Earth Planets Space, **53**, 777–791, 2001

Geostatistical approach to bayesian inversion of geophysical data: Markov chain Monte Carlo method

Seok-Hoon Oh[1] and Byung-Doo Kwon[2]

[1]*Marine Meteorology and Earthquake Research Laboratory, Meteorological Research Institute, 460-18, Shindaebang-dong, Dongjak-gu, Seoul 156-720, Korea*
[2]*Seoul National University, San 56-1, Shilim-dong, Kwanak-ku, Seoul 151-742, Korea*

(Received December 9, 1999; Revised February 26, 2001; Accepted April 13, 2001)

This paper presents a practical and objective procedure for a Bayesian inversion of geophysical data. We have applied geostatistical techniques such as kriging and simulation algorithms to acquire *a prior* model information. Then the Markov chain Monte Carlo (MCMC) method is adopted to infer the characteristics of the marginal distributions of model parameters. Geostatistics which is based upon a variogram model provides a means to analyze and interpret the spatially distributed data. For Bayesian inversion of dipole-dipole resistivity data, we have used the indicator kriging and simulation techniques to generate cumulative density functions from Schlumberger and well logging data for obtaining *a prior* information by cokriging and simulations from covariogram models. Indicator approaches make it possible to incorporate non-parametric information into the probabilistic density function. We have also adopted the Markov chain Monte Carlo approach, based on Gibbs sampling, to examine the characteristics of *a posterior* probability density function and marginal distributions of each parameter. The MCMC technique provides a robust result from which information given by the indicator method, that is fundamentally non-parametric, is fully extracted. We have used the *a prior* information proposed by the geostatistical method as the full conditional distribution for Gibbs sampling. And to implement Gibbs sampler, we have applied the modified Simulated Annealing (SA) algorithm which effectively searched for global model space. This scheme provides a more effective and robust global sampling algorithm as compared to the previous study.

1. Introduction

In the past decade, Bayesian analysis (Bayes, 1763; Tarantola, 1987) for geophysical inverse problem has been tried sometimes as test problems (Duijndam, 1988a, b) or sometimes as alternatives (Mosegaard and Tarantola, 1995; Gouveia, 1996; Moraes, 1996) to the traditional deterministic process (Lines and Treital, 1984). Although the robustness of the Bayesian analyses is well confirmed by these studies, their approaches were rather arbitrary and subjective. Therefore, within these kinds of the Bayesian frame, it is difficult to solve the geophysical inverse problems in a systematic way.

The most perplexing and significant aspect of the Bayesian frame is the preparation of *a prior* information. Duijndam (1988a, b) arbitrarily assumed the 1st and 2nd statistical moments of model parameters. Moraes (1996) defined the probabilistic structure of the model parameters in advance and then used the maximum entropy technique inversely. Gouveia (1996) incorporated various experimental noises into the model and data covariance with subjectivity. Scales and Tarantola (1994) thought of the smooth components of the P-wave sonic log as representing the geological structure and others erratic random fluctuations, but, in general, it is often difficult to distinguish between signal and noise.

A recent study given by Grandis *et al.* (1999) makes *a prior* PDF (probability density function) by digitizing the parameters over a set of values, called possible resistivity values, with no constraints. They also composed *a prior* distribution with a Markovian matrix which depends on the set of possible resistivities. This process is considerably more objective and shows good results. But the quantity of digitizing affects the *a posterior* PDF, and there is no device to include various independent data in the Bayesian analysis frame, so prompt applications of this approach to more practical and complex problems may be rather difficult.

The representation of *a prior* information may be divided into three alternative stems (Scales and Tenorio, 1998). The first strategy is a subjective Bayesian one: prior probabilities are designed to represent states of mind, prejudices or prior experiences. However, depending on the amount and type of prior information, the proper choice of probabilities may or may not be clear. Most of the previously mentioned studies of Duijndam (1988a, b), Moraes (1996) and Gouveia (1996) may be classified in this category.

The second approach attempts to make a more objective choice of priors by relying on theoretical considerations such as maximum entropy, transformation invariance, or by somehow using a large number of observations to estimate *a prior*. This approach is sometimes called empirical Bayes. We took this approach with geophysical data and geostatistical techniques as described in later sections. The empirical Bayes analysis can be seen as an approximation of a

full hierarchical Bayes analysis based on the joint probability distribution of all parameters and available data. For an introduction to empirical and hierarchical models see Carlin and Louis (1996) and Gelman *et al.* (1995).

A third strategy is to abandon Bayes altogether and use only deterministic prior information about models: wavespeed is positive (a matter of definition); velocity is less than the speed of light (a theoretical prediction). Statistical aspects of these problems are included in the look of random data uncertainties. Tokhonov's regularization is one way of implementing this strategy.

As previously mentioned, we have adopted the second approach and proposed a more practical and objective mechanism to construct two dimensional *a prior* information from various geophysical data by means of geostatistical simulations.

Actually, simple uncertainty analyses have been applied to the inversion process by many researchers. For example, Wijk *et al.* (1997) proposed some simple algorithms to apply to data domain variances in order to weigh their contributions to the inversion process, but the process was somewhat deterministic and gave no more uncertainty analysis. We import a mechanism for guessing the subspace structure to understand the noisy information and extract to a maximum extent signals that represent the feasible geologic structure. If this is fully workable, constraints to avoid instability, such as smoothing terms, are not required. Bayesian frame does all of the above.

Geostatistics is based on the variogram model to estimate and simulate properties on unsampled points or blocks with probabilistic informations. The estimation may come from one variable or combine several variables. In our numerical study, we have used the indicator kriging and simulation technique to generate cumulative probability density functions from Schlumberger and well log data for *a prior* information in interpreting geoelectrical data. Such a technique can generate probability density function independent of any statistical moments of data.

Inverted Schlumberger and well log data make a co-variogram model and present the various spatial aspects of our data. Since geostatistics concentrates on the configuration of the data locations and values, we can get spatial uncertainty as well as local uncertainty. Thus spatial consideration gives very important compensations due to the lack of statistical information of observed data and proposed model parameters itself, because most of the earth science models have very natural and reasonable spatial structure.

Bayesian approach to geophysical inversion is simply the multiplicity of two kind of PDFs, *a prior* PDF and likelihood PDF (Tarantola, 1987). But as we can see from the process of parameterization of geophysical inversion, a highly multidimensional state of *a posterior* PDF makes interpreters dizzy and doesn't give any choice in inferring the substructure. Therefore, a marginalization process is required. This process makes the Bayesian approach a hard problem to solve, because it depends on high dimensional integrals as much as the number of parameters. In addition, naturally, the kernel of the integral is highly nonlinear. So we used the Monte Carlo sampling integral technique, especially Gibbs sampling, to achieve effectiveness, which more sampling is achieved near the maximum *a posterior* values by using *a prior* information as the full conditional distribution. This process was a general and non-parametric approach, because any constraints are not imposed on *a prior*, likelihood, and *a posterior* PDF. Nonparametric and general treatment of the PDF of the data is very powerful and no information is lost in analyzing the substructure.

2. Geostatistical Approach to *a prior* Information

Geostatistics is concerned with "the study of phenomena that fluctuate in space" and/or time (Olea, 1975). Geostatistics offers a collection of deterministic and statistical tools aimed at understanding and modeling spatial variability.

The basic paradigm of predictive statistics is characterization of any unsampled (unknown) value z as a random variable Z, the probability distribution of which models the uncertainty of z. A random variable is a variable that can take a variety of outcome values according to some probability (frequency) distribution (Deutsch and Journel, 1992). The random variable Z, and more specifically its probability distribution is usually location-dependent; hence the notation $Z(\mathrm{u})$, with u being the location where coordinates vector is adequate. The random variable $Z(\mathrm{u})$ is also information-dependent in the sense that its probability distribution changes as more data about the unsampled value $z(\mathrm{u})$ become available.

Geostatistics was originally devised to estimate properties of unsampled points for delineating ore deposits. But these days those tools are used not only for estimation of unsampled points but also for inference of local and spatial uncertainty estimation (Goovaerts, 1997; Isaaks and Srivastava, 1989). Geostatistics is largely based on the random function model, whereby the set of unknown values is regarded as a set of spatially dependent random variables. Some people may ask why, the observed value being already determined, we should use such a random function model. In reply, a model is but a representation of the (unknown) reality. Although that reality is unique, it has many possible representations, depending on the information available and the goal of the study. So instead of a single estimated value for the unknown $z(\mathrm{u})$, the probabilistic approach provides a set of possible values with the corresponding probabilities of occurrence. Such presentation reflects our imperfect knowledge of the unsampled value $z(\mathrm{u})$ and, more generally, of the distribution of z within the area (Goovaerts, 1997). Our ultimate goal is to infer the uncertainty structure of an interpreted region from *a prior* informations such as well-logging and Schlumberger sounding data. This is, after all, the estimation of unsampled points or area spatially dependent on random model or variables, so it coincides with the geostatistical model.

2.1 Geostatistics: A statistical tools for spatial data

2.1.1 Kriging Kriging, the most frequently quoted terminology in the following text, is a linear, unbiased, least-squares spatial interpolation method; a weighted-average, or weighted-mean, estimator whose weights are functions of spatial covariance. The spatial covariance is derived from a simple calculation, called *semivariogram* (or traditionally variogram), given by the following equation:

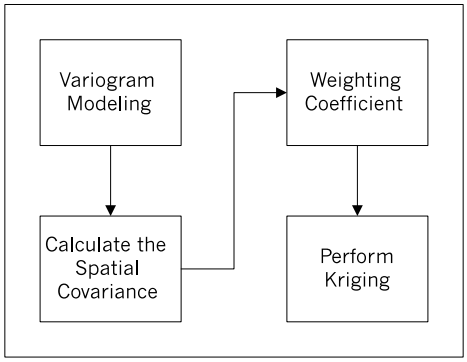

Fig. 1. The sequential stage to perform kriging for spatial data.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(x_i) - Z(x_i + h))^2. \qquad (1)$$

Variogram $\gamma(h)$ describes the variance distribution of the data given by the relative location difference h. Z is the random function attained at x. Figure 1 shows the sequential stages to perform kriging irregularly and sparsely distributed spatial data.

2.1.2 Simulation A geostatistical simulation algorithm aims to draw realizations that reflect the statistics modeled from the data. The question is how well these statistics can be reproduced. The smoothing effect of kriging, and more generally, of any low pass-type interpolation, is due to missing error components (Deutsch and Journel, 1992). Consider the random function $Z(\mathrm{u})$ as the sum of the estimator $Z^*(\mathrm{u})$ and the corresponding error $R(\mathrm{u})$:

$$Z(\mathrm{u}) = Z^*(\mathrm{u}) + R(\mathrm{u}).$$

Kriging, for example, would provide the smoothly varying estimator $Z^*(\mathrm{u})$. To restore the full variance of the random function model, one may think of simulating a realization of the random function error with zero mean and the correct variance and covariance. The simulated z-value would be the sum of the unique estimated value and a simulated error value:

$$z_c^{(l)} = z^*(\mathrm{u}) + r^{(l)}(\mathrm{u}).$$

In this process, the error component $R(\mathrm{u})$ must be independent or at least orthogonal to the estimator $Z^*(\mathrm{u})$, and the random fuction $R(\mathrm{u})$, modeling that the error must have the same spatial distribution or, at least, the same covariance as the actual error (Deutsch and Journel, 1992).

2.2 Assessment of local uncertainty

To estimate the uncertainty of a point, the most rigorous way is first to assess the uncertainty, then to deduce an estimate optimal in some appropriate sense (Goovaerts, 1997). This is significantly different from the traditional approach of first deriving the estimate then attaching to it a confidence interval. Let $Z(\mathrm{u})$ be the random value modeling the uncertainty about $z(\mathrm{u})$. The distribution function $F(\mathrm{u}; z \mid (n)) = \mathrm{Prob}\{Z(\mathrm{u}) \le z \mid (n)\}$ made conditional to the information available (n) fully models that uncertainty in the sense that probability intervals can be derived, such as

$$\mathrm{Prob}\{Z(\mathrm{u}) \in (a, b] \mid (n)\} = F(\mathrm{u}; b \mid (n)) - F(\mathrm{u}; a \mid (n)) \qquad (2)$$

or

$$\mathrm{Prob}\{Z(\mathrm{u}) > b \mid (n)\} = 1 - F(\mathrm{u}; b \mid (n)). \qquad (3)$$

Note that these probability intervals are independent of any particular estimate $z^*(\mathrm{u})$ of the unknown value $z(\mathrm{u})$. Indeed, uncertainty depends on the information available (n) (so conditional), not on the particular optimality criterion retained to define an estimate (Srivastava, 1987). Each conditional probability distribution function $F(\mathrm{u}; z \mid (n))$ provides a measure of local uncertainty in that it relates to a specific location u. Now, if we have the minimal information available about the z-value at any location u usually consists of a physical constraint interval $[z_{\min}, z_{\max}]$, indicator transform for sampled point

$$i(\mathrm{u}_\alpha; z_k) = \begin{cases} 1 & \text{if, } z(\mathrm{u}_\alpha) \le z_k \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

$$\text{where,} \quad k = 1, \ldots, K, \quad z_k \in [z_{\min}, z_{\max}]$$

divides the data to rank code by criterion z_k, and kriging for unsampled point u for respective threshold makes it possible to get a cumulative density function. This is called the indicator kriging, and it provides the local uncertainty information as well as estimation of unsampled points (Goovaerts, 1997).

As we mentioned in previous sections, kriging depends on the spatial relations of the data given by a variogram to infer the covariance and autocovariance in space. So in addition to the primary property, secondary or more various informations may be combined to generate a more informative variogram model. This is simply to add one or more terms multiplied by each coefficients making the variance minimal.

2.3 Assessment of spatial uncertainty

In the previous section, we described the way to implement the local uncertainty structure conditional to neighbour sparse sample data. But this is based on the kriging technique which makes the map smooth by the minimization of the variance, so that earth science data about which people want to get information about anomalous zone result in non-informative and uninteresting interpretations and also are confined to local information.

Geostatistical simulations used to overcome the above troubles make it possible to measure the joint uncertainty about attribute values at several locations taken together, and generate a map or a realization of z-values, say, $\{z^{(l)}(\mathrm{u}), \mathrm{u} \in \mathrm{A}\}$ with l denoting the lth realization, which reproduces statistics deemed most consequential for the problem at hand.

Typical requisites for such a simulated map are as follows (Goovaerts, 1997):

1) Data values are honored at their locations:
   $Z^{(l)}(\mathrm{u}) = z(\mathrm{u}_\alpha) \ \forall \mathrm{u} = i_\alpha, \ \alpha = 1, \ldots, n.$
   The realization is then said to be conditional (to the data values).

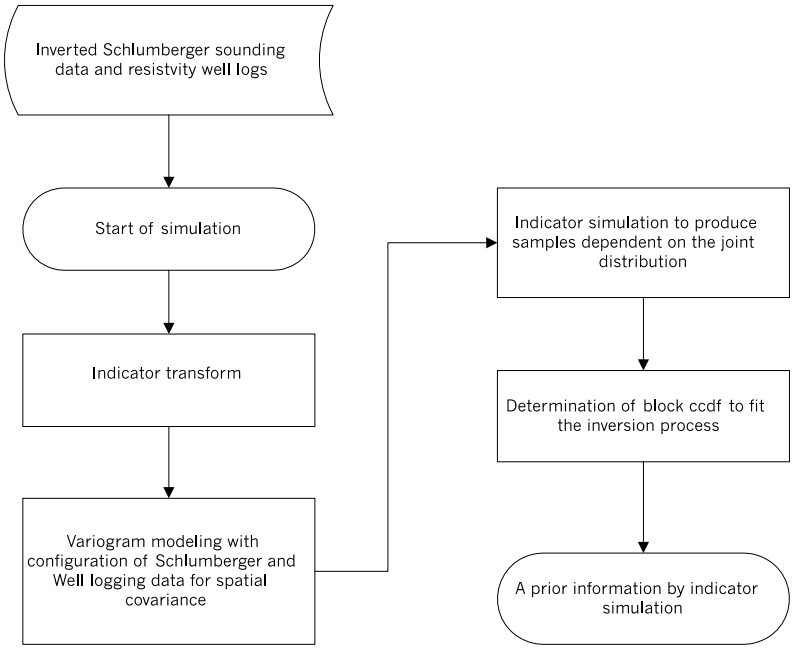

Fig. 2. Flowchart for presentation of *a prior* information by geostatistical simulation from Schlumberger and well logging data.

2) The histogram of simulated values reproduce closely the declustered sample histogram.

3) The set of indicator covariance models $C_I(h; z_k)$ for various threshold z_k are reproduced.

Although there are various geostatistical techniques, only the important characteristics are explained here. In this study, we have applied the sequential indicator simulation method (Deutsch and Journel, 1992) to describe the local and spatial uncertainty structure of resistivity for *a prior* information on the studied region. The sequential process is described in Fig. 2, and its individual implementations are followed in the next sections.

2.4 Inference of uncertainties of blocks

The points, where the simulations estimate, are denser than the number of the parameterized blocks for inversion. This is caused by the fact that we could not infer the uncertainty of the blocks parameterized for inversion. So by assessing the values on denser points, we derive values by following the indirect method.

We have considered the problem of evaluating the block ccdf (conditional cumulative density function) $F_V(\mathrm{u}; z \mid (n))$ that models the uncertainty about an average z-value over the block $V(\mathrm{u})$:

$$F_V(\mathrm{u}; z \mid (n)) = \mathrm{Prob}\{\mathrm{u} \leq z \mid (n)\}.$$

Because of the non-linearity of the indicator transform, the block ccdf cannot be derived simply as a linear combination of point ccdfs:

$$[F_V(\mathrm{u}; z \mid (n))]^* \neq \frac{1}{J}\sum_{j=1}^{J}[F(\mathrm{u}'_j; z \mid (n))]^* \qquad (5)$$

with the point-ccdf $F(\mathrm{u}'_j; z \mid (n))$ being defined at J points u'_j discretizing the block $V(\mathrm{u})$. Here $[F_V(\mathrm{u}; z \mid (n))]^*$ is the inference of ccdf for the block V conditional to n data around the point u. In the absence of block data $z_V(\mathrm{u}_\alpha)$ and corresponding block statistics and block indicator data, the block ccdf (Eq. (5)) can be numerically approximated by the cumulative distribution of many simulated block values $z_V^{(l)}(\mathrm{u})$ (Isaaks, 1990; Gomez-Hernandez, 1991; Deutsch and Journel, 1992; Glacken, 1996):

$$[F_V(\mathrm{u}; z \mid (n))]^* = \frac{1}{L}\sum_{l=1}^{L} i_V^{(l)}(\mathrm{u}; z) \qquad (6)$$

with the block indicator value defined as $i_V^{(l)}(\mathrm{u}; z) = 1$ if $z_V^{(l)}(\mathrm{u}) \leq z$, and zero otherwise. Each simulated block value $z_V^{(l)}$ is obtained by averaging a set of z-values simulated at J points u'_j discretizing the block $V(\mathrm{u})$:

$$z_V^{(l)}(\mathrm{u}) = \frac{1}{J}\sum_{j=1}^{J} z^{(l)}(\mathrm{u}'_j). \qquad (7)$$

3. MCMC Approach to Solutions

Tarantola (1987) showed that the complete solution of geophysical inverse problem is *a posterior* information given by the conjunction of the two states of information, *a prior* and theoretical or likelihood information, and this is an accordant form to the generalized Bayes' theorem. That is,

$$\sigma(\mathrm{m}) = k\rho(\mathrm{m})L(\mathrm{m}) \qquad (8)$$

where, $\sigma(m)$ is the *a posterior* PDF, $\rho(m)$ is the *a prior* PDF, and $L(m)$ is a likelihood PDF. Thus the simple multiplicity of *a prior* PDF and likelihood PDF is the solution to the geophysical inverse problem in the viewpoint of probabilistic approach. But the high-multidimensional PDF itself gives no choice, and when we are determining the most probable model for consulting or infering the characteristics of

the neighbourhood of maximum *a posterior* PDF, solving a high dimensional integral or an optimization problem called marginalization process is required. Unfortunately, the kernel of the integral is very highly nonlinear, so the analytic solution is very rare and we should attack it by numerical analysis. Of the various optimization techniques, derivative-free direct search methods such as Simulated Annealing (SA), Genetic Algorithm (GA), etc. have been known to give a global optimal parameter. Fully detailed tutorials on global optimization techniques may be found in Rothman (1985, 1986) and Sen and Stoffa (1995). This paper approaches the marginalization process by MCMC method based on the Gibbs sampler which utilizes the modified SA algorithm. The *a prior* information generated by geostatistical simulation was used as full conditional distribution in the Gibbs sampler. For MCMC, a rigorous and easy to understand explanation for geophysical applications was given by Grandis *et al.* (1999), so we will review briefly the mathematical process and some pertinent characteristics here.

3.1 MCMC method

3.1.1 Monte Carlo integration In Bayesian inversion, marginalization process by *a posterior* PDF $\pi(.)$ result in as

$$E[f(X)] = \frac{\int f(x)\pi(x)dx}{\int \pi(x)dx} \tag{9}$$

Monte Carlo integration evaluates $E[f(X)]$ by drawing samples $\{X_t, t = 1, \ldots, n\}$ from PDF $\pi(.)$ and then approximating

$$E[f(x)] \approx \frac{1}{n}\sum_{t=1}^{n} f(X_t). \tag{10}$$

So the population mean of $f(X)$ is estimated by a sample mean. When the samples $\{X_t\}$ are independent, laws of large numbers ensure that the approximation can be made as accurate as desired by increasing the sample size n. In general, drawing samples $\{X_t\}$ independently from $\pi(.)$ is not feasible, since $\pi(.)$ can be quite non-standard. However the $\{X_t\}$ need not necessarily be independent. The $\{X_t\}$ can be generated by any process which, loosely speaking, draws samples throughout the support of $\pi(.)$ in the correct proportions. One way of doing this is through a Markov chain having $\pi(.)$ as its stationary distributions. This is then Markov chain Monte Carlo.

3.1.2 Markov chains Suppose we generate a sequence of random variables, $\{X_0, X_1, X_2, \ldots\}$, such that at each time $t \geq 0$, the next state X_{t+1} is sampled from a distribution $P(X_{t+1} \mid X_t)$ which depends only on the current state of chain, X_t. That is, given X_t, the next state X_{t+1} does not depend further on the history of the chain $\{X_0, X_1, X_2, \ldots, X_{t-1}\}$. This sequence is called a Markov chain, and $P(. \mid .)$ is called the transition kernel of the chain (Gilks *et al.*, 1996). We will assume that the chain is time homogeneous: that is, $P(. \mid .)$ does not depend on t.

How does the starting state X_0 affect X_t? This is a very solicitous subject to researchers in geophysical inverse problems. Because the initial guess is always artificial and subjective, they want the final solution to have no relation with the initial one. This question concerns the distribution of X_t given X_0, which we denote $P^{(t)}(X_t \mid X_0)$. Here we are not given the intervening variables $\{X_1, X_2, \ldots, X_{t-1}\}$, so X_t depends directly on X_0. Subject to regularity conditions (Roberts, 1995), the chain will gradually forget its initial state and $P^{(t)}(. \mid X_0)$ will eventually converge to a unique stationary (or invariant) distribution, which does not depend on t or X_0. For the moment, we will denote the stationary distribution by $\phi(.)$. Thus as t increases, the sampled points $\{X_t\}$ will look increasingly like dependent samples from $\phi(.)$. Thus, after a sufficiently long burn-in of say m iterations, points $\{X_t; t = m + 1, \ldots, n\}$ will be dependent samples approximately from $\phi(.)$. We will discuss methods for determining m, the burn-in number, in the Subsection 4.2.3. Here we should explain that for the Bayesian process, the stationary distribution in a certain step is the best one that converges, that is to say, where no more information is available in the current step. When other information is incorporated, the stationary distribution is certainly different from the old one.

We can now use the output from the Markov chain to estimate the expectation $E[f(X)]$, where X has distribution $\phi(.)$. Burn-in samples are usually discarded for this calculation, giving an estimator

$$\bar{f} = \frac{1}{n-m}\sum_{t=m+1}^{n} f(X_t). \tag{11}$$

This is called an *ergodic average*.

3.1.3 The Metropolis-Hastings algorithm Then how could we construct a Markov chain such that its stationary distribution $\phi(.)$ is precisely our distribution of interest $\pi(.)$? For the Metropolis-Hastings algorithm (Hastings, 1970), at each time t, the next state X_{t+1} is chosen by first sampling a candidate point Y from a disposal distribution $q(. \mid X_t)$. Note that the proposal distribution may depend on the current point X_t. For example, $q(. \mid X_t)$ might be a multivariate normal distribution with mean X and a fixed covariance matrix. The candidate point Y is then accepted with probability $\alpha(X_T, Y)$, where,

$$\alpha(X, Y) = \min\left(1, \frac{\pi(Y)q(X \mid Y)}{\pi(X)q(Y \mid X)}\right). \tag{12}$$

If the candidate is rejected, the chain does not move, i.e. $X_{t+1} = X_t$. Remarkably, the proposal distribution $q(. \mid .)$ can have any form and the stationary distribution of the chain will be $\pi(.)$, and this can be shown from the detailed balance equation (Roberts, 1995; Tierney, 1995).

3.1.4 Single-component Metropolis-Hastings Instead of updating the whole of X in a lump, it is often more convenient and computationally efficient to divide X into components $\{X_{.1}, X_{.2}, \ldots, X_{.h}\}$ of possibly differing dimension, and then update these components one by one. This was the framework for MCMC originally proposed by Metropolis *et al.* (1953), and it is refered to as single-component Metropolis-Hastings. Let $X_{.-i} = \{X_{.1}, \ldots, X_{.i-1}, X_{.i+1}, \ldots, X_{.h}\}$, so $X_{.-i}$ comprises all of X except $X_{.i}$ (Gilks *et al.*, 1996).

An iteration of the single-component Metropolis-Hastings algorithm comprises h updating steps, as follows. Let $X_{t.i}$ denote the state of $X_{.i}$ at the end of iteration t.

For the ith step of iteration $t+1$, $X_{.1}$ is updated using Metropolis-Hastings. The candidate $Y_{.i}$ is generated from a proposal distribution $q_i(Y_{.i} \mid X_{t.i}, X_{t.-i})$, where $X_{t.-i}$ denotes the value of $X_{.-i}$ after completing the step $i-1$ of iteration $t+1$:

$$X_{t.-i} = \{X_{t+1.1}, \ldots, X_{t+1.i-1}, X_{t.i+1}, \ldots, X_{t.h}\},$$

where components $1, 2, \ldots, i-1$ have already been updated. Thus the ith proposal distribution $q_i(. \mid ., .)$ generates a candidate only for the ith component of X, and may depend on the current values of any of the components of X. The candidate is accepted with probability $\alpha(X_{t.-i}, X_{t.i}, Y_{.i})$ where

$$\alpha(X_{.-i}, X_{.i}, Y_{.i}) = \min\left(1, \frac{\pi(Y_{.i} \mid X_{.-i}) q_i(X_{.i} \mid Y_{.i}, X_{.-i})}{\pi(X_{.i} \mid X_{.-i}) q_i(Y_{.i} \mid X_{.i}, X_{.-i})}\right). \quad (13)$$

Here $\pi(X_{.i}, X_{.-i})$ is the full conditional distribution for $X_{.i}$ under $\pi(.)$. If $Y_{.i}$ is accepted, we set $X_{t+1.i} = Y_{.i}$; otherwise, we set $X_{t+1.i} = X_{t.i}$. The remaining components are not changed at step i.

Thus each updating step produces a move in the direction of a coordinate axis (if the candidate is accepted), as illustrated in Fig. 3. The proposal distribution $(q_i(. \mid ., .)$ can be chosen in any of the ways discussed earlier.

The full conditional distribution $\pi(X_{.i} \mid X_{.-i})$ is the distribution of the ith component of X conditioning on all the remaining components, where X has distribution $\pi(.)$:

$$\pi(X_{.i} \mid X_{.-i}) = \frac{\pi(X)}{\int \pi(X) dX_i}. \quad (14)$$

Full conditional distributions play a prominent role in many of the applications. That the single-component Metropolis-Hastings algorithm with acceptance probability given by Eq. (13) does indeed generate samples from the target distribution $\pi(.)$ results from the fact that $\pi(.)$ is uniquely determined by the set of its full conditional distributions (Besag, 1974).

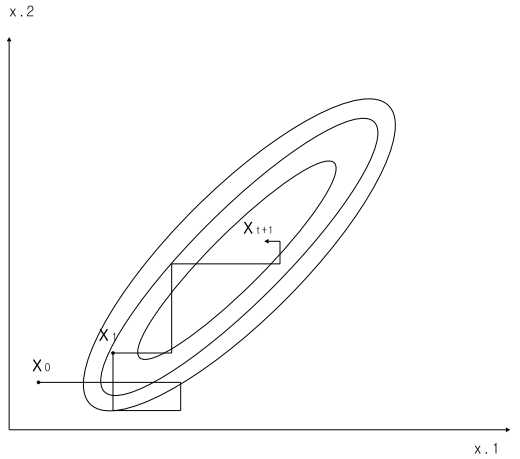

Fig. 3. Illustrating a single-component Metropolis-Hastings algorithm for a bivariate target distribution $\pi(.)$. Components 1 and 2 are updated alternately, producing alternate moves in horizontal and vertical directions (After Gilks *et al.*, 1996).

3.2 The Gibbs sampler

A special case of single-component Metropolis-Hastings is the Gibbs sampler. The Gibbs sampler was given its name by Geman and Geman (1984), who used it for analysing Gibbs distributions on lattices. The same method had been used already in statistical physics as the heat bath algorithm. Neverthless, the work of Geman and Geman (1984) led to the introduction of MCMC into the mainstream statistics via articles by Gelfand and Smith (1990) and Gelfand *et al.* (1990). To date, most statistical applications of MCMC have used Gibbs sampling.

For the Gibbs sampler, the proposal distribution for updating the ith component of X is

$$q_i(Y_{.i} \mid X_{.i}, X_{.-i}) = \pi(Y_{.i} \mid X_{.-i}) \quad (15)$$

where $\pi(Y_{.i} \mid X_{.-i})$ is the full conditional distribution Eq. (14). Substituting Eq. (15) into Eq. (13) gives an acceptance probability of 1; that is, Gibbs sampler candidates are always accepted. Thus Gibbs sampling consists purely of sampling from full conditional distributions.

In geophysical societies, Rothman (1986) used the Gibbs sampler as a nonlinear static correction algorithm. He recalculated some decade of the discrete probability on each selected parameter to be used as full conditional distribution in every stage, then picked one model to fit the acceptance criterion. However, that computational cost is high, and because he sampled the test model parameters from the unique probability density function, his method artificially discretized the model parameter space and therefore is not adaptive. Finally, his technique was totally dependent upon the data-fit alone minima, so the global search of the model space seemed to be difficult.

Recently, Grandis *et al.* (1999) clearly defined the transition distribution of the Markov chain with digitized *a prior* samples and Markovian matrix, and it gave satisfied solutions. However, our concerned problem is two dimensional, so that such a digitizing method is difficult to apply. For the purpose of including independent geophysical data, we have devised another approach to the sampling.

3.3 Sampling algorithm

To overcome the problems of the previous section, we approached the sampler with the *a prior* information. Actually, the full conditioning distribution means to select the most probable parameter conditional on remaining parameters. Therefore when we have a marginalized *a prior* PDF, we can apply it to the full conditional distribution. That is, the *a prior* cumulative density function (CDF) is created by the *a prior* PDF, then candidate sampling is conducted by the inverse CDF. Then, a component which updates the likelihood function is accepted and if not, the component is accepted with the probability,

$$P_{\text{accept}} = \exp\left(-\frac{(E_{\text{attempted}} - E_{\text{current}})}{T}\right) \quad (16)$$

where, E is an objective function generally given by discrepancy between calculated and observed data and T is a tuning parameter referred to as temperature in Simulated Annealing (Sen and Stoffa, 1995). The schedule of T followed the fast simulated annealing algorithm (Szu and Heartley, 1987).

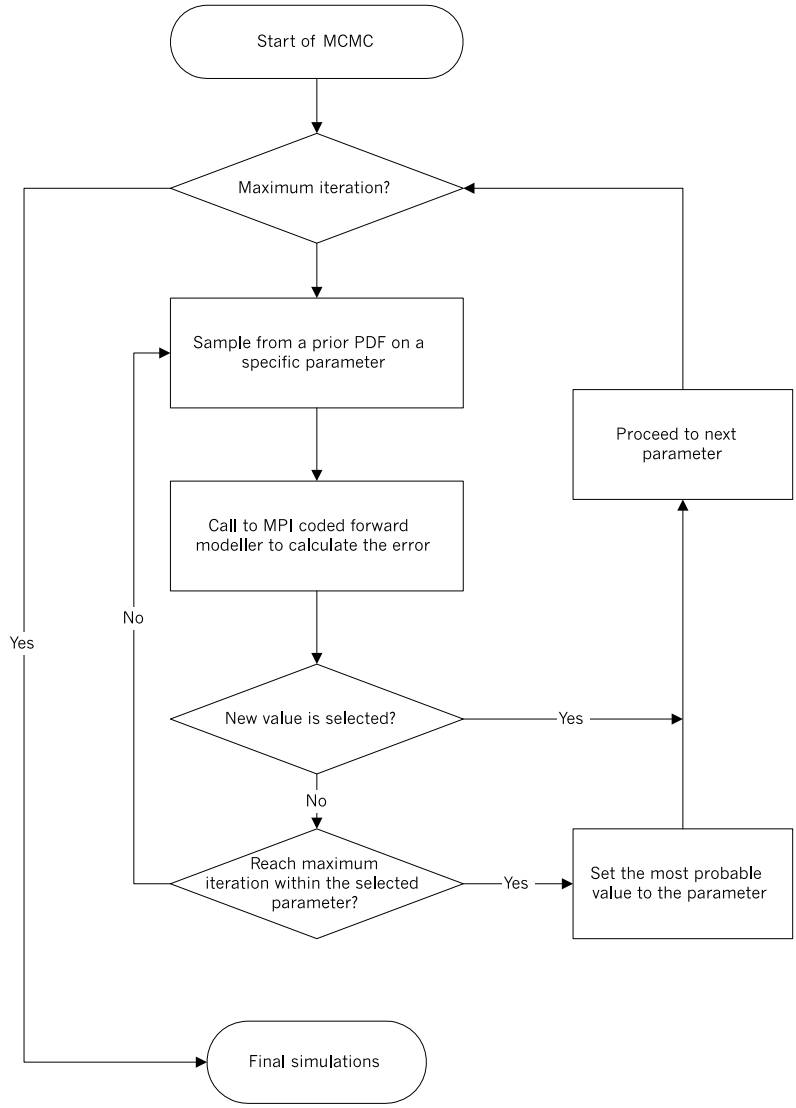

Fig. 4. Flowchart for the MCMC method by Gibbs sampling applied to perform the analysis of *a posterior* information.

And if the component is rejected, sampling is again conducted on the same CDF until the maximum number of allowances is reached.

This approach effectively samples the neighbourhood of the maximum *a posterior* PDFs which fit the *a prior* and *a posterior* PDF after burn-in trials. Figure 4 shows the flowchart for this algorithm.

Actually this approach was appropriate and adequate for our *a prior* information. That is to say, for the 2-D structure, numerous geostatistical simulations provide a tool to incorporate independent geophysical data and make objective *a prior* information. And *a prior* information is easily marginalized, so that the inverse CDF sampling is amenable. Also, random field models given by geostatistical methods are similar, or even identical, to those studied in statistical physics: this field was the original home ground for MCMC methods (Metropolis *et al.*, 1953) and physical analogies are drawn strongly in some works in the statistical analysis (Geman and Geman, 1984).

Unlike the Rothman's heat bath algorithm (1986), ours prepares the model perturbation in advance and updates the component which fits the *a prior* and *a posterior* PDFs. This makes the process remarkably effective especially when the *a prior* information reflect the likelihood functions well, though in the worst case searching the global model space should be guaranteed.

Actually no direct comparison is possibile to test convergence of the Rothman's algorithm, because his traditional heat bath alogorithm composes the transition rule after dividing the model parameter into arbitrary finite parameter sets. This means that more parameter division and sampling makes it easy to constitute the full conditional distribution, and various other Gibbs sampling studies may be identified. But the important distinction is the sampling process and state of the final parameters. The most difficult aspect of global optimization is the more the number of parameters, the more disastrous the state of parameters. This has various causes, but the main reason is insufficient model parameter search in the condition of imperfect and deficient information. Rothman's algorithm using the data-fit alone match and unique *a prior* information cannot but search restricted parameters, so that it is not adequate to the highly multidi-

Table 1. Comparison between the Rothman's algorithm and the method proposed in this study.

Comparitive index	Rothman's algorithm	The point at issue	Algorithm of this study
A prior information	unique distribution	null information	given by indepedent data
Sampling	predefined model space	non-adaptive and artificial	adaptive and effective sampling near *a prior* and *a posterior* maximum
Computational effectiveness	recalculation in each stage	high costs	predefined *a prior* information
Model search	observed data-alone model search	constrained model space	observed and independent model space

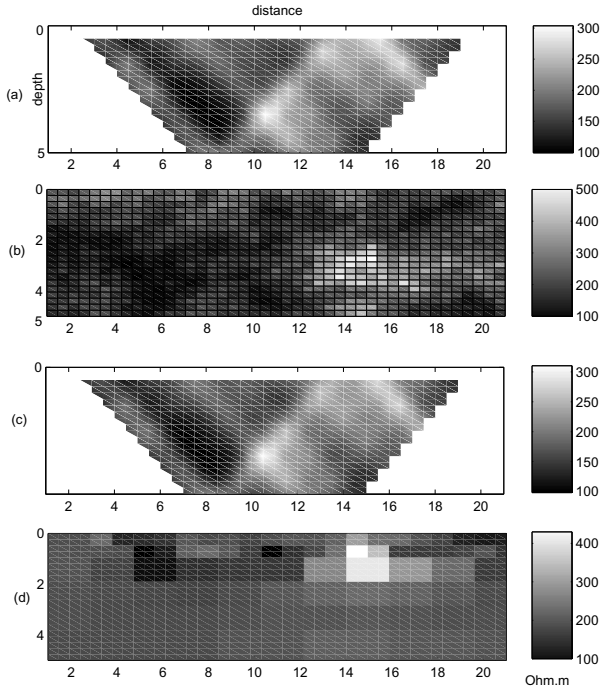

Fig. 5. Proposed model for this study generated by geostatistical simulation. (a) Observed dipole-dipole resistivity data contaminated by 10% Gaussian noise for the forward response of model (b). (c) Calculated response from the inverted model (d) by observed data-fit alone. The depth and distance are arbitrary model units.

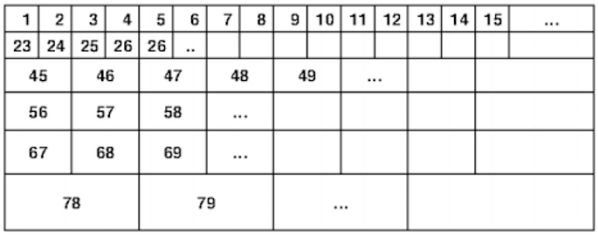

Fig. 6. Block parameterization prepared for resistivity inversion. Block numbers are displayed for following interpretations.

mensional inversion problems and it results in an absurd solution. On the other hand, we propose independent *a prior* information sustained by other geophysical data, resulting in a more effective and global model search. Moreover, our frame can easily incorporate various information acquired in the future, as we prepared *a prior* information in previous sections. In Table 1 we summarize the differences between and characteristics of the two approaches.

4. Application to Numerical Model

4.1 A model presented by geostatistical simulation

Figure 5(b) shows the resistivity model prepared to examine robustness of our algorithm. Here, we assumed that the resistivity distribution of the underground is actually continuous and has much smaller scale variation than the generally used inversion blocks such as in Figs. 5(d) and 6. First, we allocated simple blocks with some high and low resistivities, then applied geostatistical simulation to get a more plausible model reflecting the real geology and to have finer scale block variation. Based on these kind of characteristics, we may expect that the prosed model can be hardly recovered by any inversion methods using blocks of Fig. 6. In this paper, all of the resistivity model is displayed as linear scale, and this is intended to show delicate variations of the property, mainly ranged in 100 to 500 Ohm-m.

The dipole-dipole resistivity data of the above model were obtained at 135 measured points by 21 current poles, 10 layers (n) with the electrode spacing of 1 unit, and 10% Gaussian noise was added (Fig. 5(a)).

Figure 5(d) is the inversion result given by the traditional data-fit alone least-square (Lines and Treitel, 1984) method. Although the calculated resistivity pseudosection of the Fig. 5(c) from this inverted model is quite similar to the observed data of Fig. 5(a), the inversion result has almost nothing in common with the true resistivity section of Fig. 5(b). This unrealistic result might be due to the insufficient information of the field data, quite different setting of inversion blocks and the inherent non-uniqueness, and the misfit, in general, tends to be proportional to the dimensions of the data and model structure. However, this is a rather common situation actually occurred in many conventional geophysical explorations. In this study, we are going to deal with such a more complicated but realistic model to accentuate how well the uncertainties are diminished in the Bayesian framework to improve the inversion result.

Figure 7 is the configuration map where Schlumberger sounding and resistivity well logging data are modeled. Sounding data were inverted to present true resistivities by a hybrid of a global optimization technique SA and a local least-square method (Chunduru *et al.*, 1996) (Fig. 8). Re-

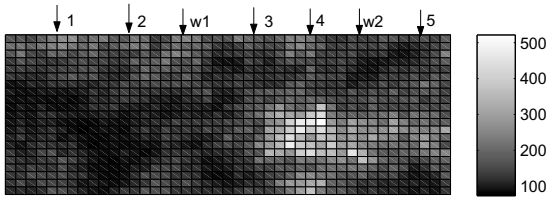

Fig. 7. Configuration map where the well logs and Schlumberger sounding simulation is performed. The prefix w means the position of well logging simulations and the others are Schlumberger sounding positions.

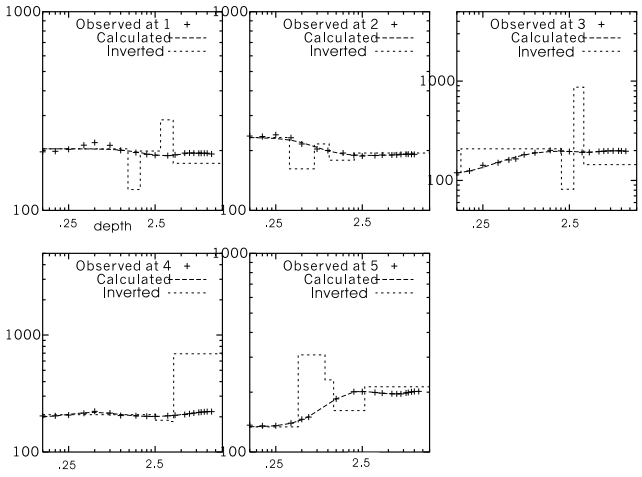

Fig. 8. Schlumberger sounding inversion simulated in Fig. 7. SA and local least-square hybrid is applied.

sistivity well logs (Fig. 9) were simulated by the five layer running average procedure, and 5% Gaussian noise, which is exponentially correlated, was added.

4.2 Implementation

4.2.1 ***A prior*** **information** We have deduced the *a prior* information from Schlumberger and well logging data by geostatistical simulation of which schematic process is shown in Fig. 10. Indicator covariogram modelings were performed for each threshold with resistivity sounding and well logging data, then on unsampled points, where the density of estimation is finer than the parameterized blocks, sequential indicator joint simulation was conducted. Many realizations were provided for assessing the uncertainty of the studied region. As described in Subsection 2.4, the simulated estimations now are used in the block CDF making process.

Because the resolution of each data is different, careful variogram modeling is required, but indicator transforming makes the steps more stable by data-classified level.

Figure 11 shows the *a prior* information given by the PDFs of the upper blocks of the parameterized region. The block numbers in Fig. 10 are defined in Fig. 6. As expected, the blocks far from the sounding and logging data had a wide-spread PDF structure that means spatially large uncertainty, and vice versa. So we may conclude that this approach reflects well the spatial information given by partial geologic and geophysical data.

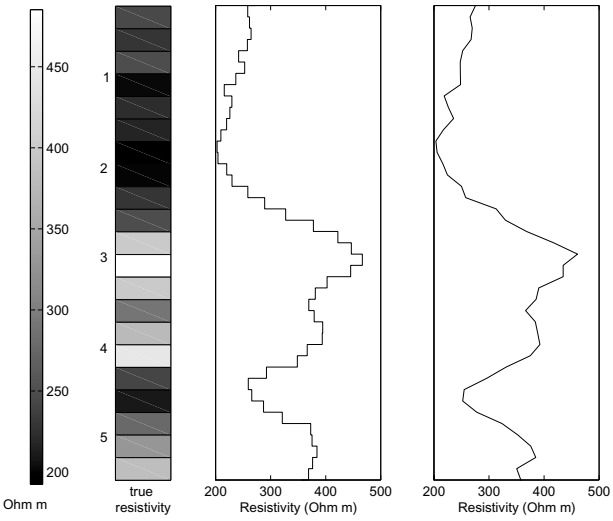

Fig. 9. Simulated resistivity well logging data at position w2 in Fig. 7. The left stripe shows the true resistivity structure and the log in the middle is generated by five-point averaging with the resistivity values read at every 0.125 unit. The well log on the right is simulated by adding correlated Gaussian noise.

4.2.2 Sampling from *a prior* information To sample a candidate from the *a prior* information for each parameterized block, cumulative density functions for the *a prior* PDF were generated and interpolated between the discrete points. Here, the more sample points which decide the degree of discretization are selected, the better the inverse sampling from the CDF is performed. However, the maximum level of discretization is constrained by the number of the conditional data and in the indicator case, the order relaxation (Isaaks and Srivastava, 1989) degrades the information for excessive discretization. We performed 500 simulations and dicretized the PDF by 20 thresholds, then random numbers between 0 and 1 were used to get sampled values.

4.2.3 Decision of burn-in number Figure 12 shows the variation of objective function given by the discrepancy between calculated and observed data.

Error was calculated by

$$\mathrm{err} = \sqrt{\frac{\sum_{i=1}^{N} (\rho_{\mathrm{obs}} - \rho_{\mathrm{cal}})^2}{N}}$$

where N is the total number of observations. We decided the burn-in trials, m, as the number making the error curve converge, about 2500.

But as we can see from Fig. 12, the Monte Carlo approach takes a heavy computational cost, so we coded the MPI (Message Passing Interface) parallelized DC forward program. Two dimensional DC problems were naturally parallelized, i.e., potentials due to each source poles may be calculated on each CPUs, so we adopted parallelized forward modeling code by MPI (Pacheco, 1997; MPI Forum, 1995). Parallel programming with MPI is very easy and portable to other systems, and heterogeneous systems may be chained via TCP/IP network, so that is very cheap.

4.2.4 Marginal distribution of *a posterior* information Figures 13(a), (b) present comparisons of *a prior* and *a posterior* PDFs produced by the above Gibbs sampling

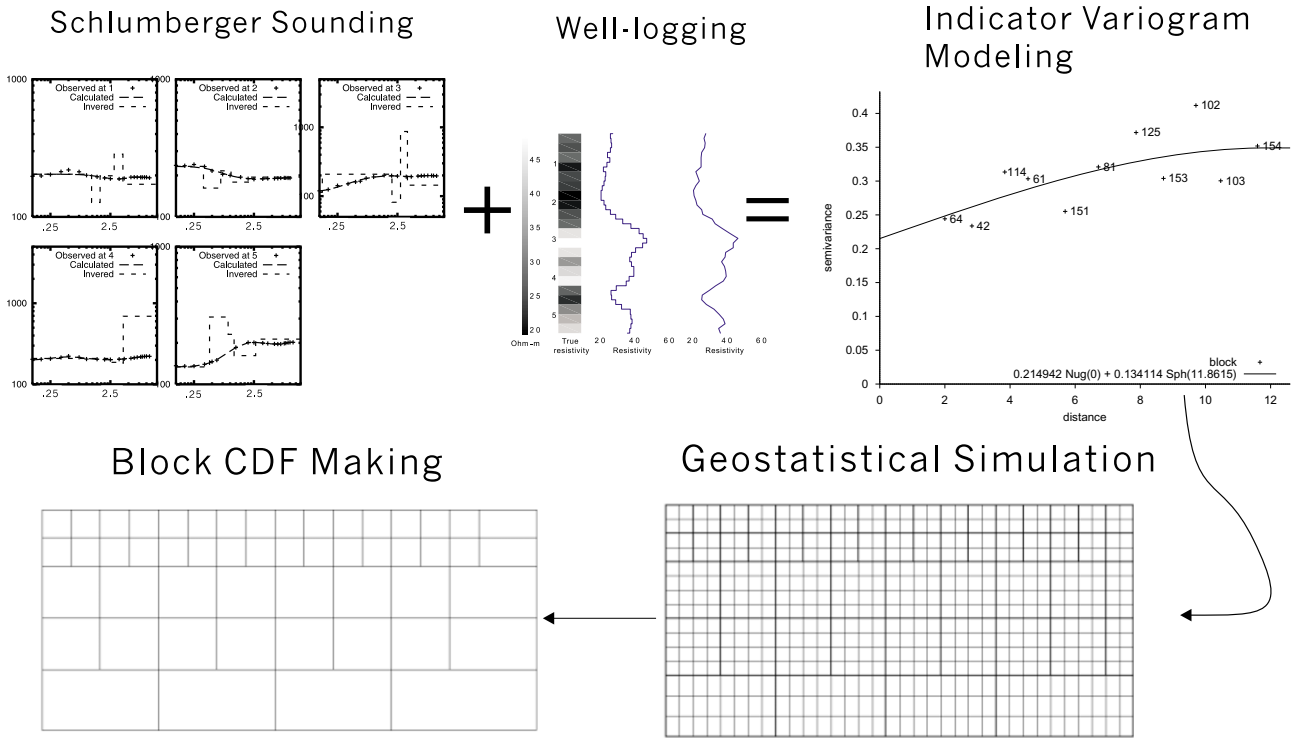

Fig. 10. Schematic process to generate *a prior* information. Schlumberger sounding and resistivity well logging data, which appear as larger picture in Figs. 8, 9, are put in covariogram modeling for a specified threshold for indicator kriging or simulation. Variogram model equation appeared in lower-right of the indicator model figure shows the data are rough from the high nugget value (Deutsch and Journel, 1992) but they have correlation and sustain a tendency. The distance is model unit. Then denser estimations than the number of blocks contribute to the block CDF making, because it is impossible to estimate the resistivity of each block directly. So indirect estimation for blocks are made by the method proposed in Subsection 2.4.

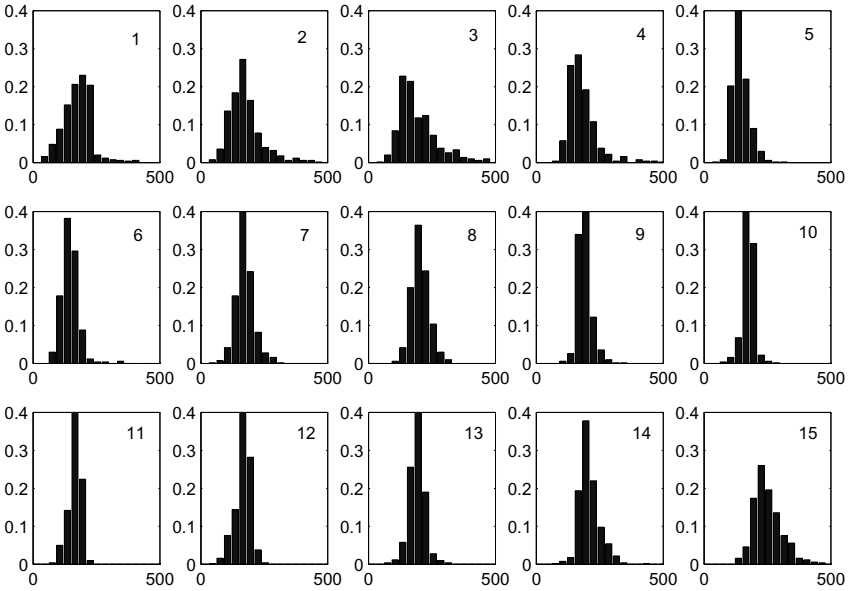

Fig. 11. *A prior* information generated by geostatistical simulation from Schlumberger and well logs data. The block numbers are defined in Fig. 6. The horizontal axes mean resistivity ($\Omega \cdot$ m), and the vertical axes present the probability.

MCMC algorithm for the upper 15 blocks. The core of the Bayesian inference is based on the point of how we can reduce the uncertainty in *a posterior* PDF as well as the error of discrepancy, because it is important to reduce the uncertainty in observed data by *a prior* information. The update of the upper blocks shown in Fig. 13 is remarkable in terms of reduction of uncertainty. But deeper blocks do not have significant reduction of uncertainty. These results suggest

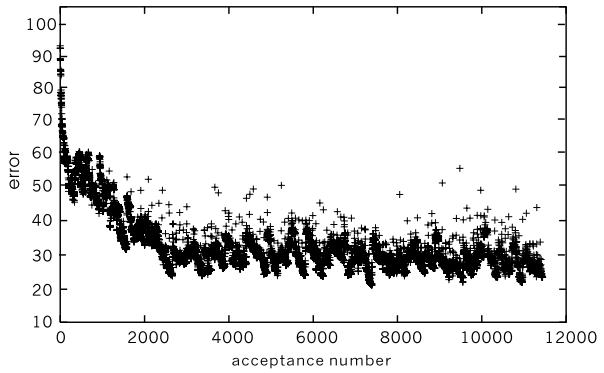

Fig. 12. Evolution of data-fit error as the function of the number of acceptance for the proposed model in Fig. 5.

that blocks having low sensitivity from observed data tend to converge to *a prior* information. But we can find that the convergence is stable, so we need not have any constraints for the stable convergence of inversion.

From these marginal distributions, various informations may be available, such as the confidence interval, mean of resistivity, etc., which were mentioned in numerous previous studies without any explanations of how to get the distribution. And *a posterior* covariance and resolution matrix may be also obtained by indirect sampling of *a posterior* PDF.

4.3 Marginal distribution analysis

Actually, the marginal distributions given in Figs. 13, 14 are the temporary solutions for the Bayesian inversion dependent on current information. That is, if we get more information for this model or data than proposed in this study, the result may be updated to include post geological or geophysical information. And the displays of only PDFs are not very informative and hard to understand. Therefore, in 1-D case, a direct overlap of PDF for true values or data is used (Grandis *et al.*, 1999). Unfortunately, in 2-D case, such displays for *a posterior* marginal PDF is impossible, so we present two kinds of displays to figure out the simple characteristics of *a posterior* marginal PDF. Two models are given by maximum values and means of each marginal PDF in Fig. 15.

Here, we should mention some reasons that the proposed *a posterior* solutions were imperfectly matched with the true section, although it conspicuously improved the data-alone inversion result in Fig. 5(d). There are couple of reasons that the recovered model is somewhat different from the true one. First, as we showed in Fig. 6, the block parameterization for inversion is much coarser than that of the true resistivity model structure (Fig. 5(d)) for the various reasons appearing in Subsection 4.1. Of course, when we chose a perfectly matched block parameterized model, a nearly complete solution to the true model was obtained excepting the parts of flanks where dipole-dipole DC data are incomplete. Although many papers have dealt with such a case, this is rather artificial and impractical in the real world. Therefore, we prepared a more realistic resistivity earth model showing more continuous and smaller scale variation than for the inversion blocks of Fig. 6. As we can see in the Fig. 5(c), the inverted resistivity section by the conventional process hardly recovers the true model. Although our result also appears to be somewhat different to the true one, it may be easily recognized to be much improved and one can delineate the general trend of resistivity variation in consideration of the intrinsic characteristics of dipole-dipole resistivity data. On the other hand, this kind of discrepancy leads to discussions on uncertainty analysis which is important to understanding of the characteristics of the probabilistic method.

As we can imagine, if we present only the observed-data supported in part by cutting off the lower parts of edges of the inverted section, such as those identified with the pseudosection (apparent section), the true section, once concerning the coarseness of the inverted blocks, is reproduced relatively well. The large discrepancy appearing on the left flank in comparison to the right one is mainly due to the lack of *a prior* information. Recognized from Fig. 7, the well data obtained at the location (w1) do not provide sufficient information on the low resistivity anomalies in the left flank. It is noted that, although there are Schlumberger data nearby, they are less reliable due to indirect characteristics as compared to well data. This means that, as for the part where well-behaved observed data is rare, *a prior* information is dominant as with the Bayesian, and the uncertainty analysis provides interpreters with such information.

Secondly, we think the Bayesian frame is an ongoing process of past, present and future states for the concerned region. As in the saying, 'Today's posterior becomes tomorrow's prior', the solution in the present state is one of solution steps, if, of course, the current information is fully extracted from a prior information and observed data. The information for the mesh structure may also be included in this category. In order to explain this aspect, we discussed the uncertainty analysis in Subsection 4.4 especially for the discrepant region of the lower left flank. Finally, as we explained in the previous section, presentation of probabilistic results for the 2-D problem is not easy. Actually, PDFs provide the user with more rigorous and informative solutions such as uncertainty analysis and flexible interpretation, but, it is hard to propose a deterministic section. So we present here a simple map with maximum and mean of PDF for each parameter. We think, of course, more studies should be made for the well-defined and reasonable presentation of such probabilistic results. Therefore, Fig. 15 should be thought of as the simple aspect result of this study, and a temporary map for the current information status of the region being interpreted.

4.4 Uncertainty ananlysis

In Figs. 15(a), (b), we found that the left high-resistivity zone of the inverted section is considerably different from that of the true resistivity section of Fig. 15(c). We might also recognize that the right high-anomalous zone has a similar appearance to that of left zone, but it represents the true resistivity aspect rather well. Then how can we discriminate and analyze this situation? We might easily deduce that both zones of the section are short of data due to the intrinsic characteristics of the dipole-dipole DC survey, resulting in the uncertain solution. Therefore we should focus on these zones for uncertainty analysis.

Figure 16 shows the resistivity section and the CDF maps for anomalous regions which were matched to the bordered

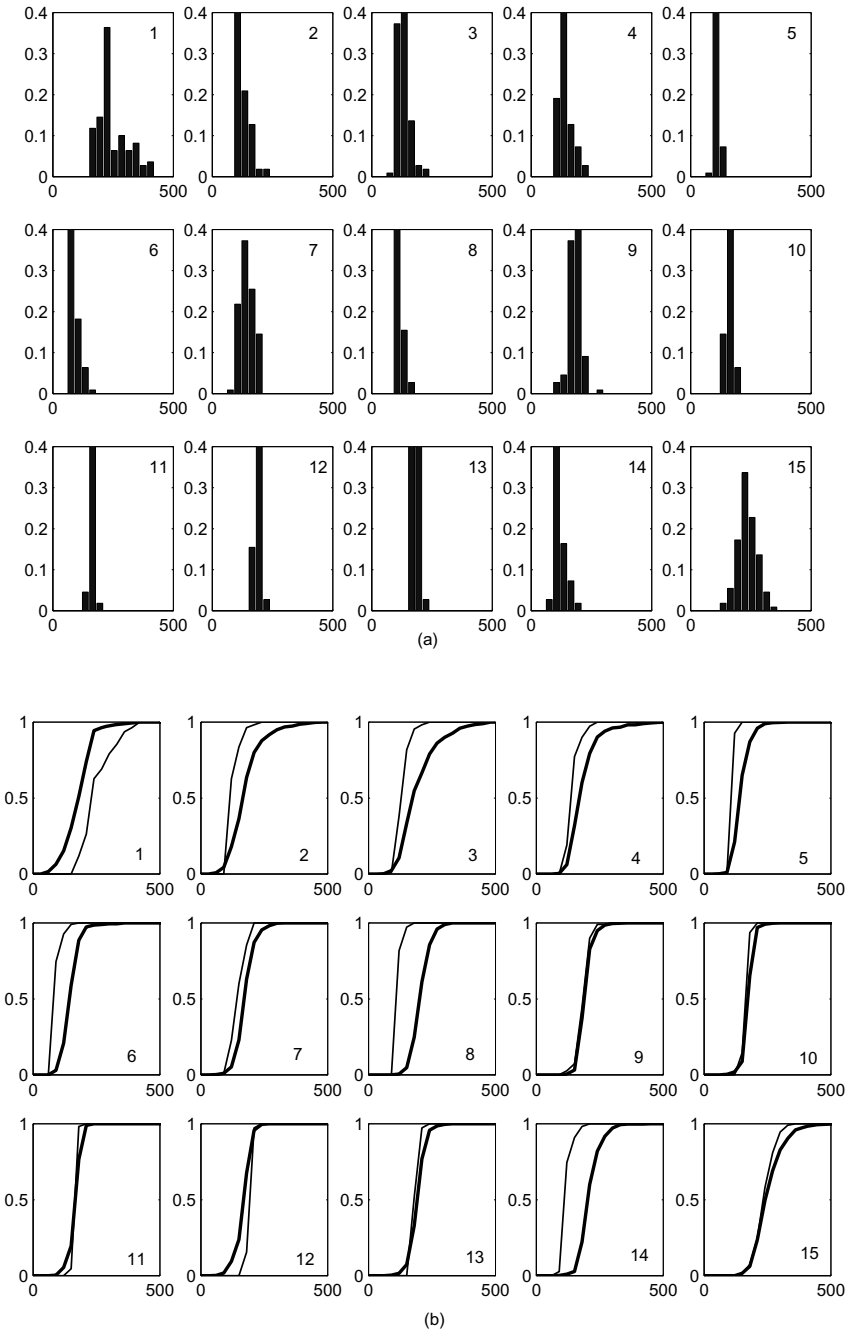

Fig. 13. (a) *A posterior* information given by probability density functions. The horizontal axes mean resistivity ($\Omega \cdot$ m), and the vertical axes present the probability. (b) Cumulative Density functions for *a prior* information (bold line) and *a posterior* information (thin line) for upper blocks. The *a posterior* CDF is concentrated on specific region representing low uncertainties. *A prior* PDFs appear as the bold line and *a posterior* PDFs as the thin line in each CDFs. The block numbers are defined in Fig. 6. The vertical axes present CDF.

blocks of the resistivity sections. The blocks showing remarkably different resistivity values from the true ones represent highly uncertain CDFs; so we may deduce the confidency of the anomalous blocks. In particular, the second column of the CDFs shows a wide range of variation, meaning large uncertainty, and reflects the unstable solution for the region. On the other hand, as we can see from Fig. 17, the right side of the section reflects relatively well the true high resistivity, except the far right column of the bordered blocks in which CDFs show high uncertainty.

Thus CDF maps tend to be well behaved in the matched regions, but show a wide range of uncertainty otherwise.

These properties fully support the probabilistic merits for the geophysical inversion problem. Not by simple deterministic results, but by informative and general PDF, we can examine the state of the interpreted region and make decisions about more explorations or opinions.

5. Discussion and Conclusion

Many stable but plausible and resolute inversion techniques have been proposed to solve various geophysical problems. Among those, the probabilistic approach has been

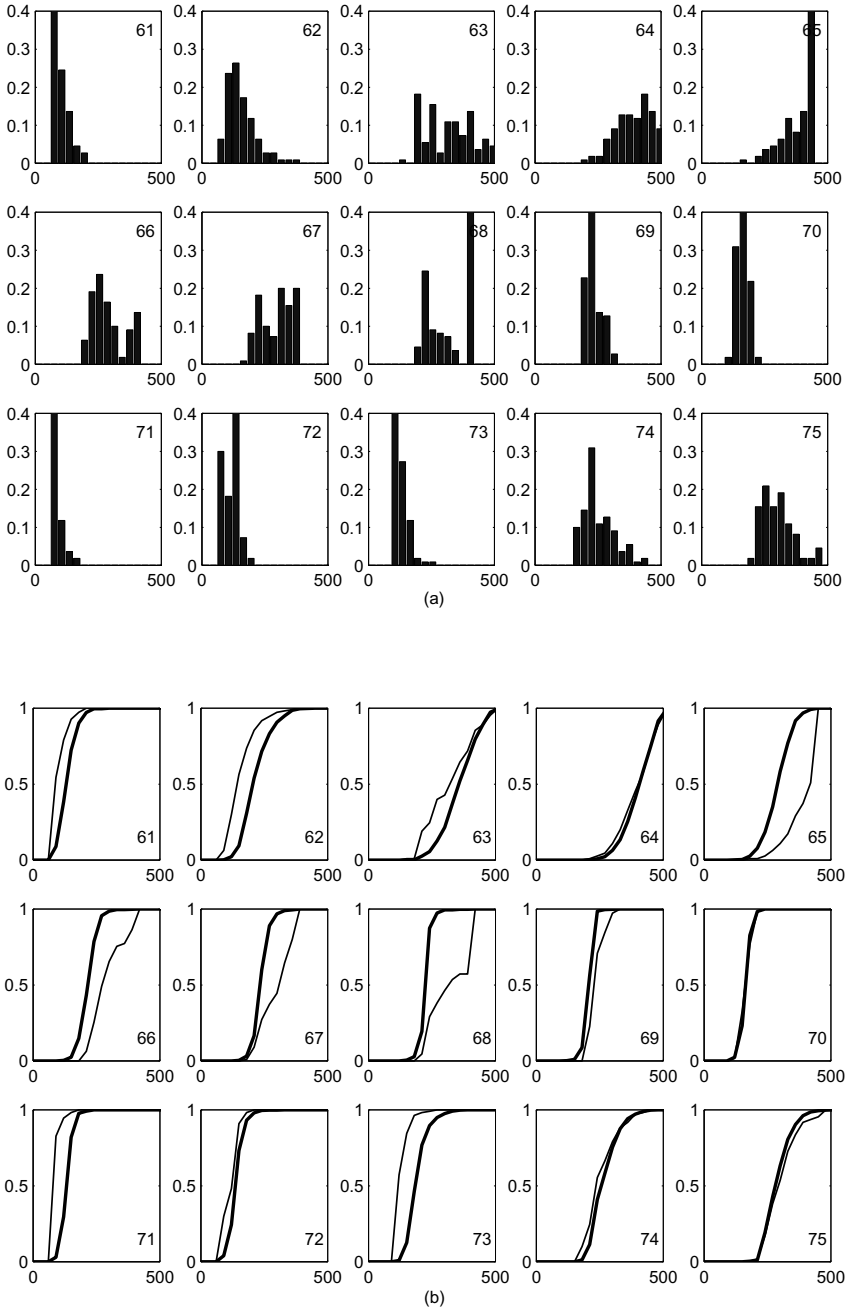

Fig. 14. (a) *A posterior* information given by probability density functions. The horizontal axes mean resistivity ($\Omega \cdot$ m), and the vertical axes present the probability. (b) Cumulative Density functions for *a prior* information (bold line) and *a posterior* information (thin line) for lower blocks. The *a posterior* CDF does not show drastic difference in uncertainties that means low information in dipole-dipole observed data at lower blocks. *A prior* PDFs appear as the bold line and *a posterior* PDFs as the thin line in each CDFs. The block numbers are defined in Fig. 6. The vertical axes present CDF.

considered to have a more generalized robust information operation frame. Especially in terms of incorporation of *a prior* information, the Bayesian inversion has been proved to give more informative and robust results (Duijndam, 1988a, b; Mosegaard and Tarantola, 1995; Loredo, 1990). But as Laplace was troubled to realize the preliminary probability distribution (Moraes, 1996), continuing disputes have been generated between Bayesians and frequentists (Scales, 1998). Finally, in the 50's, Jaynes (1994) proposed the epoch-making system to produce *a prior* information with the maximum entropy method which was based on the communication theory. But implementation of this system is very difficult, so various modified trials were proposed (Gelman *et al.*, 1995).

Thus objective and practical implementation of *a prior* information is a very important element in reaching a successful result by the Bayesian frame, and finally for geophysical inversion. So we, first, have presented a mechanism to produce *a prior* information by geostatistical simulations and kriging. It enables us to achieve various uncertainty analysis and give full information by a non-parametric approach. Also, its great ability to incorporate many informations from

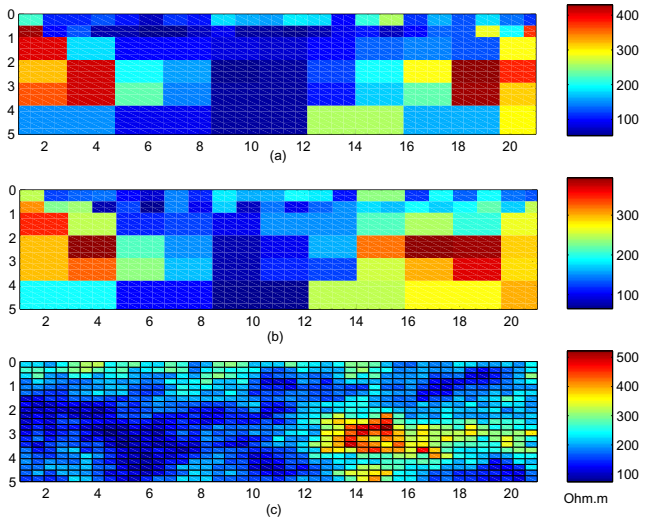

Fig. 15. Some representation from marginal PDF of *a posterior* PDF. (a) Maximum PDF resistivity section. (b) Mean resistivity section of marginal *a posterior* PDF. (c) True resistivity section plotted to be compared.

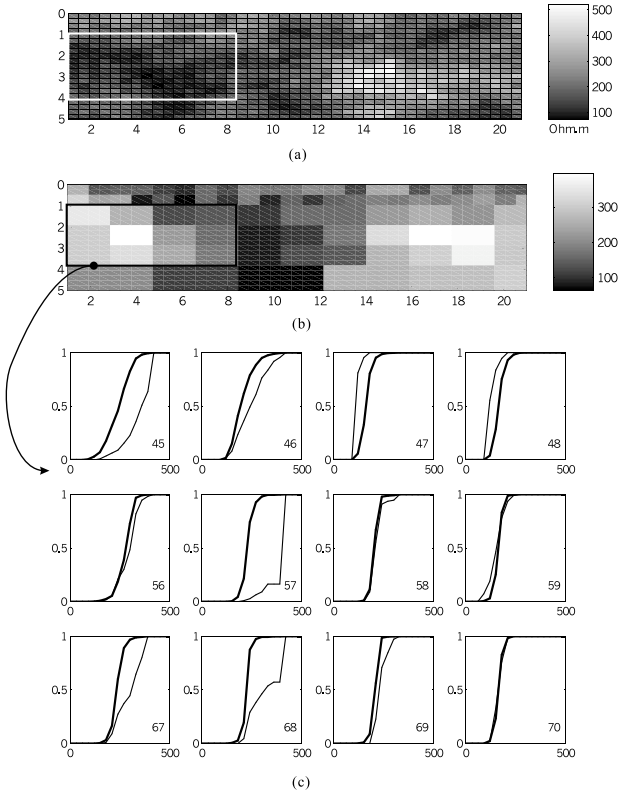

Fig. 16. CDFs for some inverted blocks. (a) True resistivity section proposed for model test. (b) Mean resistivity section of marginal *a posterior* PDF. (c) Highlighted resistivity region is matched to CDF maps to give uncertainty analysis. *A prior* PDFs are appeared as the bold line and *a posterior* PDFs are by the thin line in each CDFs. Remarkably deviated blocks (57, 68) from the true values show very uncertain aspect compared with other blocks which converge to plausible values.

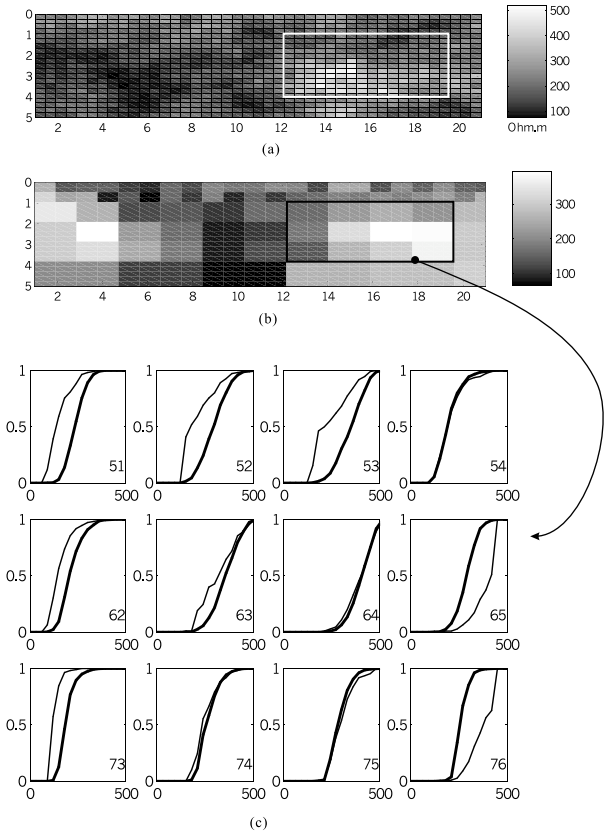

Fig. 17. CDFs for some inverted blocks. (a) True resistivity section proposed for model test. (b) Mean resistivity section of marginal *a posterior* PDF. (c) Highlighted resistivity region is matched to CDF maps to give uncertainty analysis. *A prior* PDFs appear as the bold line and *a posterior* PDFs as by the thin line in each CDFs. This region is similar to those of Fig. 16, but it delineates well the true resistivity zone excepting some blocks in the right region. CDFs show some different aspects than those of Fig. 16.

various sources provides for manifold potential applications to geophysical data processing.

Next, we have adopted MCMC approach to examine the *a posterior* PDF. MCMC was accomplished by the Gibbs sampler which was proceeded by the modified SA algorithm. The newly devised Gibbs sampler in this study is quite effective and well reflects the merit of the Bayesian frame. This global sampling technique takes advantage of the robustness and generality of the *a prior* information given by non-parametric distribution. This results in the marginal distribution of *a posterior* PDF and enables us to make various interpretations for the inversion result such as those displayed by Sambridge (1999). In this study, for simplicity, we have shown only marginal disributions.

According to our experiences in the traditional inversion scheme, the data-fit alone criteria produces an unplausible model with sudden fluctuation of parameters although the model validates the observed data, for we intended it to extract more information than was contained in the data. Therefore we have approached the geophysical inverse problem through the Bayesian paradigm. This method provides the methodology to incorporate various information to the inverse problem and is stable enough to estimate feasible solutions. And we have solved the geophysical inverse prob-

lem in terms of non-parametric conditions, so we were free from the implementation of covariance matrices, and we have presented various practical tools to implement this frame.

Acknowledgments. This study was supported by the project "The development of protection techniques against earthquake disaster" of METRI/KMA. We are grateful to Dr. T. Smith, Dr. M. Menvielle, and an anonymous reviewer for their helpful and critical reviews, which allowed us to improve this manuscript significantly. We also thank Dr. Yasuo Ogawa for his helpful comments and suggestions.

References

Bayes, T., An essay towards solving a problem in the doctrine of chances, reprinted in *Biometrika*, **45**(1958), 293–315, 1763.

Besag, J., Spatial interaction and the statistical analysis of lattice systems (with discussion), *J. R. Statist. Soc. B*, **36**, 192–236, 1974.

Carlin, B. P. and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, Champman & Hall, 1996.

Chunduru, R. K., M. K. Sen, and P. L. Stoffa, 2-D resistivity inversion using spline parameterization and simulated annealing, *Geophysics*, **61**, 151–161, 1996.

Deutsch, C. V. and A. G. Journel, *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 1992.

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

Duijndam, A. J. W., Bayesian estimation in seismic inversion, part ii: Uncertainty analysis, *Geophys. Prosp.*, **36**, 899–918, 1988b.

Gelfand, A. E. and A. F. M. Smith, Sampling-based approaches to calculating marginal densities, *J. Am. Statist. Ass.*, **85**, 398, 409, 1990.

Gelfand, A. E., S. E. Hills, A. Racine-Poon, and A. F. M. Smith, Illustration of Bayesian inference in normal data models using Gibbs sampling, *J. Am. Statist. Ass.*, **85**, 972–985, 1990.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Champman & Hall, 1995.

Geman, S. and D. Geman, Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. PAMI-6, 721–741, 1984.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Champman & Hall, 1996.

Glacken, I., *Change of Support by Direct Conditional Block Simulation*, Master's thesis, Stanford Univ., Stanford, CA., 1996.

Gomez-Hernandez, J., A *stochastic approach to the simulation of block conductivity fields conditional upon data measured at a smaller scale*, Doctoral Dissertation, Stanford University, Stanford, CA, 1991.

Goovaerts, P., *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, 1997.

Gouveia, W. P., *Bayesian seismic waveform data inversion: Parameter estimation and uncertainty analysis*, Ph.D. thesis, Colo. Sch. of Mines, Golden, Co, 1996.

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

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

Isaaks, E. H., *The application of Monte Carlo Methods to the Analysis of Spatially Correlated Data*, Doctoral Dissertation, Stanford University, Stanford, CA, 1990.

Isaaks, E. H. and R. M. Srivastava, *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 1989.

Jaynes, E. T., *Probability Theory: The Logic of Science*, `http://bayes.wustl.edu/etj/prob.html`, 1994.

Lines, L. R. and S. Treitel, Tutorial: A review of least-squares inversion and its application to geophysical problems, *Geophys. Pros.*, **32**, 159–186, 1984.

Loredo, T. J., From laplace to supernova SN 1987A: Bayesian inference in astrophysics, in *Maximum Entropy and Bayesian Methods*, edited by P. F. Fougere, pp. 81–142, Kluwer Academic Publishers, The Netherlands, 1990.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equations of state calculations by fast computing machine, *J. Chem. Phys.*, **21**, 1087–1091, 1953.

Moraes, F. S., *The application of marginalization and local distributions to multidimensional Bayesian inverse problems*, Ph.D. Thesis, Colorado School of Mines, Golden, Co, 1996.

Mosegaard, K. and A. Tarantola, Monte Carlo sampling of solutions to inverse problem, *J. Geophys. Res.*, **100**, 12,431–12,447, 1995.

MPI Forum, *MPI: A Message-Passing Interface Standard*, University of Tennesse, Knoxville, Tennesee, 1995.

Olea, R., *Optimum mapping technique: Series on spatial analysis No. 2*, Kansas Geol. Survey, Lawrence, KA, 1975.

Pacheco, P. S., *Parallel Programming with MPI*, Morgan Kaufmann Publisher, Inc., San Francisco, CA, 1997.

Roberts, G. O., Markov chain concepts related to sampling algorithms, in *Markov Chain Monte Carlo in Practice* edited by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, pp. 45–57, Champman & Hall, 1995.

Rothman, D. H., Nonlinear inversion, statistical mechanics, and residual statics estimation, *Geophysics*, **50**, 2784–2796, 1985.

Rothman, D. H., Automatic estimation of large statics corrections, *Geophysics*, **51**, 332–346, 1986.

Sambridge, M., Geophysical inversion with a neighbourhood algorithm-II. Appraising the ensemble, *Geophys. J. Int.*, **138**, 727–746, 1999.

Scales, J. A. and A. Tarantola, An example of geologic prior information in a Bayesian seismic inverse calculation, *Center for Wave Phenomena 159*, Golden, Co, 1994.

Scales, J. A. and L. Tenorio, Prior information and uncertainty in inverse problems, *Mathematical Geophysics Summer School*, Stanford University, 1998.

Sen, M. and P. L. Stoffa, *Global optimization methods in geophysical inversion*, Advances in Exploration Geophysics, v.4, Elsevier, Amsterdam, 1995.

Srivastava, R. M., Minimum variance or maximum profitability?, *CIM Bulletin*, **80**, 63–68, 1987.

Szu, H. and R. Heartley, Fast simulated annealing, *Phys. Lett.*, **122**, 157–162, 1987.

Tarantola, A., *Inverse problem theory: Methods for data fitting and model parameter estimation*, Elsevier, 1987.

Tierney, L., Introduction to general state-space Markov chain theory, in *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, pp. 59–74, Champman & Hall, 1995.

Wijk, K., J. A. Scales, W. Navidi, and K. Roy-Chowdhury, Estimating data uncertainty for least square optimization, *Center for Wave Phenomena 264*, Golden, Co, 1997.

S.-H. Oh and B.-D. Kwon (e-mail: gimul@metri.re.kr)