluric 1-D situations.

Resistivity profile			Data		Model definition	
layers	resistivity (Ωm)	Depth interface (m)			Figure 4a	
1	250	0	f_{min} (Hz)	0.1	L	94
2	25	600	f_{max} (Hz)	1000	z_1	50 m
3	100	1 000	n_F	41	z_L	6 000 m
4	10	3 000			M	81
5	1 000	3 250			$\rho_1 = \rho_{min}$	1 Ωm
					$\rho_M = \rho_{max}$	10 000 Ωm

(a)

Resistivity profile			Data		Model definition	
layers	resistivity (Ωm)	Depth interface (m)			Figure 4b	
1	250	0	f_{min} (Hz)	0.0001	L	81
2	5	50	f_{max} (Hz)	0.1	z_1	3 m
3	250	75	n_F	31	z_L	300 m
					M	81
					$\rho_1 = \rho_{min}$	1 Ωm
					$\rho_M = \rho_{max}$	10 000 Ωm

(b)

Resistivity profile			Data		Model definition	
layers	resistivity (Ωm)	Depth interface (m)			Figure 4c	
1	200	0	f_{min} (Hz)	0.000083	L	79
2	10	20 000	f_{max} (Hz)	0.001075	z_1	5 000 m
3	100	70 000	n_F	22	z_L	600 000 m
4	300	140 000			M	101
5	0.1	300 000			$\rho_1 = \rho_{min}$	0.01 Ωm
					$\rho_M = \rho_{max}$	1 000 Ωm

(c)

Data		Model definition Figure 4d		Model definition Figure 4e	
f_{min} (Hz)	0.000510	L	79	L	47
f_{max} (Hz)	0.035088	z_1	5 000 m	z_1	5 000 m
n_F	15	z_L	600 000 m	z_L	300 000 m
		M	81	M	81
		$\rho_1 = \rho_{min}$	1 Ωm	$\rho_1 = \rho_{min}$	1 Ωm
		$\rho_M = \rho_{max}$	10 000 Ωm	$\rho_M = \rho_{max}$	10 000 Ωm

(d)

their expressions given in eqs (9)–(12). It becomes

$$
\begin{aligned}
&P(X_l = \theta_{l,i} / X_k = x_k, k \neq l, \mathbf{y}) \\
&= \left(\exp\left\{ -\left[\chi^2(\mathbf{x}^{(l,i)}, \mathbf{y}) - \mathrm{Ln}[\gamma(x_1^{(l,i)})] \right.\right.\right. \\
&\quad \left.\left.\left. + \lambda \left(\sum_{k=1}^{L-1} [\log(x_k^{(l,i)}) - \log(x_{k+1}^{(l,i)})]^2 \right) \right] \right\} \right) \\
&\quad \Bigg/ \left(\sum_{j=1}^{l} \exp\left\{ -\left[\chi^2(\mathbf{x}^{(l,j)}, \mathbf{y}) - \mathrm{Ln}[\gamma(x_1^{(l,j)})] \right.\right.\right. \\
&\quad \left.\left.\left. + \lambda \left(\sum_{k=1}^{L-1} [\log(x_k^{(l,j)}) - \log(x_{k+1}^{(l,j)})]^2 \right) \right] \right\} \right),
\end{aligned}
\tag{13}
$$

where $\mathbf{x}$ and $\mathbf{x}^{\{l,j\}} \in \boldsymbol{E}$ and

$$x_k^{(l,l)} = x_k \forall k \neq 1,$$

$$x_l^{(l,j)} = \theta_{l,j}.$$

The normalization constant in eq. (13) is a sum over the set of the M possible values θ_i of the lth component x_l of the model $\mathbf{x}$.

Eq. (13) shows that the Markov chain transition distribution is a weighted sum of two functions, one measuring the fit between the model and the data, and the other the degree of smoothness of the model. The relative weight of these two functions depends on the choice of the trade-off parameter λ,

and the degree of smoothness of the *a posteriori* model can therefore be tuned by changing the λ value.

As already discussed in Section 2, the marginal distribution probabilities of Π are estimated with empirical averages of the transition probability distributions of $\mathbf{S}^{(n)}$. These transition probability distributions depend on the trade-off parameter λ and on the variances of the data y_i. We discuss the estimation of these parameters in the following sections.

3.2 Data variance estimation

Suppose that the variance $\mathrm{var}(y_i)$ of the noise on the data y_i is

$$\mathrm{var}(y_i) = \beta^2 |[\mathbf{F}(\mathbf{x})]_i|^2 .$$

An estimator of β^2 is

$$\hat{\beta}^2 = \frac{1}{2n_F} \sum_{i=1}^{n_F} \left| \frac{y_i - [\mathbf{F}(\mathbf{x})]_i}{[\mathbf{F}(\mathbf{x})]_i} \right|^2 .$$

This estimator can be incorporated in the algorithm. At step $n+1$, $\mathbf{S}^{(n+1)}$ is computed with the value of the variance provided by the estimator at the end of step n, namely

$$\hat{\beta}_n^2 = \frac{1}{2n_F} \sum_{i=1}^{n_F} \left| \frac{y_i - [\mathbf{F}(\mathbf{S}^{(n)})]_i}{[\mathbf{F}(\mathbf{S}^{(n)})]_i} \right|^2 .$$

The algorithm can then be used in two different ways, depending on whether or not the values of the variance of the noise on the data are known. If the β^2 values are given with the data, these values are used in the inversion. Otherwise, they can be estimated together with the *a posteriori* marginal probability distributions.

3.3 Smoothing factor estimation

In order to have a smoothing factor which is independent of both the number of layers and the number of possible resistivities, we use at each step n a smoothing factor

$$\lambda_n = \alpha \hat{\lambda}_n ,$$

with

$$\frac{1}{\hat{\lambda}_n} = \frac{1}{n} \sum_{k=1}^{n} \frac{2}{L} \left\{ (S_1^{(k)})^2 + \sum_{i=1}^{L-1} (S_{i+1}^{(k)} - S_i^{(k)})^2 + (S_L^{(k)})^2 \right\} .$$

When $\alpha = 1$, this estimator is the standard estimator of the parameter λ for the *a priori* model we use. It is then possible to tune the smoothing factor by choosing the appropriate value of α. It is worth noting that the meaning of α, in terms of an *a priori* degree of smoothness, does not depend on the choice of the parameters of the model. The α parameter could actually be estimated in a 'Bayesian manner', for instance with a maximum entropy criterion. However, we think that it is preferable to keep the possibility of tuning α, which enables one to investigate the influence of the degree of smoothness on the *a posteriori* pdf.

3.4 Numerical considerations

A Markov chain on $\boldsymbol{E}$ is actually a process which starts from any selected point of E and evolves, at random, according to its transition distribution. The random values are generated by means of the random function implemented with Fortran on Sun Spark Stations.

The starting point is likely to be far from the equilibrium state of the chain. As already stated, the asymptotic behaviour of the chain does not depend on the choice of the starting point. However, the very first iterations of the chain depend on that starting point, and the set of iterations, after which the chain is likely to be in equilibrium, is usually called the warm-up interval. The estimation of the marginal distributions is started once the warm-up interval has ended, using empirical averages of the transition probabilities. The computation is stopped when the expected precision in the estimates is better than a given threshold. The reliability of the results we get from the Markov chain depends on the relevance of the definition of both the warm-up interval and the stopping criteria.

3.4.1 *Warm-up interval length*

The initial value of all the components of the random variable $\mathbf{X}$ is taken to be equal to $(\rho_{\min} + \rho_{\max})/2$. The first iterations of the chain are associated with high values of the $\chi^2(\mathbf{x}, \mathbf{y})$ function defined in eq. (9). After the warm-up interval, during which it tends to decrease as time evolves, the χ^2 function starts oscillating around its expected value. The same is true for the estimated values of the data variance, thus indicating that the Markov chain has reached its stationary state. We found from experience that the warm-up interval of the Markov chain we use for 1-D MT inversion lasts a few tens of iterations. Fig. 3 shows results obtained with the model and data set described in Table 1(a).

3.4.2 *Stopping criteria*

The problem of the stopping criteria for Markov chains is not yet solved in the general case (see e.g. Fishman 1996). Here we used an empirical stopping rule based on an estimator of the variance of the estimated values of the *a posteriori* marginal probabilities of the possible resistivities. This variance varies as K/n, where n is the number of iterations for which the

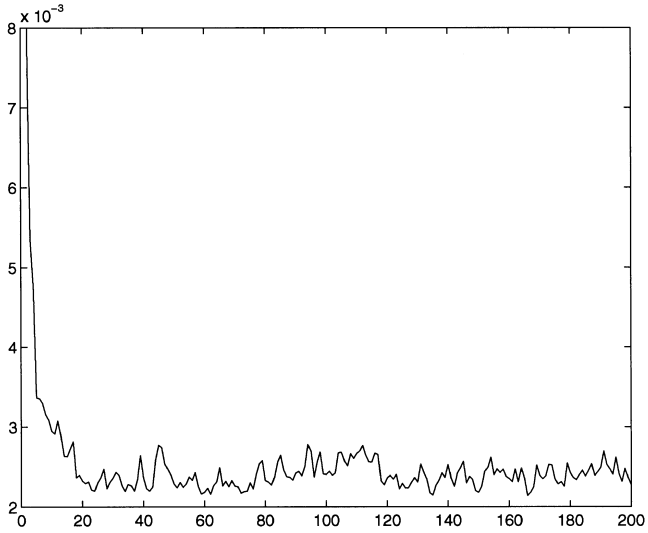

Figure 3. Evolution of χ^2 as a function of the number of iterations for the model described in Table 1(a) for a smoothing factor $\alpha = 5$. The χ^2 values are expressed in arbitrary units (y-axis); the number of iterations is indicated on the x-axis.

empirical average of the transition distributions is computed (see formula 14). For all the situations we have considered, we have found that averaging the transition distribution over 1000 iterations, after the end of the warm-up interval, ensures sufficient precision on the *a posteriori* marginal distribution estimates.

4 DATA ANALYSIS

Let us first use examples of synthetic data to discuss the results obtained with the algorithm presented in this paper. We have selected different synthetic models illustrating the wide range of magnetotelluric situations encountered in both exploration and fundamental geophysics. The results are expressed in terms of *a posteriori* probability distribution functions, described by means of the *a posteriori* marginal distributions estimated as the average over 1000 iterations of the transition distribution of the Markov chain, after a warm-up period of 10 iterations.

For all the models we are considering in this section, the *a priori* distribution $g(\mathbf{x})$ and the marginal *a posteriori* distributions are digitized over a set of possible resistivities evenly distributed on a logarithmic scale as 20 values per decade over a $[\rho_{\min}, \rho_{\max}]$ interval selected in order to cover the whole range of possible resistivity values in geophysics. Therefore, each model under consideration (see Table 1) corresponds to the most general situation, for which no relevant *a priori* knowledge on the resistivity values is available. In each case the data set we use consists of complex magnetotelluric impedances with added Gaussian noise. The results are presented hereafter as images, with the resistivities along the horizontal (x) axis, and the depths along the vertical (y). In the images, the level of grey of the (m, l) pixel corresponds to the value of the marginal *a posteriori* distribution of the parameter X_l for $X_l = \rho_m$: the darker the pixel, the higher the marginal *a posteriori* probability.

For examples in studies of the uppermost few kilometres of the crust, corresponding for instance to sedimentary cover studies, we consider the synthetic model presented in Table 1(a). It is characterized by a four-layer, 3250 m thick conductive medium overlying a resistive substratum. The vertical scales and the values for the resistivities are realistic for a sedimentary basin, although the details of the resistivity profile may not be. Their purpose is to test the ability of the algorithm to describe *a posteriori* parameter probability distribution functions in different situations. The data set consists of 41 complex impedances computed at frequencies evenly distributed on a logarithmic scale in the range 10^{-1}–10^{3} Hz; that is, 10 points per decade over the audiomagnetotelluric frequency range. Added 5 per cent Gaussian noise is of the same order as that observed in the case of natural data processed with sophisticated techniques such as robust processing (Chave *et al.* 1987).

Fig. 4(a) presents the results obtained for different values of the smoothing parameter α ($\alpha = 1$, 5 and 10). For the α values under consideration, all the *a posteriori* marginal distributions are single-peaked. For a given parameter (e.g. the resistivity of a given layer), the width of the peak of the *a posteriori* marginal distributions decreases with increasing values of α. For $\alpha = 1$, all the *a posteriori* marginal distributions are almost flat in resistivity over more than one decade, and do not actually allow a precise determination of the most likely *a posteriori* resistivity profile. However, this is not the case for larger values of α (typically $\alpha \geq 5$), for which the narrowness of the peaks almost always guarantees an accurate determination of the more likely *a posteriori* resistivity profile given the data and the selected degree of smoothness. It is worth noting that for $\alpha \geq 3$ the expected value of each parameter is not significantly different from its more likely value, thus giving a clear indication of a robust determination of the resistivity profile for these α values. The noise estimates are consistent with the intensity of the added Gaussian noise.

As a result of the smooth character of the model, the steps in the resistivity profile which mark the interfaces between two consecutive layers are accounted for in the *a posteriori* resistivity profile by transition zones, characterized by a progressive change of the resistivity values. As expected, the larger the α value, the smoother the resistivity profile, and the weaker the average slope of the resistivity profile in the transition zones.

Fig. 4(a) shows that the increasing smoothness of the resistivity profile with increasing α values results in a significant degradation of the description obtained of the actual resistivity profile. The five layers can be seen, in particular the two intermediate conductive layers for $\alpha \leq 10$. For these values, the *a posteriori* marginal distributions obtained provide a fairly good estimate of the actual resistivity profile, except for the fourth layer, which is a thin conductive layer at depth. For $\alpha > 10$ (not shown in Fig. 4), the increasing smoothness of the *a posteriori* resistivity profiles, corresponding to increasing α values, results in a progressive disappearance of the intermediate thin layers (layers 2 and 4).

We now consider studies of the uppermost few hundreds of metres of the Earth, as for instance in the case of hydrogeological studies. The synthetic model selected, presented in Table 1(b), corresponds to a 25 m thick conductive layer embedded in a resistive host. The vertical scales and the values for the resistivities are realistic for ground-water situations. The data are a set of 31 complex impedances computed at evenly distributed frequencies (10 points per decade) on a logarithmic scale in the range 10–10^{4} Hz. Again, there is added 5 per cent Gaussian noise.

Fig. 4(b) presents the results obtained with this second model for different values of the smoothing parameter α ($\alpha = 1$, 5 and 10). As observed for the previous model, the *a posteriori* marginal distributions are always single-peaked, with a peak width that decreases with increasing α values. For all the α values we considered, the conductive layer is detected. Fig. 4(b) shows that its depth and resistivity are well estimated for α values smaller than or equal to 10.

We consider also two data sets already processed with different MT 1-D inversion schemes. The first is taken from Tarits *et al.* (1994) and the second is the COPROD data set. Tarits *et al.* (1994) considered the case of stratified media with a limited number of layers, and developed a Monte Carlo inversion scheme to determine the *a posteriori* marginal distributions of the thickness and resistivity of the layers. The model simulates data which could have been recorded on the sea bottom. Because of the narrowness of the frequency range of ocean-bottom MT data [typically 10^{-6}–10^{-3} Hz; see e.g. Filloux (1987)], the information provided by the data is quite limited, as opposed to the situations simulated by the two previous models.

The synthetic model of Tarits *et al.* is described in Table 1(c). It is characterized by a resistive top layer overlying an intermediate conductive medium. The resistivity of the intermediate

Table 2. The COPROD data (from Jones & Hutton 1979).

Period (s)	$\log_{10} \rho_a$	$\sigma_{\log \rho}$	Phase (degrees)	σ_{phase} (degrees)
28.5	2.315	0.0721	57.19	22.95
38.5	2.254	0.0425	58.19	22.95
52	2.229	0.0244	61.39	4.96
70.5	2.188	0.0210	59.09	4.46
95.5	2.180	0.0164	59.89	5.96
129.0	2.162	0.0173	51.19	22.95
174.6	2.151	0.0287	46.89	22.95
236.2	2.208	0.0328	42.79	2.46
319.6	2.194	0.0193	36.89	1.65
432.5	2.299	0.0270	32.00	22.95
585.1	2.338	0.0591	44.00	6.37
791.7	2.420	0.0506	32.00	2.46
1071.1	2.405	0.0825	37.59	22.95
1449.2	2.308	0.1233	45.29	4.15
1960.7	2.397	0.0927	50.09	22.95

medium increases from layer 2 to layer 3, and the bottom half-space is conductive. The vertical scales and the values for the resistivities are realistic for a young oceanic lithosphere (see e.g. Tarits 1986), although the details of the resistivity profile may be different. The data we used are the impedances with added noise computed at 22 frequencies by Tarits *et al.* (1994) (see their Table 2).

Fig. 4(c) presents the results obtained for $\alpha = 1$, 5 and 10. The effect of the smoothing factor is more sensitive than in the two previous cases. This probably results from the more limited information available from the data, because of both the smaller number of frequencies and the increased noise.

This lack of information is clearly illustrated by the *a posteriori* distributions observed at depths greater than 350 km. For $\alpha = 1$, the more likely value of the resistivity progressively increases up to the centre of the $[\rho_{\min}, \rho_{\max}]$ interval, thus indicating that the available information comes mainly from the *a priori* marginal distribution $g(\mathbf{x})$. For $\alpha = 5$ and 10, the observed *a posteriori* marginal distributions clearly result from the smoothing-dependent downward propagation of the information provided by the data on the resistivity at depths in the range 300–350 km. At depths greater than 350 km, the model is therefore only determined by the *a priori*

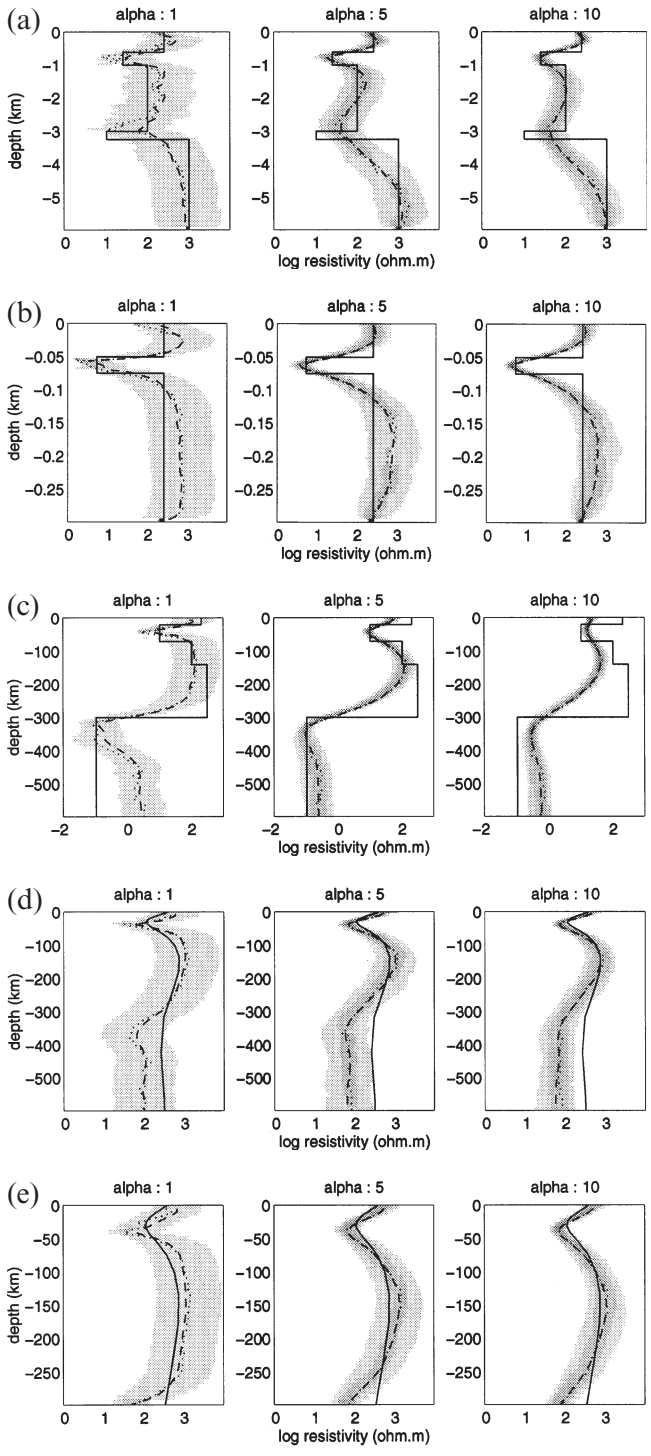

Figure 4. *A posteriori* distribution of the parameters for the models described in Table 1 for three different values of the smoothing factor: $\alpha = 1$ (left), $\alpha = 5$ (middle) and $\alpha = 10$ (right). The results are presented as images with the resistivities along the horizontal (x) axis and the depths along the vertical (y). In the images, the level of grey of the (m, l) pixel corresponds to the value of the marginal *a posteriori* distribution of the parameter X_l for $X_l = \rho_m$: the darker the pixel, the higher the marginal *a posteriori* probability. On each image, the more likely values (dotted curves) and the expected values (dashed curves) of the resistivities are also plotted. (a) *A posteriori* marginal distributions obtained for the synthetic model and parameter definition described in Table 1(a). The noise estimates (about 5 per cent for each case presented) are consistent with the intensity of the added Gaussian noise. (b) *A posteriori* marginal distributions obtained for the synthetic model and parameter definition described in Table 1(b). The noise estimates (about 5 per cent for each case presented) are consistent with the intensity of the added Gaussian noise. (c) *A posteriori* marginal distributions obtained for the synthetic model of Tarits *et al.* (1994). The synthetic model and the parameter definition are described in Table 1(c). The noise estimates are of the order of 5 per cent for $\alpha = 1$ and $\alpha = 5$, and 10 per cent for $\alpha = 10$. They are consistent with the 95 per cent error bars on the synthetic data (see Tarits *et al.* 1994). (d) *A posteriori* marginal distributions obtained for the COPROD data with the model definition given in the left column of Table 1(d) (note that there are 79 layers in the uppermost 600 km). The OCCAM model (Constable *et al.* 1987) is also plotted (continuous line) on each image. The noise estimates are of the order of 10 per cent for the three α values considered. (e) *A posteriori* marginal distributions obtained for the COPROD data with the model definition given in the right column of Table 1(d) (note that there are 47 layers in the uppermost 300 km). The OCCAM model (Constable *et al.* 1987) is also plotted (continuous line) on each image. The noise estimates are of the order of 10 per cent for the three α values considered.

knowledge, thus indicating that it can actually be chosen freely without affecting the consistency of the model and the data. The results presented in Fig. 4(c) provide an experimental determination of the critical depth z_c topping the region inaccessible to this magnetotelluric sounding (Parker 1982).

Fig. 4(c) clearly shows that the resistivity profiles obtained for $\alpha = 5$ provide quite a good balance between the smoothness hypothesis inherent in our method and the best possible use of the information contained in the data set. The *a posteriori* variances are in fact quite large for $\alpha = 1$, whilst the information is partly hidden by smoothness requirements for $\alpha = 10$. It is worth noting that the results we obtained for $\alpha = 1$ and 5 provide fairly clear information on the depths where sharp variations in the resistivity actually exist, whilst the results obtained by Tarits *et al.* (1994) did not, even in the case of a five-layer *a priori* model. This may stem from the existence of significant covariances between some of the parameters used in the case of Tarits *et al.* (layer resistivities and interface depths), whilst it is likely not to be the case between the parameters used in the present case.

The second data set already processed that we consider is the COPROD data set. It is a field data set circulated by Dr Alan Jones, collected at a site near Newcastleton in Britain and described in Jones & Hutton (1979). This data set, already used by Constable *et al.* (1987) for testing their OCCAM inversion algorithm, is presented in Table 2.

Fig. 4(d) presents the results obtained with the COPROD data for $\alpha = 1$, 5 and 10, together with the smallest first-derivative model obtained from this data set with the OCCAM inversion (Constable *et al.* 1987). In the uppermost 150 km, the results obtained with the two methods are consistent. The comparison between our results and those obtained using the OCCAM inversion gives a clear illustration of the influence of the smoothing parameter α. For α values small enough (in this case typically $\alpha \leq 5$), the variance of the *a posteriori* distribution of the parameters decreases with increasing α values, without any significant change in their more likely value. Furthermore, the results obtained for $\alpha = 1$ clearly suggest the existence of a conductive layer limited by fairly sharp resistivity contrasts. For larger α values, the increasing influence of the smoothing *a priori* distribution of the parameters results in changes in both the variance and the more likely value of the parameters. For depths smaller than 150 km, Fig. 4(d) shows that the OCCAM inversion leads to a model corresponding to an α value between 5 and 10.

At greater depths, our results are significantly different from those obtained by Constable *et al.* (1987) because of different estimates of the more likely resistivity of the terminating half-space: about 400 Ω m for the OCCAM model versus about 100 Ω m for our models. As already observed for the previous model, the determinations of the more likely value of this resistivity correspond to the middle of the $[\rho_{\min}, \rho_{\max}]$ interval. In this case, the observed *a posteriori* marginal distributions suggest for Parker's critical depth z_c a range between 300 and 350 km, in accordance with the estimates made by Parker (1982) for the same data set.

We modified the model by reducing the thickness z_L of the stack of layers to 300 km and reprocessed the COPROD data with our algorithm. Fig. 4(e) presents the results obtained for $\alpha = 1$, 5 and 10. As expected, the estimated *a posteriori* marginal distributions corresponding to depths smaller than z_c do not depend on z_L, provided it is on the order of or larger than z_c. Figs 4(d) and (e) thus provide a striking illustration of the robustness of the *a posteriori* marginal distribution estimate with respect to an overestimate of the thickness of the region accessible to the magnetotelluric sounding.

5 CONCLUSIONS

This paper clearly shows that MCMC methods provide a very efficient tool for solving non-linear geophysical inverse problems in a Bayesian context. The Markov chain used is a homogeneous aperiodic ergodic one, the transition probability of which is the conditional probability of the parameters given the statistical distribution of the data and the *a priori* knowledge. The invariant distribution of such a Markov chain is the *a posteriori* marginal distributions of the parameters. According to the ergodic theorem, this invariant distribution can be estimated from the observed sequence of images (that is, the trajectory of the chain; see Section 2).

The *a priori* knowledge on the parameters and the *a posteriori* marginal distributions are digitized on a set of selected possible values of the parameters. The Markov chain we used is a Gibbs sampler on the set $\boldsymbol{E}$ of possible images that considers each element successively and updates its parameter values.

We used the well known MT 1-D inverse problem situations to illustrate the use of MCMC methods in geophysical imaging. We considered models consisting of a stack of thin, homogeneous layers with fixed thicknesses. The parameters are then the resistivity of the layers. Because of the non-uniqueness of the MT 1-D inverse problem with a finite number of noisy observations, using the simplest uniform *a priori* distribution leads to marginal *a posteriori* distributions almost uniform over a wide range of resistivity values. Overcoming this problem requires a better specification of the *a priori* distribution to introduce a dependence between the parameters which select a particular class of models. It is widely accepted that adding a further smoothing constraint provides an efficient solution to obtain a stable and realistic solution (see e.g. Constable *et al.* 1987; Smith & Booker 1988). This was done by introducing an *a priori* distribution which tends to favour smooth models. The *a priori* distribution involves a smoothing factor α which controls the smoothness of the model (the greater α, the smoother the model).

The results we present in this paper give a striking illustration of the efficiency of such an MCMC approach.

(1) The resistivity profile, which can be deduced from the *a posteriori* marginal distributions, remains generally stable for a fairly wide range of α values, and in particular for quite small α values. This then permits effective scanning of different classes of models with different smoothnesses. Smoothing factors α of the order of 1 in particular would allow one to decide whether or not sharp variations in resistivity are likely to exist.

(2) In some situations, the *a posteriori* marginal distributions provide fairly clear information on the depths where sharp variations in the resistivity actually exist, whilst *a priori* simple layered models (models with a limited number of homogeneous layers allowing the existence of sharp resistivity contrasts between the layers) do not. This may result from the existence of significant covariances between some of the parameters used in the case of the simple layered model (layer resistivities and interface depths), whilst this is unlikely to be the case between the parameters used in the present case.

(3) The analysis of the *a posteriori* marginal distributions and of their dependence on α gives an indication of the information available in the data: the more limited the information available in the data, the more sensitive the *a posteriori* distribution to changes in the smoothing factor. On the other hand, analysing the *a posteriori* distributions observed at large depths permits an experimental determination of Parker's critical depth z_c (Parker 1982) topping the region inaccessible to magnetotelluric sounding.

ACKNOWLEDGMENTS

This work started during the Thèse d'Université of Hendra Grandis, prepared in France in the frame of the French–Indonesian cooperation program. It was partly supported by the French Institut National des Sciences de l'Univers (MM, contract N°96/TOM/14).

REFERENCES

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

Bailey, R.C., 1970. Inversion of the geomagnetic induction problem, *Proc. R. Soc. Lond.,* **315,** 185–194.

Bayes, T., 1763. An essay toward solving a problem in the doctrine of chances, *Phil. Trans. R. Soc. Lond.,* **53,** 370–418.

Berger, J.O., 1985. *Statistical Decision Theory and Bayesian Analysis,* Springer-Verlag, New York.

Box, G. & Tiao, G., 1973. *Bayesian Inference in Statistical Analysis,* Addison-Wesley, Reading.

Chave, A.D., Thomson, D.J. & Ander, M.E., 1987. On the robust estimation of power spectra, coherences and transfer functions, *J. geophys. Res.,* **92,** 633–648.

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.

Feller, W., 1970. *An Introduction to Probability Theory and its Application,* Willey, New York.

Filloux, J.H., 1987. Instrumentation and experimental methods for oceanic studies, in *Geomagnetism,* pp. 143–248, ed. Jacobs, J.A., 1, Academic Press, San Diego.

Fishman, G.S., 1996. Monte Carlo: concepts, algorithms, and applications, *Springer Series in Operations Research,* Springer, New York.

Grandis, H., 1994. Imagerie electromagnetique Bayesienne par la simulation d'une chaîne de Markov, *Doctorat d'Université,* Université Paris VII.

Jones, A.G. & Hutton, R., 1979. A multi-station magnetotelluric study in southern Scotland: I. Fieldwork, data analysis and results, *Geophys. J. R. astr. Soc.,* **56,** 329–349.

Jouanne, V., 1991. Application des techniques statistiques Bayesiennes à l'inversion de données électromagnétiques, *Doctorat d'Université,* Université Paris VII.

Langer, R.E., 1933. An inverse problem in differential equations, *Bull. Am. Math. Soc.,* ser. 2, **29,** 814–820.

Lavielle, M., 1991. 2-D Bayesian deconvolution, *Geophysics,* **56,** 2008–2018.

Mackie, R.L., Bennett, B.R. & Madden, T.R., 1988. Long period magnetotelluric measurements near the central California coast: a land-locked view of the conductivity structure under the Pacific Ocean, *Geophys. J.,* **95,** 181–194.

Madden, T.R. & Nelson, P., 1964. A defense of Cagniard's magnetotelluric methods, in *Magnetotelluric Methods,* pp. 371–401, ed. Vozoff, K., Geophysics Reprint Series, Vol. 5, SEG Publishing.

Oldenburg, D.W., 1990. Inversion of electromagnetic data: an overview of new techniques, *Surv. Geophys.,* **11,** 231–270.

Parker, R.L., 1980. The inverse problem of electromagnetic induction: existence and construction of solutions based on incomplete data, *J. geophys. Res.,* **85,** 4421–4428.

Parker, R.L., 1982. The existence of a region inaccessible to magnetotelluric sounding, *Geophys. J. R. astr. Soc.,* **68,** 163–170.

Parker, R.L., 1983. The magnetotelluric inverse problem, *Surv. Geophys.,* **6,** 5–25.

Parker, R.L. & McNutt, M.K., 1980. Statistics for the one-norm misfit measure, *J. geophys. Res.,* **85,** 4429–4430.

Press, S.J., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T., 1989. Numerical Recipes: the Art of Scientific Computing, Cambridge University Press, Cambridge.

Raiche, A., 1994. Modelling and inversion—progress, problems and challenges, *Surv. Geophys.,* **15,** 159–208.

Robert, C., 1996. *Méthodes de Monte Carlo par Chaînes de Markov,* Economica, Paris.

Roussignol, M., Jouanne, V., Menvielle, M. & Tarits, P., 1993. Bayesian electromagnetic imaging, in *Computer Intensive Methods,* pp. 85–97, eds Hardle, W. & Siman, L., Physical Verlag.

Smith, J.T. & Booker, J.R., 1988. Magnetotelluric inversion for minimum structures, *Geophysics,* **53,** 1565–1576.

Tarits, P., 1986. Conductivity and fluids in the oceanic upper mantle, *Phys. Earth planet. Inter.,* **42,** 215–226.

Tarits, P., 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.

Tarlowski, Z., 1982. Direct and inverse problems in local electromagnetic induction, *Surv. Geophys.,* **4,** 395–404.

Ward, S.H. & Hohmann, G.W., 1988. Electromagnetic theory for geophysical applications, in *Electromagnetic Methods in Applied Geophysics,* pp. 131–312, ed. Nabighian, M.N., SEG Publishing.

Weidelt, P., 1972. The inverse problem of geomagnetic induction, *Z. Geophys.,* **38,** 257–289.

Wiggins, R.A., 1969. Monte Carlo inversion of body wave observations, *J. geophys. Res.,* **74,** 3171–3181.

Wiggins, R.A., 1972. The general linear inverse problem: implication of surface waves and free oscillations for earth structure, *Rev. Geophys. Space Phys.,* **10,** 251–285.