STUDIES IN GEOPHYSICAL INVERSION

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

GEOLOGY AND GEOPHYSICS

AUGUST 1994

By

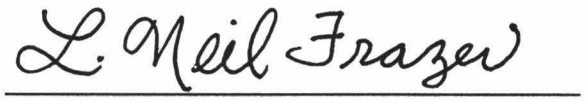
Atanu Basu

Dissertation Committee:

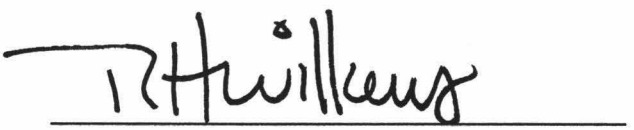
L. Neil Frazer, Chairperson
Eduard Berg
Frederick K. Duennebier
Roy H. Wilkens
Narendra K. Saxena

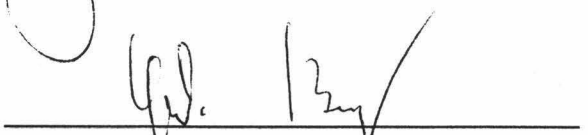
We certify that we have read this dissertation and that, in our opinion, it is satisfactory in scope and quality as a dissertation for the degree of Doctor of Philosophy in Geology and Geophysics.

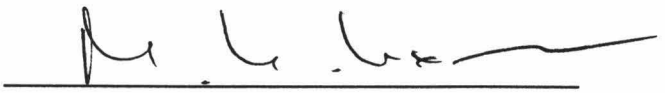
DISSERTATION COMMITTEE



Chairperson







ACKNOWLEDGMENTS

After years of study and research now is the time to express the appreciation and acknowledge everyone who helped me in different ways to bring my Ph.D program to its end successfully. I thank my advisor, Neil Frazer, from whose teaching, advice and encouragement I benefited most during my stay at UH. He allowed me to explore the vast unknown *solution* space with a little *a priori* knowledge on my own but was always present to lend a hand when I was lost. My gratitude goes to all the members of my committee, Dr's Eduard Berg, Frederick K. Duennebier, Roy H. Wilkens and Narendra K. Saxena. I thank them all for their help and cooperation.

I thank my friend Alakabha Datta for all his help, encouragement and editorial assistance. I also thank my other friends, especially Bertram Nolte and Banibrata Chakravarty for their help over the years. I thank Dr. Kabir Roychaudhury for helping me with his knowledge of real seismic data. I thank my wife Crystalyn for her encouragement and my newborn daughter Isabella for sweet smiles.

Finally I thank my parents late Chuni Lal and Ruby Basu for their ever present support from half the world away.

ABSTRACT

The mapping from the data space to the model space is known as the inverse problem. Nonlinear inverse problems are in general more difficult and computationally expensive to solve than linear problems. For nonlinear inverse problems when *a priori* information in the form of a starting model is not available, then a gradient based algorithm may converge to a local solution rather than to the global one. Global optimization methods such as simulated annealing (SA) have been applied recently to several geophysical inverse problems.

SA resembles the thermodynamic process of annealing to form crystals from a melt. The minimum energy state may be viewed as corresponding to the minimum of the cost function. It is well known that determining the 'critical temperature' is one of the most important factors regarding the efficiency of the SA algorithm. Here we determined the 'critical temperature' by executing the SA process for a fixed number of sweeps at a fixed temperature for different temperatures and then calculating the average energy for those sweeps. We constructed the a posteriori probability density function (PPD) and then determined the best model from it.

We performed inversion using the SA method at a fixed temperature on two-offset VSP data, and on cross-borehole data to determine the slowness of the layers between the source well and the receiver well. The results of our inversion suggest that SA has the potential to solve nonlinear inverse problems even when the solution space is large and multimodal.

We also performed inversion to compute bathymetry from shipboard free-air gravity anomaly data. Our results show an error of less than 1 % of depth, which is within the acceptable error range for measuring depth according to International Hydrographic Bureau standards.

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

Chapter 1

Inverse Theory

1.1 INTRODUCTION

The main goal of seismic inversion is to estimate the values of the parameters that describe an accurate earth model. In a typical geophysical inverse problem (Menke 1984) we assume that, given any model $\mathbf{m}$, we can calculate the corresponding data $\mathbf{d}$ by the forward theory $\mathbf{d} = \mathbf{g}(\mathbf{m})$, where $\mathbf{g}$ is the operator that describes the forward modelling. All vectors $\mathbf{m}$ belong to a functional space known as the 'model space' M and similarly all vectors $\mathbf{d}$ belong to the 'data space' D. The mapping from the model to the measured fields (e.g., elastic wavefields, magnetic fields) is called the forward problem, while the mapping from the measured fields to the (unknown) model is known as the inverse problem. In seismic experiments the vector $\mathbf{m}$ describes a model of the real earth which normally includes P and S-wave slowness (inverse wavespeed) and density.

In the inversion problem it is important to determine whether the proposed model is unique or not. If the model is not unique, then many models can satisfy the observed data. This nonuniqueness of the solution could be due to inherent nonuniqueness of the problem, or uncertainty in the data and parametrization, or a combination of both. For example, gravity, magnetic, and electrical potential field

data are all inherently nonunique. However, this nonuniqueness can be reduced if we impose some constraints or *a priori* information on the model parameters. Even if an inverse problem is known to have a unique solution, the solution space may have many local minima (Fig. 1.1).

In view of the above one can classify inverse problems on the basis of whether the equation $\mathbf{d} = \mathbf{g}(\mathbf{m})$ provides enough information to specify uniquely the model parameters or not. If this equation doesn't provide enough information then the problem is said to be *underdetermined.* If this equation provides exactly enough information to determine the model parameters, then it is said to be *even-determined.* When this equation provides too much information then it is said to be *overdetermined.* Overdetermined problems typically have more data than unknowns.

1.2 TYPES OF INVERSE PROBLEMS

1.2.1 Linear Inverse Problems

Among all types of inverse problems, this is the simplest kind of inverse problem to solve. In this case, instead of writing $\mathbf{d} = \mathbf{g}(\mathbf{m})$, we can write $\mathbf{d} = \mathbf{G}\mathbf{m}$, where $\mathbf{G}$ represents a linear operator acting from the model space into the data space. The discrete form of this relation can be written as

$$d_i = \sum_{j=1}^{M} G_{ij} m_j \tag{1}$$

1.2.2 Linearized Inverse Problems

In real situations most of the inverse problems can not be written in explicit linear form. But if the perturbation of the model from the initial guess is small in some sense,

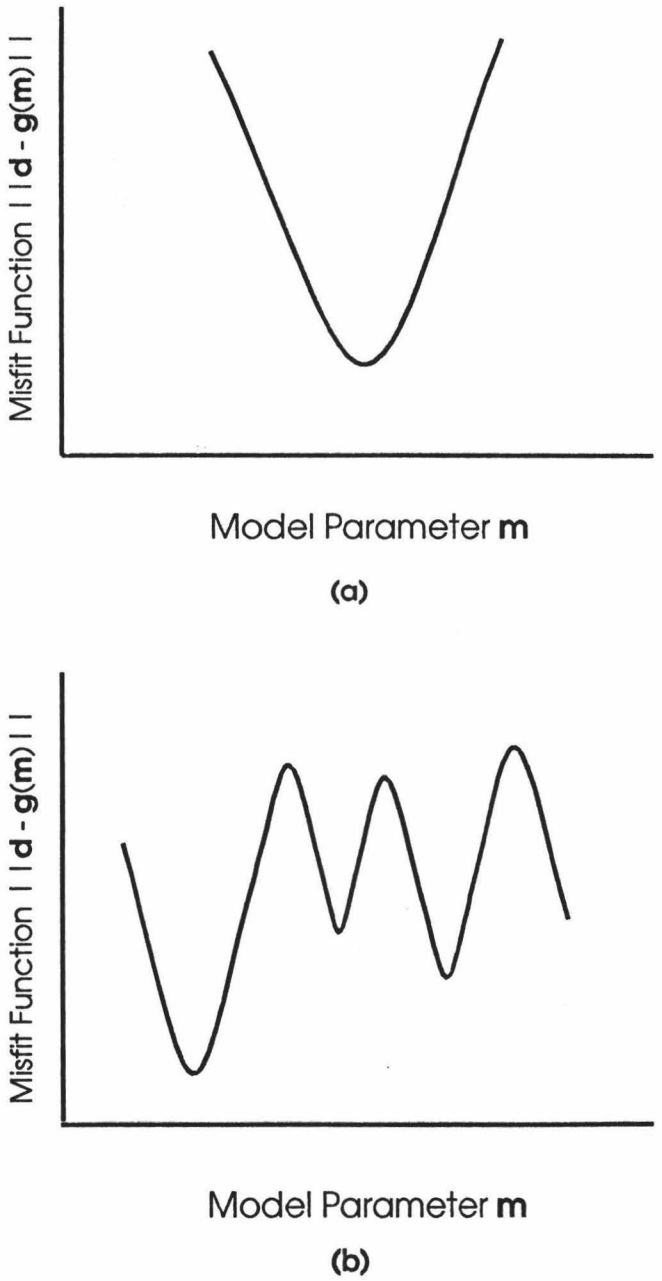


Figure 1.1: Minima of misfit function (a) Single minimum (b) Several local minima and one global minimum

then the problem can be linearized. Such problems are known as weakly nonlinear. The forward equation $\mathbf{d} = \mathbf{g}(\mathbf{m})$ can be linearized around some initial guess, $\mathbf{m}_{prior}$ by the Taylor series expansion as

$$\mathbf{d} = \mathbf{g}(\mathbf{m}_{prior}) + \left(\frac{\partial \mathbf{g}_i}{\partial \mathbf{m}_i}\right)_{\mathbf{m}_{prior}} (\mathbf{m} - \mathbf{m}_{prior}) + \left(\frac{\partial^2 \mathbf{g}_i}{\partial \mathbf{m}_i^2}\right)_{\mathbf{m}_{prior}} (\mathbf{m} - \mathbf{m}_{prior})^2 + \ldots \quad (2)$$

Neglecting second and higher derivative terms,

$$\mathbf{d} = \mathbf{g}(\mathbf{m}_{prior}) + \left(\frac{\partial \mathbf{g}_i}{\partial \mathbf{m}_i}\right)_{\mathbf{m}_{prior}} (\mathbf{m} - \mathbf{m}_{prior}) \quad (3)$$

or,

$$\mathbf{d} - \mathbf{g}(\mathbf{m}_{prior}) = \left(\frac{\partial \mathbf{g}_i}{\partial \mathbf{m}_i}\right)_{\mathbf{m}_{prior}} (\mathbf{m} - \mathbf{m}_{prior}) \quad (4)$$

or,

$$\mathbf{Y} = \mathbf{A}\mathbf{X} \quad (5)$$

where $\mathbf{Y} = \mathbf{d} - \mathbf{g}(\mathbf{m}_{prior})$ is the vector difference between the observed data and the computed data for the starting model, $\mathbf{X}$ is the vector difference between the new model parameter and the initial model parameter, and $\mathbf{A} = \left(\frac{\partial \mathbf{g}_i}{\partial \mathbf{m}_i}\right)_{\mathbf{m}_{prior}}$ is a matrix.

Multiplying equation (5) by $\mathbf{A}^T$ gives

$$\mathbf{A}^T\mathbf{Y} = (\mathbf{A}^T\mathbf{A})\mathbf{X} \quad (6)$$

where T denotes a transpose.

Then, $\overline{\mathbf{X}}$, the estimate of $\mathbf{X}$, is given by

$$\overline{\mathbf{X}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y} \tag{7}$$

where $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is called the pseudo inverse.

1.2.3 Nonlinear Inverse Problems

In this type of problem, $\mathbf{d} = \mathbf{g}(\mathbf{m})$ is not a linear function of $\mathbf{m}$ and the probability density function (PDF) of uncertainty of data values is not Gaussian. The more nonlinear $\mathbf{g}(\mathbf{m})$ is, the more the PDF differs from a Gaussian function. In this case $\mathbf{g}(\mathbf{m})$ has to be computed at each iteration without using any linear approximation. As a result, nonlinear problems are in general more expensive to solve than linearizable problems.

1.3 TYPES OF INVERSION ALGORITHMS

1.3.1 Gradient Methods

The simplest type of gradient method is the steepest descent method. In this method a minimum of the misfit function is sought for in the direction of steepest descent. The direction of steepest descent is locally optimal and if infinitely many steps are taken, each one being infinitely small, this would simulate the motion of a raindrop on the slope of a mountain (Tarantola 1987). On the other hand, Nolet (1987) compared this method as 'a blind man walking in the mountains who wants to find his way to the beach.' This method converges very slowly and sometimes it takes hundreds of iterations to obtain a small change in the misfit function.

Let $\mathbf{m}_n$ represent the current point, $\mathbf{m}_{n+1}$ the updated point and λ_n the direction of steepest descent. The method connects the updated point and the current point by the relation

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \mu_n \lambda_n \tag{8}$$

where μ_n is an arbitrarily positive real number small enough so that the misfit function of updated point is smaller than the current point. Both analytical and trial and error methods can be used to chose μ_n.

In the Newton method, another variation of the gradient search method, the updated point is given by

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \mathbf{G}_n \lambda_n \tag{9}$$

where $\mathbf{G}_n$ is the curvature operator.

In the quasi-Newton method the relationship is

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \mathbf{H}_n \lambda_n \tag{10}$$

where $\mathbf{H}_n$ is the Hessian operator which is given as

$$\mathbf{H}_n = \left(\frac{\partial^2 \mathbf{S}}{\partial x^2} \right)_{x_n} \tag{11}$$

where $\mathbf{S}$ is the misfit function.

There are many other gradient methods, e.g. preconditioned steepest descent, the variable matrix method, the conjugate method (Hestenes and Stiefel 1952), etc. In all these methods the basic idea is the same, and differences are due to stability and convergence rates. All gradient methods usually converge to the nearest local minimum rather than the global minimum. A detailed review of this method has been

given by Krarup (1968), Moritz (1970), Hillier & Lieberman (1980), Scales (1985) and many others.

1.3.2 Random Search Method

The major drawback of any kind of gradient search method is that, for highly non-linear problems, these methods converge to a local minimum, rather than the global minimum. One alternative to this problem is to use a random search method, known as the Monte Carlo method, which uses the generation of pseudo-random numbers. With the help of modern fast computer millions of models can be tried, but it is still a very expensive method. This method was used in geophysics by Press (1968, 1970) and many others.

1.3.3 Global Optimization Methods

Recently, two global optimization methods were successfully applied in geophysics to solve nonlinear inverse problems. One of them is called simulated annealing (Kirkpatrick et al. 1983). This method mimics the thermodynamic process of annealing, which is the process of finding low energy states of a solid by initially melting the substance, and then lowering the temperature slowly and observing the formation of crystals. In geophysical exploration, Rothman (1985, 1986) first introduced this technique to estimate large residual statics corrections, and thereafter it has been used in various other problems in seismology (Landa et al. 1989, Sen and Stoffa 1991, Nørmark and Mosegaard 1993).

The second method is known as the Genetic Algorithm. It was originally developed by Holland (1975) in the field of artificial intelligence. This method has a

similarity to the genetic evolution of biological systems. It simulates natural genetics by following an entire population of trial models encoded as finite-length strings (chromosomes), wherein model parameters are part of the string. In geophysics this method has also been applied in seismic waveform inversion (Stoffa & Sen 1991, Nolte & Frazer 1993).

1.4 REFERENCES

Hestenes, M. R. & Stiefel, E., 1952. Methods of conjugate gradients for solving linear systems, *J. of Research of the National Bureau of Standards*, **49**, 409-436.

Hillier, F. S. & Lieberman, G. J., 1980. *Introduction to Operations Research*, Holden-Day, Inc., California, U.S.A.

Holland, J. H., 1975. *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Harbor, Michigan, U.S.A.

Krarup, T., 1968. A framework for least-squares determination of the potential of the earth, *Report to the members of IAG Special Study Group*, **5**, 31.

Landa, E., Beydoun, W. & Tarantola, A., 1989. Reference velocity model estimation from prestack waveforms: Coherency optimization by simulated annealing, *Geophysics*, **54**, 984-990.

Menke, W., 1984. *Geophysical Data Analysis: Discrete Inverse theory*, Academic Press, New York.

Moritz, H., 1970. Least-squares estimation in physical geodesy, *Dept. of Geodetic Science Report, Ohio State Univ.*, **130**.

Nolet, G., 1987. *Waveform tomography, In: Seismic Tomography*, Reidal Publ. Com.

Nolte, B. & Frazer, L. N., 1993. VSP inversion with genetic algorithms, *Geophys. J. Int* (in press).

Nørmark, E. & Mosegaard, K., 1993. Residual statics estimation: scaling temperature schedules using simulated annealing, *Geophysical Prospecting*, **41**, 565-578.

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

Press, F., 1970. Earth models consistent with geophysical data, *Phy. Earth and Plan. Int.*, **3**, 3-22.

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

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

Scales, L. E., 1985. *Introduction to nonlinear optimization*, Macmillan.

Sen, M. K. & Stoffa, P. L., 1991. Nonlinear one-dimensional seismic waveform inversion using simulated annealing, *Geophysics*, **56**, 1624-1638.

Stoffa, P. L. & Sen, M. K. 1991. Nonlinear multiparameter optimization using genetic algorithms: inversion of plane wave seismograms, *Geophysics*, **56**, 1794-1810.

Tarantola, A., 1987. *Inverse Problem Theory*, New York: Elsevier Science Publishing Company Inc.

Chapter 2

Bathymetry Computation From Free-air Anomaly Data

2.1 INTRODUCTION

Detailed bathymetric mapping of the seafloor is expensive, difficult, and time consuming. Every year more and more area of the ocean is surveyed using high resolution bathymetric mapping systems, but still the coverage speed is very slow compared to the total area of the oceans. The airborne scanning laser bathymeter can survey a large area in relatively short periods of time, but because of its low depth penetration capability it is used only in shallow water mapping (Estep 1993). As a result we have thoroughly surveyed only a few areas of the world's oceans. There are many areas where historical data are available, but old bathymetric sounding systems and old methods of navigation incurred significant errors in these data. The gap in bathymetry data can be filled, and the accuracy of the historical data can be improved, if we can predict bathymetry from other types of data collected in that area. This prediction depends on the quality of the other type of data and the correlation between bathymetry and that data.

It is well known that there is some correlation between gravity observations obtained at sea and sea floor topography (Khan et al. 1971; McKenzie and Bowin 1976;

Watts 1978; Watts et al. 1985). It is also known that short-wavelength bathymetry produces short-wavelength gravity anomalies, while long wavelength bathymetry are compensated (Talwani et al. 1972; McKenzie and Bowin 1976). In the continental data set Forsyth (1985) showed that the coherence between Bouguer anomalies and topography is very high at long wavelengths and decreases with decreasing wavelength. For continents the compensation signature at wavelengths shorter than 20 km is not seen because the compensation is so deep that its gravity signal is too severely attenuated to be measured at the surface (Lewis and Dorman 1970; McNutt and Parker 1978). In the ocean the compensation depth is shallower and therefore the short wavelength gravity signal is not too attenuated to be measured (McNutt 1979).

The relationship between the gravity field and the sea floor topography is analyzed using a linear transfer function, called admittance (McKenzie and Bowin 1976; McNutt 1979). Basically, admittance is the ratio of the Fourier transforms of gravity and bathymetry. The computation of admittance function from the data and comparing it with the theoretical value is now a common approach to finding the compensation depth of the surveyed area (Dalloubeix et al. 1988; Kister and Wilhelm 1988; Maia et al. 1990; Ashalatha et al. 1991). The admittance function has also been used in predicting the bathymetry from Seasat altimeter and gravity data (Goslin and Diament 1987; Black and McAdoo 1988/89; Jung and Vogt 1992).

In this chapter we investigate the feasibility of computing bathymetry in an area where only the shipboard gravity data are available and both gravity and bathymetry data are available in the surrounding region. We use the concept of admittance in the prediction of bathymetry for very short wavelengths (< 12 km) using closely spaced seabeam and gravity data. From the theoretical formula of admittance we derive an objective function that is independent of knowledge of the density and the

compensation depth value of the area as long as they are constant. We look for the unknown bathymetry of a point by minimizing the objective function. To perform an exhaustive search of the solution space, we make this space small by restricting the range of the model parameters. This gives the limitations in the accuracy of the predicted bathymetry. We apply our inversion technique to the data collected south of the island of Hawaii by Scripps Institution of Oceanography in 1987.

2.2 ADMITTANCE ANALYSIS

2.2.1 Observed Admittance

As explained by several authors (McKenzie and Bowin 1976; Watts 1978), the forward model of predicting the bathymetry from the gravity can either be constructed in the space domain or in the wave number domain. In the space domain, the idea is to find a filter that convolves with the bathymetry observations to produce a series that resembles the gravity observations:

$$g(x) = f(x) * b(x) + n(x) \tag{1}$$

where g(x) is the gravity profile, f(x) is the impulse response, b(x) is the bathymetry profile, and n(x) is the noise that includes the part of gravity which is not linearly related with the bathymetry. In the wave number domain this is equivalent to a simple multiplication:

$$G(k_n) = Z(k_n)B(k_n) + N(k_n) \tag{2}$$

where $G(k_n)$, $Z(k_n)$, $B(k_n)$, and $N(k_n)$ are the discrete Fourier transforms of g(x), f(x), b(x), and n(x), respectively, and k_n is the wave number ($k_n = 2\pi/\lambda$).

If the noise is not significantly correlated with the isostatic gravity anomaly, then an estimate of admittance, $Z(k_n)$ is given by

$$Z(k_n) = G(k_n)/B(k_n) \tag{3}$$

In the presence of significant noise, a better estimate of admittance was given by McKenzie and Bowin (1976):

$$Z(k_n) = [G(k_n)B(k_n)^*]/[B(k_n)B(k_n)^*] \tag{4}$$

where * indicates the complex conjugate. In this case the admittance is given by the cross spectrum of bathymetry and gravity divided by the power spectrum of bathymetry.

2.2.2 Theoretical Admittance

Parker (1973) derived how the gravity effect of a layer of material can be expressed as an infinite series of Fourier transforms. Considering only the first term in the Fourier series expansion, the simplest model of admittance (McKenzie and Bowin 1976) can be written as

$$Z(k_n) = 2\pi G(\rho_2 - \rho_w)e^{-k_n d} \tag{5}$$

where G is the gravitational constant, ρ_2 is the density of the seafloor topography, ρ_w ($= 1.03$ g/$cm^3$is the density of seawater, and d is the mean depth of sea. In this case we assume that the gravity anomalies are caused only by the sea floor topography and that the topography is uncompensated. This simple model works well for short wavelength values for admittance.

In the simplest compensation model the mass associated with the depth variations is compensated either by corresponding density variations (Pratt's hypothesis) or by low-density roots (Airy's hypothesis). In the Airy compensation model, the bathymetry is compensated by variations in the thickness of a constant density layer of mean thickness t. The admittance expression for this model (McKenzie and Bowin, 1976) is

$$Z(k_n) = 2\pi G(\rho_2 - \rho_w)e^{-k_n d}(1 - e^{-k_n t}) \tag{6}$$

The other parameters are the same as in equation (5).

2.3 BATHYMETRY COMPUTATION

2.3.1 Method

The idea of finding bathymetry values from gravity observations is based on the fact that we have both bathymetry and gravity values at the ends of a traverse line or at the outer segment of a gridded area but only the gravity values in the middle of the line or in the inner segment of the gridded area. In the case of the traverse line we start from one end and continue along the line. Our initial inputs to the model are the bathymetry and gravity values of some fixed number of consecutive points at the near end of the traverse line and the output is the predicted bathymetry of the next point where only the gravity value is known. We then use the bathymetry just found, and the gravity as an input to find the bathymetry of the next point. We continue this process until we reach the far end of the traverse line. Similarly, we start from the far end of the traverse line and move in the opposite direction. The final predicted bathymetry is the average of these two bathymetry values. In the case of a gridded area we move along the grid in four different directions, i.e., left,

right, up, and down and then take the average of these four set of bathymetry values. The correlation between the observed bathymetry and the calculated bathymetry is better in the case of gridded data than in case of the traverse line. One of the reason is that in the case of gridded data we take the average of four sets of bathymetry values as compared to the two sets for the traverse line case, so the random errors in the predicted bathymetry get cancelled more in the gridded data. Following Basu and Saxena (1993) the mathematical derivation of this method is described in the following section.

2.3.2 Objective Function and Inversion Algorithm

The method described in the previous section can be regarded as an inverse problem wherein the goal is to minimize an objective function to find the bathymetry. In this section we derive the objective function from the observed and theoretical admittance formulas.

Let us consider a traverse line which has n sample points at regular intervals. The first and last m points have both gravity and bathymetry values, and the rest have only gravity values. Consider the following notations:

$Z_{1,r}$ = admittance for the points 1 to r

$Z_{2,r+1}$ = admittance for the points 2 to r+1

$Z_{3,r+2}$ = admittance for the points 3 to r+2

$d_{1,r}$ = mean depth for the points 1 to r

$d_{2,r+1}$ = mean depth for the points 2 to r+1

$d_{3,r+2}$ = mean depth for the points 3 to r+2

We assume that the points 1 to r+2 are close enough to neglect the lateral crustal density variations between them. So, in the admittance formula (5) the admittance

of these points varies because of the mean depth variation between them. Now if we take natural logarithm of equation (5), we get

$$lnZ(k_n) = lnC - k_n d \tag{7}$$

where C is constant. Writing the above equation for $Z_{1,r}$, $Z_{2,r+1}$, $Z_{3,r+2}$, $d_{1,r}$, $d_{2,r+1}$, $d_{3,r+2}$ and doing some algebra we get

$$\frac{lnZ_{1,r} - lnZ_{2,r+1}}{lnZ_{2,r+1} - lnZ_{3,r+2}} = \frac{d_1 - d_{r+1}}{d_2 - d_{r+2}} \tag{8}$$

and

$$\frac{lnZ_{1,r} - lnZ_{2,r+1}}{lnZ_{1,r} - lnZ_{3,r+2}} = \frac{d_1 - d_{r+1}}{(d_1 - d_{r+1}) + (d_2 - d_{r+2})} \tag{9}$$

Let us denote the expression of L.H.S. of equation (8) as $Z_{12,23}(k_n)$ and that of R.H.S. as $d_{12,23}$. Similarly, we write $Z_{12,13}(k_n)$ and $d_{12,13}$ for equation (9). Now, joining equation (8) and (9) we get

$$Z_{12,23}(k_n) + Z_{12,13}(k_n) = d_{12,23} + d_{12,13} \tag{10}$$

Since there is no wavenumber term involved in the right hand side of equation (10) it is true for all wavenumbers in the left hand side. In the compensation model (equation (6)) if we assume that all the extra terms are also constant for the points 1 to r+2, then we can also write the above equation (10) for this model. We set up the objective function ϵ_j as follows:

$$\epsilon_j = \sum_{n=1}^{p} [Z_{12,23}(k_n) + Z_{12,13}(k_n) - d_{12,23} - d_{12,13}]_j^2 \tag{11}$$

where ϵ_j corresponds to different bathymetry values d_{r+2} for this point. Here we assume that we have both the bathymetry and gravity values for points 1 to r+1 and only the gravity values for point r+2. We calculate $Z_{12,23}(k_n)$, $Z_{12,13}(k_n)$ using equation (4) and $d_{12,23}$, $d_{12,13}$ for the points 1 to r+3. We search for the minimum of the objective function to find the true d_{r+2} in the range $d_{r+2} \pm d_{range}$. In this case

there is only one parameter in the model space, $\mathbf{m} = \{d_{r+2}\}$. Considering the errors in the known gravity values for points 1 to r+2, and in the bathymetry values for points 1 to r+1, the model space can be written as

$$\mathbf{m} = \{g_1, g_2, ..., g_{r+1}, g_{r+2}, d_1, d_2, ..., d_{r+1}, d_{r+2}\} \tag{12}$$

Suppose the known gravity values have the error $\pm g_{error}$ and the known bathymetry values have the error $\pm d_{error}$; we then minimize the objective function (11) in the solution space $\{g_{1...r+2} \pm g_{error}, \{d_{1...r+1} \pm d_{error}, d_{r+2} \pm d_{range}\}$. The $\pm g_{error}$, $\pm d_{error}$, and $\pm d_{range}$ are divided in steps of Δg_{error}, Δd_{error}, Δd_{range}, respectively. After we find the bathymetry of the point d_{r+2} we use the gravity values of the points 2 to r+3 and the bathymetry values of the points 2 to r+2 to compute the bathymetry value of the point r+3. We proceed this way till we cover n-2m points.

In a similar fashion, we start from the other end of the traverse line and move in the opposite direction. The average of these two profiles gives the final calculated bathymetry profile.

In the case of gridded data, we follow the above mentioned algorithm along a grid line starting from one corner. After we calculate the bathymetry values along one grid line then we move to the next grid line in the upward or downward direction if the grid line runs from left to right and in the left and right direction if the grid line runs from top to bottom. Here we have four sets of bathymetry profiles to calculate average and so the noise elimination in the calculated bathymetry is better in this case.

2.4 APPLICATIONS

In this section we show two examples of bathymetry computation from free air anomaly along a traverse line and one example of gridded data. As it was explained by McKenzie and Bowin (1976) and McNutt (1979), we can use free air anomaly data in equation (4) instead of Bouguer anomaly data because the free air anomaly is directly related to the stress within the earth and is also easily available. In all the following examples we use the following values for inversion:

r = 3

g_{error} = 1 mgal, Δg_{error} = 0.1 mgal

d_{error} = 5 m, Δd_{error} = 1 m

d_{range} = 100 m, Δd_{range} = 1 m

p, number of wave numbers = 9

For the starting value of the unknown depth we use the mean depth of the previous r (r=3) points.

2.4.1 Data

The data used here were collected south of the island of Hawaii by Scripps Institution of Oceanography (SIO) from April 30 to June 3, 1987 on the R/V *T. Washington.* The part of the track line is shown in Fig. 2.1. Sea Beam was used to measure the bathymetry and a Bell Aerospace BGM-3 meter was used to measure gravity. For navigation primarily Global Positioning System (GPS) was used and where it was unavailable then old satellite navigation system was used.

2.4.2 First Example

We selected a small section of bathymetry and free air anomaly data from the SIO data set along the traverse line from longitude 156.54903W and latitude 15.20519N to longitude 156.37729W and latitude 15.51802N. The length of this portion of traverse line is 37.177 km. The depth of sea water varies from 5550 m to 5418 m, and the free air anomaly varies from -11.1 mgals to 1.0 mgals. Along the traverse line we sampled every alternate observation point; there are 60 sample points having an average distance of 630 m between them. The free air anomaly and observed bathymetry are shown in Fig. 2.2a and 2.2b respectively. The correlation between them is 0.7064. We calculated the admittance using equation (4) after removing means of both the bathymetry and free air anomaly data, and computed equation (12) for inversion. We did not use any smoothing technique of the admittance function (Bowin and Milligan 1985) before inversion. The results of the computed bathymetry along with the observed bathymetry and the residuals (observed minus computed bathymetry) are shown in Fig. 2.3a and 2.3b respectively. The correlation value between them is 0.9550. The standard deviation of error is 6.8760 m and the maximum deviation of error is 25 meters. Here the wavelength range of admittance considered is 7.8 - 11.7 km. In Fig. 2.3a we notice that in steep-slope regions the computed bathymetry is in good agreement with the observed profile, but in the flat region the computed profile is oscillatory.

2.4.3 Second Example

In this example we show the computed bathymetry for a smaller portion of the traverse line. This part of the line covers the points from longitude 156.48164W and

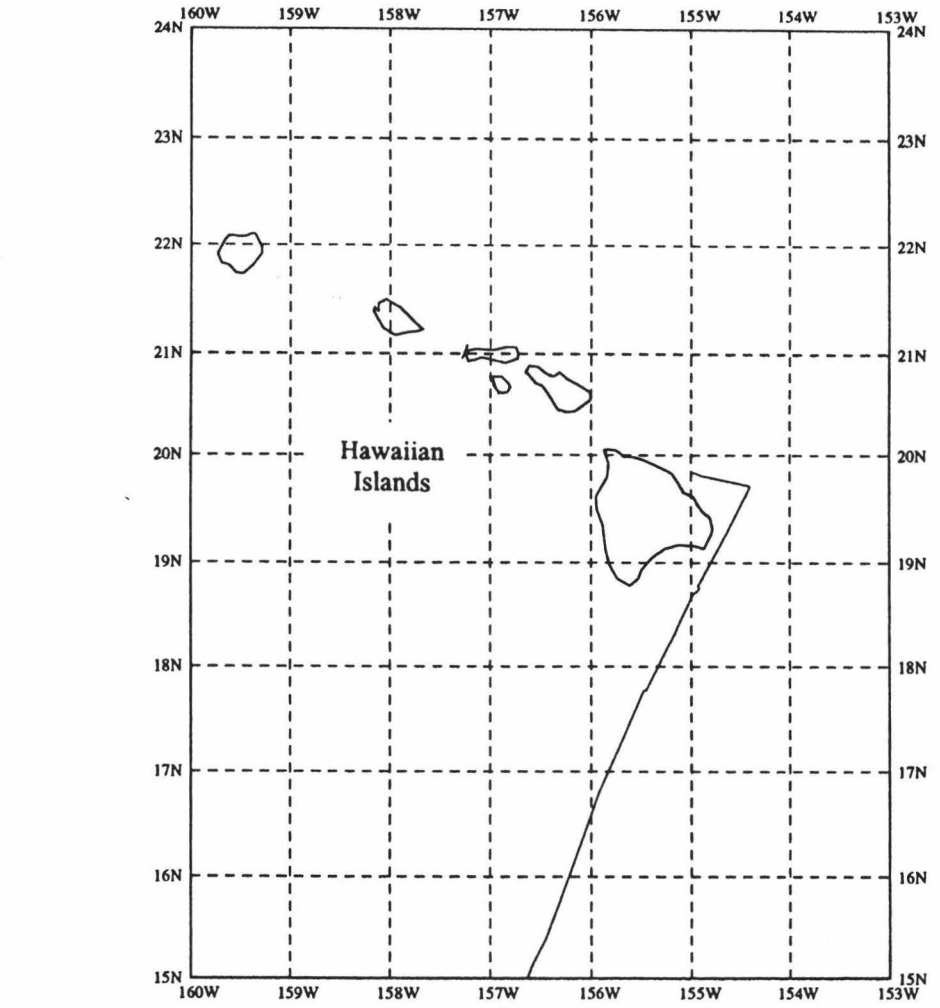
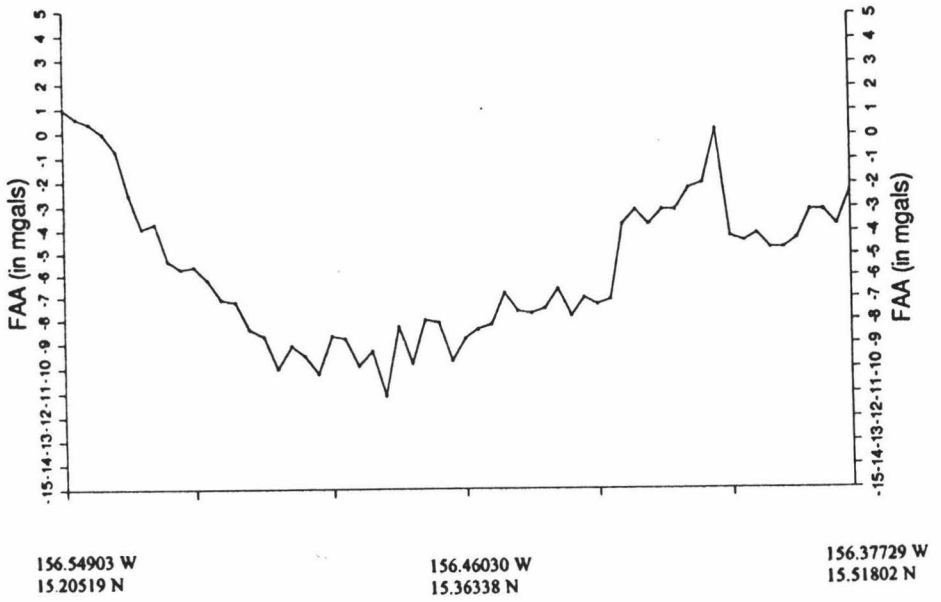


Figure 2.1: Trackline of the part of the cruise that covers the study area

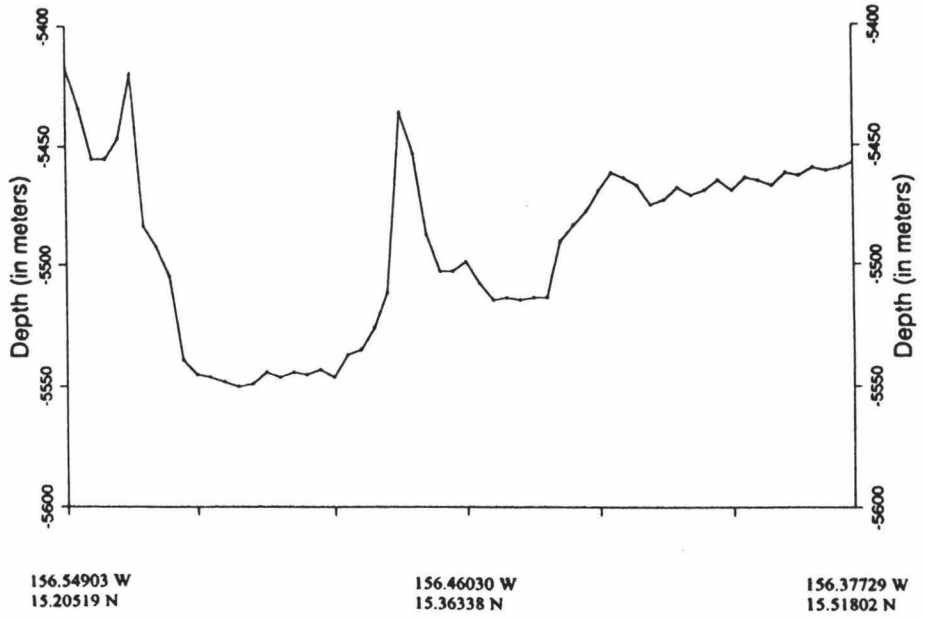
latitude 15.32417N to longitude 156.45332W and latitude 15.37665. The distance between the end points is 6.209 km. Here the sea water depth varies from 5529 m to 5437 m and the free air anomaly varies from -11.1 mgals to -6.8 mgals. We sampled every observation point, and there are 21 sample points having an average distance of 310 m between them. The free air anomaly and observed bathymetry are shown in Fig. 2.4a and 2.4b respectively. The correlation between them is 0.2232. The correlation here is very poor as compared to the previous example. The results of the computed bathymetry along with the true bathymetry and the residuals are shown in Fig. 2.5a and 2.5b respectively. The correlation value between them is 0.9437. The standard deviation of error is 6.4488 m and the maximum deviation of error is 21 m. Here the wavelength range of admittance considered is 3.9 - 5.9 km. Even though the correlation between the observed bathymetry and the free air anomaly is very low in this case, the computed profile has good correlation with the observed one.

2.4.4 Third Example

In this example the selected bathymetry and free air anomaly data cover the area from longitude 155.852W to 155.312W and from latitude 17.0002N to 17.9978N. Here the depth of sea water varies from 5125 m to 4579 m and free air anomaly varies from -28.47 mgals to 11.45 mgals. We gridded the data using the SURFER (Surfer 1990) software package by selecting inverse distance gridding method with the grid size 50 x 50. We computed bathymetry of the inner segment of this grid for 42 x 42 points starting from outermost 4 points on each side. The free air anomaly and observed bathymetry 3D pictures are shown in Fig. 2.6 and 2.7 respectively. The correlation between them is 0.5683. The 3D picture of the computed bathymetry is shown in Fig. 2.8. The correlation between them is 0.9976. The correlation obtained here is

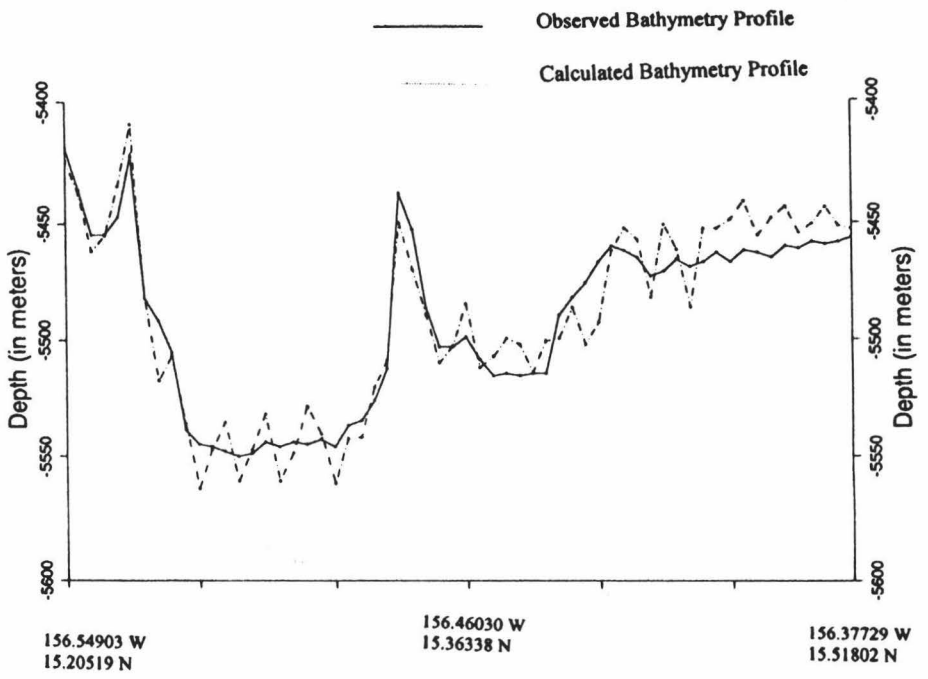


(a)

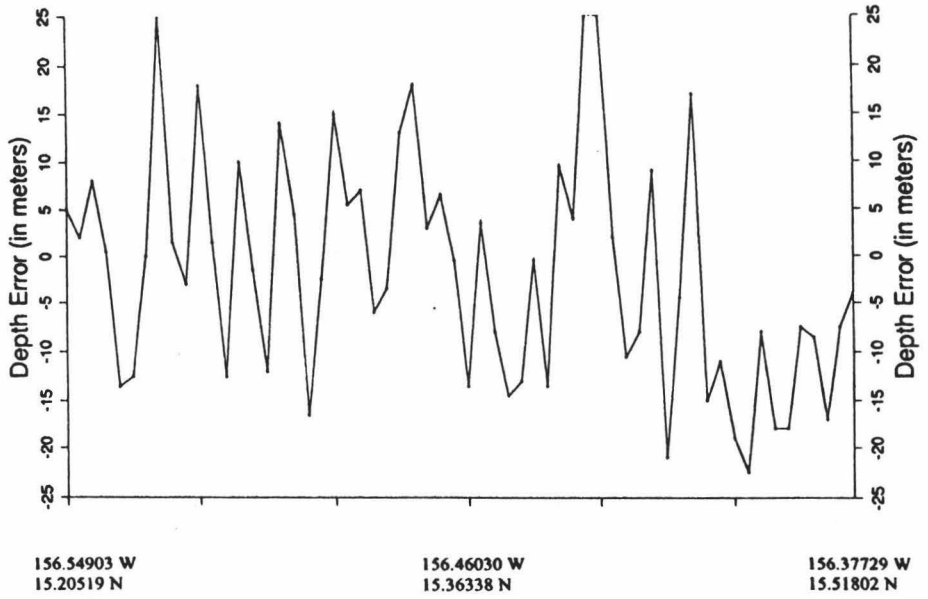


(b)

Figure 2.2: Profiles of (a) free-air anomaly, (b) observed bathymetry for longer traverse line

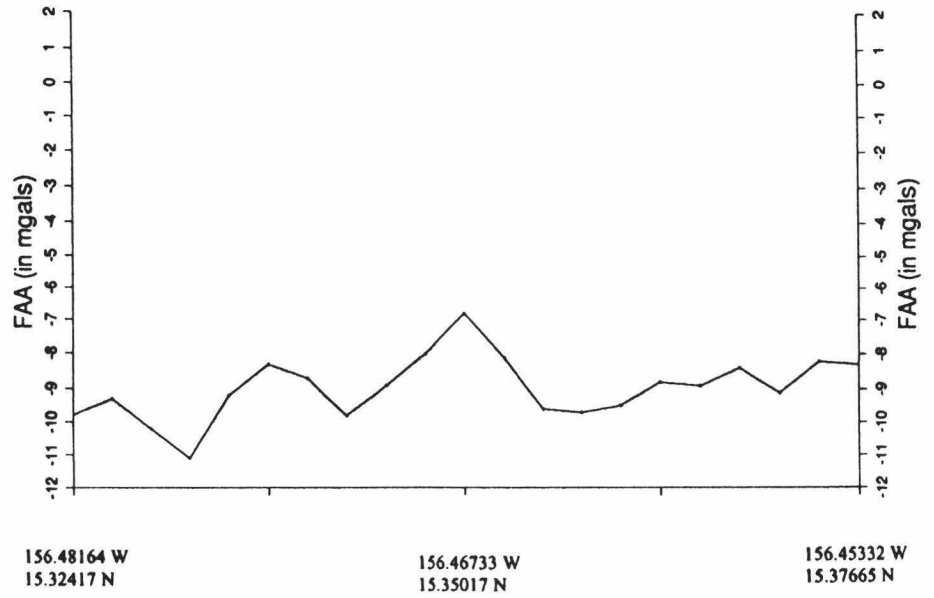


(a)

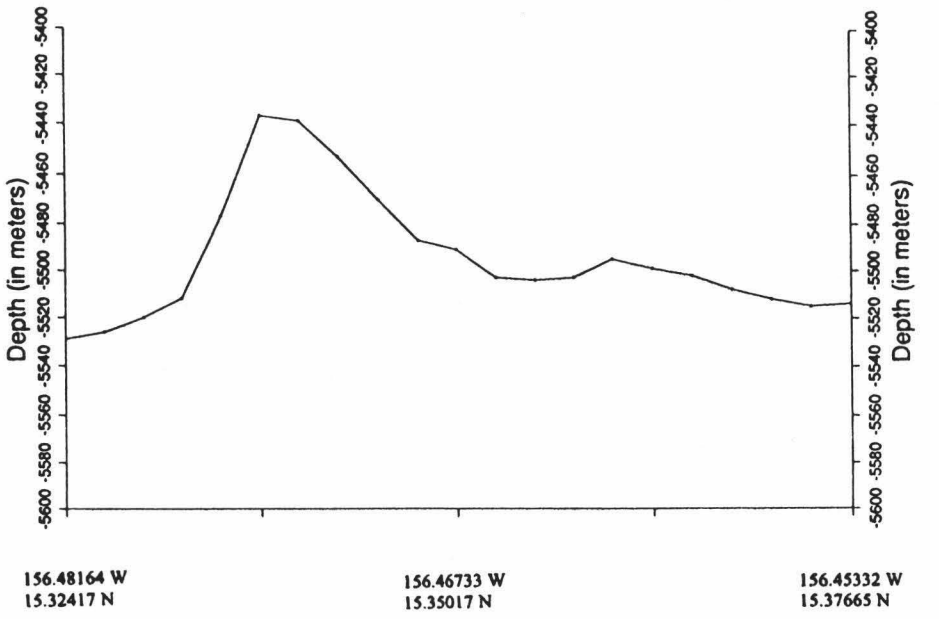


(b)

Figure 2.3: Profiles of (a) calculated bathymetry along with observed bathymetry, (b) residuals (observed minus computed) bathymetry for longer traverse line

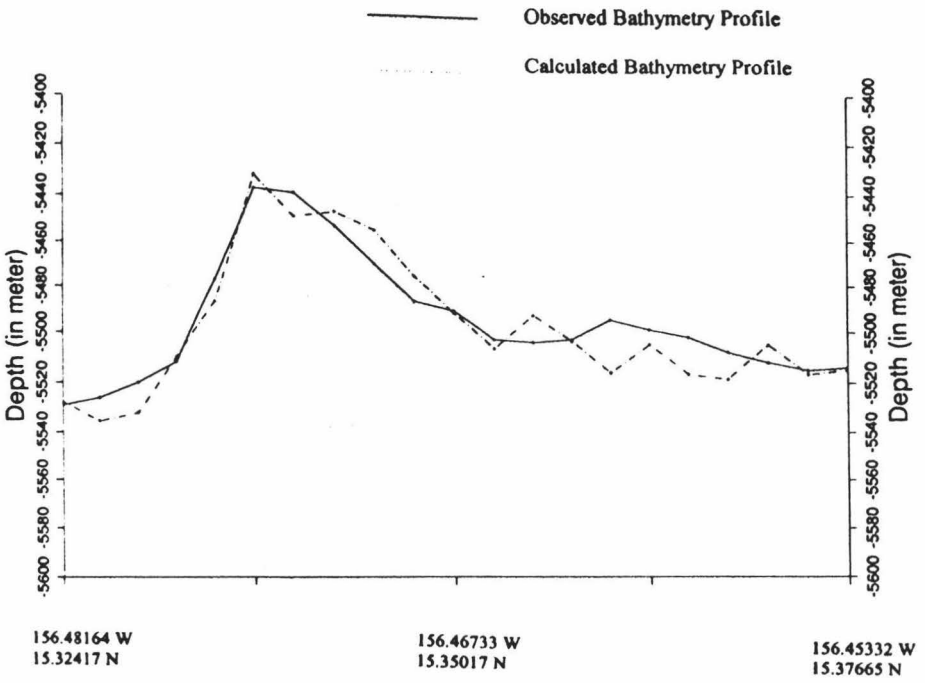


(a)

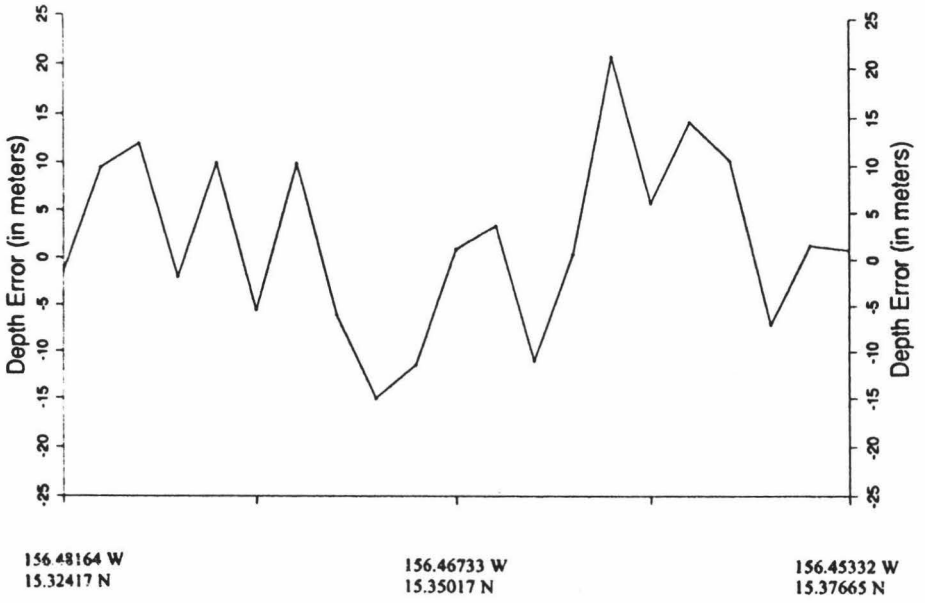


(b)

Figure 2.4: Profiles of (a) free-air anomaly, (b) observed bathymetry for smaller traverse line



(a)



(b)

Figure 2.5: Profiles of (a) calculated bathymetry along with observed bathymetry, (b) residuals (observed minus computed) bathymetry for smaller traverse line

much higher than the previous two examples. The standard deviation of error is 5.3616 m and the maximum deviation of error is 24 m. In calculating the admittance we use 1D fast fourier transform (FFT) along the grid line instead of using the 2D FFT technique used by McNutt (1979), and Goslin and Diament (1987). Comparing Fig. 2.7 and 2.8 we notice that the computed bathymetry follows the trend of the true profile in an oscillatory way. The noise is more prominent where the true profile is more or less flat. This is because we have data only along the track line and in the flat area of the observed profile there are no data. The gridding method is used here to help cancel the random errors in the computed bathymetry.

2.5 DISCUSSION

In this chapter we have tested the feasibility of computing bathymetry from shipboard gravity data. The result shows an error of less than 1 % of depth, which is within the acceptable error range for measuring depth according to International Hydrographic Bureau standard. The smaller the number of computation points the better result we obtained. The result was also improved with gridded data because the random errors were canceled more in computation.

For inversion we developed an objective function from theoretical admittance formulas for different models. The objective function was derived assuming that the crustal density and other parameters are constant in the study area. The result shows that this constraint in the parameters works well in this area because it is a small area and it is not tectonically active. In tectonically active regions (e.g., midoceanic ridge, subduction zone) we may not assume that density, compensation depth, and other parameters are constant for the whole area.

We performed inversion to compute bathymetry of one point and then move on

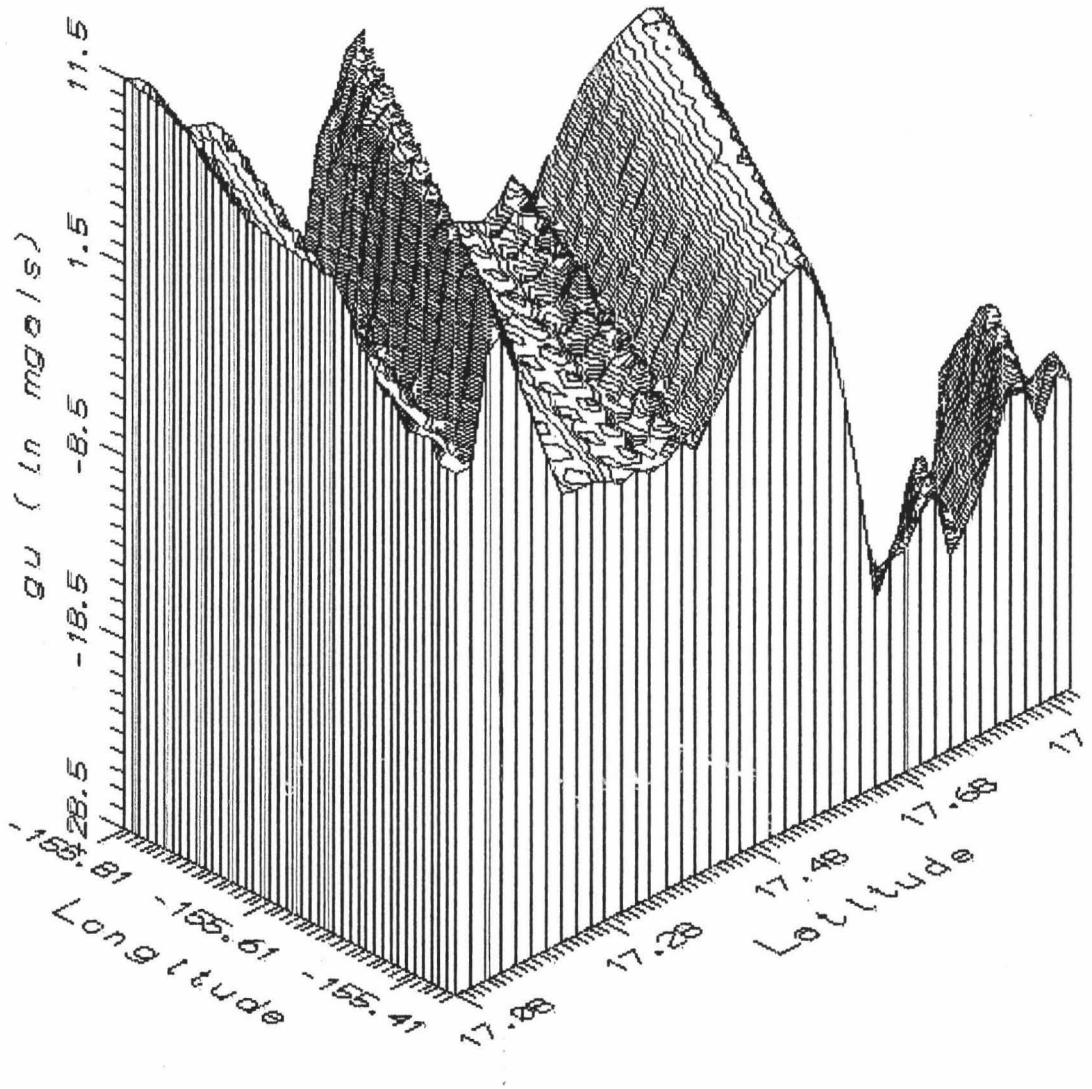


Figure 2.6: 3D plot of observed free-air anomaly

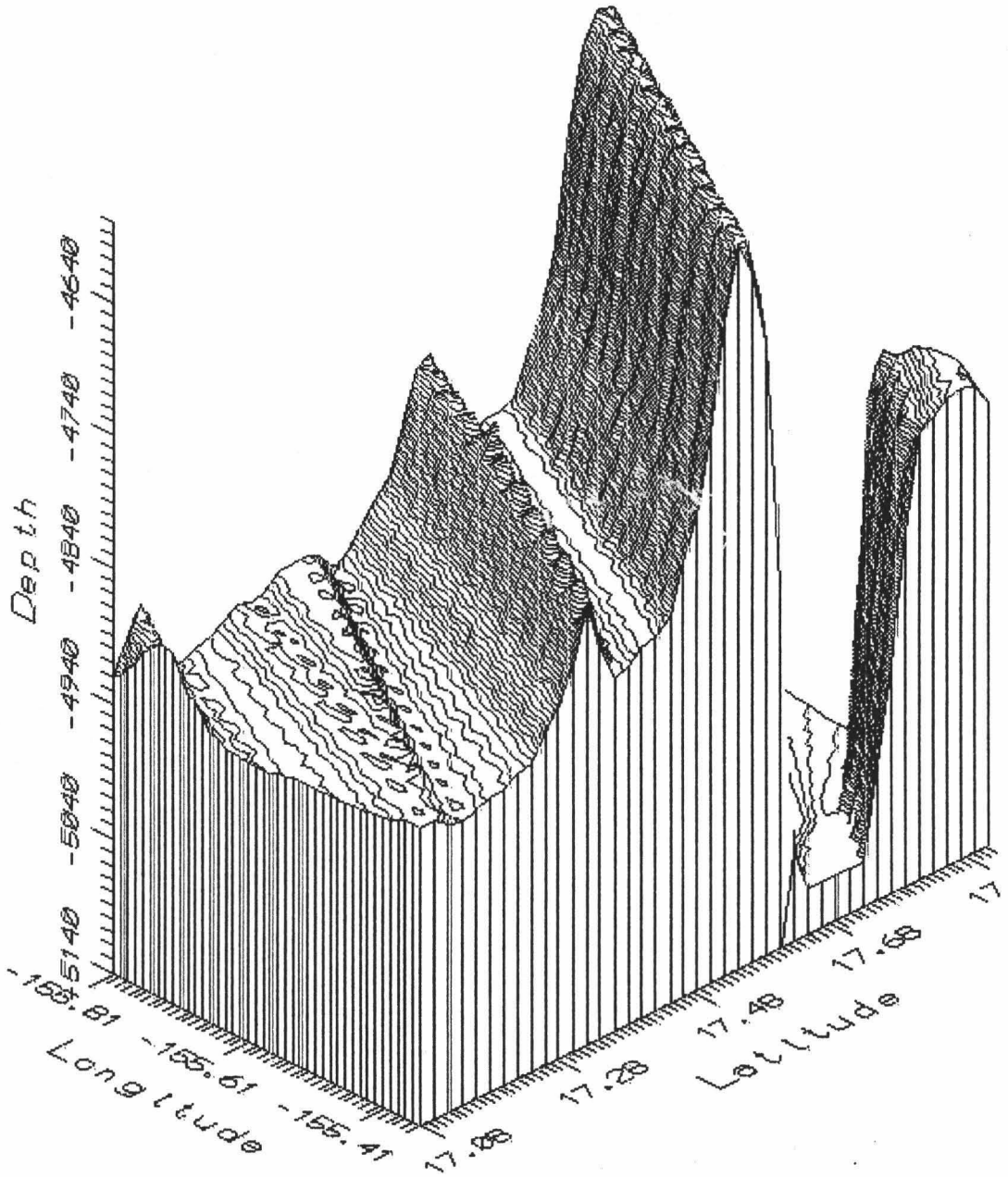


Figure 2.7: 3D plot of observed bathymetry

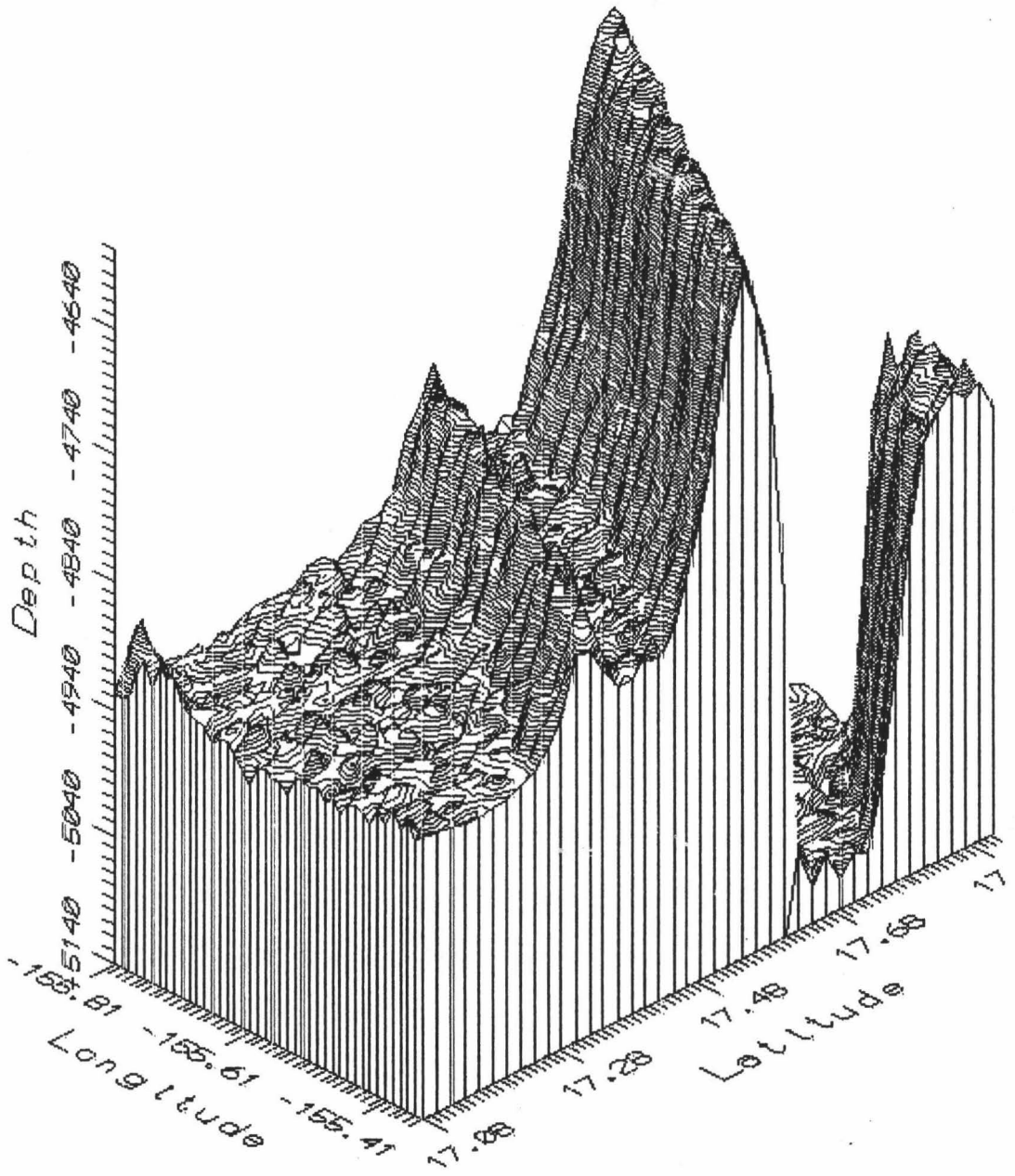


Figure 2.8: 3D plot of calculated bathymetry

to the next point and so on. This process incurred cumulative errors in the computed bathymetry, and the minimum value of the objective function became larger as we moved further from the starting point. We minimized this effect by starting at two ends of the traverse line and moved in the opposite directions; the final bathymetry was the average of these two profiles. In the case of gridded data this effect was minimized further because we could move in four different directions and the average consisted of four different profiles.

The minimization of objective function was performed by exhaustive search of the solution space. We could afford the CPU time for this search method because the search space is not too large. It took 48 hours of CPU time on a SUN SPARC Station II to compute the gridded data example. In some other data sets for larger solution space, one has to use a non-linear inversion technique (Tarantola 1987). There is a question whether the objective function we derived here has an unique solution or not. Following the exhaustive search of the solution space for this data set and another data set (near Guam, the result of which we have not shown here), we found a unique solution for every computation point.

In the future we plan to develop an inversion algorithm wherein the inversion is performed for all points with unknown bathymetry at the same time. This method will eliminate the cumulative error problem, but the solution space will be too large to be searched by a linear optimization technique. We may use a non-linear optimization technique such as simulated annealing (Aarts and Korst 1989) or genetic algorithm (Goldberg 1989).

2.6 REFERENCES

Aarts, E. & Korst, J., 1989. *Simulated Annealing and Boltzmann Machines*, John Wiley and Sons, New York.

Ashalatha, B., Subrahmanyam, C. & Singh, R. N., 1991. Origin and compensation of Chagos-Laccadive ridge, Indian ocean, from admittance analysis of gravity and bathymetry data, *Earth Planet. Sci. Lett.*, **105**, 47-54.

Basu, A. & Saxena, N.K., 1993. Bathymetry computation from free-air anomaly data, *Marine Geodesy*, **16**, 325-336.

Black, M.T. & McAdoo, D.C., 1988/89. Spectral analysis of marine geoid heights and ocean depths: Constraints on models of lithospheric and sublithospheric processes, *Marine Geophys. Res.*, **10**, 157-180.

Bowin, C. & Milligan, J., 1985. Negative gravity anomaly over spreading rift valleys: Mid Atlantic ridge at 26° N., *Tectonophysics*, **113**, 233-256.

Dalloubeix, C., Fleitout, L. & Diament, M., 1988. A new analysis of gravity and topography data over the Mid Atlantic Ridge: non compensation of axial valley, *Earth Planet. Sci. Lett.*, **88**, 308-320.

Estep, L., 1993. A Review of Airborne Lidar Hydrography (ALH) Systems, *Hydrography Journal.*, **67**, 25-42.

Forsyth, D.W., 1985. Subsurface loading and estimates of the flexural rigidity of continental lithosphere, *J. Geophys. Res.*, **90**, 12623-12632.

Goldberg, D. E., 1989. *Genetic Algorithms in search, optimizations, and machine learning*, Addision Wesley Publishing Co., New York, U.S.A.

Goslin, J. & Diament, M., 1987. Mechanical and thermal isostatic response of the Del Cano Rise and Crozet Bank (Southern Indian Ocean) from altimetry data, *Earth Planet. Sci. Lett.*, **84**, 285-294.

Jung, W. & Vogt, P. R., 1992. Predicting bathymetry from Geosat-ERM and shipborne profiles in the South Atlantic Ocean, *Tectonophysics*, **210**, 235-253.

Khan, M. A., Woollard, G. P. & Daugherty, K. I., 1971. Statistical analysis of the relation of marine gravity anomalies to bathymetry. *Hawaii Inst. Geophys. Rept.*, **HIG 71-20**.

Kister, B. & Wilhelm, H., 1988. A regional compensation model for the Great Meteor Seamount using the response function technique, *Tectonophysics*, **147**, 1-17.

Lewis, B.T.R. & Dorman, L. M., 1970. Experimental isostasy, 2, An isostatic model for the U.S.A. derived from gravity and topographic data, *J. Geophys. Res.*, **75**, 3367-3386.

Maia, M., Diament, M., & Recq., M., 1990. Isostatic response of the lithosphere beneath the Mozambique Ridge (SW Indian Ocean) and geodynamic implications, *Geophys. J. Int.*, **100**, 337-348.

McKenzie, D. & Bowin, C., 1976. The relationship between bathymetry and gravity in the Atlantic ocean, *J. Geophys. Res.*, **81:11**, 1903-1915.

McNutt, M.K., & Parker, R. L., 1978. Isostasy in Australia and the evolution of the compensation mechanism, *Science*, **199**, 773-775.

McNutt, M., 1979. Compensation of oceanic topography: An application of the response function technique to the Surveyor area, *J. Geophys. Res.*, **84:B13**, 7589-7598.

Parker, R.L., 1973. The rapid calculation of potential anomalies, *Geophys. J. Roy. Astron. Soc.*, **31**, 447-455.

Surfer, 1990. Golden Software, Inc., Golden, Colorado, U.S.A.

Talwani, M., Poppe, H. R., & Rabinowitz, P. D., 1972. Gravimetrically determined geoid in the western North Atlantic, Sea surface topography from space, *NOAA Tech. Rep.*, **ERL-228-AOML-7-2**, 1-34.

Tarantola, A., 1987. *Inverse Problem Theory*, New York: Elsevier Science Publishing Company Inc.

Watts, A. B., 1978. An analysis of isostasy in the world's oceans, 1. Hawaiian-Emperor seamount chain, *J. Geophys. Res.*, **83:B12**, 5989-6004.

Watts, A. B., McKenzie, D. P., Parsons, B. E., & Roufosse, M., 1985. The relationship between gravity and bathymetry in the Pacific Ocean, *Geophys. J. R. Astron. Soc.*, **83**, 263-298.

Chapter 3

Global Optimization Methods

In the past seismic waveform inversion has often been based on the least-squares (or more generally, least error or maximum likelihood) principle; and, even though it is successful in many cases (Mora 1987, Pica et al. 1990), it fails in others. Specifically, when a priori information for a good starting model is not available, then this technique, coupled with a local, gradient based algorithm can converge to a local minimum of the misfit function rather than to the global minimum. Global (non gradient) optimization methods such as Monte Carlo search (Press et al. 1988), simulated annealing (Laarhoven & Aarts 1987), or genetic algorithms (Goldberg 1989) have been applied recently to several geophysical inverse problems. In this chapter we discuss these three methods and examine the simulated annealing method in detail.

3.1 MONTE CARLO SEARCH

The Monte Carlo search is truly global since it is virtually a memoryless random walk in the solution space. When it generates a new solution it neglects the information gained from the previous samples. As a consequence, this search is not biased towards the true solution, it relies totally on random exploitation of the model space. This results in spending too much computation time in exploring unfavorable regions of

the model space. As a result, this search is too slow to be very efficient for large-scale problems.

3.2 GENETIC ALGORITHMS

Genetic algorithms (GA) were originally developed by Holland (1975) as models for learning in biological and artificial systems. This method, originally confined to problems of artificial intelligence (AI), has recently been used for various optimization problems in engineering, computer science, and other fields (Davis 1991). This method has also been applied in seismic waveform inversion (Stoffa & Sen 1991, Nolte & Frazer 1994).

The GA method has a similarity to the genetic evolution of biological systems. It simulates natural genetics by following an entire population of trial models encoded as finite-length strings (chromosomes), wherein model parameters are part of the string. A typical GA uses three operators (reproduction, cross-over, and mutation) on the elements of the population (models) to find the best fit model. Reproduction defines the process by which an interim population is generated from a randomly selected initial population by selecting models from the original group, with the likelihood of selection determined by some measure of the cost functions. Crossover defines the outcome, as gene exchange, whose specific values are called *alleles*. From the parent population we create a new generation, each of which is derived from a mixing (crossover) of the bit strings from two parents. Initially, all parents are randomly paired off to produce couples and then each pair of parents is considered for a possible crossover. In practice, a random number is generated between 0 and 1 to determine whether the current selected pair is to be crossed over. If the value is less than the predetermined constant probability of crossover, then a position is chosen at random

and by cutting and transposing the two segments two new strings are created. This step must be modified and particularized to different types of problems. Finally, mutation is the introduction of a random element, often used to amend the result of a gene exchange when the outcome does not successfully meet appropriate restrictions. A small mutation probability is used to control the likelihood of an individual bit in each model being altered in parity.

These three processes perform different roles in the GA. The reproduction step affects the survival of the fittest between generations. The crossover step controls the degree of mixing and sharing of information that occurs between the models. The mutation step is necessary for maintaining diversity in the population. Without mutation certain *alleles* might be irretrievably lost in the selection process. In this respect, mutation is more helpful for the success of GA in solving a problem, than which is found in nature. With these three steps (reproduction, crossover and mutation) GA is able to explore the model space very efficiently, which makes it superior to the random search Monte Carlo technique.

3.3 SIMULATED ANNEALING

Simulated annealing (SA) is another global optimization method. The name is derived from the process of annealing, which is the process of finding low energy states of a solid by initially melting the substance, and then lowering the temperature slowly and observing the formation of crystals. The minimum energy state or the ground state may be viewed as corresponding to the minimum of the cost function (or energy function) in the optimization problem. Since its introduction in 1983 (Kirkpatrick et al. 1983) SA has been applied to a fairly large number of different problems including the travelling salesman problem, image processing, flow jobshop scheduling, pollution

control, coastguard deployment, and DNA mapping (Aarts & Korst 1989). In the geophysical exploration problem, Rothman (1985, 1986) first introduced this technique to estimate large residual statics correction, and thereafter it has been used in various other problems in seismology (Landa et al. 1989, Sen and Stoffa 1991, Nørmark and Mosegaard 1993).

3.3.1 Overview

Simulated Annealing (SA) is an algorithmic approach to solving optimization problems. This technique, originally applied in the physical sciences (Metropolis et al. 1953), has been the subject of a great deal of interest for obtaining 'good' solutions to a wide range of problems. Metropolis et al. (1953) attempted to simulate the behavior of an ensemble of atoms in equilibrium at a given temperature. They constructed a mathematical model of the behavior of such a system that contained a method for minimizing the total energy of the system. Later Kirkpatrick et al. (1983) and Černý (1985) independently demonstrated the potential of SA for combinatorial or discrete optimization problems. They have used SA in the Travelling Salesman Problem (Liu 1968, Press et al. 1988), in which the goal is to find the shortest route that touches every one of a number of cities once and returns one to the point of departure. The problem is similar to finding a Hamiltonian circuit that has the minimal sum of distances. Vanderbilt & Louie (1984) have shown that SA can be used to obtain global solutions for continuous optimization problems. As a result, the SA approach has been the subject of intensive study by mathematicians, statisticians, physicists, computer scientists, engineers, and it has also been applied to numerous other areas.

3.3.2 Method

The name, *simulated annealing algorithm*, derives from an analogy between solving optimization problems and simulating the annealing of solids as proposed by Metropolis et al. (1953). In condensed matter physics, 'annealing' is a thermal process by which a solid at a high temperature is brought to a low temperature crystalline state by gradual reduction of temperature. During the annealing process, if the temperature is reduced very quickly, or quenched then the crystal will be defective. In particular, the cooling rate must be very slow near the freezing temperature.

From statistical physics (e.g. Landau and Lifshitz 1980) we know that for a system in thermal equilibrium at a temperature T the probability $P(E_i)$ of a state having energy E_i is governed by the Boltzmann distribution:

$$P(E_i) = \frac{\exp(-E_i/k_B T)}{\sum_{j=1}^{k} \exp(-E_j/k_B T)} \tag{1}$$

where k_B is the Boltzmann constant. One can note here that at high temperatures all energy states are equally likely, while at low temperatures the system is more likely to be in states of low energies.

In the optimization problem, the objective function to be minimized corresponds to the energy (E) of the states of the solid. In the most common version of the SA technique, 'downhill' (i.e. improving) perturbations of the computed solution are always accepted but an 'uphill' (i.e. degrading) perturbation of step ΔE is accepted with probability $\exp(-\Delta E/T)$, where T is a control parameter corresponding to temperature in the analogy. This mechanism saves the SA algorithm from being trapped in a local minimum in its search for the global minimum.

3.3.3 Algorithm

There are two popular SA algorithms that have been successfully applied in various problems. One of them is the Metropolis algorithm (Metropolis et al. 1953, Kirkpatrick et al. 1983), and the other one is the 'heat bath' method (Rothman 1986, Miyatake 1986). We follow the 'heat bath' method in our application of SA algorithm for inversion. This method involves the following steps:

step 0: Choose a random starting model $\mathbf{m_0}$ with 'energy' or objective function value E_0. Select an initial value T_0 for the 'temperature'.

step 1: For each temperature stage do the following:

step 1.1: To update m_1, calculate the energies $E_{11}, E_{12}, \cdots, E_{1k}$ for all the possible values of m_1 keeping the other components of $\mathbf{m_0}$ fixed at their current values.

step 1.2: Choose the new value of m_1 by sampling from the following distribution:

$$P(m_1 = m_{1i}) = \frac{\exp(-E_i/k_B T)}{\sum_{j=1}^{k} \exp(-E_j/k_B T)} \tag{2}$$

step 1.3: Update the remaining components of $\mathbf{m_0}$ the same way as described in steps 1.1 and 1.2.

step 1.4: If 'thermal equilibrium' is not reached, go back to step 1.1. Otherwise go to step 2.

step 2: If the 'stopping criterion' is not satisfied, reduce the temperature and go back to step 1.

The basic structure of the Metropolis algorithm is almost the same as the 'heat bath' algorithm except that step 1.1 and step 1.2 are replaced by

step 1.1: To update m_1, generate a 'feasible' value of m_1 by a small random perturbation from the current configuration. Evaluate the difference in energy ΔE between these two models.

step 1.2: If $\Delta E \leq 0$ (i.e. the candidate configuration has a lesser energy than the current configuration), accept this solution and replace the current configuration with the newly accepted configuration.

If $\Delta E > 0$ (i.e. the candidate configuration has a higher energy than the current configuration), accept the solution with a probability given by

$$P(\Delta E) = \frac{\exp(-E_i/k_B T)}{\sum_{j=1}^{k} \exp(-E_j/k_B T)} \tag{3}$$

One notes that, in the SA algorithm, every state of the system is accessible from every other state in one iteration or sweep. This is analogous to the transitions of a Markov chain. So, SA can be represented by a Markov chain, whose transition probabilities depend on the 'temperature' parameter. For the homogeneous Markov model of SA, the number of iterations must be large at a given temperature to bring the system into 'thermal equilibrium'. For the inhomogeneous Markov model, it is necessary that the cooling is carried out very slowly from one iteration to the next iteration.

3.3.4 Finer Detail

In any implementation of the algorithm, a 'cooling schedule' must be specified that determines when and by how much the 'temperature' is to be reduced. Initially the temperature is set at a relatively high value, so that most changes are accepted and

there is little chance of the algorithm being unable to move out of a local minimum during the early stages. A scheme is then required for reducing the 'temperature' parameter at the end of each stage and also for deciding the number of repetitions to be performed at each value of the 'temperature'. As discussed before, if we follow the inhomogeneous Markov model of SA, then we perform only one iteration at each 'temperature', but the 'temperature' reduction should be very small. It is important to mention here is that reducing the 'temperature' too quickly tends to get the algorithm trapped at a local minimum. On the other hand, reducing the 'temperature' very slowly consumes more CPU time, which is undesirable. Finally, a stopping criterion is required to terminate the process. One can stop after performing a fixed number of iterations, or if the 'temperature' goes below a pre-set final value, or if the energy changes are small at successive temperatures. Clearly, there is no well-defined 'cooling schedule' that may be employed. The judicious selection of a particular 'cooling schedule' is more of an art rather than a science. The efficiency of the SA algorithm for a particular problem depends on the experience and the ingenuity of the user. If the algorithm is applied properly then it will converge to a globally optimal solution. The convergence of the SA algorithm has been studied extensively (e.g. Geman and Geman 1984, Gidas 1985, Mitra et al. 1986, Haario and Saksman 1991) in the last few years.

3.3.5 Critical Temperature

As discussed before the efficiency of the SA algorithm depends primarily on the cooling schedule. During the cooling process one should spend a long time at the 'temperature' close to the freezing point. In the physical process of producing crystals from the molten substance, crystallization starts occurring near this 'temperature', which

we will call the 'critical temperature' of the system. From our experience we have seen that determining the critical 'temperature' is one of the most important factors regarding the efficiency of the SA algorithm. To determine the 'critical temperature' we followed the procedure used by Basu and Frazer (1989, 1990), which involves executing the SA process for a fixed number of sweeps at a fixed 'temperature' and then calculating the average energy for those sweeps. This routine is performed at different 'temperatures' ranging from a very high 'temperature' where almost all the states are equally probable to a very low 'temperature' where the system is almost frozen.

The above mentioned procedure is described in the following algorithm:

step 0: Choose a random starting model $\mathbf{m_0}$ with 'energy' or objective function value E_0 and the number of sweeps N_S, which is fixed for each 'temperature'. Select a high 'temperature' T_h and a low 'temperature' T_l, and divide this temperature range equally on the logarithmic scale.

step 1: For each temperature stage using the same set of random numbers do the following:

step 1.1: Perform SA algorithm N_S times and calculate the energy at the end of each sweep.

step 1.2: Calculate the average energy $\overline{E}$ by adding all the energies of N_S sweeps and then dividing the sum by N_S.

step 1.3: Plot this value of $\overline{E}$ in the graph of $\overline{E}$ vs. log T.

step 2: Repeat step 1 for different set of random numbers.

The plot of $\overline{E}$ vs. log T is a curve that has a minimum at some temperature which we consider as the 'critical temperature' of the system. At high 'temperature' $\overline{E}$ is high because virtually all transitions are accepted and SA is unbiased enough to pick up only the good solutions. At low 'temperature' $\overline{E}$ is also high because the

probability of accepting solutions worse than the current one is small and SA picks up the good solutions close to the starting model only. At 'temperatures' in between these two extremes, SA behaves more effectively, and at the 'critical temperature' SA is most effective. The number of sweeps N_S to carry out at each temperature should be large enough so that the graphs of $\overline{E}$ vs. log T for each set of random numbers are not very different and also that $\overline{E}$ has a smooth well-defined minimum.

3.3.6 SA at Constant Temperature

A detailed account of this procedure has been discussed by Frazer and Basu (1992) and therefore only a brief summary of the method will be given here. The goal of a typical SA in an optimization problem is to find the best model that minimizes or maximizes the cost function. When SA is applied to an inversion problem then its goal should be not only to find the best model, but also other good models. From our experience we have observed that at the critical temperature SA consistently picks up good models; and this is also the reason for the average energy function to be the highest at this temperature as we have seen in the previous section. We can also compute the a posteriori probability density function from the chosen models and then construct the best fit model.

Suppose A is a set of chosen models at the critical temperature, $\sharp(A)$ is the number of models in A, and $A_{i\mu}$ is the subset of models $\mathbf{m}$ whose i^{th} element m_i is equal to μ,and $\sharp(A_{i\mu})$ is the number of models in $A_{i\mu}$. Then the probability that m_i is equal to μ is given by the following equation:

$$\sigma_i(m_i = \mu) = \frac{\sharp(A_{i\mu})}{\sharp(A)} \tag{4}$$

The above equation is the i^{th} marginal distribution of the a posteriori density function

$(\sigma_M)^{1/T_{cr}}$. The best fit model $\check{\mathbf{m}}$ denotes the array $(\check{m}_1, \check{m}_2, ...)$ where $\check{m}_i$ is the component value of $\check{\mathbf{m}}$ at depth z_i that maximizes σ_i.

We also estimate the a posteriori model covariance matrix given by the following equation:

$$cov(\mathbf{m})_{ij} = \frac{1}{\sharp(A)} \sum_{\mathbf{m} \in A} [m_i - \check{m}_i][m_j - \check{m}_j] \tag{5}$$

The covariance matrix shows the uncertainty in the evaluation of model parameters. The correlation matrix is computed from the covariance matrices by the following equation

$$cor(\mathbf{m})_{ij} = \frac{cov(\mathbf{m})_{ij}}{\sqrt{cov(\mathbf{m})_{ii} cov(\mathbf{m})_{jj}}} \tag{6}$$

The off-diagonal elements of the cor($\mathbf{m}$) matrix show the dependency of model parameters on each other. The larger the value of an off-diagonal element the more dependent the corresponding parameters are on one another. Thus in a SA run containing a number of sweeps we construct the best fit model $\mathbf{m}$, the a posteriori probability density (PPD) function $\sigma(\mathbf{m})$, and the correlation matrix cor($\mathbf{m}$). In the next chapters we will use the SA method on two-offset VSP data, and on cross-borehole data to determine the slowness of the layers between the source well and the receiver well.

3.4 REFERENCES

Aarts, E. & Korst, J., 1989. *Simulated Annealing and Boltzmann Machines*, John Wiley and Sons, New York.

Basu, A. & Frazer, L. N., 1989. Rapid determination of critical temperature in simulated annealing, *Soc. Industrial and Appl. Math. Workshop on Geophysical Inversion, Final Program*, Sept. 27-29, Houston, Texas.

Basu, A. & Frazer, L. N., 1990. Rapid determination of critical temperature in simulated annealing inversion, *Science* , **249**, 1409-1412.

Černý, V., 1985. Thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm, *J. Opt. Theory Appl.*, **45**, 41-51.

Davis, L., 1991. *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York, U.S.A.

Frazer, L. N. & Basu, A., 1992. Bayesian inversion by statistical physics with an application to offset VSP, submitted to *Geophys. J. Int.*.

Geman, S. & Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian distribution of images, *IEEE Trans. Patt. Anal. Mach. Int.*, **6**, 721-741.

Gidas, B., 1985. Non-stationary Markov chains and convergence of the annealing algorithm, *J. Statistical Phys*, **39**, 73-131.

Goldberg, D. E., 1989. *Genetic Algorithms in search, optimizations, and machine learning*, Addision Wesley Publishing Co., New York, U.S.A.

Haario, H. & Saksman, E., 1991. Simulated annealing process in general state space, *Adv. in applied probability*, **23**, 866-893.

Holland, J. H., 1975. *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Harbor, Michigan, U.S.A.

Kirkpatrick, S., Gelatt, C. D. Jr. & Vecchi, M. P., 1983. Optimization by simulated annealing, *Science*, **220**, 671-680.

Laarhoven, P. J. M. Van & Aarts, E. H. L., 1987. *Simulated Annealing: Theory and applications*, D. Reidel, Boston, U.S.A.

Landa, E., Beydoun, W. & Tarantola, A., 1989. Reference velocity model estimation from prestack waveforms: Coherency optimization by simulated annealing, *Geophysics*, **54**, 984-990.

Landau, L. D. & Lifshitz, E. M., 1980. *Statistical Physics*, Pergamon, Oxford.

Liu, C. L., 1968. *Introduction to Combinatorial Mathematics*, McGraw-Hill

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

Mitra, D., Romeo, F. & Sangiovanni-Vincentelli, A., 1986. Convergence and finite time behavior of simulated annealing, *Advances in applied probability*, **18**, 747-771.

Miyatake, Y., Yamamoto, M., Kim, J. J., Toyonaga, M. & Nagai, O., 1986. On the implementation of the 'heat bath' algorithms for Monte Carlo simulations of classical Heisenberg spin systems, *J. Phys. C. Solid State Physics*, **19**, 2539-2546.

Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data, *Geophysics*, **52**, 1211-1228.

Nolte, B. & Frazer, L. N., 1994. Vertical seismic profile inversion with genetic algorithms, *Geophys. J. Int*, **117**, 162-178.

Nørmark, E. & Mosegaard, K., 1993. Residual statics estimation: scaling temperature schedules using simulated annealing, *Geophysical Prospecting*, **41**, 565-578.

Pica, A., Diet, J. P. & Tarantola, A., 1990. Nonlinear inversion of seismic reflection data in a laterally invariant medium, *Geophysics*, **55**, 284-292.

Press, H. W., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T., 1988. *Numerical Recipes*, Cambridge University Press.

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

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

Sen, M. K. & Stoffa, P. L., 1991. Nonlinear one-dimensional seismic waveform inversion using simulated annealing, *Geophysics*, **56**, 1624-1638.

Stoffa, P. L. & Sen, M. K. 1991. Nonlinear multiparameter optimization using genetic algorithms: inversion of plane wave seismograms, *Geophysics*, **56**, 1794-1810.

Vanderbilt, D. & Louie, S. G., 1984. A Monte Carlo simulated annealing approach to optimization over continuous variables, *Journal of Computational Physics*, **56**, 259-271

Chapter 4

Offset VSP Inversion using Simulated Annealing

4.1 INTRODUCTION

Vertical Seismic Profiling (VSP) data are often more reliable to determine the subsurface geology than sonic logs, because VSP data are not as sensitive to the borehole conditions as sonic logs. Contrary to surface seismic profiling, in VSP, the receivers are placed in the well to record the data. There are two kinds of VSP data depending on the position of the source. In one case, the surface source is close to the well head (zero offset VSP) and in the other case, the source is away from the well head (offset VSP). A comprehensive reference of this technique can be found in the book by Hardage (1983). Here we perform inversion using the SA method at a fixed temperature on a two-offset VSP data sets to determine the slowness of the layers between the two boreholes.

4.2 ENERGY FUNCTION

The actual geometry of the experiment consists of two shots and one borehole containing the receivers. Assuming the earth is horizontally stratified near the shots and

the borehole, we replace this geometry by one shot and two boreholes. This extra borehole is at the same position as of one of the shots (Fig. 4.1).

4.2.1 Correlation Function

The forward modelling in our inversion consists of migrating the data recorded in the first borehole to the second borehole using different slowness profiles. We then compare the migrated data and the actual data of the second borehole. If our velocity estimation is correct then the correlation between these two data sets will be the maximum. Let's denote the seismic data in the first and second borehole by $U_1(t,z)$ and $U_2(t,z)$ respectively, in which t is time and z is receiver depth. Let's denote the migrated data by $U_2^M(t,z)$. The correlation between $U_2(t,z)$ and $U_2^M(t,z)$ are given as

$$\phi(U_2^M, U_2) = \frac{1}{N_z} \sum_{j=1}^{N_z} \left(\frac{\sum_{t=1}^{N_t} U_2^M(t,z_j) U_2(t,z_j)}{\sqrt{\sum_{t=1}^{N_t} U_2^M(t,z_j)^2} \sqrt{\sum_{t=1}^{N_t} U_2(t,z_j)^2}} \right) \tag{1}$$

in which N_t is the number of time sample points and N_z is the number of receivers. The correlation value would be 1.0 if our migration algorithm is perfect, the velocity profile is true, and there is no noise.

4.2.2 Penalty Function

In some cases, especially with many layers, the energy function value for an oscillatory velocity profile is the same as for a smoother velocity profile. Since we want to select a smoother velocity profile we use a penalty function that gives a high penalty for an oscillatory velocity profile and no penalty for constant profile. Following Basu &

Frazer (1990) the penalty function $p(\mathbf{m})$ is written as

$$p(\mathbf{m}) = \frac{\sum_{i=2}^{N-1} |s_{i+1} - 2\,s_i + s_{i-1}|}{2\,(N-2)(s_{max} - s_{min})} \tag{2}$$

where s_i is the slowness of the i^{th} layer, s_{max} is the maximum allowed slowness, s_{min} is the minimum allowed slowness, and N is the number of layers.

4.2.3 Objective Function

Our objective function or the energy function consists of both the correlation function and the penalty function. We write the energy function as

$$E(\mathbf{m}) = -\phi(U_2^M, U_2) + \epsilon p(\mathbf{m}) \tag{3}$$

where ϵ is the penalty weight. The method used to determine ϵ is described in the following algorithm:

step 1: Find the critical temperature of the system with $\epsilon = 0$. Choose a random starting model $\mathbf{m_0}$ with 'energy' or objective function value E_0 and the number of sweeps N_S, which is fixed for each ϵ . Select a high penalty value ϵ_h and a low penalty value ϵ_l, and divide this range equally in the linear scale.

step 2: For each ϵ using the same set of random numbers do the following:

step 2.1: Perform the SA algorithm for N_S times and calculate E($\mathbf{m}$), $\phi(U_2^M, U_2)$, p($\mathbf{m}$) at the end of each sweep.

step 2.2: Calculate the average of E($\mathbf{m}$), $\phi(U_2^M, U_2)$, p($\mathbf{m}$).

step 2.3: Plot $\overline{p(\mathbf{m})}$ / $\overline{-\phi(U_2^M, U_2)}$ vs. ϵ.

The plot of $\overline{p(\mathbf{m})}$ / $\overline{-\phi(U_2^M, U_2)}$ vs. ϵ is a curve with a maximum at some ϵ which is the chosen value of the penalty weight.

4.3 EXAMPLE: SYNTHETIC DATA

The synthetic data for this study were computed by frequency-wavenumber integration using the computer program by Mallick and Frazer (1988). Fig. 4.1(a) illustrates the experimental layout for the offset VSP which consists of two shots 100 m apart and a 400 m deep borehole which is 150 m apart from the nearest shot. Assuming the earth is stratified near the shots and the borehole, we replace this geometry by Fig. 4.1(b) which outplaces the nearest shot to the borehole by another borehole equally deep. The sources are positioned at 5 m depth from the ground and the receivers are placed at 40 levels spaced every 10 m from 5 m deep. The data were recorded with a sample interval of 1 ms. Fig. 4.2(a) and (b) displays the synthetic data recorded in borehole 1 and borehole 2 respectively.

We perform inversion for the slowness of the layers between the two boreholes. We divide 400 m thickness into 40 evenly thick layers with the receiver located in the middle of the layer. In our inversion we have 40 model parameters, which represent the slowness of each layer, and each model parameter can have any of the 5 possible values. Therefore, there are $5^{40} \approx 10^{28}$ possible solutions in the solution space, which is large enough for using the exhaustive search method to find the best solution.

The energy function which we use here is given in chapter 3. To find the critical temperature we followed the algorithm given above in section 2.5. We search from log T = -2.5 to log T = 1.5 in steps of 0.5 performing 25 iterations at each temperature using five different sets of random numbers. Fig. 4.3(a)-(e) shows the results of these runs. The average of these five plots is shown in Fig. 4.3(f) and the

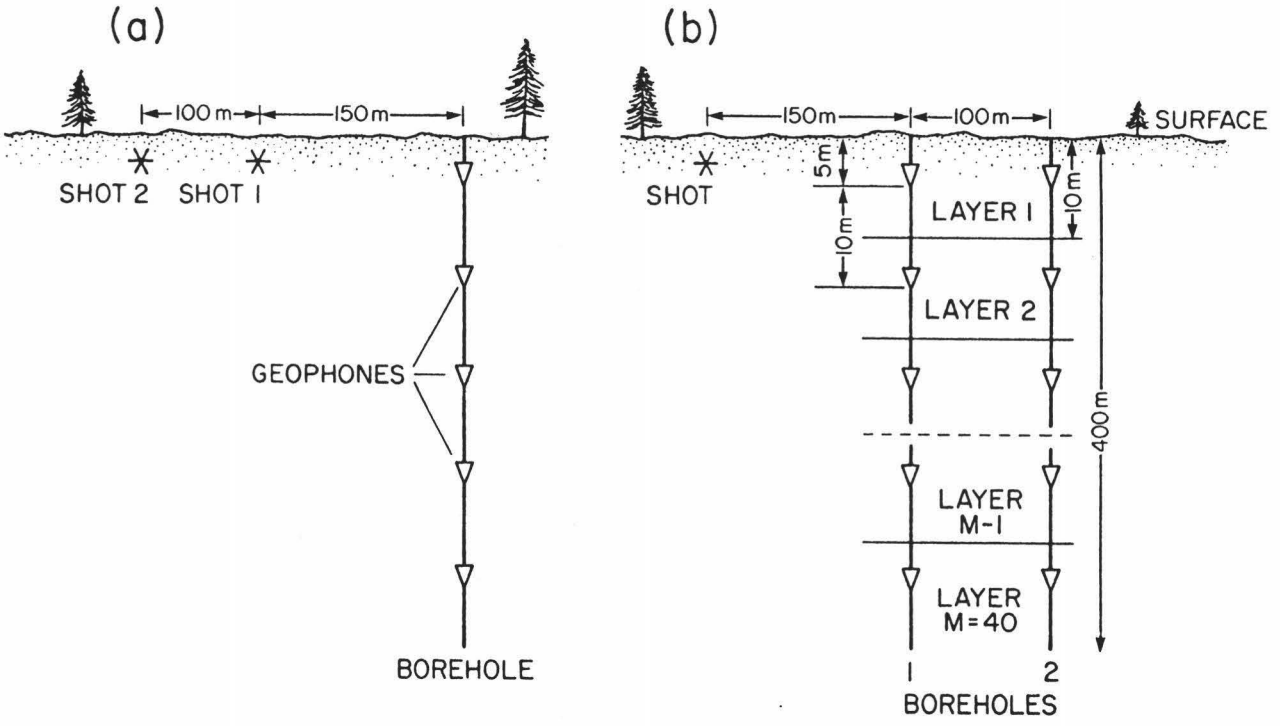


Figure 4.1: (a) Two-offset VSP geometry (b) Mathematically equivalent geometry

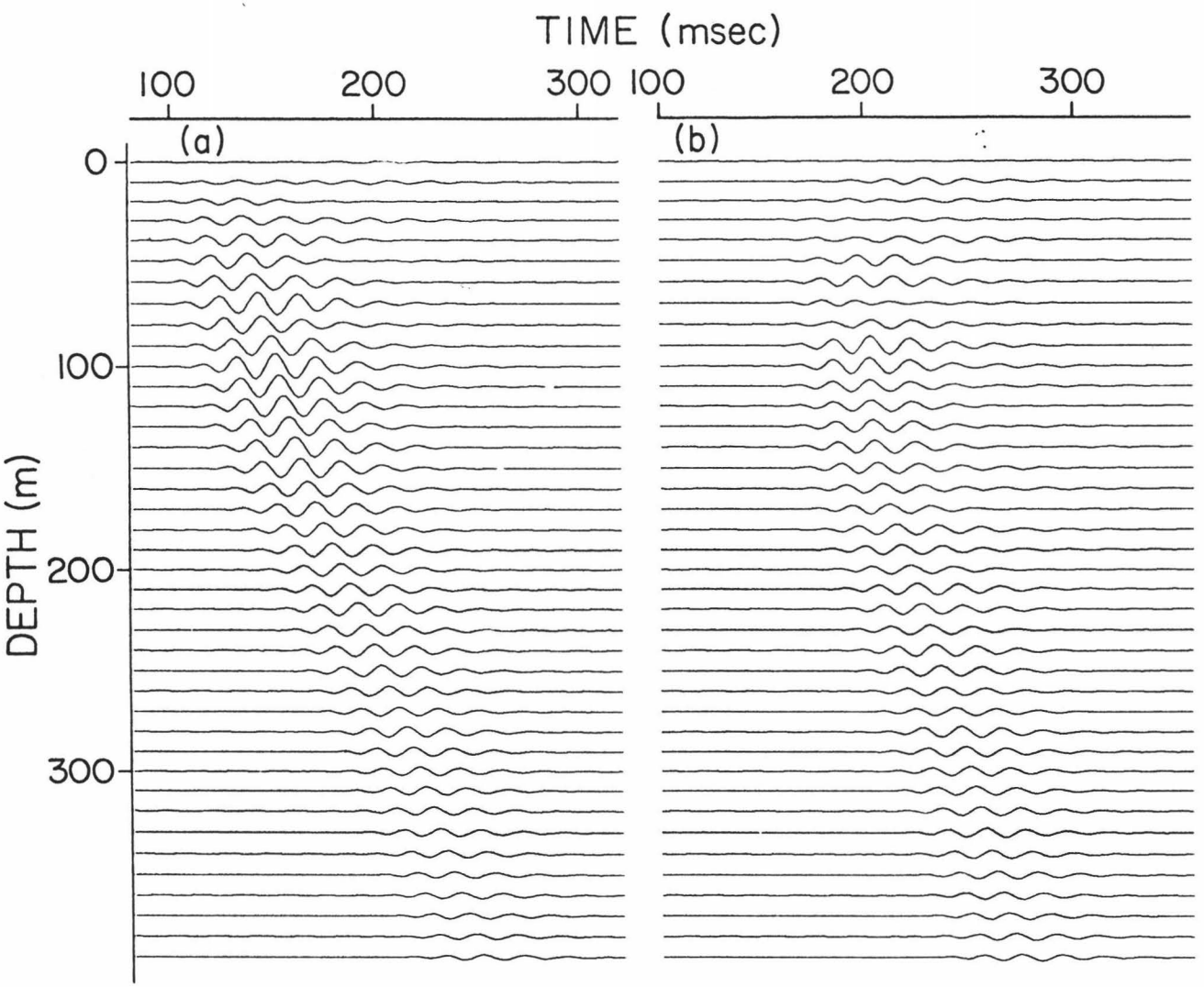


Figure 4.2: (a) Seismograms from borehole 1 (b) Seismograms from borehole 2

graph has a minimum at log T = -2.0. To obtain a better estimate of critical temperature we repeated the above mentioned procedure near this minimum, searching from log T = -2.5 to log T = -1.6 in steps of 0.1. The result of these runs is shown in Fig. 4.4(a)-(e) and the average of these plots is shown in Fig. 4.4(f). From Fig. 4.4(f) we estimate the critical temperature as log T_{cr} = -2.1 or T_{cr} = 0.00794.

After determining T_{cr} we find the penalty weight ϵ by following the algorithm given in section 4.2.3. The value of T_{cr} mentioned in the above paragraph was found with $\epsilon = 0$. We search from $\epsilon = 0$ to 35 in steps of 5 performing 25 iterations at each penalty weight. From this search we chose $\epsilon = 25$. The energy function E($\mathbf{m}$) now becomes $-\phi(U_2^M, U_2) + 25p(\mathbf{m})$.

After finding T_{cr} and ϵ we compute the PPD and Fig. 4.5 shows the result of this computation for all the profiles in Fig. 4.4. Fig. 4.5(a)-(e) shows the PPD after 25 iterations for five different sets of random numbers and we see no significant difference between these plots. The final PPD diagram is constructed using all 125 (5*25) profiles and is shown in Fig. 4.5(f). The blow up of this diagram is shown in Fig. 4.5(g). The true velocity profile is shown by the bold line. In this diagram we see that except for the top and bottom layers the PPD is narrow and has well defined peaks. The best fit slowness profile can be constructed by joining the peak of the marginal density function of each layer from the PPD diagram. We repeated the above mentioned procedure considering that each model parameter can take 13 possible values instead of 5. The final PPD diagram is shown in Fig. 4.6(a). To demonstrate the usefulness of the penalty function, we computed the PPD in this case without using any penalty function (i.e. $\epsilon = 0$) and the result is shown in Fig. 4.6(b). Comparing Fig. 4.6(a) and (b) we see that the penalty function helped to make the PPD to be well defined.

We estimate cor($\mathbf{m}$) matrix for each set of profiles in Fig. 4.6(a) and 4.6(b).

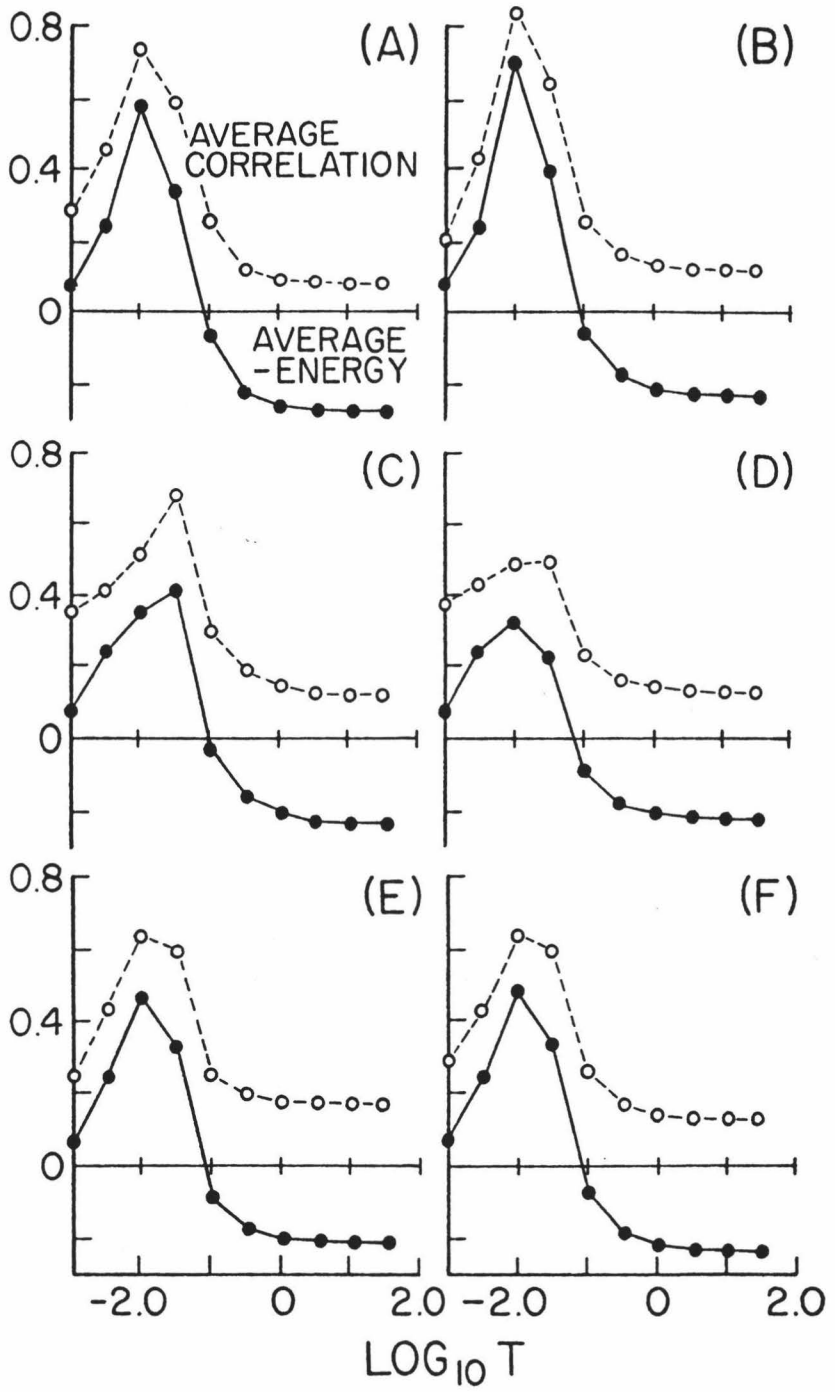


Figure 4.3: (a) - (e) Determination of T_{cr} in the wider scale for five different sets of random numbers. (f) The average of the five plots.

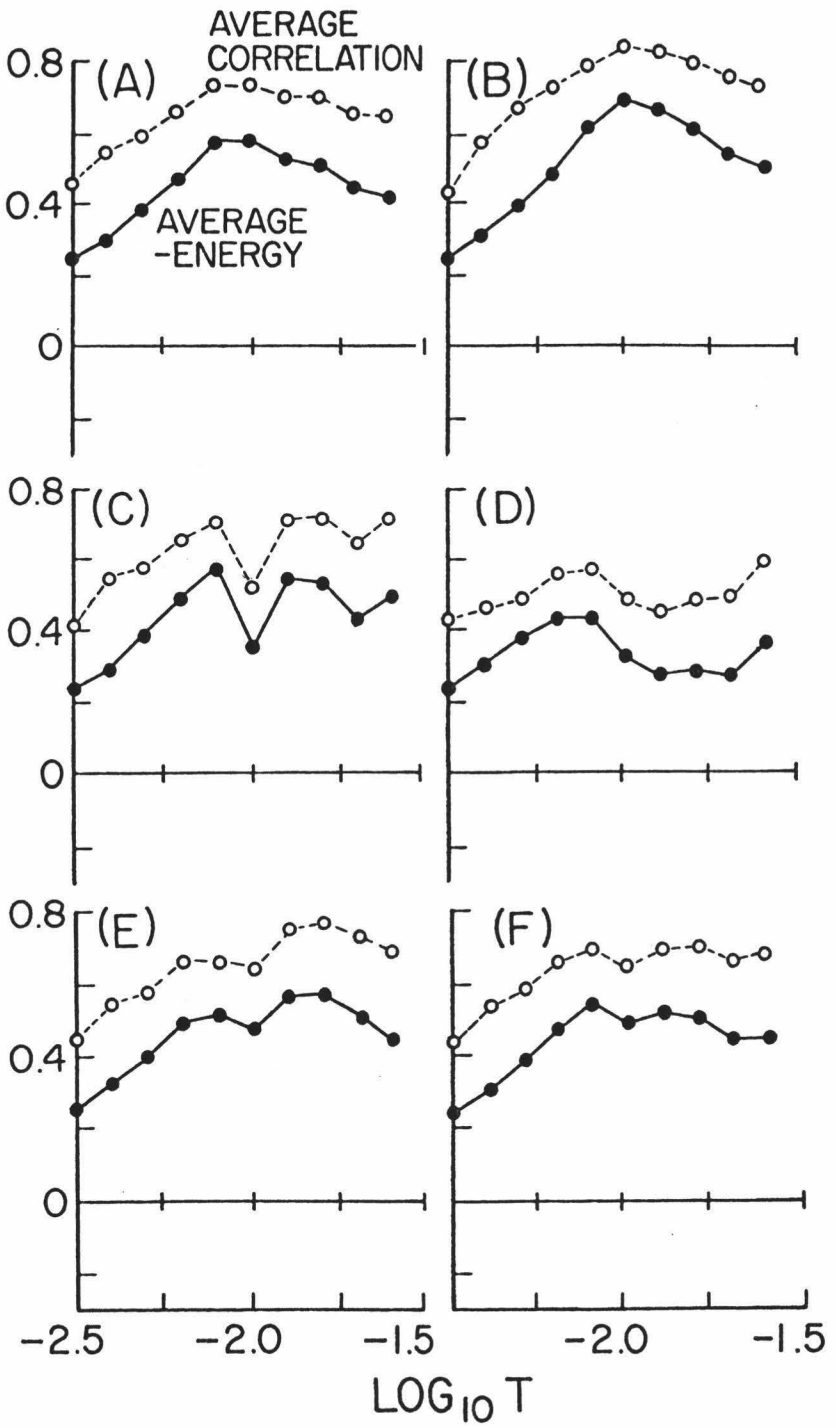


Figure 4.4: (a) - (e) Determination of T_{cr} in the narrower scale for five different sets of random numbers. (f) The average of the five plots.

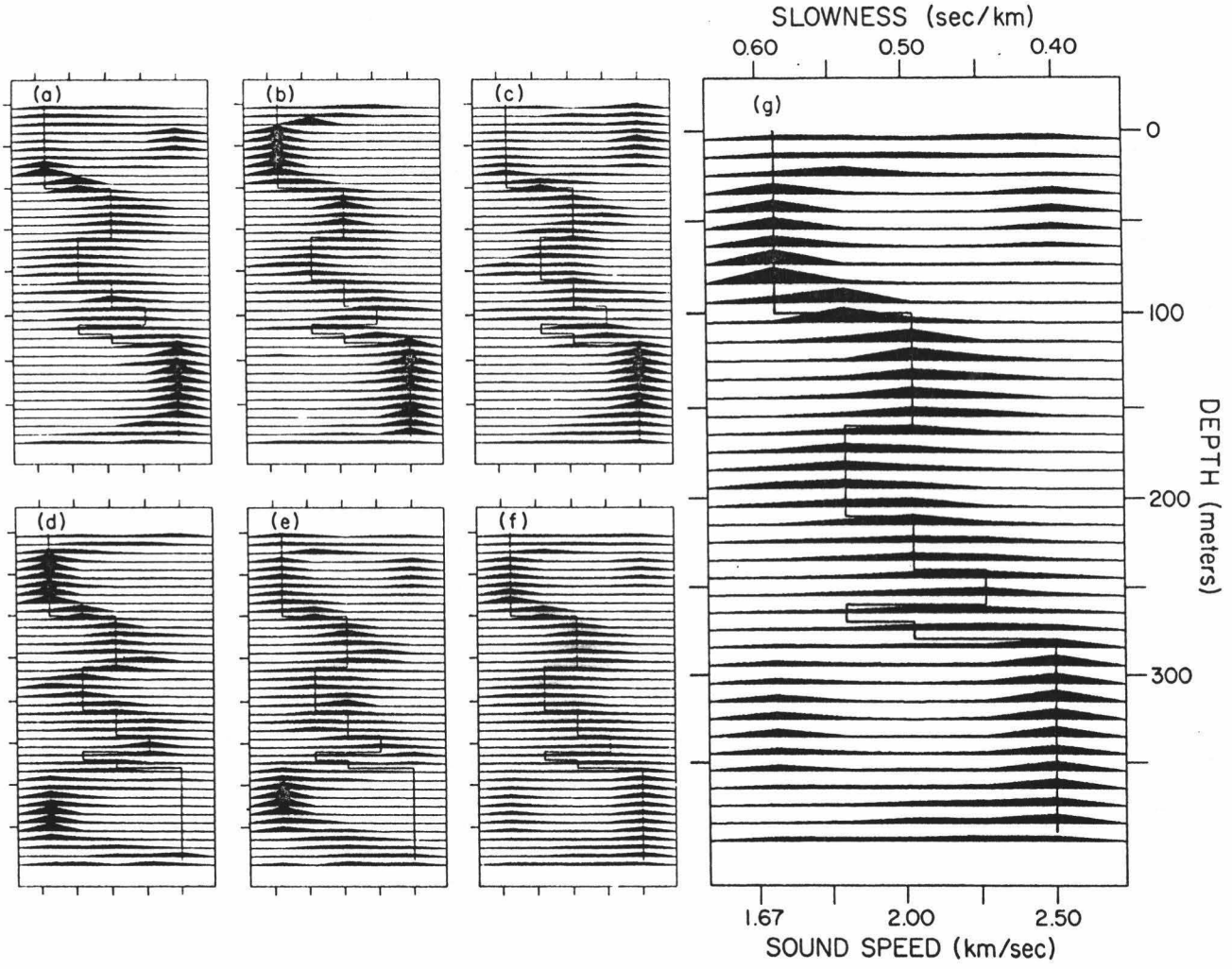


Figure 4.5: (a) - (e) A posteriori probability density function (PPD) for five different sets of random numbers. (f) The PPD constructed using all the profiles in (a) - (e). (g) The blow up of figure (f).

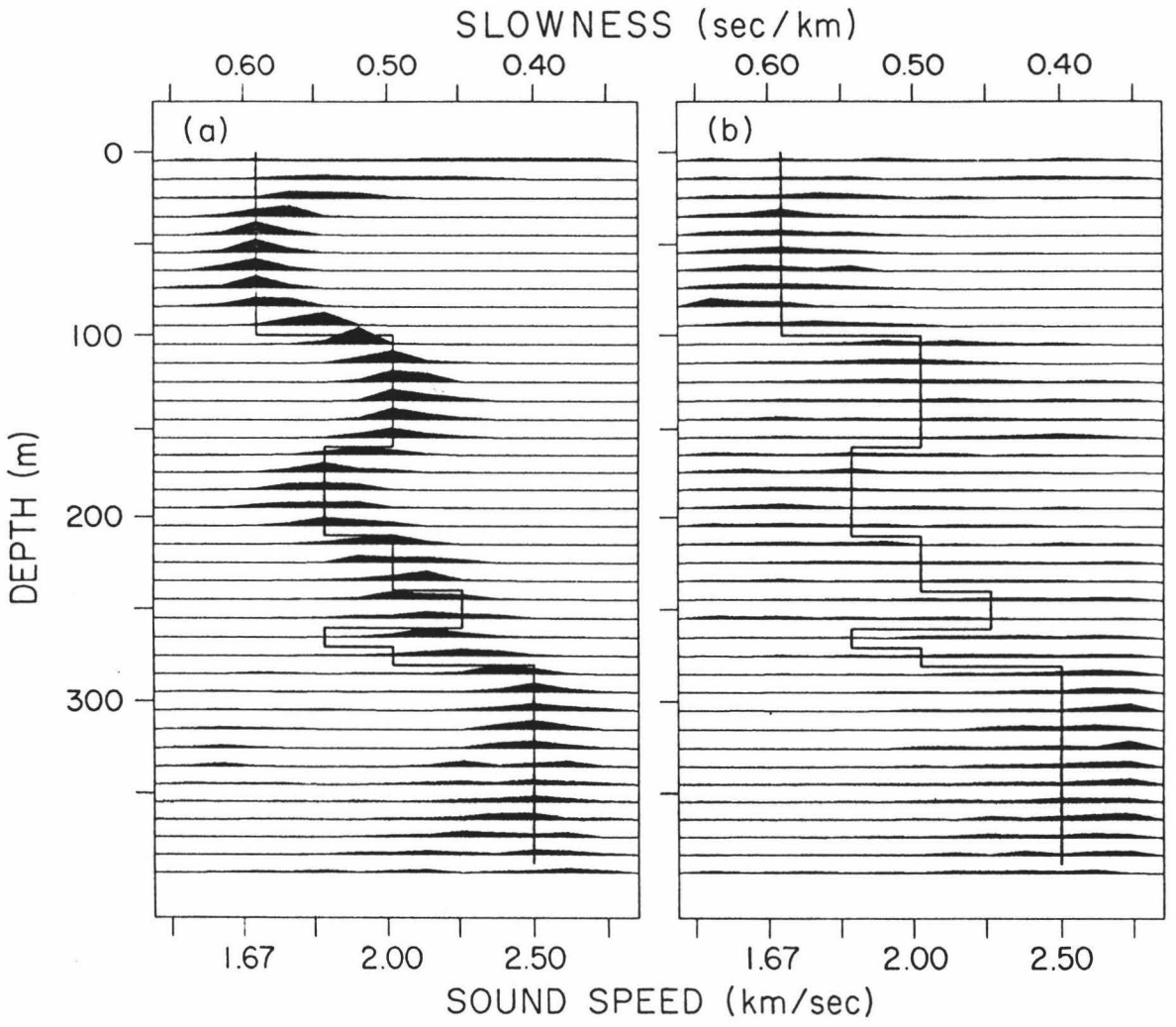


Figure 4.6: A posteriori probability density function (PPD) when the velocity was allowed to take any of 13 possible values in each layer. (a) Penalty function was used (b) Penalty function was not used.

using equation (6) of chapter 3. The part of these matrices are shown in Fig. 4.7(a) and 4.7(b) respectively. Comparing these two diagrams we see that when we don't use any penalty function, then we pick up oscillatory velocity profiles which give rise to small correlation in the neighborhood of the diagonal. The correlation value along the diagonal is 1.0 and is represented by white color.

4.4 DISCUSSION

The results of this chapter suggests that simulated annealing has the potential to solve global optimization problems. The SA algorithm has several attractive features. First of all, this algorithm is very easy to implement. Secondly, SA doesn't depend on the starting model. So, even if we have poor knowledge on *a-priori* information of model parameters, we can use SA algorithm. This characteristics of SA makes it much more efficient than the gradient search method to solve non-linear inverse problems. Thirdly, SA can provide good solutions if not the best solution to many problems.

In addition to the advantages discussed above, SA has some practical difficulties which need to be considered. One of the major problems is to choose an appropriate cooling schedule. The quality of the obtained solutions and the required CPU time depends on the choice of the cooling schedule. Our way of implementing SA bypass this problem by finding the critical temperature of the system and performing SA at that temperature (Frazer & Basu 1992).

Our way of implementing SA algorithm also deviates from the conventional SA technique by giving emphasis on constructing the PPD and then determine the best model from it rather than selecting the best model having the lowest energy. Even though our model space is vast, we still can pick up good models to construct the PPD by performing SA at the critical temperature.

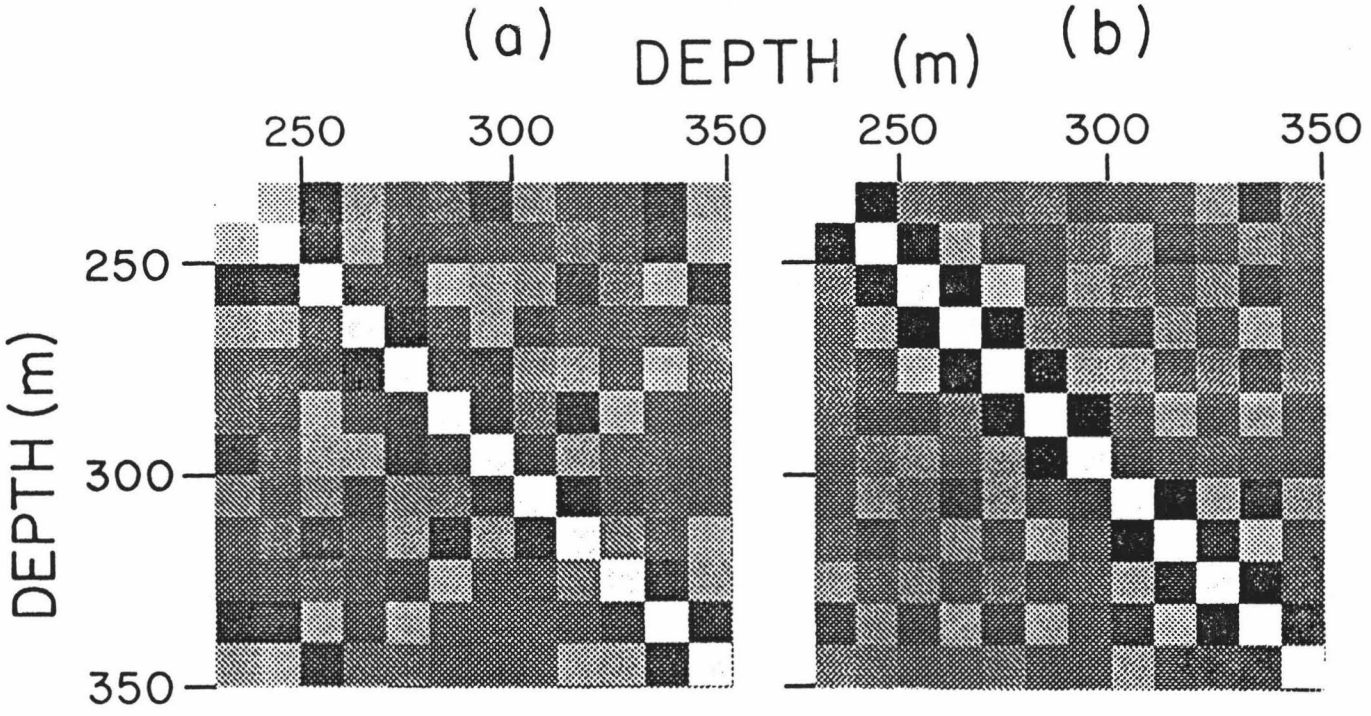


Figure 4.7: Part of correlation matrix cor(**m**) constructed (a) with using penalty function (b) without using penalty function. White is 1.0 and black is -1.0.

We have assumed that the earth is horizontally stratified and it has layers of equal thickness and the number of layers are the same as the number of receivers. This assumption has caused the problem of overparametrization, which resulted in picking up oscillatory velocity profiles because the energy values of those models are as high as the non-oscillatory velocity models. We have taken care of this problem by applying a penalty function and finding the optimum penalty weight. We have seen how penalty function has helped to make well defined PPD.

4.5 REFERENCES

Basu, A. & Frazer, L. N., 1990. Rapid determination of critical temperature in simulated annealing inversion, *Science* , **249**, 1409-1412.

Frazer, L. N. & Basu, A., 1992. Bayesian inversion by statistical physics with an application to offset VSP, submitted to *Geophys. J. Int.*.

Hardage, B. A., 1983. *Vertical seismic profiling*, Geophysical Press, London-Amsterdam.

Mallick, S. & Frazer, L. N., 1988. Rapid computation of multi-offset VSP synthetic seismograms, *Geophysics*, **53**, 479-491.

Chapter 5

Cross-borehole Inversion Using Simulated Annealing

5.1 INTRODUCTION

Cross-borehole seismic surveys provide methods for estimating velocity and density between the two boreholes. In this case, a source is placed in one borehole and the receivers record the energy propagation in an adjacent borehole. The source is then placed at different depth and the energy propagation is recorded again. This process is continued until the region between these two boreholes are well covered by the ray paths. In this chapter we provide another example of the SA method by performing inversion of a cross-borehole data at a fixed temperature to determine the slowness of the layers between the source well and the receiver well.

5.2 ENERGY FUNCTION

The forward modelling in our inversion procedure consists of migrating the data recorded in the receiver well to the shot position in the source well using different slowness profiles. For each shot in the source well we migrate the corresponding recorded seismograms in the receiver well after windowing the data with a specified

length around the first arrivals. This migration produces a wavelet at the shot position and its width is the length of the time window. If our velocity estimation is correct then the power of the wavelet will be the maximum. We define the energy function as the sum of the powers of all the wavelets and it can be written as

$$E(\mathbf{m}) = -W(\mathbf{m}) \tag{1}$$

where

$$W(\mathbf{m}) = \sum_{i=1}^{N} W_i. \tag{2}$$

Here $\mathbf{m}$ is the model vector, W_i is the power of the wavelet for i^{th} source, and N is the number of sources. The energy function is a function of the slownesses of the model and our goal here is to find the model vector $\mathbf{m}$ that minimizes $E(\mathbf{m})$. In the cases, where the energy function value for an oscillatory velocity profile is the same as for a smoother velocity profile, the energy function becomes

$$E(\mathbf{m}) = -W(\mathbf{m}) + \epsilon p(\mathbf{m}) \tag{3}$$

where ϵ is the penalty weight. The method used to determine ϵ is described in chapter 4.

5.3 EXAMPLES: REAL CROSS-HOLE DATA

The data for this study were collected by EXXON Production and Research Corporation who conducted the 'Loudon Experiment' to compare results from seismic, VSP, cross-borehole and well-logs data. Fig. 5.1 illustrates the experimental layout for the cross-borehole which consists of two boreholes 292 ft apart and drilled to a depth of 1600 ft. The data were collected with a 48-channel hydrophone streamer deployed in the receiver well using dynamite charges in the source well. Sources were positioned

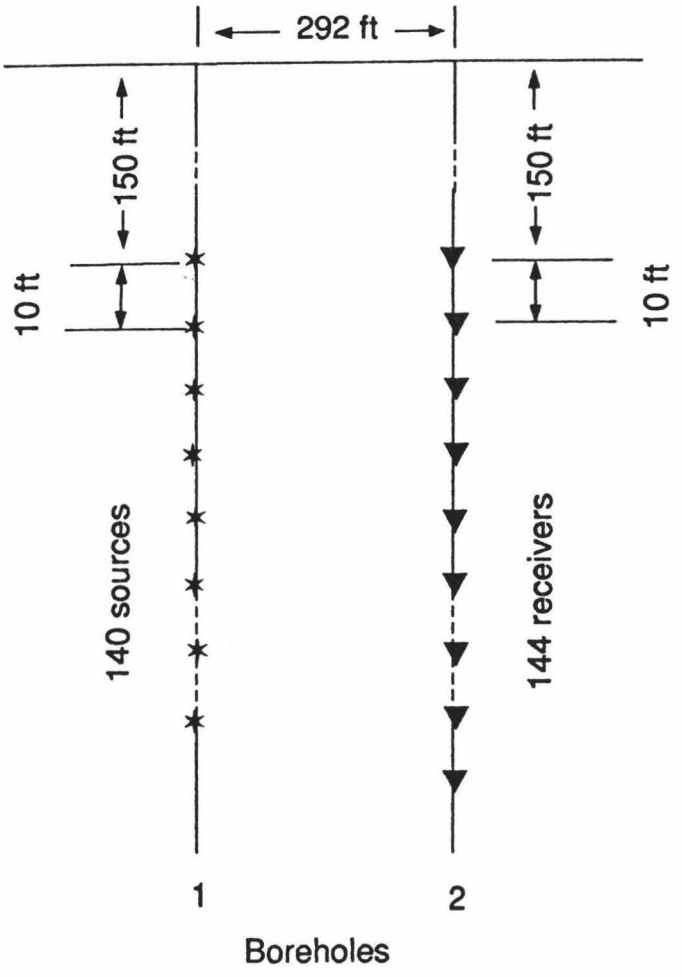


Figure 5.1: Cross borehole geometry.

at 140 levels spaced every 10 ft from 150 ft to 1540 ft (45 m to 469 m) deep in source well and the receivers were placed at 144 levels also spaced every 10 ft from 150 ft to 1580 ft (45 m to 481 m) deep in receiver well. Sources and receivers were aligned so that every source has one hydrophone at the same depth. The data were recorded with a sample interval of 0.25 ms. More details about the site and geometry are given by Zimmerman and Chen (1993).

Fig. 5.2 displays an example of the part of a raw data set for a source located at 1330 ft deep in the source well and receivers located between 890 ft and 1580 ft in the receiver well. It can be seen that the data are dominated by tube waves that have a velocity of about 5000 ft/s, even though it is not very difficult to pick up the first arrivals. Fig. 5.3 shows the comparison of the acquired sonic velocity log and tomographic velocity obtained by EXXON from the receiver well. The velocity here varies from 8000 ft/s to 18000 ft/s. The rock types in the area are shale, siltstone, sandstone, and limestone.

The data are processed with a 100-600 Hz band-pass filter to minimize the effect of low-frequency tube-wave noise. After processing the data the first arrivals are picked using a 2.5 ms window with 1 ms cosine tapers at either end.

We perform inversion for the slowness of the layers between the source well and the receiver well. The following two examples show the inversion results using a smaller (11 sources and 11 receivers) subset and a larger subset (30 sources and 30 receivers) of the data.

5.3.1 Smaller Data Subset

In this example we have selected 11 sources and 11 receivers, which are located from 1230 ft to 1330 ft (375 m to 405 m) deep in the corresponding source and receiver

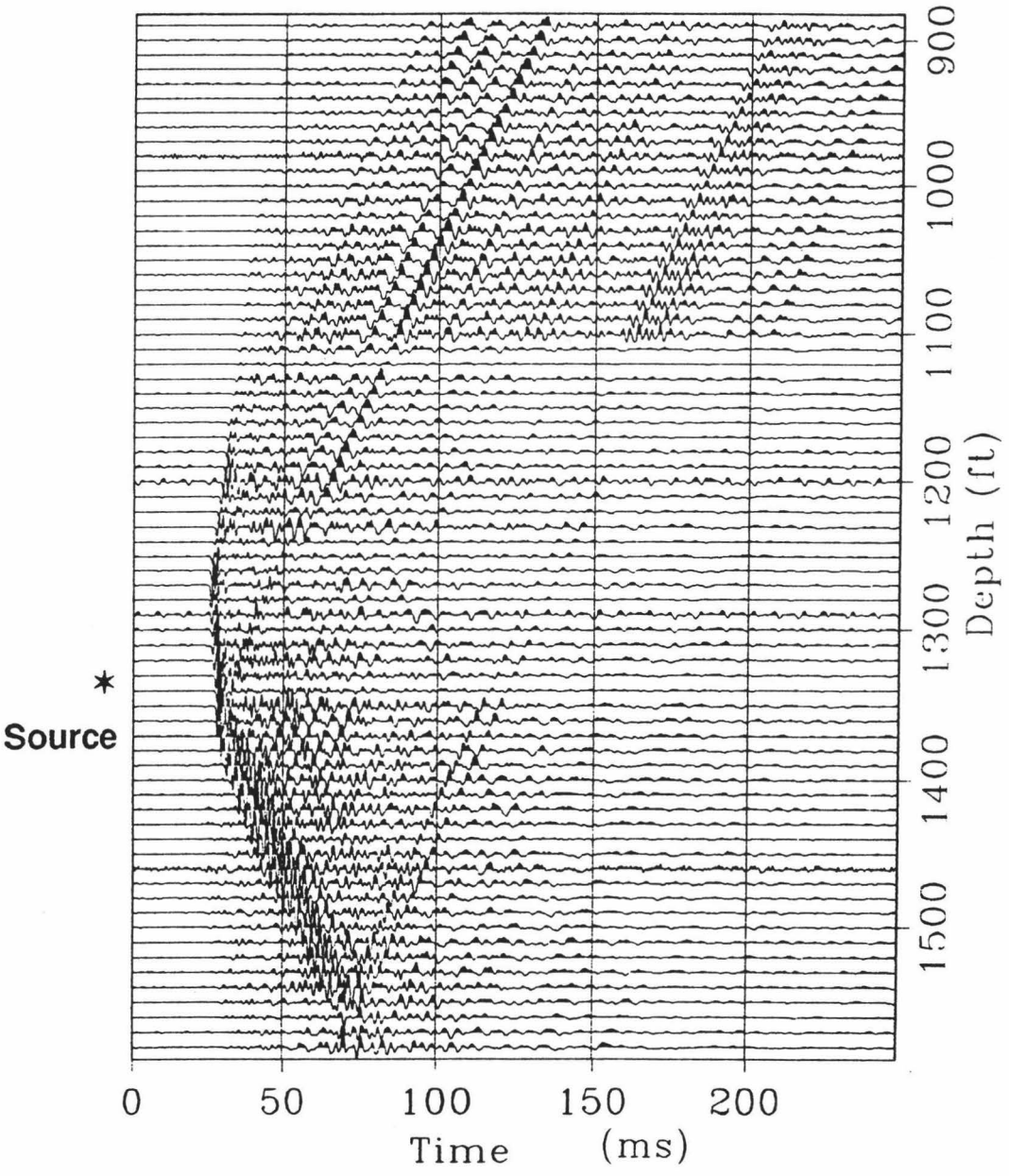


Figure 5.2: An example of the part of a raw data set.

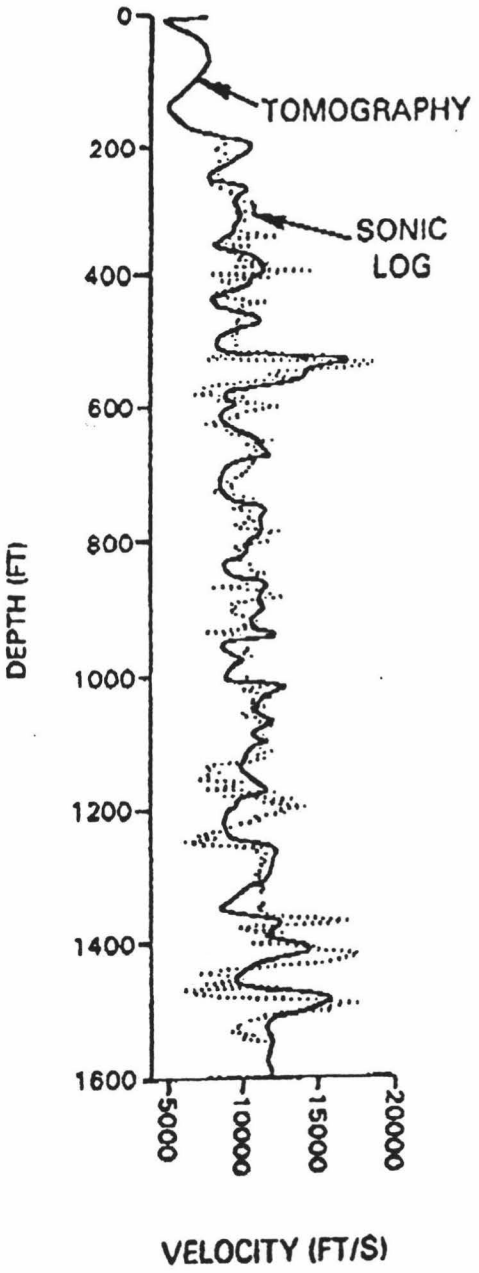


Figure 5.3: Sonic velocity log and the tomographic velocity at the receiver well.

wells at 10 ft depth intervals. The velocity at this depth varies from 9250 ft/s to 12250 ft/s. We invert for each layer using 110 traces (11 traces for each source and there was one bad source). We divide 110 ft thickness into 11 evenly thick layers with the source and the receiver located in the middle of the layer. In our inversion we have 11 model parameters, which represent the slowness of each layer, and each model parameter can have any of the 41 possible values. Therefore there are $41^{11} \approx 10^{18}$ possible solutions in the solution space, which is large enough for using the exhaustive search method to find the best solution.

The energy function used here is given by equation (3) and it includes the penalty function. To find the critical temperature we followed the algorithm given above in chapter 3. We search from log T = -2.0 to log T = 2.0 in steps of 0.5 performing 50 iterations at each temperature. We use four different sets of random numbers and Fig. 5.4(a) shows the results of these runs. The average of these four plots is shown in Fig. 5.4(b) and the graph has a minimum at log T = -0.5. To obtain a better estimate of critical temperature we repeated the above mentioned procedure near this minimum, searching from log T = -0.8 to log T = -0.1 in steps of 0.1. The result of these runs is shown in Fig. 5.5(a) and the average of these plots is shown in Fig. 5.5(b). From Fig. 5.5(b) we estimate the critical temperature as log $T_{cr} = -0.6$ or $T_{cr} = 0.2512$. It may be interesting to note that the value of T_{cr} is less than unity.

After determining T_{cr} we find the penalty weight ϵ by following the algorithm given in chapter 4. The value of T_{cr} mentioned in the above paragraph was found with $\epsilon = 0$. We search from log ϵ = -3.0 to log ϵ = 5.0 in steps of 1.0 performing 50 iterations at each penalty weight. We used four different sets of random numbers and Fig. 5.6(a) shows the results. The average of these plots is shown in Fig. 5.6(b) and from this plot we chose log $\epsilon = 2.0$ or $\epsilon = 100.0$. The energy function E($\mathbf{m}$) now becomes -W($\mathbf{m}$) + 100.0 p($\mathbf{m}$).

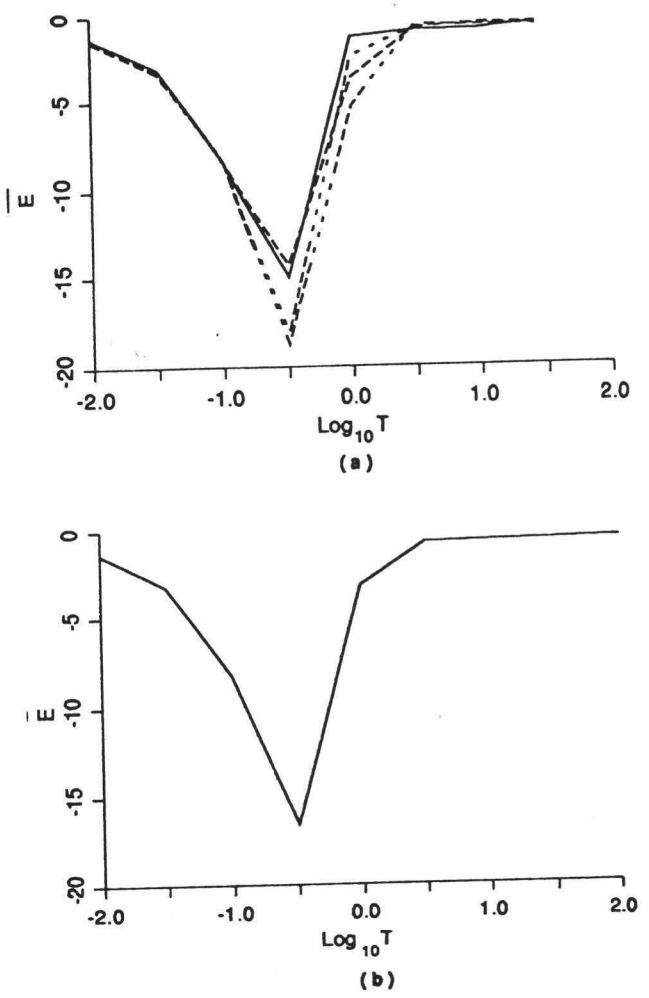


Figure 5.4: Determination of T_{cr} in the wider scale for the smaller data subset. (a) plot of $\overline{E}$ vs. log T for four different sets of random numbers. (b) the average of the four plots in (a).

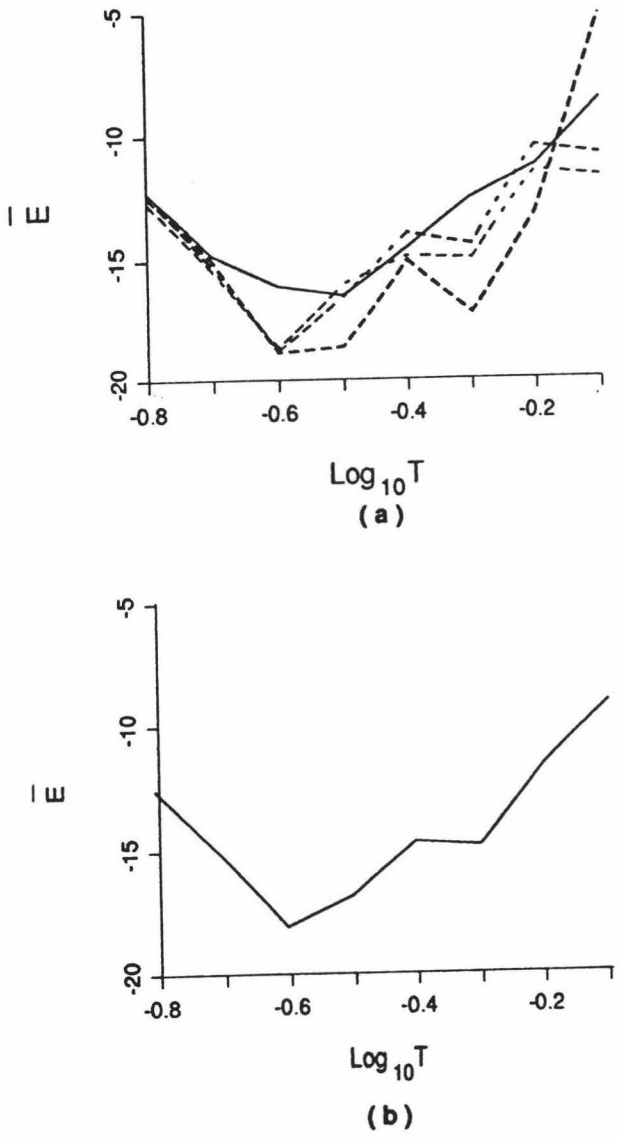


Figure 5.5: Determination of T_{cr} in the narrower scale for the smaller data subset. (a) plot of $\overline{E}$ vs. log T for four different sets of random numbers. (b) the average of the four plots in (a).

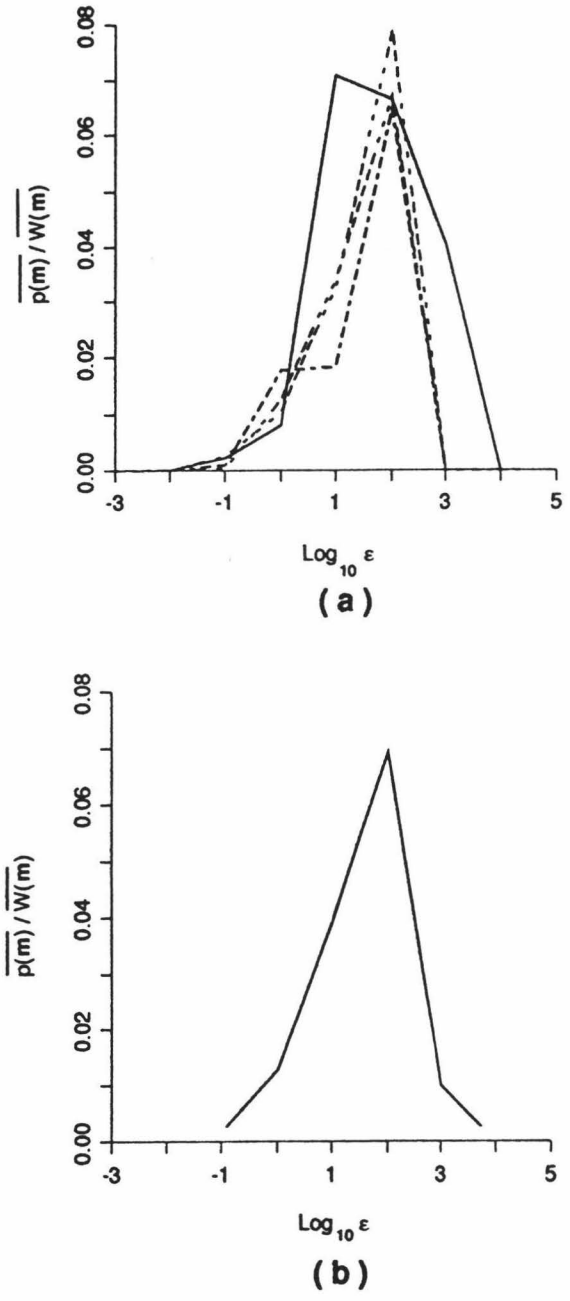


Figure 5.6: Determination of ϵ for the smaller data subset. (a) plot of $\overline{p(\mathbf{m})}$ / $\overline{W(\mathbf{m})}$ vs. log ϵ for four different sets of random numbers. (b) the average of the four plots in (a).

After finding T_{cr} and ϵ we compute the PPD after generating the slowness profiles at T_{cr}. To find the number of profiles needed for the convergence of the PPD we plot the PPD after 25, 75, 150 and 200 iterations for one set of random numbers. The result is shown in Fig. 5.7(a)-(d) and it can be seen that the PPD has converged at 200 iterations. Fig. 5.8(a)-(d) shows the PPD after 300 iterations for four different sets of random numbers and we see no significant difference between these plots. The final PPD diagram is constructed using all 1200 (4*300) profiles and is shown in Fig. 5.9. In this diagram we see that except for the top two layers the PPD is narrow and has well defined peaks. The best fit slowness profile can be constructed by joining the peak of the marginal density function of each layer from the PPD diagram. The PPD is shown along with the sonic log slowness profile at the receiver well in Fig. 5.10(a) and with the tomographic slowness profile in Fig. 5.10(b) which were provided by EXXON. The PPD matches more closely with the tomographic slowness than the sonic log slowness. The PPD shows the average slowness profile between the source and the receiver wells whereas the tomographic and sonic slowness profiles represent the slowness near the receiver well.

We estimate cov($\mathbf{m}$) matrix for each set of profiles in Fig. 5.9 using equation (5) of chapter 3. From cov($\mathbf{m}$) we generate cor($\mathbf{m}$) matrix using equation (6) of the same chapter and the matrix is shown in Fig. 5.11. We see in this diagram that the model parameters in the middle layers are more dependent on each other than any other layers.

5.3.2 Larger Data Subset

In this example we have selected 30 sources and 30 receivers which are located from 1150 ft to 1440 ft (351 m to 439 m) deep in the corresponding source and receiver

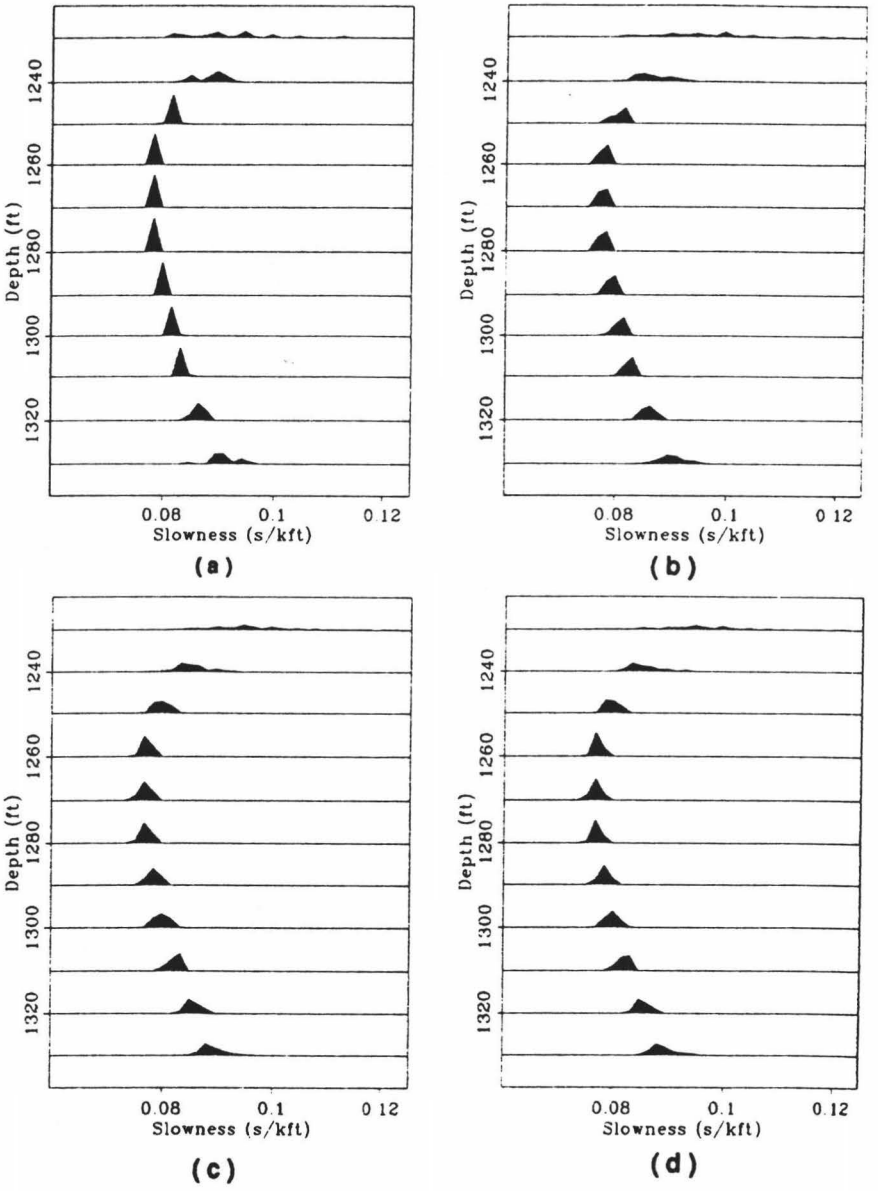


Figure 5.7: A posteriori probability density function (PPD) after (a) 25 iterations, (b) 75 iterations, (c) 150 iterations, (d) 200 iterations

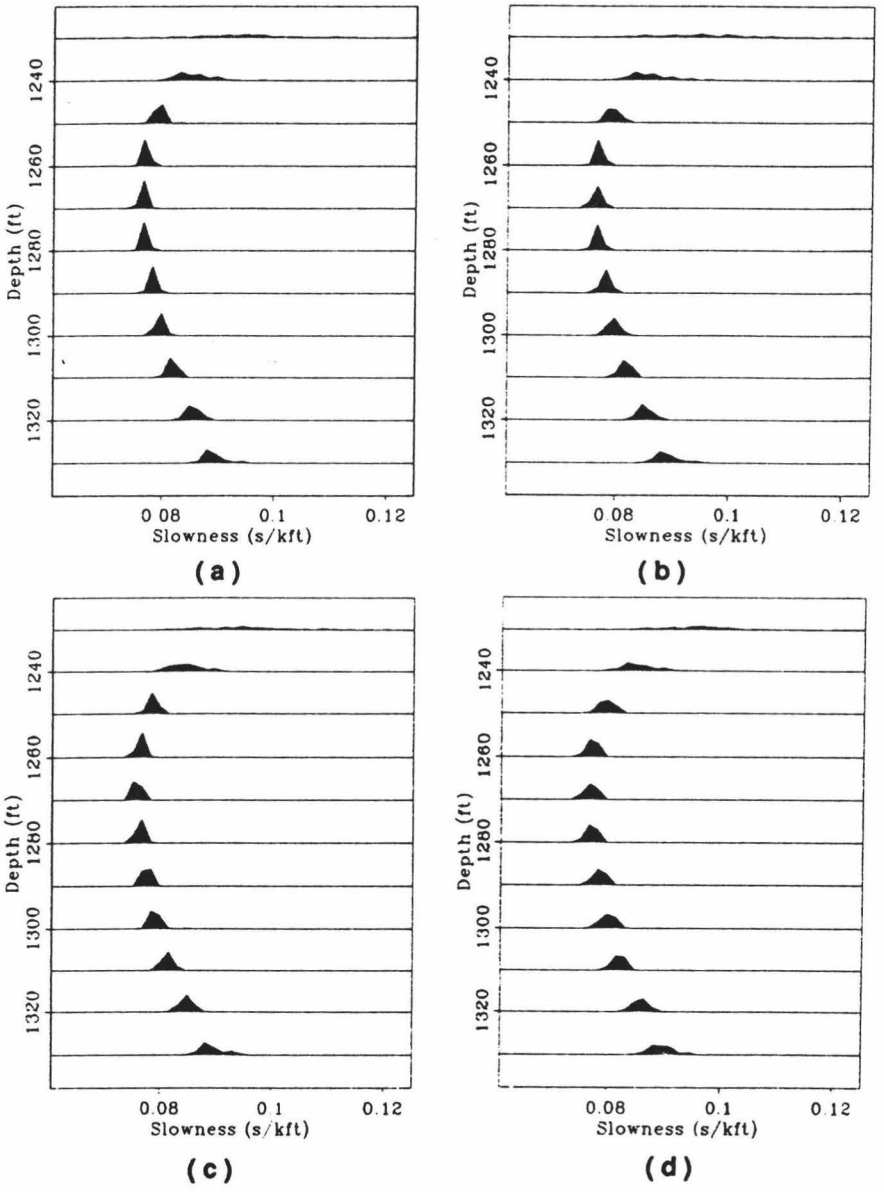


Figure 5.8: The PPD after 300 iterations. (a)-(d) are for four sets of random numbers.

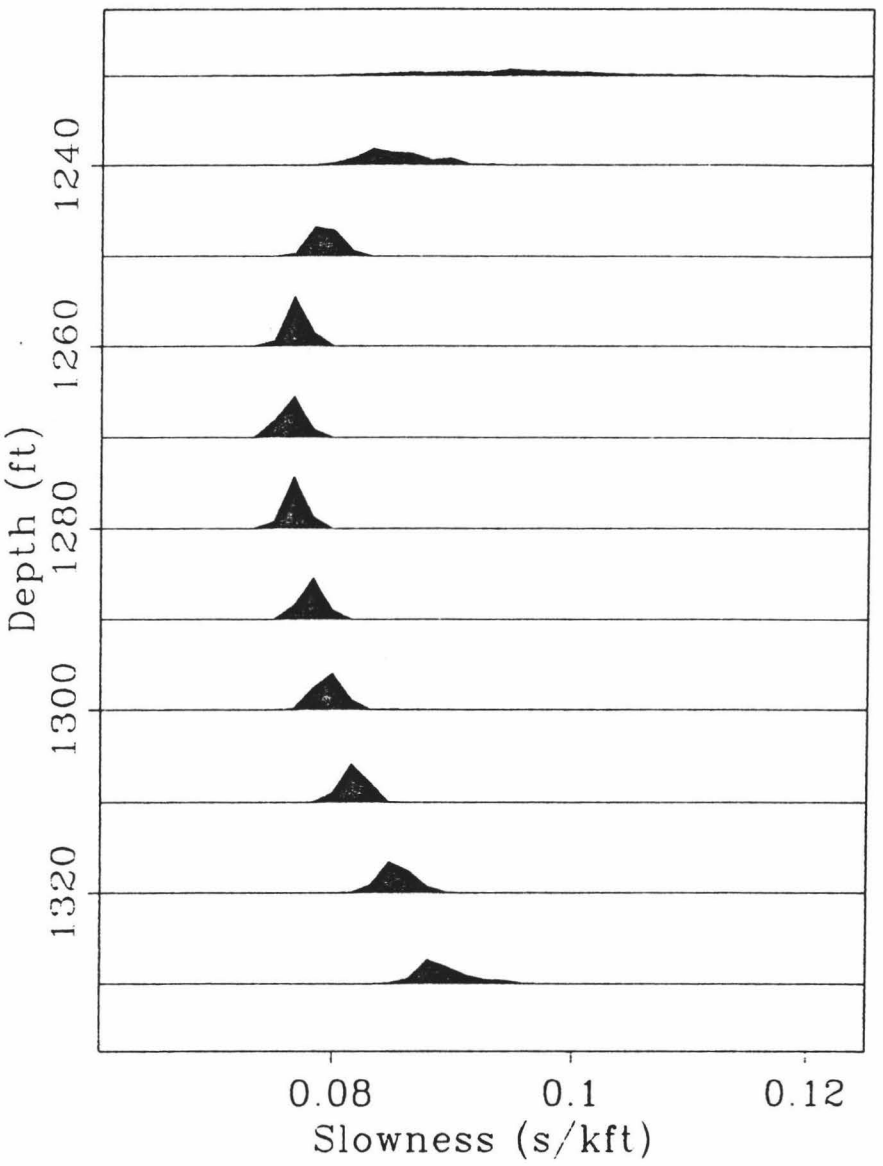


Figure 5.9: The PPD constructed using all 1200 iterations of Fig. 5.8(a)-(d).

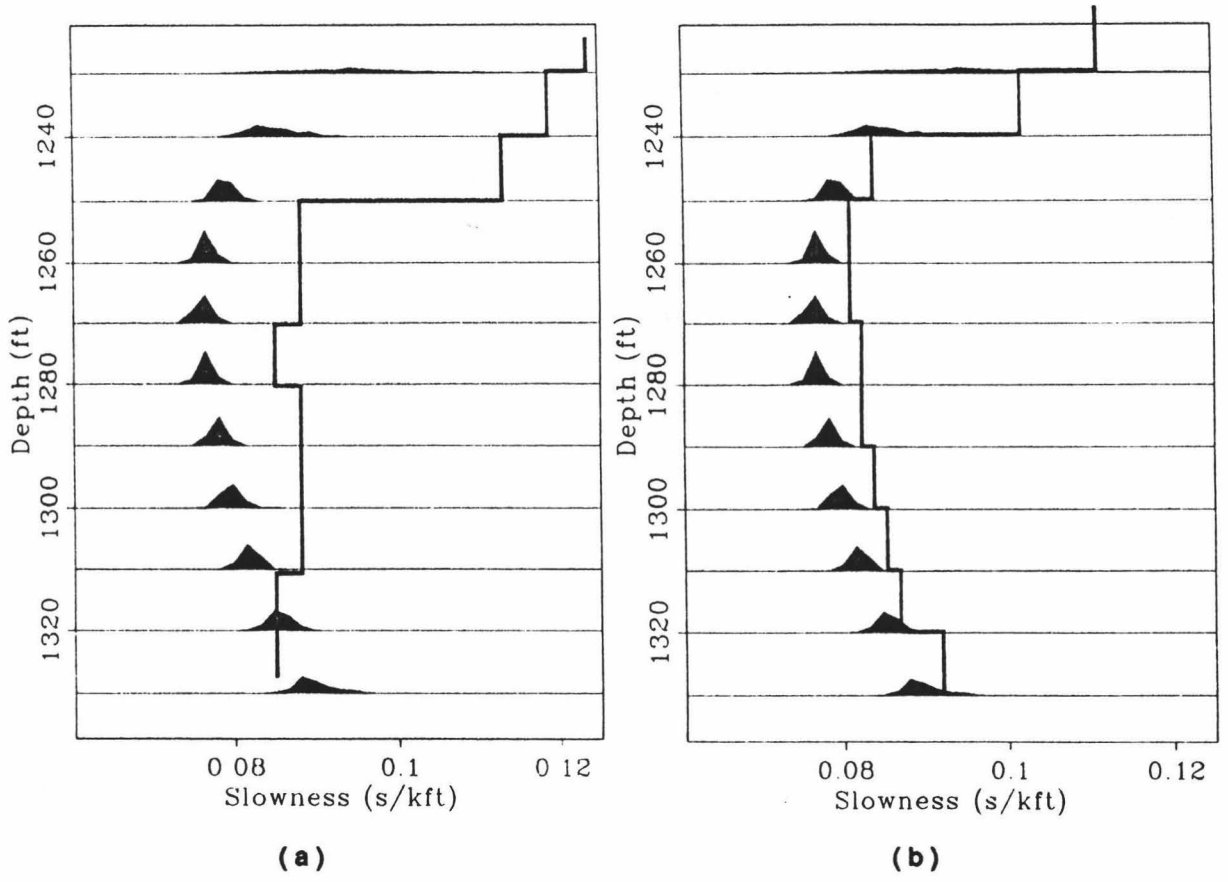


Figure 5.10: The PPD along with the (a) sonic slowness profile and the (b) tomographic slowness profile.

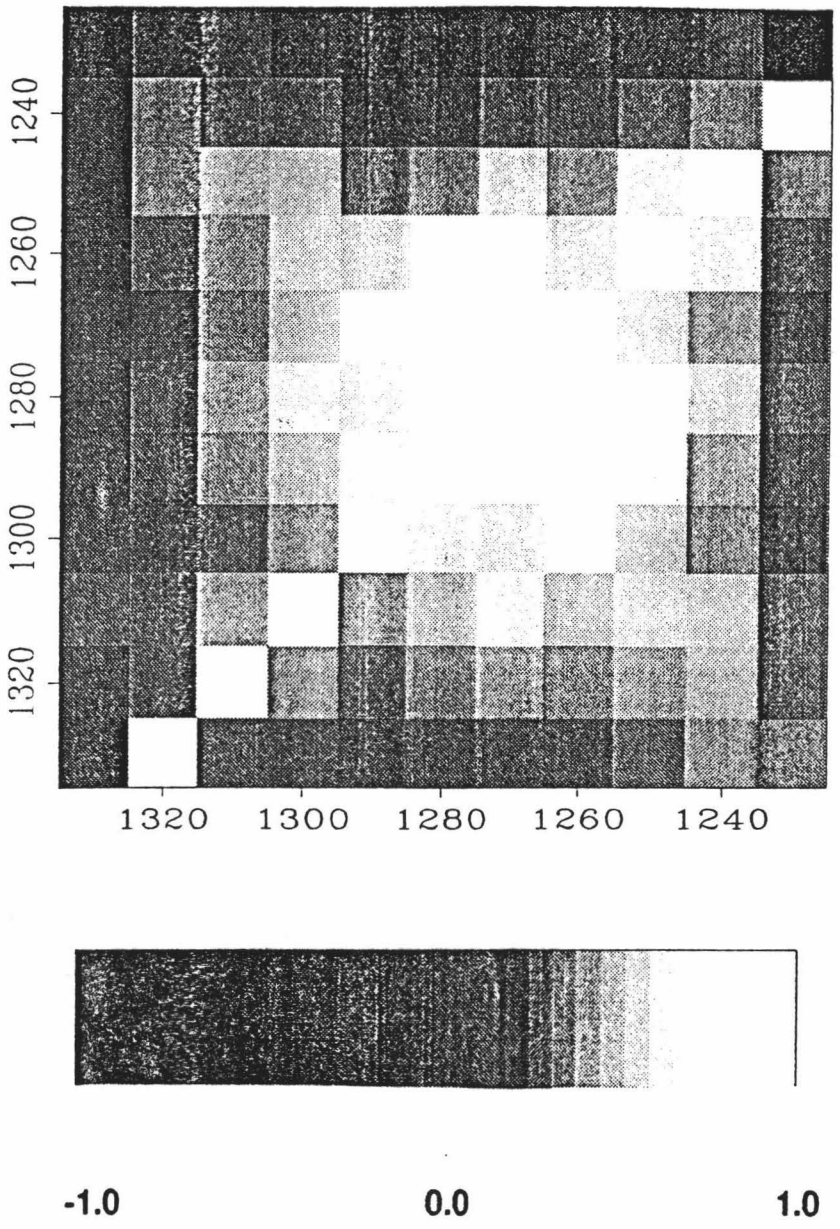


Figure 5.11: The correlation matrix. White is 1.0 and black is -1.0 respectively.

wells at 10 ft depth interval. The velocity in this region varies from 8750 ft/s to 14500 ft/s. Here we encounter two bad sources, so we use 840 remaining good traces for inversion. We divide 300 ft thickness into 30 evenly thick layers with the source and the receiver located in the middle of the layer. Here we have 30 model parameters representing the slownesses of each layer and each model parameter can have any of the 41 possible values, and therefore there are $41^{30} \approx 10^{48}$ possible solution in the solution space.

The energy function which we use here is given by equation (3). As before we search for the critical temperature from log T = -2.0 to log T = 2.0 in steps of 0.5 performing 50 iterations at each temperature. We use four different sets of random numbers and Fig. 5.12(a) shows the results of these runs. The average of these four plots is shown in Fig. 5.12(b) and the graph has a minimum at log T = 0.0. We again search from log T = -0.4 to log T = 0.4 in steps of 0.1 and the result of these runs is shown in Fig. 5.13(a) and the average of these plots is shown in Fig. 5.13(b). From Fig. 5.13(b) we estimate the critical temperature as log T_{cr} = 0.1 or T_{cr} = 1.26.

It may be interesting to note that T_{cr} is higher than the smaller data subset and also it is more than unity. As explained by Frazer and Basu (1992), T_{cr} is a measure of the energy barrier between the local minima. Since the value of the energy function is higher for the larger data subset than for the smaller data subset so we expect T_{cr} to be higher for the larger data subset.

After determining T_{cr} we find the penalty weight ϵ by searching from log ϵ = -3.0 to log ϵ = 5.0 in steps of 1.0 performing 50 iterations at each penalty weight. We used four different sets of random numbers and Fig. 5.14(a) shows the results. The average of these plots is shown in Fig. 5.14(b) and from this plot we chose log ϵ = 3.0 or ϵ = 1000.0. The energy function E($\mathbf{m}$) now becomes -W($\mathbf{m}$) + 1000.0 p($\mathbf{m}$).

Similar to the previous example we find the number of profiles needed for the

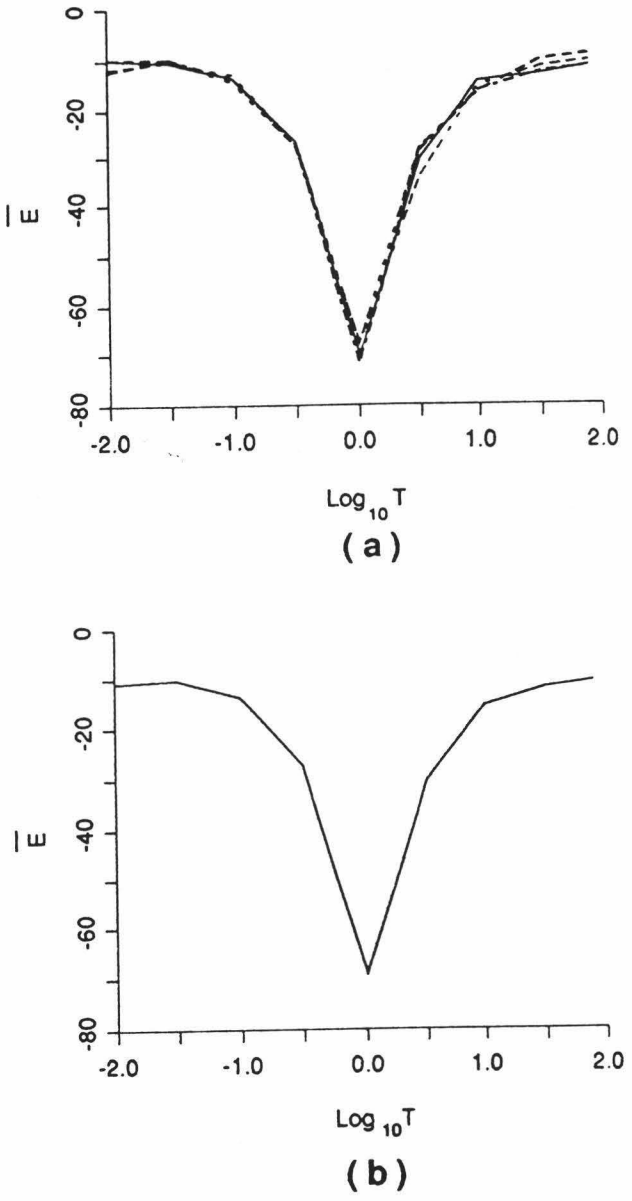


Figure 5.12: Determination of T_{cr} in the wider scale for the larger data subset. (a) plot of $\overline{E}$ vs. log T for four different sets of random numbers. (b) the average of the four plots in (a).

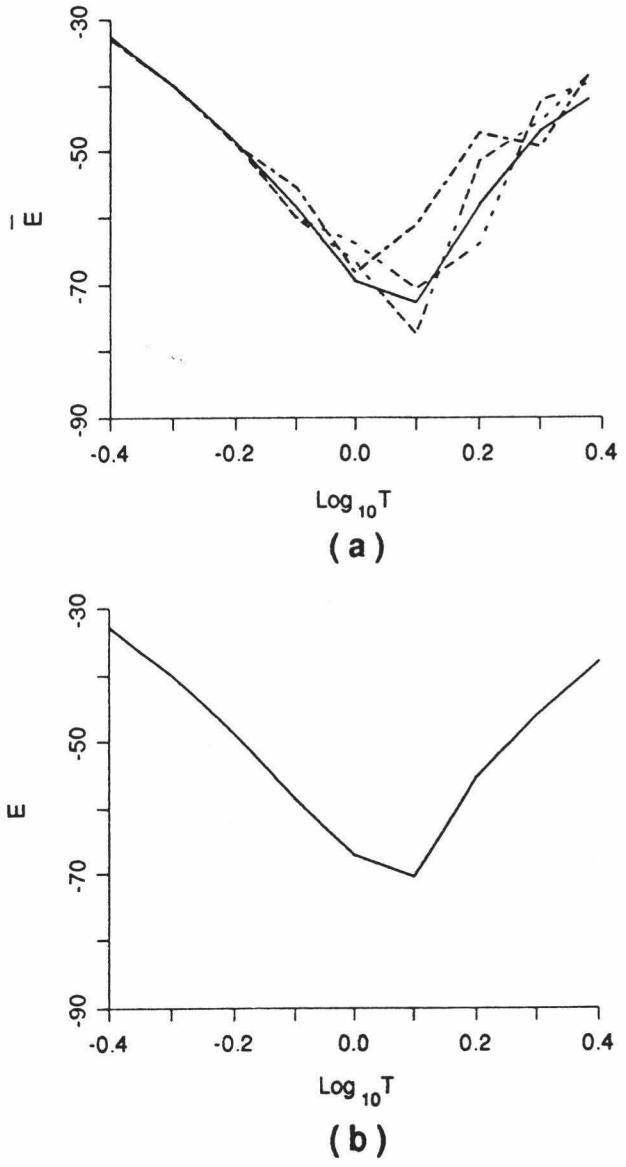


Figure 5.13: Determination of T_{cr} in the narrower scale for the larger data subset. (a) plot of $\overline{E}$ vs. log T for four different sets of random numbers. (b) the average of the four plots in (a).

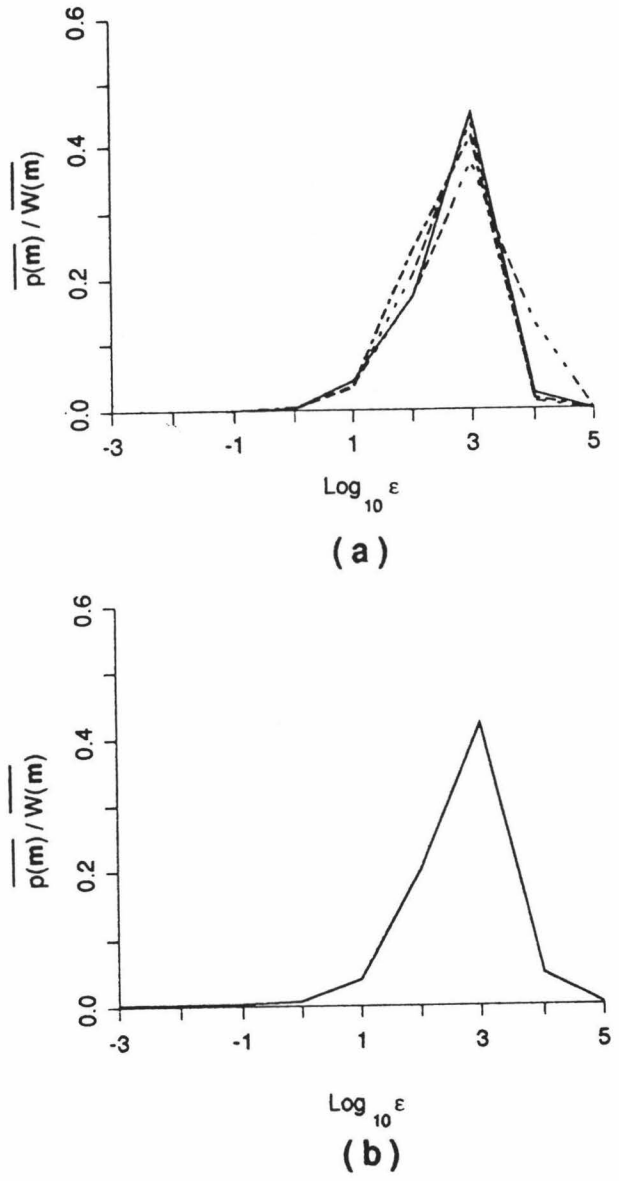


Figure 5.14: Determination of ϵ for the larger data subset. (a) plot of $\overline{p(\mathbf{m})}$ / $\overline{W(\mathbf{m})}$ vs. log ϵ for four different sets of random numbers. (b) the average of the four plots in (a).

convergence of the PPD by plotting the PPD after 50, 200, 300 and 500 iterations for one set of random numbers. The result is shown in Fig. 5.15(a)-(d) and it can be seen that the PPD has been converged at 500 iterations. Fig. 5.16(a)-(d) shows the PPD after 600 iterations for four different sets of random numbers and we see no significant difference between these plots. The final PPD diagram is constructed using all 2400 (4*600) profiles and is shown in Fig. 5.17. In this diagram we see that the PPD of the middle layers have well defined peaks and the topmost and the bottommost layers have very wide PPD. Since more rays pass through the middle layers, there the model parameters are well resolved. The PPD is shown along with the sonic slowness profile at the receiver well in Fig. 5.18(a) and with the tomographic slowness profile in Fig. 5.18(b). The PPD matches with the tomographic slowness profile more in the middle layers and in other layers it generally has peaks at higher slowness values than the corresponding tomographic slowness.

As before we estimated cor($\mathbf{m}$) matrix and it is plotted in Fig. 5.19. The layers in the middle are found to be more dependent on each other than the end layers. This inter dependency is more prominent in between 1210 ft and 1250 ft.

5.4 DISCUSSION

In this chapter we have investigated the use of a global optimization method, simulated annealing to solve a non-linear seismic inverse problem assuming a 1-D acoustic earth model. The result of this paper suggests that SA is more efficient than exhaustive search, gradient search, and Monte Carlo search methods to solve non-linear inverse problem when the solution space is large and multimodal. A simple local search method (a descent algorithm) is rapid to execute but it can get trapped in a local minimum. One way to improve the solution in this method could be to run the

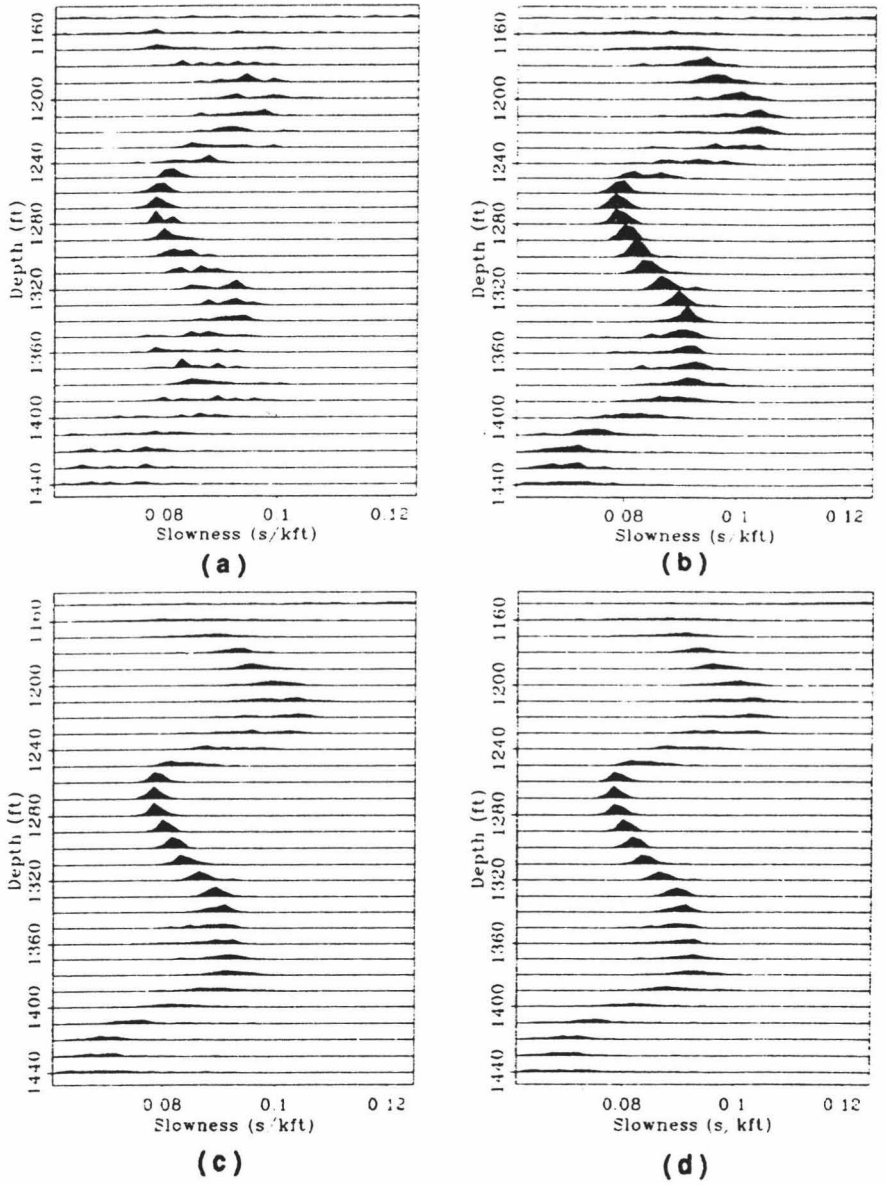


Figure 5.15: A posteriori probability density function (PPD) after (a) 50 iterations, (b) 200 iterations, (c) 350 iterations, (d) 500 iterations

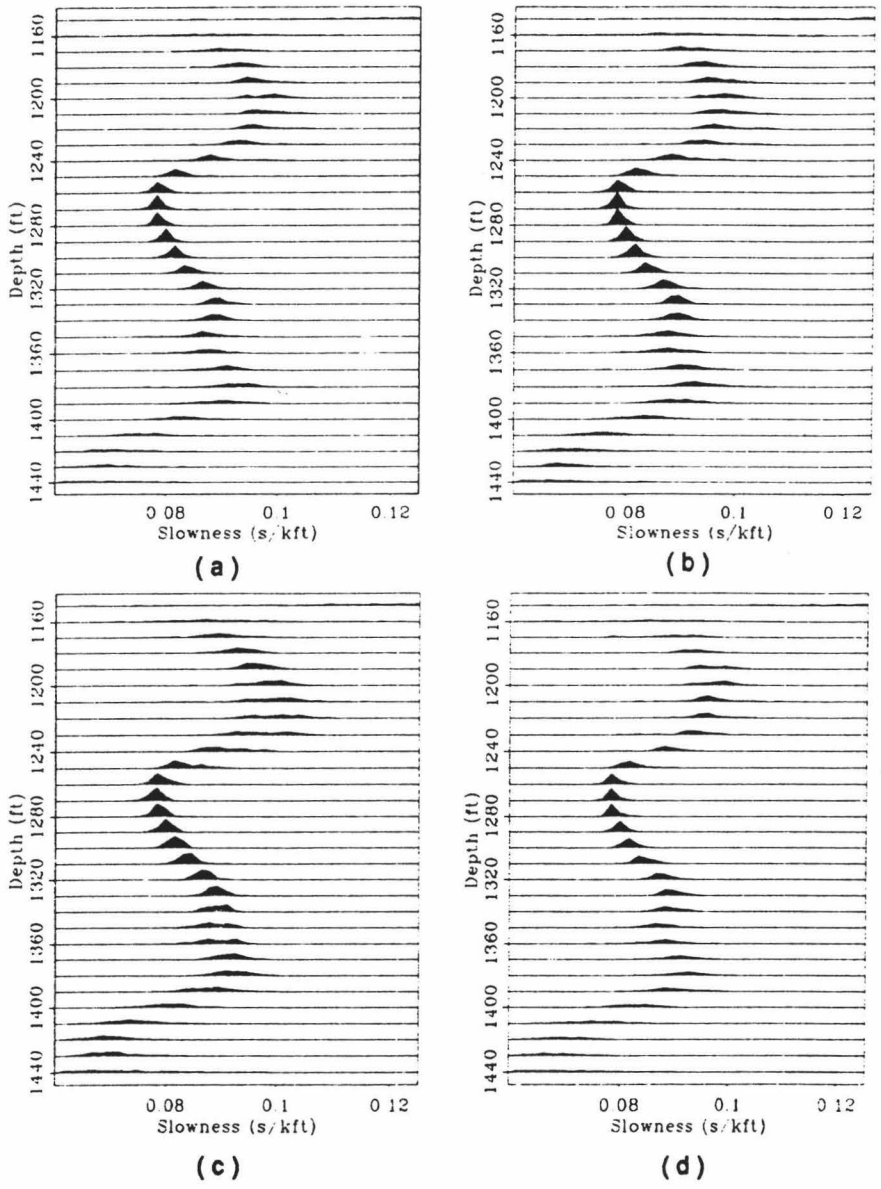


Figure 5.16: The PPD after 600 iterations. (a)-(d) are for four sets of random numbers.

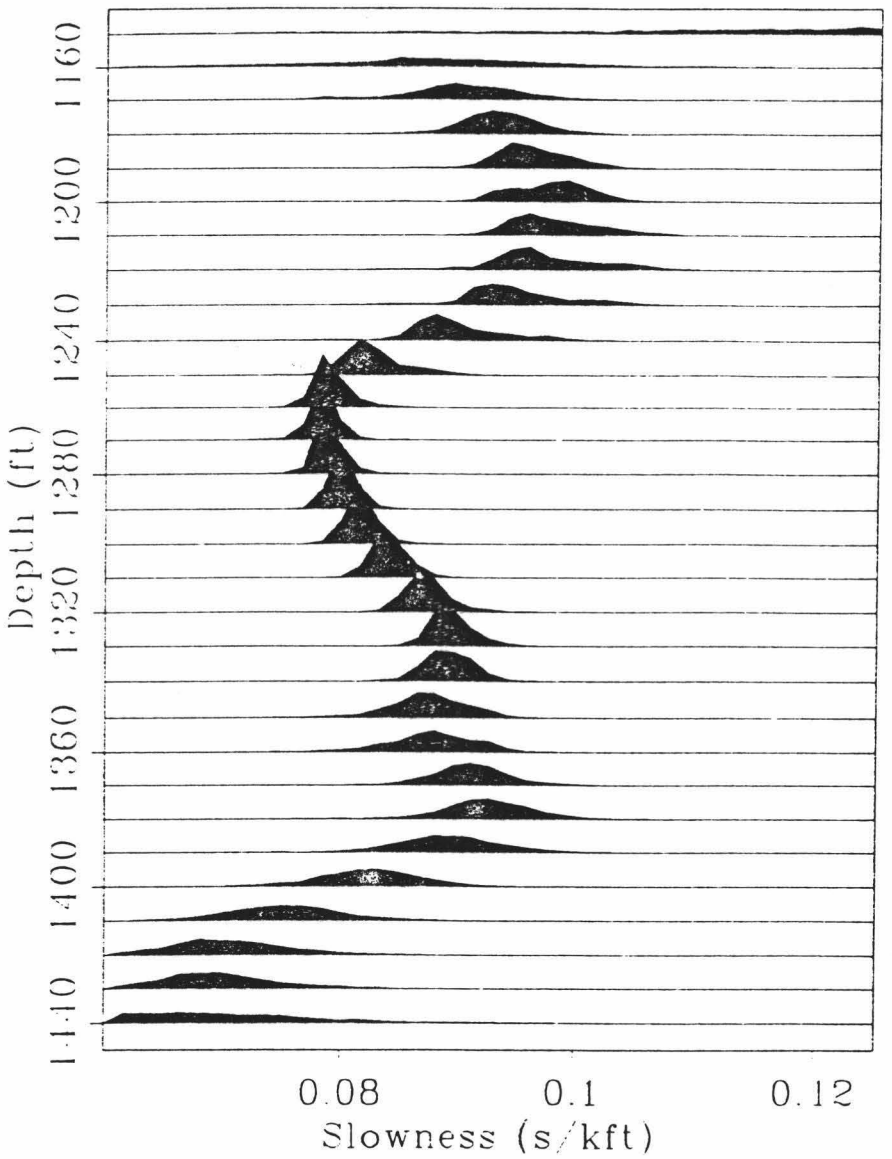


Figure 5.17: The PPD constructed using all 2400 iterations of Fig. 5.16(a)-(d).

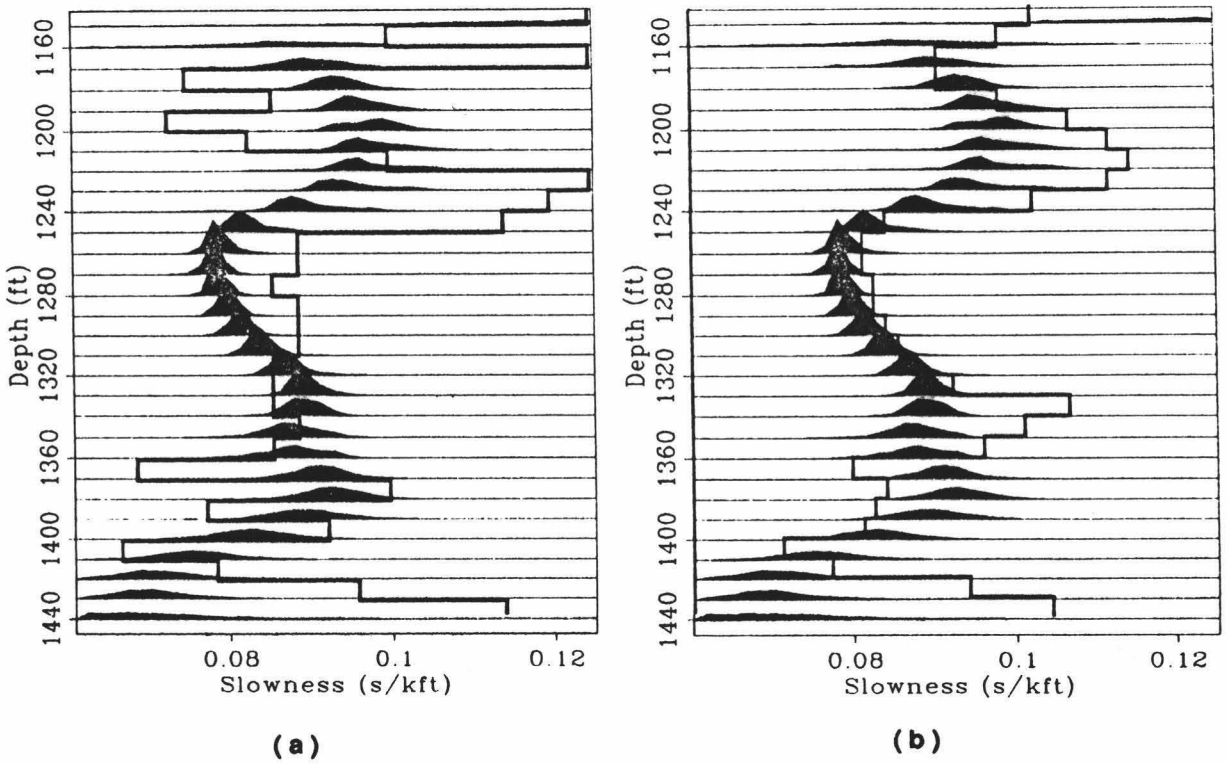


Figure 5.18: The PPD along with the (a) sonic slowness profile and the (b) tomographic slowness profile.

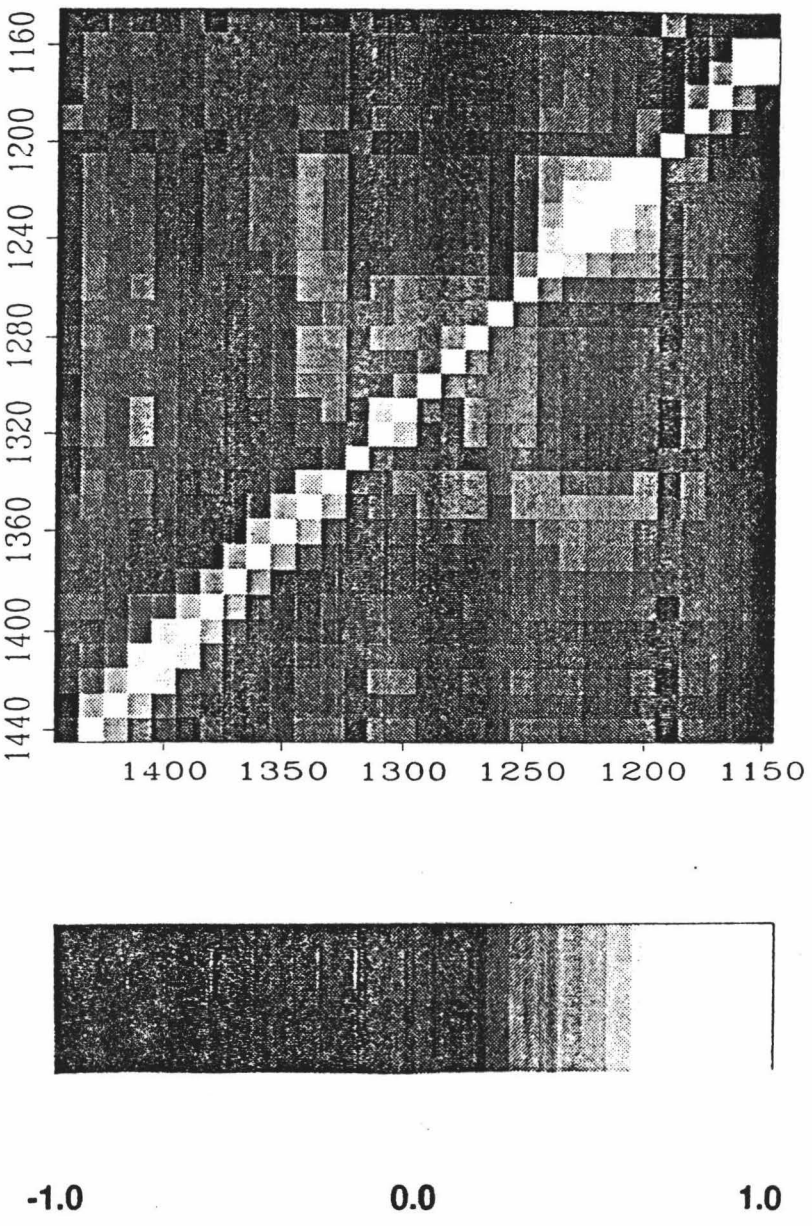


Figure 5.19: The correlation matrix. White is 1.0 and black is -1.0.

descent algorithm several times starting from different initial solutions and consider the best solution found. Studies of a number of different problems (Lundy & Mees 1986, van Laarhoven & Aarts 1987) have shown that in a given amount of computing time SA can obtain better solutions than repeating a descent algorithm using different starting models. Even though SA is proved superior than the descent algorithm in several different problems, if the problem has no local minima but only a global minimum then one run of descent algorithm will reach that minimum faster than SA.

Some of the drawbacks of SA are that it is computationally expensive and the best choice of cooling schedule is critical. One approach to speed up SA is to implement the parallel version of the algorithm (Aarts & Korst 1989). For the cooling schedule it can be said that whatever cooling schedule is chosen, it is important not to spend too long at high temperatures as this can waste CPU time, and also the temperature should be lowered as slowly as possible. Here we have skirted these problems by finding the critical temperature of the system and performing SA at that temperature which we called 'freeze bath method' (Frazer & Basu 1992). Even though determination of critical temperature of a system poses another big problem, by using the method applied by Basu & Frazer (1990) we can rapidly determine the critical temperature without following any trial and error process.

In our technique we have deviated from conventional SA method by giving emphasis on constructing the PPD and determining the best model from it rather than picking up the model having the lowest energy after the SA runs. The vastness of our model space makes the construction of the actual PPD impossible, but we can construct a PPD which closely resembles the actual one using the models which significantly contribute to the actual PPD. Our method of SA actually tries to find those good models in the least amount of time. For this reason we gave a lot of emphasis on finding the critical temperature of the system because at this temperature

we have pretty high chances of finding good models scattered over the whole solution space.

Here we have assumed that the earth is stratified and it has layers of equal thickness and the number of layers are the same as the number of sources or receivers. This assumption has caused the overparametrization problem, which resulted in picking up oscillatory velocity profiles because the energy values of those models are as high as the non-oscillatory velocity models. This problem has been taken care of by applying a penalty function and finding the optimum penalty weight. Fig. 5.20(a)-(d) showed the PPD for the larger data set after generating 600 profiles at T_{cr} for four sets of random numbers without using any penalty function. We see in this diagram that 600 profiles are not enough for the PPD to be converged as compared to the previous example. And also the PPD here is mostly flat in most of the layers.

Some results have been published which compare SA with other global optimization methods. Frazer et al. (1990) and Scales et al. (1992) have applied SA and GA to geophysical inversion problem. Since both methods are developing rapidly and research is continuing to find the best way of implementing them in different problems, so it is perhaps premature to make relative judgements. In conclusion we can say that in any problem if one of these two methods can be implemented successfully, then the other method can also be implemented with success.

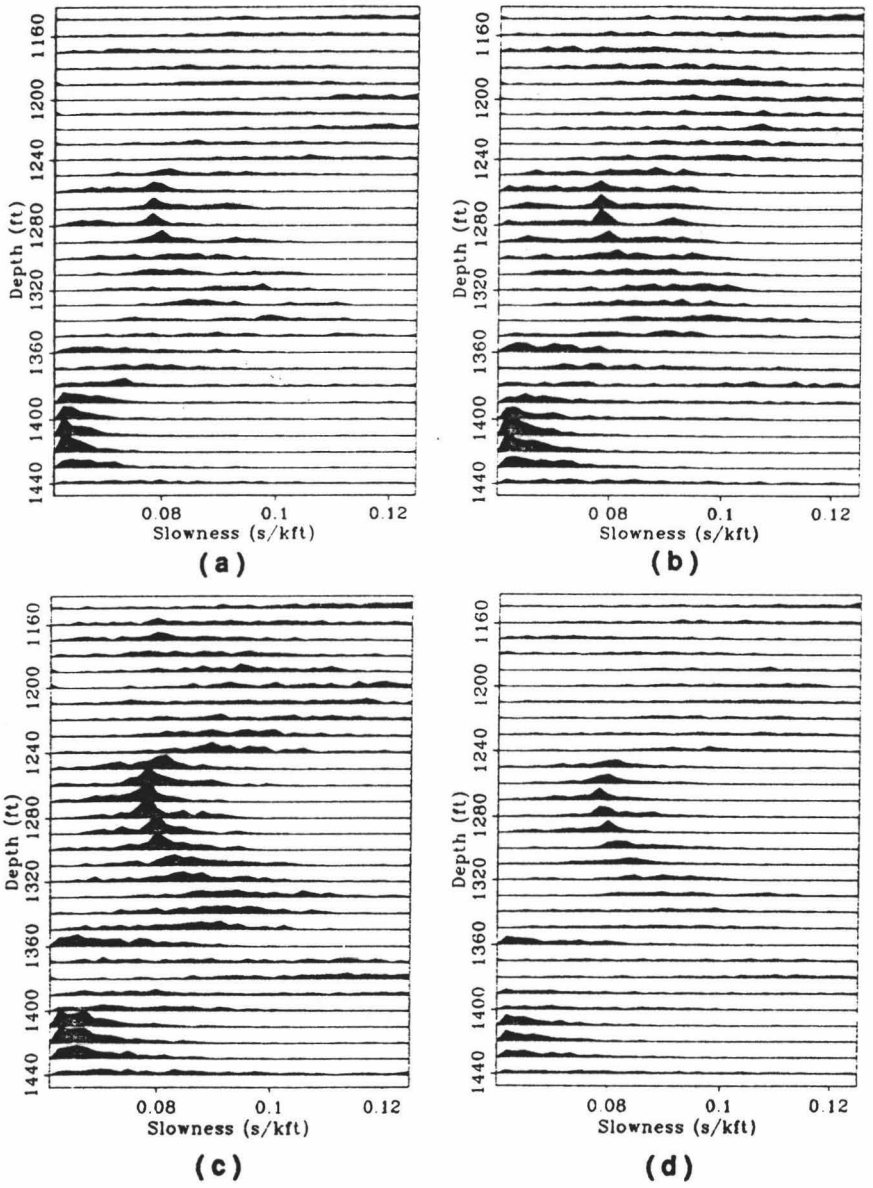


Figure 5.20: The PPD after 600 iterations without using any penalty function. (a)-(d) are for four sets of random numbers.

5.5 REFERENCES

Aarts, E. & Korst, J., 1989. *Simulated Annealing and Boltzmann Machines*, John Wiley and Sons, New York.

Basu, A. & Frazer, L. N., 1990. Rapid determination of critical temperature in simulated annealing inversion, *Science* , **249**, 1409-1412.

Frazer, L. N., Basu, A. & Low, J. D., 1990. Geophysical inversion with simulated annealing and genetic algorithms, *EOS, Transactions, Am. Geophysical Union*, **71**, 1477.

Frazer, L. N. & Basu, A., 1992. Bayesian inversion by statistical physics with an application to offset VSP, submitted to *Geophys. J. Int.*.

Laarhoven, P. J. M. Van & Aarts, E. H. L., 1987. *Simulated Annealing: Theory and applications*, D. Reidel, Boston, U.S.A.

Lundy, M. & Mees, A., 1986. Convergence of an annealing algorithm, *Mathematical Programming*, **34**, 111-124.

Scales, J. A., Smith, M. L. & Fischer, T. L., 1992. Global optimization methods for multimodal inverse problems, *J. Comp. Phys.*, **103**, 258-268.

Zimmerman, L. J. & Chen, S. T., 1993. Comparison of vertical seismic profiling techniques, *Geophysics*, **58**, 134-140.