Nonlinear 2D elastic inversion of multi-offset seismic data

Peter Mora

ABSTRACT

The treatment of multi-offset seismic data as an acoustic wavefield is becoming increasingly disturbing to many geophysicists who see a multitude of wave phenomena such as amplitude-offset variations or shear wave events, which can only be explained by using the more correct physical equation, namely the elastic wave equation. Not only are these phenomena ignored by acoustic theory, but they are also treated as undesirable noise when they should rather be used to provide extra information about the subsurface such as S-velocity.

The problems of using the conventional acoustic wave equation approach can be eliminated by starting afresh with an elastic approach. One framework has been provided by Tarantola (1984) who described how to do an elastic inversion of seismic data in very general situations for the Lamés parameters and density. In this paper, new equations have been derived to perform an inversion for P-velocity, S-velocity and density as well as the P-impedance, S-impedance and density since these are better resolved than the Lamé parameters. The inversion is based on nonlinear least squares and proceeds by iteratively updating the earth parameters until a good fit is achieved between the observed data and the modeled data corresponding to these earth parameters. The iterations are based on the preconditioned conjugate gradient algorithm. The fundamental requirement of such a least squares algorithm is the gradient direction which tells how to update the model parameters. This can be derived directly from the wave equation and it may be computed by several wave propagations. Although any scheme could in principle be chosen to perform the wave propagations, the elastic finite difference method is used because it directly simulates the elastic wave equation and can

handle complex and thus realistic distributions of elastic parameters. This method of inversion is costly since it is similar to an iterative prestack shot-profile migration but has greater power than any migration since it solves for the P-velocity, S-velocity and density and can handle very general situations including transmission problems.

Three main weaknesses of this technique are that it requires fairly accurate knowledge of the low frequency velocity model beforehand, it assumes Gaussian model statistics and that it is very computer intensive. All these problems seem surmountable. The low frequency information can be obtained either by a prior tomographic step, or by the conventional NMO method or by adding an additional inversion step for low frequencies to each iteration; the statistics can be altered by preconditioning the gradient direction perhaps to make the solution blocky in appearance like well logs; and with some improvements to the algorithm and more parallel computers, it is hoped the technique will soon become routinely feasible.

INTRODUCTION

Elastic waves or acoustic waves?

Seismic data contains many features such as shear reflections and amplitude-offset variation that provide useful information about the P-velocity, S-velocity and density. These are becoming increasingly evident as the use of longer cables and multi-component recording becomes more widespread. Conventional seismic processing and inversion based on the acoustic wave equation do not take these features into account and they therefore treat this useful information as undesirable coherent noise. The problem is that the acoustic wave equation (which assumes the earth is a liquid) was used to derive the algorithms rather than the more physically correct elastic wave equation (which assumes the earth is a solid). Therefore, if we are to use all the information contained in seismic data effectively we must commence to use the elastic wave equation and build up a new framework (or expand the old one) for the treatment of seismic data. Once the elastic wave equation is used as a basis for handling seismic data the various seismic events usually considered to be noise such as ground roll, multiples, mode conversions and S-wave events can be treated as signal and should actually be helpful (I include multiples in this list for completeness and because the elastic wave equation handles their amplitudes more correctly than the acoustic wave equation). In fact, almost all seismic "noise" ceases to be bothersome.

Inversion or conventional processing?

Another question often debated is whether inversion or standard processing methods should be applied to seismic data. Most practicing geophysicists would not hesitate to say that while the philosophy of inversion, to get a quantitative estimate of the physical properties is appealing, they would prefer to use standard methods because they do not believe inversion can work well in practice. Actually, the devout inverters would reply that standard methods are inversion too because they try to obtain a picture of the subsurface. The main difference is that standard methods are one step processes and that the result is not a quantitative estimate of the physical properties, for instance, migration gives a relative reflectivity picture and not a P-velocity picture. In general, standard methods are not required to pay heed to the exact properties of the earth but concentrate on obtaining a good image of the subsurface. In practice, the product of conventional processing would usually be a quantitative estimate of the low frequency part the P-velocities (from velocity analysis) and a relative estimate of the high frequency part of the reflectivities (from migration). Ultimately, it will be shown in this paper that the picture that can be obtained by inversion and that obtained by the conventional approach (migration) are very similar. The strength (and also weakness) of the inversion philosophy is that it tries to account for the seismic data in terms of the earth properties using the known equations of physics. The reason it is also a weakness is that it often leads to impractical algorithms.

In this paper, the method of least squares is invoked on a grand scale and by using the elastic wave equation it is possible to make use of all the amplitude information in the seismic data and perform an inversion for the P-velocity, S-velocity and density. The result of the inversion is the most likely set of physical parameters that could have given rise to the observed data (provided any energy not accountable by the elastic wave equation is uncorrelated Gaussian noise). It is believed that with the rapid development of more parallel computers that this method will soon be routinely feasible.

Inversion philosophies

Inversion is more complicated than the forward problem which attempts to simulate the equations of physics. This is because it attempts to solve a problem which is inherently unstable. For instance, in seismic exploration we attempt to predict rock properties based on the appearance of waves which in the past have propagated through some distant rocks. If the effect of the rocks on the waves is not great then in the presence of any noise we cannot hope to predict the properties of the rocks. In other words, it is sensible to only try to predict those properties which can be resolved given the

quality of the data. If a parameter is not well resolved then we want to restrict the solution in a logical way using our a priori knowledge.

To make matters more difficult, the seismic inverse problem is nonlinear (i.e. no linear relation exists between the seismogram amplitudes and the earths properties). Despite the nonlinear nature of the seismic inverse problem, most literature on seismic inversion concerns linearized methods (see Cohen and Bleistein, Stolt and Weglein, Lailly, Clayton and Stolt). The reason for this is that through functional analysis, the linearized problem is well understood. The resolution problem has been studied and largely solved by Backus and Gilbert in the linear case. To advance, we will first consider the ideal inversion, namely, mapping out the probability functions of the desired physical properties based on all available information (Tarantola and Valette, 1982). However, this is completely impractical for most inverse problems because it would take an almost infinite amount of time on any computer. Sometimes the Monte Carlo methods may be utilized to map partially the probability functions or solve for the single most probable set of model parameters (see Rothman, 1984). Unfortunately, such methods require the forward problem to be solved many times so for most seismic inverse problems where the forward problem consists of a computationally intensive wave propagation, Monte Carlo methods are too costly. Hence, we must compromise between practicality and elegance of the inversion. This paper constitutes my compromise at the present time. The inversion is based on the relatively fast and tractable least squares optimization method. It does *not* use a linearized wave equation for the solution of the forward problem but rather uses the full elastic wave equation. The iterative least squares algorithm requires a gradient direction which can be derived directly from the wave equation (see also Tarantola, 1984, for the elastic case or Lailly, 1984 for the acoustic case).

INVERSE THEORY

Introduction

In general, the process of inversion can be considered as the location of the single most probable set of model parameters $\mathbf{m}$ given some data observations $\mathbf{d}$ and knowledge of the probability distributions of $\mathbf{m}$ and $\mathbf{d}$ (see Tarantola and Valette, 1982). In the case of multi-offset seismic data inversion for the elastic properties of rocks $\mathbf{m}$, the computation of $\mathbf{d}(\mathbf{m})$, the common shot gathers, represents elastic wave propagation which consumes large amounts of computer time. Therefore, the linear function space based inversion schemes which require relatively few forward problems (wave

propagations) to be solved are preferred over the more general Monte Carlo schemes which would require many forward simulations. The reason is that linear function space methods assume local linearizability of $\mathbf{d}(\mathbf{m})$ and so can progress steadily uphill on the probability function until the nearest peak is located. Unfortunately, this may not be the biggest peak (most probable solution) which is the price paid for the requirement of only a few forward simulations. For tractability, the least squares optimization method which assumes Gaussian probability distributions for model parameters and data errors has been chosen. This should often turn out to be a reasonable assumption at least for the data errors when there are many independent sources of noise considering the central limit theorem which states that the sum of independent noise tends to be Gaussian distributed. However, the model parameters, i.e. the physical properties of the earth are in general non-Gaussian distributed but this can be at least partially handled, though not rigorously, by allowing the model mean to vary with iteration (see Mora, 1986) or by modifying the solution at each iteration by using some statistical arguments (Harlan, 1984).

Least squares and the preconditioned conjugate gradient method

The conjugate gradient (c.g.) method of nonlinear least squares has been chosen because of its simplicity yet good convergence properties. I begin with a description of nonlinear least squares and the preconditioned gradient method and later extend this to make use of the so called conjugate directions which helps to speed convergence.

Consider the Gaussian probability density function

$$P(\mathbf{d},\mathbf{m}) \quad = \quad constant \ \exp - \frac{1}{2}\left(\Delta\mathbf{d}^* \mathbf{C}_\mathbf{d}^{-1} \Delta\mathbf{d} + \Delta\mathbf{m}^* \mathbf{C}_\mathbf{m}^{-1} \Delta\mathbf{m} \right) \ , \tag{1a}$$

where $\Delta\mathbf{d} = \mathbf{d} - \mathbf{d}_0 = \mathbf{d}(\mathbf{m}) - \mathbf{d}_0$, $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$, $\mathbf{d}$ is the data vector, $\mathbf{d}_0$ is the data observations, $\mathbf{m}$ is the model vector, $\mathbf{m}_0$ is the a priori model, and $\mathbf{C}_\mathbf{d}$ and $\mathbf{C}_\mathbf{m}$ are the covariance matrices for data and model spaces respectively. (Note that * indicates conjugate transpose, normally the data consists of real values so $^* \ = \ ^T$, unless the inversion is carried out in Fourier space (or some other complex space)). Clearly, the maximum probability solution occurs when the least squares functional

$$S(\mathbf{d},\mathbf{m}) \quad = \quad \Delta\mathbf{d}^* \mathbf{C}_\mathbf{d}^{-1} \Delta\mathbf{d} + \Delta\mathbf{m}^* \mathbf{C}_\mathbf{m}^{-1} \Delta\mathbf{m} \ , \tag{1b}$$

is minimized so the least squares solution is equal to the maximum a posteriori solution when the prior distributions are Gaussian.

In the following development, I have chosen to review the least squares theory in terms of vectors and matrices though this development also applies to linear function

spaces in the L2 norm. In the continuous case, vectors would become elements of infinite-dimensional vector spaces and matrices linear operators. The vector/matrix description applies to discretized spaces which is more easily translated into computer code. One further point is that the final algorithm will be iterative and so all vectors and matrices in the following section refer to the n-th iteration though subscript n is not given explicitly to avoid cluttering in the development.

By taking the derivative of the square error functional $S(\mathbf{d},\mathbf{m})$ with respect to the model vector $\mathbf{m}$ we obtain the gradient vector $\mathbf{g}$ defined to be -1/2 of the steepest descent vector

$$\mathbf{g} = \frac{1}{2}\frac{\partial S}{\partial \mathbf{m}} = \mathbf{D}^* \mathbf{C}_\mathbf{d}^{-1}\Delta\mathbf{d} + \mathbf{C}_\mathbf{m}^{-1}\Delta\mathbf{m} \quad , \tag{2}$$

where $\mathbf{D} = \dfrac{\partial \mathbf{d}}{\partial \mathbf{m}}$ is the Frechet derivative matrix (later, $\mathbf{D}$ for the elastic wave equation will be derived as an operator because the derivation is more straightforward). Note that the factor 1/2 is introduced to simplify later expressions. One way to solve for a minimum of the least squares functional is by substituting the linearization for $\mathbf{d}(\mathbf{m})$,

$$\delta\mathbf{d} \;=\; \mathbf{d}(\mathbf{m}') - \mathbf{d}(\mathbf{m}) \;=\; \frac{\partial \mathbf{d}}{\partial \mathbf{m}}(\mathbf{m}' - \mathbf{m}) \;=\; \mathbf{D}\delta\mathbf{m} \quad , \tag{3}$$

into equation (2) and solving $\mathbf{g} = 0$. This leads to the Newton algorithm,

$$\mathbf{m}' \;=\; \mathbf{m} - \mathbf{H}\mathbf{g} \quad , \tag{4}$$

where

$$\mathbf{H} \;=\; \left(\mathbf{D}^* \mathbf{C}_\mathbf{d}^{-1}\mathbf{D} + \mathbf{C}_\mathbf{m}^{-1}\right)^{-1} \quad , \tag{5}$$

is the inverse Hessian matrix. This yields the maximum a posteriori solution in one iteration for linear $\mathbf{d}(\mathbf{m})$ (see also Tarantola and Valette, 1982). The inverse Hessian matrix modifies the gradient direction and chooses a magnitude of the model parameter update such that the best fit solution is located in one iteration for linear functions $\mathbf{d}(\mathbf{m})$.

The size of the seismic inverse problem is usually too large to handle using matrices (see Mora, 1986, for an exception) but we will see that the gradient direction $\mathbf{g}$ can be calculated by only two forward modeling runs in very general situations (i.e. elastic waves, exotic survey specifications and complex geologic models). Then by approximating the inverse Hessian $\mathbf{H}$ by $\hat{\mathbf{H}} = \eta\mathbf{C}_\mathbf{m}$ where η is called the steplength we arrive at an preconditioned gradient type iterative least squares formula,

$$\mathbf{m}' \;=\; \mathbf{m} - \hat{\mathbf{H}}\mathbf{g} \;=\; \mathbf{m} - \eta\mathbf{C}_\mathbf{m}\mathbf{g} \quad . \tag{6}$$

If this algorithm were iteratively applied the solution should converge to the same solution as that of equation (4) so the gradient iterations effectively apply the true inverse Hessian $\mathbf{H}$. The minimum located by the least squares iterations may not be the global minimum of the error functional in the case of nonlinear functions $\mathbf{d}(\mathbf{m})$ where many local minima may exist (such as in the seismic problem). Whether or not the global minima is located depends on the proximity of the starting point on the functional surface to the global minimum. This problem of local minima is intrinsic to nonlinear inverse problems and cannot be easily avoided. Actually, there are some different techniques that attempt to handle local minima including: (1) doing several inversions from different starting locations, (2) bounce methods which try to bounce out of local minima, (3) adjustment of the a priori information, (4) adding new kinds of a priori constraints, (5) careful choice of model parameters in order to minimize the degree of nonlinearity of the function $\mathbf{d}(\mathbf{m})$ in the region of interest. Of these, the most efficient and meaningful are (4) and (5). I do not include Monte Carlo methods in this list because for the situation under study, where the forward simulation is currently very CPU intensive, the Monte Carlo methods are impractical.

Often, by using some a priori knowledge or constraint, it is possible to modify the gradient direction $\mathbf{g}$ in some sensible way thus greatly speeding convergence (this may also help avoid local minima and resolve the null space problem). Another way to modify the gradient would be to derive an operator corresponding to the true inverse Hessian $\mathbf{H}$ to perform a spatial deconvolution boosting more poorly resolved low frequencies as well and unscrambling between the different model parameters. However, this would not be easy because it would be a model-dependent operator that is spatially variable. In general, a modification to the gradient is termed preconditioning and it need not be a linear operation such as in equation (6) (see for example Harlan, 1984). Let the new preconditioned gradient be denoted

$$\mathbf{p} = P(\mathbf{g}) \quad . \tag{7}$$

In order to perform the inversion by iteratively updating the model in the direction $\mathbf{p}$ we must determine the steplength η. Denote the updated model

$$\mathbf{m}' = \mathbf{m} - \eta\mathbf{p} \quad . \tag{8}$$

Assuming linearity of $\mathbf{d}(\mathbf{m})$ in the vicinity of the current model $\mathbf{m}$, the new data after the model perturbation is

$$\mathbf{d}' = \mathbf{d}(\mathbf{m}') = \mathbf{d} - \eta\mathbf{p} \quad . \tag{9}$$

Now, solve for the η which minimizes the new error functional S' corresponding to the

perturbed model.

$$\min_{\eta} S' = \min_{\eta} \left[(\mathbf{d}' - \mathbf{d}_0)^* \mathbf{C}_\mathbf{d}^{-1} (\mathbf{d}' - \mathbf{d}_0) + (\mathbf{m}' - \mathbf{m}_0)^* \mathbf{C}_\mathbf{m}^{-1} (\mathbf{m}' - \mathbf{m}_0) \right] . \quad (10)$$

Solving for the derivative of S' with respect to η yields

$$\frac{\partial S'}{\partial \eta} = -2 \left[\mathbf{p}^* \mathbf{D}^* \mathbf{C}_\mathbf{d}^{-1} (\Delta\mathbf{d} - \eta \mathbf{D}\mathbf{p}) + \mathbf{p}^* \mathbf{C}_\mathbf{m}^{-1} (\Delta\mathbf{m} - \eta \mathbf{p}) \right. . \quad (11)$$

Finally, by setting this derivative to zero we obtain the solution for η

$$\eta = \frac{\mathbf{p}^* \mathbf{g}}{\mathbf{p}^* \mathbf{D}^* \mathbf{C}_\mathbf{d}^{-1} \mathbf{D}\mathbf{p} + \mathbf{p}^* \mathbf{C}_\mathbf{m}^{-1} \mathbf{p}} . \quad (12)$$

This value for η may not be very good for highly nonlinear functions $\mathbf{d}(\mathbf{m})$ so it may be necessary to do a line search using the above value as a starting point.

Preconditioned conjugate gradient algorithm.

The use of conjugate directions helps to speed convergence by choosing a direction that is a linear combination of the past and current steepest decent directions (Luenberger, 1973). Actually, this is not very important if the preconditioning operator is well chosen but is worthwhile because it can be included at no extra cost to the algorithm (it can be considered as an additional preconditioning). One choice for the conjugate direction $\mathbf{c}_n$ is the Polak-Ribiere method

$$\mathbf{c}_n = \mathbf{p}_n + \frac{\mathbf{p}_n^* (\mathbf{g}_n - \mathbf{g}_{n-1})}{\mathbf{p}_{n-1}^* \mathbf{g}_{n-1}} \mathbf{c}_{n-1} .$$

This choice is preferred over the Fletcher-Reeves method because it tends to revert to the simple gradient algorithm in situations where the function is very nonlinear and so is more reliable (Powell, 1981).

Thus, the final algorithm for nonlinear least squares by the preconditioned conjugate gradient method (c.g. algorithm) is

for $n = 1$ *to* ∞

$\mathbf{d}_n = \mathbf{d}(\mathbf{m}_n)$	*data calculation*
$\Delta\mathbf{d}_n = \mathbf{d}_n - \mathbf{d}_0$, $\Delta\mathbf{m}_n = \mathbf{m}_n - \mathbf{m}_0$	*compute residuals*
$S(\mathbf{d},\mathbf{m}) = \Delta\mathbf{d}^* \mathbf{C}_\mathbf{d}^{-1} \Delta\mathbf{d} + \Delta\mathbf{m}_n^* \mathbf{C}_\mathbf{m}^{-1} \Delta\mathbf{m}_n$	*square error functional*
exit if converged	

$$\mathbf{g}_n = \mathbf{D}_n^* \mathbf{C}_\mathbf{d}^{-1} \Delta\mathbf{d} + \mathbf{C}_\mathbf{m}^{-1} \Delta\mathbf{m}_n \qquad \textit{gradient}$$

$$\mathbf{p}_n = P(\mathbf{g}_n) \approx \mathbf{C}_\mathbf{m} \mathbf{g}_n \qquad \textit{preconditioning}$$

$$\mathbf{c}_n = \mathbf{p}_n + \frac{\mathbf{p}_n^* \left(\mathbf{g}_n - \mathbf{g}_{n-1} \right)}{\mathbf{p}_n^* \mathbf{g}_n} \mathbf{c}_{n-1} \quad , \quad \mathbf{c}_1 = \mathbf{p}_1 \qquad \textit{conjugate direction}$$

$$\eta_n = \frac{\mathbf{c}_n^* \mathbf{g}_n}{\mathbf{c}_n^* \mathbf{D}_n^* \mathbf{C}_\mathbf{d}^{-1} \mathbf{D}_n \mathbf{c}_n + \mathbf{c}_n^* \mathbf{C}_\mathbf{m}^{-1} \mathbf{c}_n} \qquad \textit{calculate steplength}$$

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \eta_n \mathbf{c}_n \qquad \textit{update model}$$

end (13)

The reason n starts at 1 is because I used subscripts 0 to indicate the a priori model $\mathbf{m}_0$ and field data $\mathbf{d}_0$.

Computational aspects of the conjugate gradient algorithm

It will become clear in the next section that the operation of $\mathbf{D}_n^*$ on a vector can be achieved by a forward simulation (wave propagation) so the c.g. equations may be calculated with three forward simulations, one to compute the synthetic data $\mathbf{d}_n = \mathbf{d}(\mathbf{m}_n)$, one to make the gradient as described in the next section (i.e. to compute the $\mathbf{D}_n^* \mathbf{C}_\mathbf{d}^{-1} \Delta\mathbf{d}_n$ term of $\mathbf{g}_n$), and one to compute the steplength η (i.e. the $\mathbf{D}_n \mathbf{c}_n$ term). All other computations are simple dot products. If a line search is required, then several more forward simulations may be necessary in order to optimize the steplength η. In the examples in this paper, I carried out such a line search for the optimal η which tended to be within a factor of two of that predicted by equation (12). Because the predicted and optimal values of η were so close, only two additional forward simulations were required to do the line search. For a particular problem it may be found that the optimal η is always approximately the same so maybe a line search is not always necessary.

There is a complication in the computation of η. Because the elastic wave equation is nonlinear and we never explicitly obtain the Frechet derivatives $\mathbf{D}_n$, we cannot compute $\mathbf{D}_n c_n$ directly. Instead, it must be computed using the forward problem $\mathbf{d}(\mathbf{m})$ with the following formula

$$\mathbf{D}_n \xi\mathbf{c}_n = \mathbf{D}_n (\mathbf{m}_n + \xi\mathbf{c}_n) - \mathbf{D}_n \mathbf{m}_n = \mathbf{d}(\mathbf{m}_n + \xi\mathbf{c}_n) - \mathbf{d}(\mathbf{m}_n) \quad , \qquad (14)$$

where ξ is chosen such that the model perturbation $\xi\mathbf{c}_n$ is sufficiently small relative to the model parameters $\mathbf{m}_n$ (say 1%) that the function $\mathbf{d}(\mathbf{m})$ is about linear. Therefore, in practice, the formula for η is given by

$$\eta_n = \frac{\mathbf{c}_n^* \mathbf{g}_n}{\frac{1}{\xi^2}\left[\mathbf{c}_n^* \xi \mathbf{D}_n^* \mathbf{C}_\mathbf{d}^{-1} \mathbf{D}_n \, \xi \mathbf{c}\right] + \mathbf{c}_n^* \mathbf{C}_\mathbf{m}^{-1} \mathbf{c}_n} \quad , \tag{15}$$

where $\mathbf{D}_n \, \xi \mathbf{c}_n$ is computed using equation (14).

A further complication is the covariance matrices. If these are assumed to be arbitrary they would require excessive computer storage and CPU time so in practice diagonal matrices are taken. Though diagonal, the covariance matrices need not be constant but may vary along the diagonal which assumes independent but non-constant noise and independent model parameters. In order that this assumption be reasonable, it may be necessary to choose carefully the model parameters (and perhaps data parameters). A viable alternative to the choice of diagonal covariance matrices is when $\mathbf{C}_\mathbf{d}^{-1}$ and $\mathbf{C}_\mathbf{m}^{-1}$ are constant in the diagonal direction and so represent simple filtering operations.

GRADIENT CALCULATION

Introduction and overview

If we assume the earth is perfectly elastic, then the seismic forward problem $\mathbf{d}(\mathbf{m})$ may be computed by solving the elastic wave equation (E.W.E.) (see Aki and Richards, 1980),

$$\rho \ddot{u}_i - \partial_j \, C_{ijkl} \, \partial_l \, u_k = f_i \quad , \tag{16a}$$

$$C_{ijkl} \, \partial_l \, u_k \, n_j = T_i \quad , \tag{16b}$$

$$u_i = 0 \quad , \quad t < 0 \quad , \tag{16c}$$

$$\dot{u}_i = 0 \quad , \quad t < 0 \quad , \tag{16d}$$

where $u_i = u_i(\mathbf{x}_S, \mathbf{x}, t)$ is the i-th component of displacement resulting from a shot (i.e. body force f_i and/or traction T_i) located at $\mathbf{x}_S$. If we have receivers located at $\mathbf{x}_R$ then the data, $\mathbf{d}(\mathbf{m})$, is given by

$$\mathbf{d}(\mathbf{m}) = u_i(\mathbf{x}_S, \mathbf{x}_R, t) \quad . \tag{17}$$

In order to perform inversion using the c.g. algorithm equations (13) and (15), we require the gradient $\mathbf{g}$ defined by equation (2) corresponding to model parameters $\mathbf{m}$. For the elastic wave equation, it is clear that the most obvious choices of model parameters $\mathbf{m}$ are the Hooke tensor C_{ijkl}, the density ρ, the body force f_i and the traction T_i (for an isotropic medium the Hooke tensor can be described by the two Lame parameters, λ and μ). However, there are other choices of model parameters which are more

physically meaningful and (as it turns out) better resolved (see Tarantola et al. 1985). In particular, we could choose the P- and S-wave velocities and density or the P- and S-wave impedances and density as model parameters. The gradient in terms of the different model parameter choices are related and so the development will proceed by deriving the gradient in terms of the simplest choice (λ, μ and ρ). Subsequently, the expressions for the gradients in terms of the other model parameter choices will be derived and cast in terms of the gradient for λ, μ and ρ.

The derivation is based on finding an expression equivalent to the linearized forward problem

$$\delta\mathbf{d} \quad = \quad \mathbf{D}\delta\mathbf{m} \quad . \tag{18a}$$

or in continuous form

$$\delta\mathbf{d}(\mathbf{D}) \quad = \quad \int_M dM \frac{\partial \mathbf{d}(D)}{\partial \mathbf{m}(M)} \delta\mathbf{m}(M) \quad , \tag{18b}$$

where M indicates the model space. This expression indicates how to calculate a small perturbation in the wavefield $\delta\mathbf{d}$ resulting from a small perturbation in the model parameters $\delta\mathbf{m}$ by integrating over the model space, i.e. it is the linearized Green function representation of the forward problem or Born approximation. Once we have a representation equivalent to equation (18b) we can identify the Frechet kernel $\mathbf{D} = \dfrac{\partial \mathbf{d}}{\partial \mathbf{m}}$ and hence we can compute the adjoint operation

$$\delta\hat{\mathbf{m}} \quad = \quad \mathbf{D}^* \delta\mathbf{d} \quad , \tag{19a}$$

or in continuous form

$$\delta\hat{\mathbf{m}}(M) \quad = \quad \int_D dD \left(\frac{\partial \mathbf{d}(D)}{\partial \mathbf{m}(M)} \right)^* \delta\mathbf{d}(D) \quad . \tag{19b}$$

Note that the hat is used to make it clear that $\delta\mathbf{m}$ and $\delta\hat{\mathbf{m}}$ are not the same (in fact $\delta\hat{\mathbf{m}}$ does not even have the same units as $\delta\mathbf{m}$; it has units of $\Big(units\,(D) \Big)^2 / units\,(M)$ rather than simply $units\,(M)$).

Once expression (19b) has been derived we simply replace $\delta\mathbf{d}$ by $\mathbf{C}_\mathbf{d}^{-1}\Delta\mathbf{d}$ and add this result to $\mathbf{C}_\mathbf{m}^{-1}\Delta\mathbf{m}$ to obtain the desired gradient,

$$\mathbf{g} \quad = \quad \mathbf{D}^* \mathbf{C}_\mathbf{d}^{-1}\Delta\mathbf{d} + \mathbf{C}_\mathbf{m}^{-1}\Delta\mathbf{m} \quad .$$

Adjoint of the E.W.E. parametrized by Lamés parameters and density

This derivation of the adjoint of the elastic wave equation (E.W.E.) follows Tarantola 1984. For the seismic problem, the linearized forward problem equivalent to equation (18b) is of form

$$\delta u_i(\mathbf{x}_S,\mathbf{x}_R,t) = \int_V dV(\mathbf{x}) \frac{\partial u_i(\mathbf{x}_S,\mathbf{x}_R,t)}{\partial \mathbf{m}(\mathbf{x})} \delta \mathbf{m}(\mathbf{x}) \quad , \tag{20}$$

where $u_i(\mathbf{x}_S,\mathbf{x}_R,t)$ represents the seismogram located at receiver location $\mathbf{x}_R$ which records the i-th component of displacement of an elastic wavefield due to a shot at $\mathbf{x}_S$ and $\mathbf{m}(\mathbf{x}) = \left[C_{ijkl}(\mathbf{x}),\rho(\mathbf{x}),f_i(\mathbf{x}),T_i(\mathbf{x})\right]^T$ at location $\mathbf{x}$ in the earth. Therefore, the adjoint problem corresponding to equation (19b) is

$$\delta \hat{\mathbf{m}}(\mathbf{x}) = \sum_S \int dt \sum_R \frac{\partial u_i(\mathbf{x}_S,\mathbf{x}_R,t)}{\partial \mathbf{m}(\mathbf{x})} \delta u_i(\mathbf{x}_S,\mathbf{x}_R,t) \quad , \tag{21}$$

i.e. the integral over the data space of the data residuals multiplied by the Frechet kernel. From the above two expressions, it is clear that in order to obtain the adjoint operation (21), all that is required is the integral expression (20) for the forward problem which gives the perturbation in displacement δu_i corresponding to some perturbations in the model parameters $\delta \mathbf{m}$. This expression would define the Frechet kernel so the adjoint operation (21) could then be written by identifying terms in equations (20) and (21). The linearized solution of the elastic wave equation in terms of Green's functions (the Born approximation) supplies the appropriate integral.

To avoid cluttering the equations with the shot location $\mathbf{x}_S$, the following development refers to a single shot. Therefore, the required adjoint operation can be obtained from the result of this development by simply including the sum over shots. Now consider the elastic wave equation

$$\rho \ddot{u}_i - \partial_j C_{ijkl} \partial_l u_k = f_i \quad , \tag{22a}$$

$$C_{ijkl} \partial_l u_k n_j = T_i \quad , \tag{22b}$$

$$u_i = 0 \quad , \quad t < 0 \quad , \tag{22c}$$

$$\dot{u}_i = 0 \quad , \quad t < 0 \quad , \tag{22d}$$

In order to find the integral corresponding to (20), we make the following substitutions and subsequently solve using the Green's functions.

$$u_i \rightarrow u_i + \delta u_i \quad , \tag{23a}$$

$$\rho \rightarrow \rho + \delta \rho \quad , \tag{23b}$$

$$C_{ijkl} \rightarrow C_{ijkl} + \delta C_{ijkl} \quad , \tag{23c}$$

$$f_i \rightarrow f_i + \delta f_i \quad , \tag{23d}$$

and

$$T_i \rightarrow T_i + \delta T_i \quad . \tag{23e}$$

These substitutions yield a new elastic wave equation describing the displacement perturbation δu_i as a function of new force and traction terms Δf_i and ΔT_i:

$$\rho \delta \ddot{u}_i - \partial_j C_{ijkl} \partial_l \delta u_k = \Delta f_i \quad , \tag{24a}$$

$$C_{ijkl} \partial_l \delta u_k n_j = \Delta T_i \quad , \tag{24b}$$

$$\delta u_i = 0 \quad , \quad t < 0 \quad , \tag{24c}$$

$$\delta \dot{u}_i = 0 \quad , \quad t < 0 \quad , \tag{24d}$$

where the new force and traction terms are

$$\Delta f_i = \delta f_i - \delta \rho \ddot{u}_i + \partial_j \delta C_{ijkl} \partial_l u_k + O_i(\delta \rho, \delta C_{ijkl}, \delta f_i, \delta T_i)^2 \quad , \tag{25a}$$

and

$$\Delta T_i = \delta T_i - \delta C_{ijkl} \partial_l u_k n_j + O_i(\delta \rho, \delta C_{ijkl}, \delta f_i, \delta T_i)^2 \quad . \tag{25b}$$

The new wave equation given in equation (24) has the same form as the elastic wave equation and hence its solution can be obtained in terms of the Green's functions of the elastic wave equation. The solution in terms of the elastic Green's functions (Aki and Richards, 1980) is

$$\delta u_i(\mathbf{x}_R, t) = \int_V dV(\mathbf{x}) G_{ij}(\mathbf{x}_R, t; \mathbf{x}, 0) * \Delta f_j(\mathbf{x}, t) + \int_S dS(\mathbf{x}) G_{ij}(\mathbf{x}_R, t; \mathbf{x}, 0) * \Delta T_j(\mathbf{x}, t) \quad . \tag{26}$$

This equation has the form of the desired expression for the forward problem, equation (20), and so it defines the Frechet kernel $\partial u_i(\mathbf{x}_S, \mathbf{x}_R, t)/\partial \mathbf{m}(\mathbf{x})$ from which we may obtain the adjoint expression, equation (21). A simplified adjoint expression can be obtained with some algebra.

Substituting the force and traction terms given in equations (25) into equation (26) and neglecting the O^2 terms (i.e. assume small perturbations in order to obtain the Frechet derivatives) and dropping the arguments of the various functions to avoid cluttering, and using notation $u_{m,l} = \partial_l u_m$ yields

$$\delta u_i = \int_V dV G_{ij} * \left[\delta f_j - \delta \rho \ddot{u}_j + \partial_k \delta C_{jklm} u_{l,m} \right] + \int_S dS G_{ij} * \left[\delta T_j - \delta C_{jklm} u_{l,m} n_k \right] \quad . \tag{27}$$

This equation may be greatly simplified by using the following three results

(i) $f\,(t)*\dot{g}\,(t) = \dot{f}\,(t)*g\,(t)$, a property of convolution,

(ii) $\int_V dV\,\partial_k F = \int_S dS\;n_k F$, the divergence theorem, and

(iii) $\partial_k \left[G_{ij} * \left(\delta C_{jklm}\, u_{l,m} \right) \right] = G_{ij} * \left(\partial_k\, \delta C_{jklm}\, u_{l,m} \right) + \left(\partial_k\, G_{ij} \right) * \left(\delta C_{jklm}\, u_{l,m} \right)$,

by the chain rule.

Applying these results to equation (27) yields the formula

$$\delta u_i = \int_V dV G_{ij} * \delta f_j + \int_S dS G_{ij} * \delta T_j$$

$$- \int_V dV \dot{G}_{ij} * \dot{u}_j\, \delta\rho - \int_V dV \left(\partial_k\, G_{ij} \right) * \left(\delta C_{jklm}\, u_{l,m} \right) . \tag{28}$$

Finally, assuming isotropy, we have

$$\delta C_{jklm} = \delta\lambda \delta_{jk}\, \delta_{lm} + \delta\mu(\delta_{jm}\, \delta_{kl} + \delta_{jl}\, \delta_{km}) , \tag{29}$$

and therefore we get

$$\delta u_i = \int_V dV G_{ij} * \delta f_j + \int_S dS G_{ij} * \delta T_j$$

$$- \int_V dV \dot{G}_{ij} * \dot{u}_j\, \delta\rho - \int_V dV G_{ij,j} * u_{m,m}\, \delta\lambda - \int_V dV G_{ij,k} * (u_{j,k} + u_{k,j})\delta\mu . \tag{30}$$

This equation has exactly the same form as equation (20) so clearly it defines the Frechet kernel $\frac{\partial u}{\partial \mathbf{m}}$ for model parameters ρ, λ, μ, f_j and T_j. Use of this equation to solve the forward problem is referred to as the Born approximation. In this paper the Born approximation is not used to solve the forward problem but rather the full elastic wave equation (22). By integrating the Frechet kernel defined by equation (30) over the data space (see equation (21)) we obtain the adjoint operation $\delta\hat{\mathbf{m}} = [\delta\hat{\rho}\ ,\ \delta\hat{\lambda}\ ,\ \delta\hat{\mu}\ ,\ \delta\hat{f}_j\ ,\ \delta\hat{T}_j\ ]^T$ where

$$\delta\hat{\rho} = -\int dt \sum_R \dot{G}_{ij} * \dot{u}_j\, \delta u_i \quad , \tag{31a}$$

$$\delta\hat{\lambda} = -\int dt \sum_R G_{ij,j} * u_{m,m}\, \delta u_i \quad , \tag{31b}$$

$$\delta\hat{\mu} = -\int dt \sum_R G_{ij,k} * (u_{j,k} + u_{k,j})\delta u_i$$

$$= -\int dt \sum_R \frac{1}{\sqrt{2}} \left(G_{ij,k} + G_{ik,j} \right) * \frac{1}{\sqrt{2}} \left(u_{j,k} + u_{k,j} \right) \delta u_i \quad , \tag{31c}$$

$$\delta \hat{f}_j = \int dt \sum_R G_{ij} * \delta u_i \quad , \tag{31d}$$

$$\delta \hat{T}_j = \int dt \sum_R G_{ij} * \delta u_i \quad . \tag{31e}$$

These expressions may be simplified by using the commutativity of convolution (i.e. $f*g = g*f$) and the following property in order to shift the location of the convolution

$$\int dt \; f(t) * g(t) h(t) = \int dt \; f(t) g(t) * h(-t) \quad . \tag{32}$$

The resulting integrals corresponding to equations (22) have the form

$$\delta \hat{\mathbf{m}}(\mathbf{x}) = \int dt \sum_R \Big(\Omega u_j(\mathbf{x},t) \Big) \Big(\Omega G_{ij}(\mathbf{x}_R, t; \mathbf{x}, 0) * \delta u_i(\mathbf{x}_R, -t)$$

$$= \int dt \Big(\Omega u_j(\mathbf{x},t) \Big) \Big(\Omega \sum_R G_{ij}(\mathbf{x}, 0; \mathbf{x}_R, -t) * \delta u_i(\mathbf{x}_R, -t) \Big)$$

$$= \int dt \Big(\Omega u_j(\mathbf{x},t) \Big) \Big(\Omega \psi_j(\mathbf{x},t) \Big) \quad . \tag{33a}$$

Following Tarantola 1984, I have made use of the properties of the elastic Green's functions, namely reciprocity (i.e. interchangeability of $\mathbf{x}$ with $\mathbf{x}_R$ and time translational invariance (i.e. $G_{ij}(\mathbf{x}_R, t; \mathbf{x}, 0) = G_{ij}(\mathbf{x}, 0; \mathbf{x}_R, -t)$ and introduce the new wavefield

$$\psi_j(\mathbf{x},t) = \sum_R G_{ij}(\mathbf{x}, 0; \mathbf{x}_R, -t) * \delta u_i(\mathbf{x}_R, -t) \quad , \tag{33b}$$

to be interpreted later. Note that I also introduced an operator Ω which is dependent on the type of model parameter at each location $\mathbf{x}$. Specifically, from equations (32) we have

$$\Omega_\rho = \partial_t \quad , \quad \Omega_\lambda = \partial_j \quad , \quad \Omega_\mu = \frac{1}{\sqrt{2}} \left(\delta_{jk} \partial_i + \delta_{ji} \partial_k \right) \quad , \tag{34a}$$

and

$$\Omega_{f_j} = \Omega_{T_j} = \sqrt{\frac{\delta(t - t')}{u_j}} \quad . \tag{34b}$$

This operator is described in more detail in the section that compares inversion to migration.

Interpreting the adjoint operation

Consider equation (33a) which is the adjoint to the elastic wave equation. There are only two unknowns, the wavefields u_j and ψ_j. The wavefield u_j can be computed by forward modeling (i.e. doing a shot simulation). Now consider the definition of ψ_j of equation (33b). The Green's function G_{ij} gives the displacement at location $\mathbf{x}$ and time 0 due to an impulsive force at $\mathbf{x}_R$ and time $-t$. Hence, the wavefield ψ is generated applying $\delta u_i(\mathbf{x}_R, -t)$ as the forcing term in the wave equation and therefore can be calculated by another forward modeling run. Since δu_i is the time reversed residuals (difference between two recorded wavefields), we call ψ_j the *back propagated* residual wavefield. From this interpretation it is clear that the adjoint can be calculated with the following steps:

(i) Simulate elastic wave propagation to solve for $\mathbf{u}$.

(ii) Compute the data residuals $\delta\mathbf{u} = \mathbf{u}-\mathbf{u}_0$.

(iii) Back propagate the data residuals (i.e. compute the wavefield resulting from a forcing function equal to $\delta\mathbf{u}$). Simultaneously apply $\mathbf{\Omega}$ and compute the time integral which represents the computation of the correlation between $\mathbf{\Omega}$ operated on the forward propagated field and Ω operated on the back propagated residual wavefield.

The gradient direction

The c.g. inversion algorithm (13) requires the gradient $\mathbf{g} = \mathbf{D}^* \mathbf{C_d}^{-1}\Delta\mathbf{d} + \mathbf{C_m}^{-1}\Delta\mathbf{m}$ but equation (33) only represents the adjoint operation, i.e. $\mathbf{D}^*\delta\mathbf{d}$. Assuming a diagonal data covariance function (uncorrelated noise) for simplicity of computation and using equation (33) to define the adjoint operator $\mathbf{D}^*$, we obtain for the gradient

$$\mathbf{g}(\mathbf{x}) = \sum_S \int dt\, \mathbf{C_d}^{-1}(t) \Big(\mathbf{\Omega} u_j(\mathbf{x},t) \Big) \Big(\mathbf{\Omega} \psi_j(\mathbf{x},t) \Big) + \mathbf{C_m}^{-1}\Delta\mathbf{m} \quad , \tag{35a}$$

where I have now reintroduced the shot sum that was dropped in the development to avoid cluttering and ψ is now defined by

$$\psi_j(\mathbf{x},t) \;=\; \sum_R \mathbf{C_d}^{-1}(\mathbf{x}_R)\, G_{ij}(\mathbf{x},0;\mathbf{x}_R,-t) * \Delta u_i(\mathbf{x}_R,-t) \quad , \tag{35b}$$

so it is the back propagated residual wavefield created by taking the difference Δu_i between the observed data u_{i_0} and the forward modeled wavefield u_i and using this as a forcing function in the wave equation. This assumes that the data covariances can be represented as

$$\mathbf{C_d}^{-1}(\mathbf{x}_R, t) \;=\; \mathbf{C_d}^{-1}(\mathbf{x}_R)\, \mathbf{C_d}^{-1}(t) \quad .$$

Finally, rewriting the gradient (35) using the definition of the operator Ω from equation (25) we obtain for the components of the gradient now denoted $\delta\hat{\hat{\rho}}$, $\delta\hat{\hat{\lambda}}$, $\delta\hat{\hat{\mu}}$, $\delta\hat{\hat{f}}_j$ and $\delta\hat{\hat{T}}_j$, (i.e. $\mathbf{g}^T = \left[\delta\hat{\hat{\rho}}\ ,\ \delta\hat{\hat{\lambda}}\ ,\ \delta\hat{\hat{\mu}}\ ,\ \delta\hat{\hat{f}}_j\ ,\ \delta\hat{\hat{T}}_j\ \right]$)

$$\delta\hat{\hat{\rho}} = -\sum_S \int dt\ \mathbf{C}_{\mathbf{d}}^{-1}\dot{u}_j\,\dot{\psi}_j\ + \mathbf{C}_{\rho\rho}^{-1}(\rho-\rho_0) + \cdots$$

$$= \delta\rho'\ + \mathbf{C}_{\rho\rho}^{-1}(\rho-\rho_0) + \cdots\ , \tag{36a}$$

$$\delta\hat{\hat{\lambda}} = -\sum_S \int dt\ \mathbf{C}_{\mathbf{d}}^{-1}u_{m,m}\,\psi_{j,j}\ + \mathbf{C}_{\lambda\lambda}^{-1}(\lambda-\lambda_0) + \cdots$$

$$= \delta\lambda'\ + \mathbf{C}_{\lambda\lambda}^{-1}(\lambda-\lambda_0) + \cdots\ , \tag{36b}$$

$$\delta\hat{\hat{\mu}} = -\sum_S \int dt\ \mathbf{C}_{\mathbf{d}}^{-1}\frac{1}{\sqrt{2}}\left(u_{k,j}+u_{j,k}\right)\frac{1}{\sqrt{2}}\left(\psi_{k,j}+\psi_{j,k}\right) + \mathbf{C}_{\mu\mu}^{-1}(\mu-\mu_0) + \cdots$$

$$= \delta\mu'\ + \mathbf{C}_{\mu\mu}^{-1}(\mu-\mu_0) + \cdots\ , \tag{36c}$$

$$\delta\hat{\hat{f}}_j = \sum_S \mathbf{C}_{\mathbf{d}}^{-1}\psi_j\ + \mathbf{C}_{f_j f_j}^{-1}(f_j-f_{j0}) + \cdots$$

$$= \delta f_j'\ + \mathbf{C}_{f_j f_j}^{-1}(f_j-f_{j0}) + \cdots\ , \tag{36d}$$

$$\delta\hat{\hat{T}}_j = \sum_S \mathbf{C}_{\mathbf{d}}^{-1}\psi_j\ + \mathbf{C}_{T_j T_j}^{-1}(T_j-T_{j0}) + \cdots$$

$$= \delta T_j'\ + \mathbf{C}_{T_j T_j}^{-1}(T_j-T_{j0}) + \cdots\ . \tag{36e}$$

These equations define $\delta\lambda'$, $\delta\mu'$ and $\delta\rho'$ which will be used later in the expressions for the gradient in terms of the other choices of model parameters. The dots at the end of each equation represent cross-covariance terms between the different types of model parameters. These cross covariances are usually considered to be negligible for the sake of simplicity and due to lack of knowledge. Furthermore, constant diagonal covariance functions are normally assumed, i.e. $\mathbf{C}_{\rho\rho} = \sigma_{\rho\rho}^2\mathbf{I}$ etc. (for an exception see Mora, 1986). This means that the physical parameters are assumed to be spatially uncorrelated. This is generally false so biases can be expected in the solutions. Perhaps it would be best to at least include diagonal cross covariance functions $\mathbf{C}_{\rho\lambda} = \sigma_{\rho\lambda}^2\mathbf{I}$ etc. if reasonable values can be estimated.

Although the force and traction terms of the gradient have been derived as if they are spatially variable, in seismic exploration we can assume that the source locations are known so the force terms need to be solved only at the true source locations. (i.e. the force and traction covariance functions $\mathbf{C}_{f_j}$ and $\mathbf{C}_{T_j}$ are Dirac functions of form $\delta(\mathbf{x}-\mathbf{x}_S)$. Actually, it would be best to know the source functions rather than invert for them because for reflection data, the source time history cannot be well resolved from

the near surface geology and reverberation effects. However, if the direct wave is known such as is the case for VSP data, the source is much better determined and can be inverted simultaneously with the geology, see Harlan 1984. Normally, with only reflection data it is best to use a preprocessing step to find the source or more generally, to use relaxation. That is, to invert for the source only followed by inversion for the geology only and so on.

The time dependent data covariance function can be approximated as

$$\mathbf{C}_{\mathbf{d}}^{-1}(t) = \frac{t^{2p}}{\sigma_{\mathbf{d}}^2} \quad , \tag{37}$$

which allows for data errors which decay with time (i.e. errors are assumed to be unaccounted seismic waves which diverge with time and hence diminish in amplitude). Typically, for a 2D problem a value of $p \geq .5$ is reasonable while in a 3D problem $p \geq 1$ would be appropriate.

The receiver dependent inverse covariance function $\mathbf{C}_{\mathbf{d}}^{-1}(\mathbf{x}_R)$ is usually chosen to be unity except at the edges of the shot gather where a taper can be applied to decrease artifacts resulting from a sudden edge in the recorded data.

Different model parameters

A theoretical study by Tarantola et al. 1985 indicates that the choice of model parameters is especially important in the elastic case where there are three independent parameters at each earth location. The reason is that for some choices, three elastic parameters can be easily resolved from one another while for other choices they cannot. The results of Tarantola's study indicated that for seismic surveys where the source is dominantly P-waves, the choice of the Lamé parameters was much worse than either the choice of the velocities or impedances. Furthermore, the results indicated that the density was not very well resolved (especially the from P-wave velocity) and that in practice it is probable that only two parameters such as P- and S-velocity or P- and S-impedance will be resolvable. Therefore, I will derive expressions for the gradient in terms the velocities and impedances.

The equations for the gradient in terms of the P-velocity α, S-velocity β and density ρ and the P-impedance Z_P, S-impedance Z_S and density ρ can be easily derived in terms of the gradient for the Lamé parameters and density by changing variables. For example, consider the gradient

$$\mathbf{g_m} \quad = \quad \frac{\partial \mathbf{d}}{\partial \mathbf{m}} \mathbf{C}_{\mathbf{d}}^{-1} \Delta \mathbf{d} + \mathbf{C}_{\mathbf{m}}^{-1} \Delta \mathbf{m} \quad ,$$

where the subscript $\mathbf{m}$ on $\mathbf{g}$ indicates that the gradient is in terms of model parameters $\mathbf{m}$. The gradient in terms of different model parameters $\mathbf{m}'$ is

$$\mathbf{g}_{\mathbf{m}'} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}'} \mathbf{C}_{\mathbf{d}}^{-1} \Delta \mathbf{d} + \mathbf{C}_{\mathbf{m}'}^{-1} \Delta \mathbf{m}' = \frac{\partial \mathbf{d}}{\partial \mathbf{m}} \frac{\partial \mathbf{m}}{\partial \mathbf{m}'} \mathbf{C}_{\mathbf{d}}^{-1} \Delta \mathbf{d} + \mathbf{C}_{\mathbf{m}'}^{-1} \Delta \mathbf{m}' \quad . \quad (38)$$

Therefore, all that is needed to evaluate the gradient in terms of different model parameters is the Jacobian $\frac{\partial \mathbf{m}}{\partial \mathbf{m}'}$.

For example, the gradient for the P-wave velocity is

$$\delta \hat{\hat{\alpha}} = \frac{\partial \lambda}{\partial \alpha} \delta \lambda' + \frac{\partial \mu}{\partial \alpha} \delta \mu' + \frac{\partial \rho}{\partial \alpha} \delta \rho' + \mathbf{C}_{\alpha\alpha}^{-1} (\alpha - \alpha_0) + \cdots \quad , \quad (39)$$

where $\delta\lambda'$, $\delta\mu'$ and $\delta\rho'$ are defined by equations (36). The required elements of the Jacobian $\frac{\partial \lambda}{\partial \alpha}$ etc. can be obtained from the relationships for α and β

$$\alpha = \sqrt{\frac{\lambda + 2\mu}{\rho}} \quad , \quad \beta = \sqrt{\frac{\mu}{\rho}} \quad , \quad (40)$$

or solving for λ and μ,

$$\lambda = \rho\alpha^2 - 2\rho\beta^2 \quad , \quad \mu = \rho\beta^2 \quad . \quad (41)$$

Using these equations we obtain the gradient in terms of the P- and S-wave velocities and density as

$$\delta \hat{\hat{\alpha}} = 2\rho\alpha\delta\lambda' + \mathbf{C}_{\alpha\alpha}^{-1} (\alpha - \alpha_0) + \cdots \quad , \quad (42a)$$

$$\delta \hat{\hat{\beta}} = -4\rho\beta\delta\lambda' + 2\rho\beta\delta\mu' + \mathbf{C}_{\beta\beta}^{-1} (\beta - \beta_0) + \cdots \quad , \quad (42b)$$

$$\delta \hat{\hat{\rho}}_{\alpha} = (\alpha^2 - 2\beta^2)\delta\lambda' + \beta^2\delta\mu' + \delta\rho' + \mathbf{C}_{\rho\rho}^{-1} (\rho - \rho_0) + \cdots \quad . \quad (42c)$$

The equations relating the impedances Z_P and Z_S to the Lame parameters are

$$Z_P = \rho\alpha = \sqrt{\rho(\lambda + 2\mu)} \quad , \quad Z_S = \rho\beta = \sqrt{\rho\mu} \quad , \quad (43)$$

or solving for λ and μ

$$\lambda = \frac{1}{\rho} \left(Z_P^2 - 2 Z_S^2 \right) \quad , \quad \mu = \frac{1}{\rho} Z_S^2 \quad , \quad (44)$$

so the components of the gradient with respect to the impedances are

$$\delta \hat{\hat{Z}}_P = 2\alpha\delta\lambda' + \mathbf{C}_{Z_P Z_P}^{-1} (Z_P - Z_{P_0}) + \cdots \quad , \quad (45a)$$

$$\delta \hat{\hat{Z}}_S = -4\beta\delta\lambda' + 2\beta\delta\mu' + \mathbf{C}_{Z_S Z_S}^{-1} (Z_S - Z_{S_0}) + \cdots \quad , \quad (45b)$$

$$\delta \hat{\hat{\rho}}_Z = -(\alpha^2 - 2\beta^2)\delta\lambda' - \beta^2\delta\mu' + \delta\rho' + \mathbf{C}_{\rho\rho}^{-1} (ro - \rho_0) + \cdots \quad . \quad (45c)$$

Note that aside from a factor of ρ, the gradient in terms of the velocities (equations (42)) and the gradient in terms of the impedances (equations (45)) are identical except that the density gradient terms have some sign differences. The main effect of these sign differences is to alter the magnitude of the density gradient depending on which choice of model parameters is used.

Comparison of inversion to migration

The gradient to be used in the conjugate gradient inversion algorithm has the form

$$\delta\hat{\mathbf{m}} \quad = \quad \sum_S \int dt\, \mathbf{C_d^{-1}} \Omega u_j\, \Omega \psi_j \; + \; \mathbf{C_m^{-1}} \Delta \mathbf{m} \quad . \tag{46}$$

If Ω and $\mathbf{C_d^{-1}}$ were set to unity and $\mathbf{C_m^{-1}}$ was set to zero then the gradient of equation (46) does an elastic prestack shot-profile migration by correlation of the upgoing reflected waves ψ_j with the downgoing direct wave u_j. Where the correlation is great, there is a reflector or in other words, the reflected waves intersect the direct wave at the reflecting interface. This is also similar to Claerbout's U/D concept (Claerbout, 1976) that states that the reflectivity is the ratio of the amplitude of upgoing (reflected) waves U to the downgoing (direct) wave D. Note that sometimes the correlation of U and D are used as a stable approximation to the U/D method. If in addition, the two wavefields u_j and ψ_j represented pressure wavefields, then we would have a conventional acoustic prestack shot-profile migration. Hence, conventional migration is equivalent to the computation of the least squares gradient direction of inversion except that it does not include covariances or the operator Ω (see also Lailly, 1984).

It is clear that Ω is a very important operator. It unravels the wavefield information into the desired model parameters using the amplitude information (i.e. the overall amplitude of the reflected waves as well as the amplitude-incidence angle information). Another main difference between the inversion algorithm described here and migration is that it is iterative. The iterations tend to allow for the fact that the modeling function $\mathbf{d(m)}$ is nonlinear as well as iteratively applying the true inverse Hessian of equation (5) which would correct all the relative amplitudes of velocities and densities to their most likely values (i.e. the best fit values). Note that the iterations also help to regain some of the more poorly resolved low frequencies. Without iterations, the inversion algorithm, like migration, would have little hope of obtaining any of the lower frequencies except perhaps if a very clever preconditioning were constructed. The reason some of the low frequencies can be regained is that the inversion tries to match the entire shot profile which consists of approximately hyperbolic events that can only be matched if the low frequency velocity model is correct. The iterations are necessary because a gradient

algorithm is used and such algorithms obtain the best resolved components of the model (high frequencies) first and the poorest resolved components (low frequencies) last. It will be a challenge of future research to allow for the resolution of low and high frequencies simultaneously.

Speedup by data contraction

One reason that the inversion method is so costly is that it requires several elastic wave propagations per shot for each iteration and the number of shots is typically large for seismic surveys. However, the number of shots can in theory be reduced by about an order of magnitude thereby bringing the computer time required by this method to a more reasonable figure. This can be done by changing the concept of the experiment so that instead of doing say 100 shot simulations we do 10 super-shot simulations. In this case, a super shot profile would be obtained by firing 10 shots (every 10-th shot) simultaneously. Almost the same information is contained in the 10 super-shots but in a slightly scrambled form so the inversion should give almost the same results but with some additional noise. In practice, to avoid having to reshoot seismic lines as super shots, we could simply stack every 10-th shot profile. Some analysis is yet required to evaluate whether super shot inversions would be significantly corrupted by artifacts and noise.

Low frequencies

The method of inversion presented here is very powerful and can be used for almost any form of seismic data including reflection data (conventional surface seismics) or transmission data (well to well or VSP data). When used for transmission problems it can resolve the low frequency components of the model and does a kind of wave equation tomography (see Gauthier and Tarantola, 1985 for an acoustic transmission example). However, when used to invert reflection seismic data the low frequency model is not well resolved by the algorithm. Two possible solutions to this problem are

(i) Use of a clever preconditioning operator (i.e. approximate inverse Hessian) to enhance low frequency velocity perturbations. This could either be a boosting of the low frequency components which would be the small eigenvalues of the problem or an imposition of non-Gaussian statistics such as blockiness (see Harlan, 1984).

(ii) Derive a secondary inverse problem that attempts to resolve the low frequencies. In this case, one iteration of the "total" inversion algorithm would consist of one iteration for the low frequency part followed by one iteration for the high frequency part. This method is not unreasonable because the low and high frequencies are fairly well decoupled. That is, the high frequency model perturbations cause amplitudes but not shapes

of the hyperbolas to change while the low frequency model perturbations cause the shapes but not amplitudes of hyperbolas to change.

As presented in the paper, the low frequency a priori model $\mathbf{m}_0$ is assumed to be known from a prior "inversion" step (such as NMO) and thereafter, it is never changed. However, if we replace NMO with least squares, i.e. finding a low frequency model that produces hyperbolas that best match the data, and include this as an extra step at every iteration, then we have a "total" inversion algorithm. One way to do the extra step is by observing that the size of the high frequency model perturbations found in the current algorithm is maximized when the best low frequency model is used (because the kinematics is correct). This is analogous to doing NMO and stacking at different velocity functions. The best quality stack is achieved when the best low frequency velocity model is used. All that would be required for the extra inversion step is the operator $\mathbf{L}$ that relates the low frequency model perturbations $\Delta\mathbf{m}_{low}$ to the high frequency perturbations $\Delta\mathbf{m} = \Delta\mathbf{m}_{high}$, i.e.

$$\Delta\mathbf{m}_{high} \quad = \quad \mathbf{L}\Delta\mathbf{m}_{low} \tag{47}$$

Then, as for the high frequency derivation, this can be used to define the adjoint operation and hence the gradient direction for a least squares inversion. Note that we wish to maximize the perturbations $\Delta\mathbf{m}_{high}$ so we would minimize a functional of form

$$S(\mathbf{m}_{high},\mathbf{m}_{low}) \quad = \quad -\Delta\mathbf{m}_{high}^{*}\,\mathbf{C}_{\mathbf{m}_{high}}^{-1}\Delta\mathbf{m}_{high} + \Delta\mathbf{m}_{low}^{*}\,\mathbf{C}_{\mathbf{m}_{low}}^{-1}\Delta\mathbf{m}_{low} \tag{48}$$

in the extra inversion step (c.f. equation (1b)). This method seems preferable to use of (i) because it is uses the low frequency information in the seismic data (hyperbolic events) to restrict the low frequency model rather than only relying on some additional a priori statistics. Each iteration of the low frequency inversion step is really an elastic wave equation based NMO. Perhaps, the best method of obtaining the low frequencies would be a combination of (i) and (ii).

RESULTS

Introduction

The following are some examples to demonstrate the algorithm applied to synthetic reflection seismic data. The synthetic data was calculated with elastic finite differences using an algorithm similar to that of Kosloff et al., 1984 (see also Cerjan et al. 1985). The inversion was carried out with the conjugate gradient algorithm (equations (13) and (15)) and using the gradient parametrized in P- and S-wave velocities and density

denoted α, β and ρ. The gradient was calculated with equations (42a) through (42c) using definitions from equation (35b) and equations (36a) through (36c). To concentrate on the resolvability of α, β and ρ, the source was assumed to be known and so was *not* varied with equations (36d) and (36e) but was fixed at its true value throughout the inversions. The method of elastic finite differences was used to compute the forward modeled wavefield u_j and the back propagated residual wavefield ψ_j. Equation (37) was used to define the data covariance matrix (this is equivalent to saying that a time varying gain of t^p was applied) where the value of p was .5 or .75 depending on the example. The data variance was small so the inversion was not heavily damped and hence the solution was not required to stay close to the a priori model $\mathbf{m}_0$ (i.e. little heed was paid in the c.g. algorithm (equation (13)) to $\Delta\mathbf{m}_n^* \mathbf{C}_\mathbf{m}^{-1} \Delta\mathbf{m}_n$, $\mathbf{C}_\mathbf{m}^{-1}\Delta\mathbf{m}$ or $\mathbf{c}_n^* \mathbf{C}_\mathbf{m}^{-1} \mathbf{c}_n$). A preconditioning of $\mathbf{C}_\mathbf{m}$ was applied in algorithm (13) where constant diagonal model covariance matrices were used, $\mathbf{C}_{\alpha\alpha} = \alpha_{ave}^2 \mathbf{I}$, $\mathbf{C}_{\beta\beta} = \beta_{ave}^2 \mathbf{I}$ and $\mathbf{C}_{\rho\rho} = \rho_{ave}^2 \mathbf{I}$ and cross-covariances between different model parameters were assumed to be zero. A band limited source wavelet was used (either 2-nd or 4-th derivative of a Gaussian curve depending on the example).

Spatial resolution test, 4-diffractor model

In order to test the spatial resolution of the algorithm, an inversion was carried out on data generated from model consisting of four shallow diffractors (Figures 1a, 1c, and 1e). The leftmost diffractor is perturbation in P-velocity only, the second from the left is a perturbation in S-velocity only, the third is in density and the last is a perturbation in all three. A vertical force and a 4-th derivative Gaussian far field wavelet with fundamental frequency of about 25 hertz was used to generate the data, a split spread two-component shot profile shown in Figure 2. An inversion was performed using the correct constant background model as the initial guess. Figures 1b, 1d and 1f show the result of the inversion for P-velocity, S-velocity and density after 5 iterations plotted at the same scale as the model. Clearly, the diffractors have been very well located. Furthermore, the inversion result for each of the 3 parameters have two main anomalies of the correct sign at the true locations of the diffractors. The main limitation is poor resolution between P-wave velocity and density. This agrees with the study of model parameter choice (Tarantola et al. 1985) which indicated that probably only 2 parameters can be resolved, say P- and S-wave velocity or P- and S-impedance. However, considering that only one shot was used in the inversion it has done a remarkable job of locating the diffractors and giving an indication of the relative magnitudes of the three physical

properties. With more shots, it is expected that both the spatial resolution and the resolution between the different properties would slightly improve. The residual (difference between the synthetic data of Figure 2 and the synthetic data computed with the elastic properties obtained after 5 iterations) is shown in Figure 3 at the same scale as the data. After only 5 iterations most of the energy in the synthetic data has been accounted for by the inversion algorithm. Note that the absolute magnitudes of the inversion result shown in Figure 1 are not large enough due to problems associated with band limitations which leads to a null space. That is, for band limited data, more than one physical model have almost identical seismic responses.

Resolution versus depth and the effect of iterations

A similar example to the spatial resolution test was performed but using three sets of diffractor perturbations at three different depths. The point of this example is to study whether deeper diffractors are more poorly resolved and whether resolution between the different parameters decreases significantly at greater depths where there are smaller incidence angles. A different source wavelet was used, a second derivative Gaussian with a fundamental of about 20 hertz. The synthetic data (with the direct waves removed) is shown in Figure 4, the residual after 5 iterations is in Figure 5. The true model and inversion result after 1 and 5 iterations is shown in Figure 6 at a relative scale (the previous example has already illustrated that the absolute magnitudes of the inversion result cannot be well resolved and therefore a relative scale is used to display all the following results). From the results of Figure 6 it is clear that spatial resolution is only slightly worst at greater depths while resolution between the different parameters is about the same at all depths. This is because the incidence angle for the deepest diffractors is still quite large (around 45 degrees). The spatial resolution is worse than in the first example probably because of the different source wavelet which contains more low frequency content. The main result of iterations is to decrease the size of artifacts relative to the desired point perturbations. Actually, it is the size of the point perturbations that increases while the artifacts generated from the first iteration remain about constant. Another main effect of the iterations is to correct the relative sizes of the S-wave velocity solution. This illustrates one advantage of iterative elastic inversion over elastic migration which would give a result similar to the first iteration of the inversion.

Inversion in the presence of Gaussian noise

The data of the previous example was contaminated with independent Gaussian noise (Figure 7). An inversion was carried out of this noisy data. The residual after 5 iterations is shown in Figure 8 and the inversion result in Figure 9. There is some noise in the solution the result is very similar to the noise free solution except in areas around the edge of the model which are more poorly resolved and hence more susceptible to noise contamination.

Single component data inversion

The examples so far have been inversions of two component data. To illustrate the importance of two component data which contains significant shear wave energy in the form of P-S and S-S diffractions, an inversion was carried out on the synthetic vertical component data of Figure 2a only. The residual after 5 iterations is shown in Figure 10 and the true model and the inversion results after 1 and 5 iterations are plotted in Figure 11. Compared with the inversion results of the two-component data shown in Fig. 6, the resolution between P- and S-wave velocity is poorer especially in the early iterations. However, after 5 iterations there is still fair resolution between these parameters though it is certainly much worse than when two component data was used. Probably, in a more realistic example with very slow near surface velocities and hence even less P-S and S-S energy than in this example, the resolution between P- and S-wave velocity would be even worse.

Both vertical and horizontal shots and two component receivers

Is it worth doing shear wave surveys when there is plenty of shear wave energy contained in normal surveys using large offsets and two component geophones? With conventional methods of processing, expensive shear surveys are the only way to get good shear velocity images because the processing assumes only one wave type (i.e. conventional shear processing mainly uses the old P-wave processing techniques). However, the elastic inversion algorithm uses the elastic wave equation and can correctly handle the presence of both P-waves and S-waves to produce two separate images. In this case is the resolution significantly improved by including shear sources (i.e. horizontal forces) as well as vertical sources?

Figure 12 shows data generated from a horizontal source which was used simultaneously with the vertical source data of Figure 4 to do an inversion. The residual after 5 iterations is shown in Figure 13 and the inversion result is given in Figure 14. Clearly, the result after 5 iterations is almost identical to the result when only vertical shots were

used. The only significant difference is that when the horizontal source was included, the shear velocities had the correct relative magnitudes from the start (i.e. iteration 1). This is because there is now significant S-S reflections even at zero incidence angle (in the horizontal source shot profile) and this helps resolve the S-velocity. Previously, the S-wave velocity was better resolved where there were more P-S and S-S conversions, namely in the shallow part of the model where the incidence angle was higher. Therefore, it seems that by using the iterative inversion approach, we can obtain good S-wave velocity images even when there is no horizontal (shear) source data provided there is two component data recorded to fairly long offsets. Perhaps this means that by basing algorithms on the elastic wave equation we can do away with expensive horizontal source surveys. Of course, the inversion approach is much more expensive than conventional processing and in any case the final test will be on field data. One further point is that even with two component shots and receivers the density is still not perfectly resolved from the from the P- and S-wave velocities. Compared to the vertical source result, P-wave velocity and density are better resolved from one another but S-wave velocity and density are more poorly resolved from one another. Therefore, it appears that only two parameters (say P- and S-wave velocities) can be resolved from seismic data even under almost ideal circumstances. Furthermore, P-wave and S-wave velocities are better resolved from one another whereas there is always fairly poor resolution between density and at least one of the other two parameters. These results indicate that acoustic inverters who do inversions for P-wave velocity and density are wasting their time when a better resolved parameter (S-wave velocity) exists.

A plane layered example

The following example demonstrates that the algorithm can resolve some of the low frequency model perturbations but that this takes many iterations. Therefore, the algorithm is very inefficient as a method of obtaining the low frequencies and hence some improvements are required (see the section titled "Low frequencies" for some possible solutions to this problem). The slightly noise contaminated data shown in Figure 15 was generated from modeling over some plane layers. The model used to generate this data is shown in Figure 17 in the solid line. A linear fit of this model was used as the starting guess in the inversion. There is still significant energy (especially due to S-waves) in the residuals after 15 iterations (Figure 16) illustrating slow convergence. The reason for the slow convergence is that the low frequency model perturbations are fairly poorly resolved and so take many iterations of a gradient algorithm to appear in the solution. The inversion result at various iterations is shown in Figure 17 again illustrating the gradual

creeping of low frequencies into the solution as iterations proceed. Clearly, the iterations help and the inversion result is much better than the migration result (which would essentially be the same as the first iteration). However, the low frequencies only very gradually build up so the algorithm requires either a better low frequency starting model or some improvements which help resolve the low frequencies. One technique to enhance the low frequencies is by finding the main layer boundaries from the solution and using these to obtain a blocky solution (i.e. a "blockiness" preconditioning).

Complicated synthetic data inversion, horst model

Data was generated for the very complicated geologic model (Figure 18) consisting of a horst structure and reef complex amidst a multitude of larger layers each with random fine layer structure within them. P-velocity varied by about 7.5% and S-velocity by about 15%. This range of velocity variation was deliberately chosen to be fairly small so the gradient inversion algorithm would converge rapidly in this test case and to avoid problems of local minima in the least squares objective function. A vertical force was applied and a fourth derivative Gaussian wavelet was used to generate the synthetic shot gather shown in Figure 19. Note that high dip events on the vertical component shot gather are S-S and P-S reflections and not finite difference boundary artifacts (the boundaries are almost perfectly absorbing). An initial model consisting of a linear fit to the true model was used and an inversion was performed. The residual after 3 iterations shown in Figure 20 still has significant energy indicating the algorithm has not yet converged to the best fit solution. The low frequencies could not be resolved in just 3 iterations (results from the layered example indicate that further iterations help regain at least some of the low frequencies) so only the well resolved high frequency component of the inversion result is shown (Figure 21). To aid the comparison of these results, the true model was filtered appropriately and zeroed in the regions near the boundaries where few reflections could occur (Figure 22). Many features can be seen on the inversion result which coincide with the main boundaries of the true model. Another obvious feature is the lack of energy in the S-velocity result of Figure 21 directly below the shot. This is because the S-velocity can only be reconstructed where significant S-S, P-S or S-P reflections occur (i.e. not beneath the shot since it is a vertical force). Considering the complexity of the data and model, the inversion had done a good job of imaging (inverting for the higher frequency model variations). If many shots were used then the inversion should be able to reconstruct complete P-velocity, S-velocity and density images instead of a partial image in the region directly below the shot location where most energy was reflected for this shot.

DISCUSSION

A method of elastic inversion was derived using the most obvious choice of model parameters given the elastic wave equation, that is the Lamé parameters. Equations were also given in terms of two other choices of model parameters, velocities and impedances. The choice of model parameters should be made, where possible, in order to attempt to make the model parameters as independent as possible and hence better resolved (see Tarantola et al., 1985, who study resolution of different physical parameter choices or Harlan, 1984, who carries out inversion for the depth derivative of impedance in 1D inversion thus avoiding problems of knowing a spatially variable mean). Also, the data was parametrized by the displacement amplitude leading to the Born linearization (i.e. the Frechet derivatives were derived from a Born equation). This has limitations and perhaps better linearizations exist. For example, in transmission problems, the Rytoff linearization which derives Frechet derivatives from phase rather than amplitude data is better (i.e. has a larger linear region). Also, perhaps the model parameter space can be more heavily restricted in early iterations in order to minimize those components of the model which are clearly nonlinear. For instance, in the Born approximation the kinematics of the wavefield must be approximately correct or the linearization is bad. Therefore, if all frequency components of the model were included in an inversion which started relatively far from the solution, false scatterers would be introduced by the inversion in its early iterations thereby making convergence difficult if not impossible (probably the false scatterers would aggravate the local minima problem). However, if the inversion proceeded by first inverting only for the lower spatial frequency components of the model and progressively increasing the frequency range upward, then no such problem would occur.

Because of the cost of the inversion algorithm, the big area for research and improvements relates to the study of the preconditioning operator. Preconditioning should do many things including source wavelet deconvolution and restriction of the estimated model to lie in the part of the null space which intersects with the true model statistics (which are usually non- Gaussian). It should greatly speed convergence if it is well chosen. Note that preconditioning is often considered to be an approximation of the inverse Hessian operator but it can actually do much more. This is because the inverse Hessian is a least squares concept while the preconditioning operator can be chosen with no regard for least squares but paying more heed to other factors such as the desired model statistics.

The algorithm outlined is for 2D inversion. It would be impractical for 3D at the present time since no one knows how to solve rapidly the 3D forward problem for

relatively arbitrary distributions of physical properties. However, in principle it could be extended to 2.5 D relatively easily by making some modifications in the computations to allow for out of plane wave divergence.

CONCLUSIONS

Two-dimensional elastic inversion of multi-offset seismic data has been described and with some speedups it should be feasible though costly on todays super computers. Synthetic data tests indicate that the elastic inversion works on realistic sized problems with excellent spatial resolution of the high frequency components of the elastic parameters. In the case of reflection seismic data where there is no direct arrival, the low frequency components are at best partially resolved though they are better resolved than in a migration. The degree of resolution of the low frequencies depends mainly on the number of iterations, the degree of noise contamination and the statistics of the earth model. The resolution between P-velocity and S-velocity is quite good while the resolution between P-velocity and density is poor. The method has done at least as much as an elastic multi-offset migration would be expected to do (no prestack elastic migration algorithm currently exists so it is difficult to make any comparisons other than through analysis of the equations). It provides understanding for a framework of elastic processing and inversion that is badly needed with the increase in multi-component recordings and wide-angle experiments. It is very powerful because it can in principle be used for almost any situation including inversions of reflected wavefields (e.g. shot profile inversion as in the examples) and inversions of transmitted wavefields such as in well-to-well tomography. Limitations are that it is a very computer intensive process (being an order of magnitude more costly than a prestack migration) and that like migration, it requires fairly accurate knowledge of the very low frequency velocity model (so that the kinematics of events are approximately correct). Future work is required to test the algorithm more thoroughly under different circumstances, study the effect of the choice of model parameters on resolution, and to invent a better preconditioning operator to speed convergence and regain the poorly resolved low frequencies and/or an additional low frequency (model) inversion step.

ACKNOWLEDGMENTS

Thanks to Albert Tarantola for many stimulating discussions and making possible an 8 month stay at the Institut de Physique du Globe where much of this work was carried out. Particular thanks to Alexandre Nercessian for priceless computer aid. Thanks to Odile Gauthier and Antonio Pica who are working on the acoustic equivalent of the algorithm presented and helped me to understand many aspects. Finally, thanks to my fellow students at Stanford, particularly, John Toldi, Dan Rothman, Kamal Al-Yahya and Jos Van Trier, for provocative and interesting discussion.

REFERENCES

Aki K. and Richards P., 1980, Quantitative Seismology: Freeman Press.

Cerjan C., Kosloff D., Kosloff R., and Reshef M., 1985, Short note: A nonreflecting boundary condition for discrete acoustic and elastic wave equations: Geophysics, **50**, 705-708.

Claerbout, J.F., 1976, Fundamentals of geophysical data processing: McGraw-Hill.

Clayton, R.W. and Stolt, R.H., 1981, A Born-WKBJ inversion method for acoustic reflection data: Geophysics, **46**, 1559-1567.

Cohen, J.K. and Bleistein, N., 1979, Velocity inversion procedures of acoustic waves: Geophysics, **44**, 1077-1087.

Fletcher R., and Reeves, C.M., 1964, Function minimization by conjugate gradients: The computer journal, **7**, 149-154.

Gauthier O., and Tarantola A., 1985, Nonlinear inversion of two-dimensional seismic wavefields: Preliminary results: Geophysics, submitted for publication.

Harlan B., 1984, Robust inversion of VSP's: Tech. Rep., Stanford Exploration Project report 41 , Stanford university.

Kosloff D., Reshef M., and Loewenthal D., 1984, Elastic wave calculations by the Fourier method: Bulletin of the Seismological Society of America, **74**, 875-891.

Lailly P., 1984, Migration methods: partial but efficient solutions to the seismic inverse problem, in: Inverse problems of acoustic and elastic waves, edited by: F. Santosa, Y.H. Pao, W. Symes, and Ch. Holland, SIAM, Philadelphia.

Luenberger D., 1973, Introduction to linear and nonlinear programming: Adison- Wesley.

Mora P., 1986, Elastic inversion of multi-offset seismic amplitude data for P-wave velocity, S-wave velocity and density, in: SEG special volume on shear wave exploration, edited by: Steve Danbom and Norman Domenico.

Polak, E. et Ribiére, G., 1969, Notes sur la convergence de méthodes de directions conjugées: Revue Fr. Inf. Rech. Oper, **16-R1**, 35-43.

Powell, M.J.D., 1981, Approximation theory and methods: Cambridge university press, Cambridge.

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

Stolt, R.H., and Weglein, A.B., 1985, Migration and inversion of seismic data: Geophysics, **50**, 2458-2472.

Tarantola A., and Valette, B., 1982, Inverse problems = quest for information: Journal of geophysics, **50**, 159-170.

Tarantola A., 1984, The seismic reflection inverse problem, in: Inverse problems of acoustic and elastic waves, edited by: F. Santosa, Y.H. Pao, W. Symes, and Ch. Holland, SIAM, Philadelphia.

Tarantola A., Fosse I., Mendes M., and Nercessian A., 1985, The choice of model parameters for linearized elastic inversion of seismic reflection data: Tech. Rep., Equipe de Tomographie du Geophysique report 2, Institut de Paris du Globe.

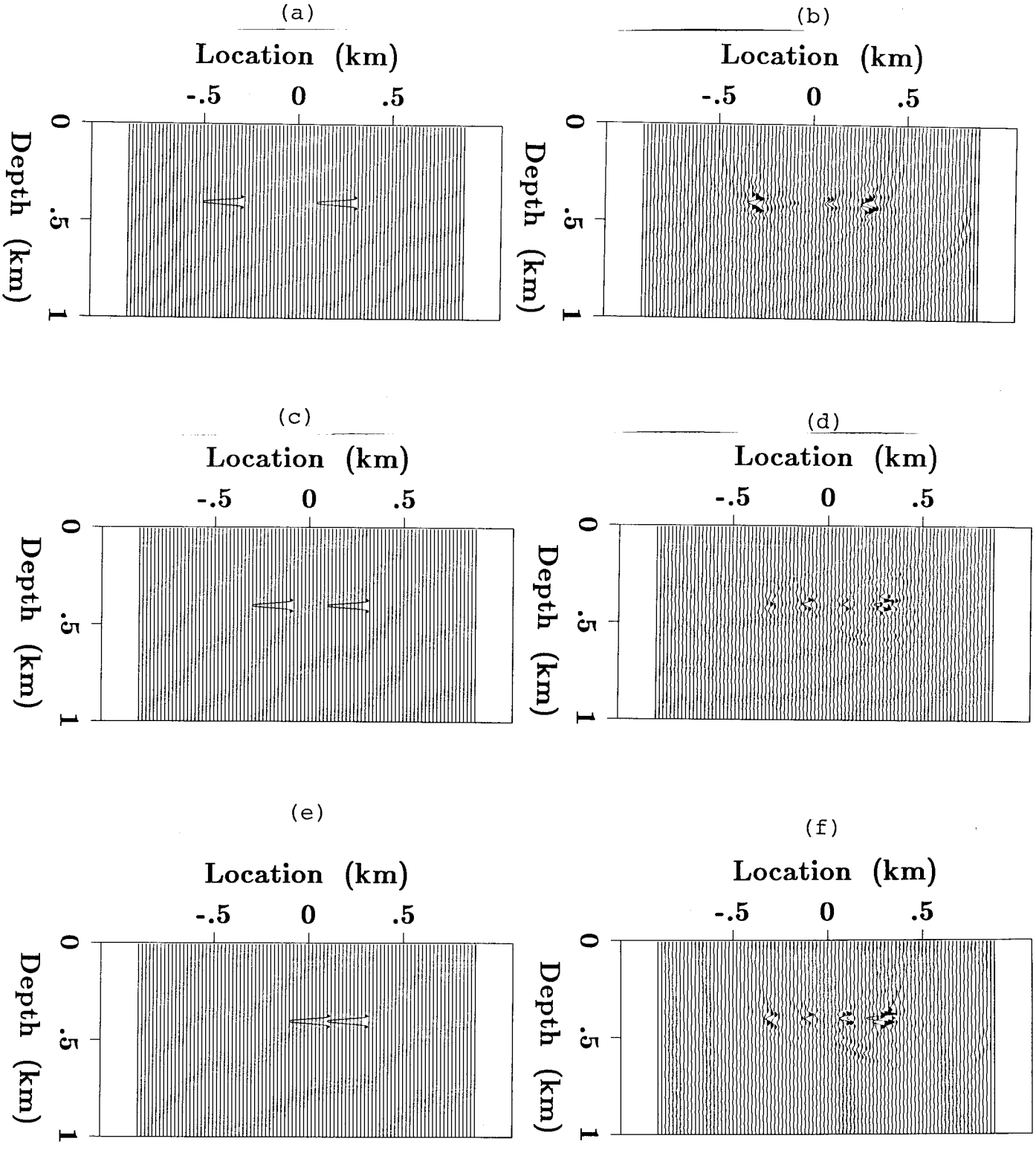


Figure 1: True 4 diffractor model and inversion result after 5 iterations plotted at the same scale. (a) True P-wave velocity, (b) P-wave velocity result, (c) true S-wave velocity, (d) S-wave velocity result, (e) true density, (f) density result.

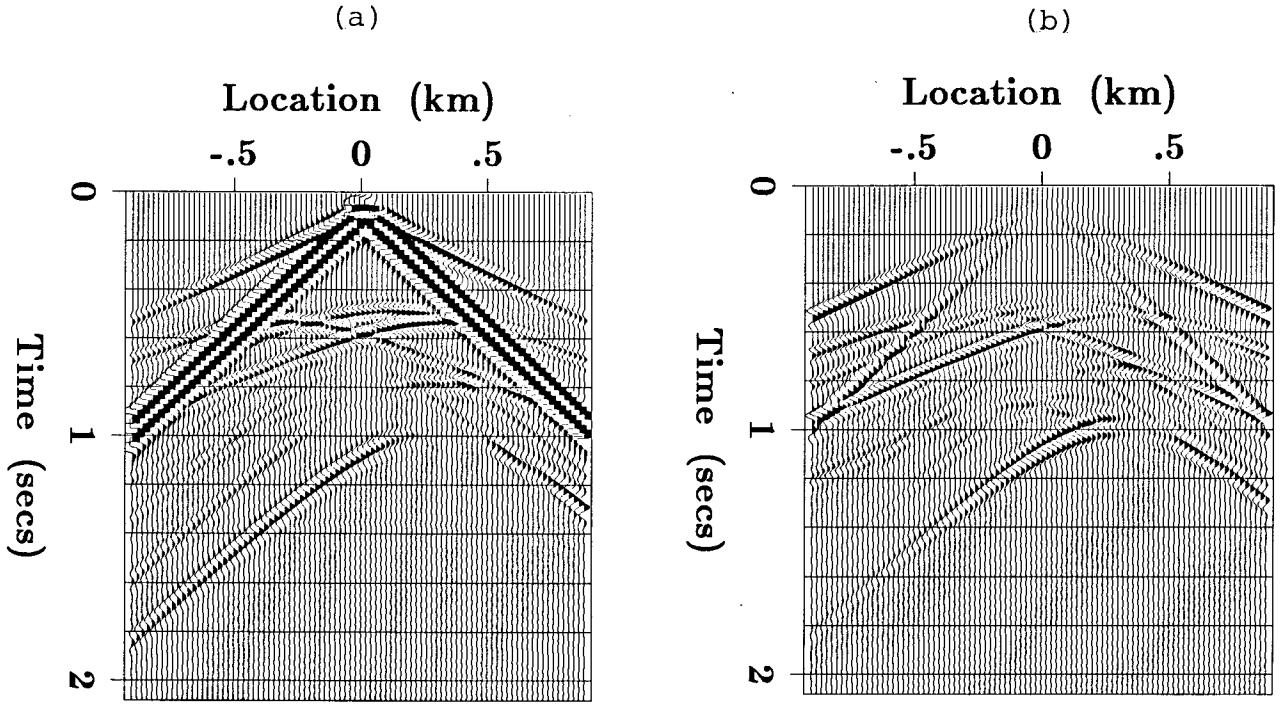


Figure 2: Vertical source synthetic shot gather generated from the 4 diffractor model shown in Fig. 1. (a) Vertical component, and (b) horizontal component.

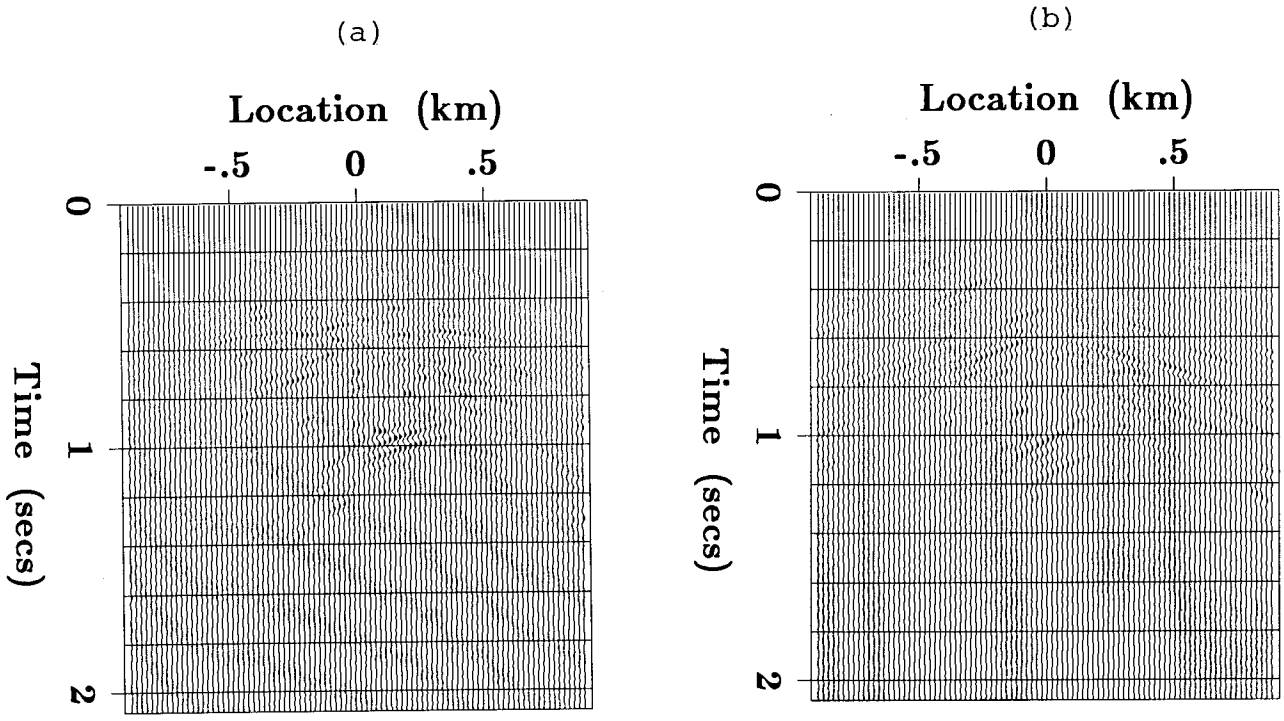


Figure 3: Residual after 5 iterations for the 4 diffractor example. (a) Vertical component, and (b) horizontal component.

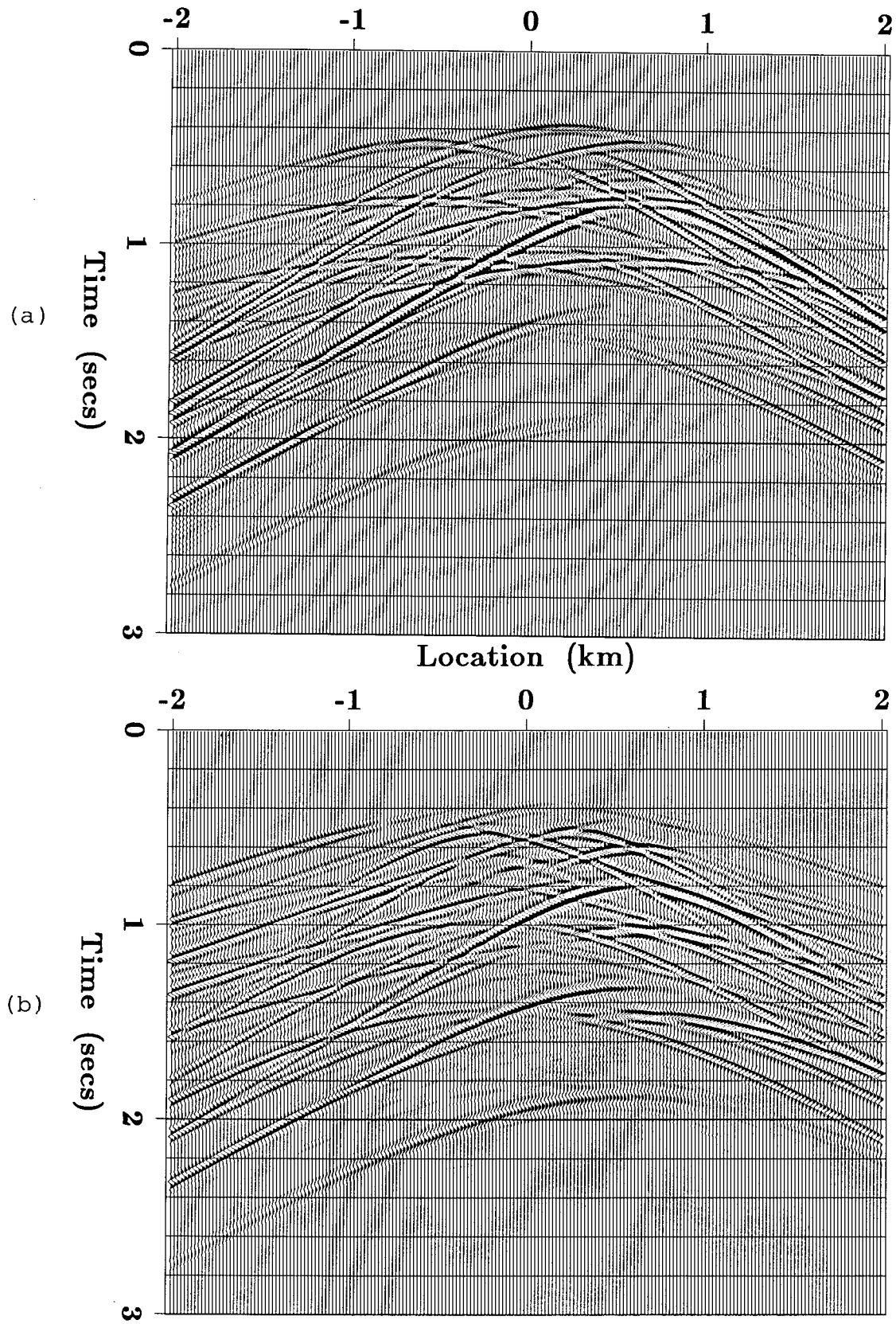


Figure 4: Vertical source synthetic shot gather for the 12 diffractor model (the direct wave has been removed). (a) Vertical component, (b) horizontal component.

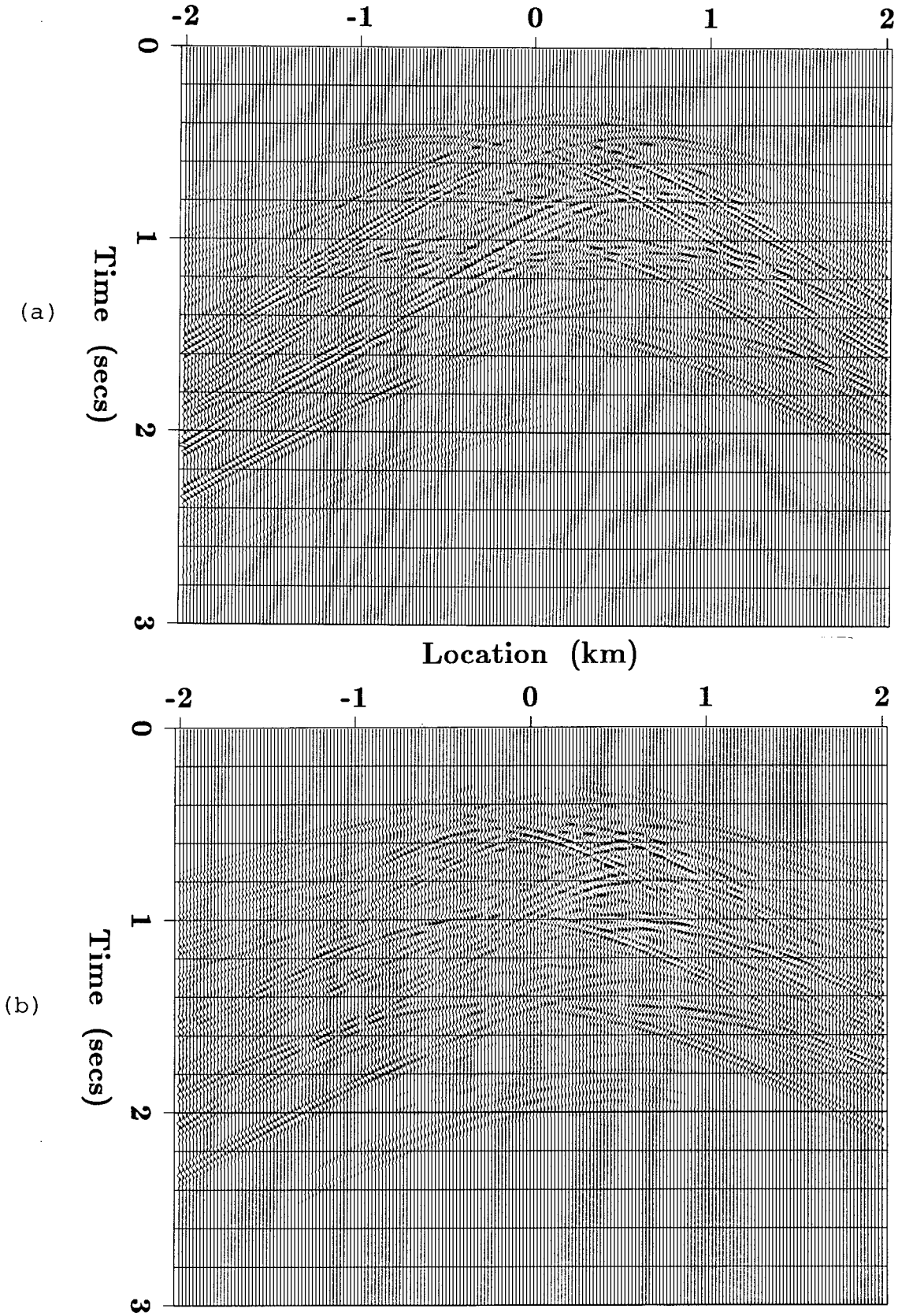
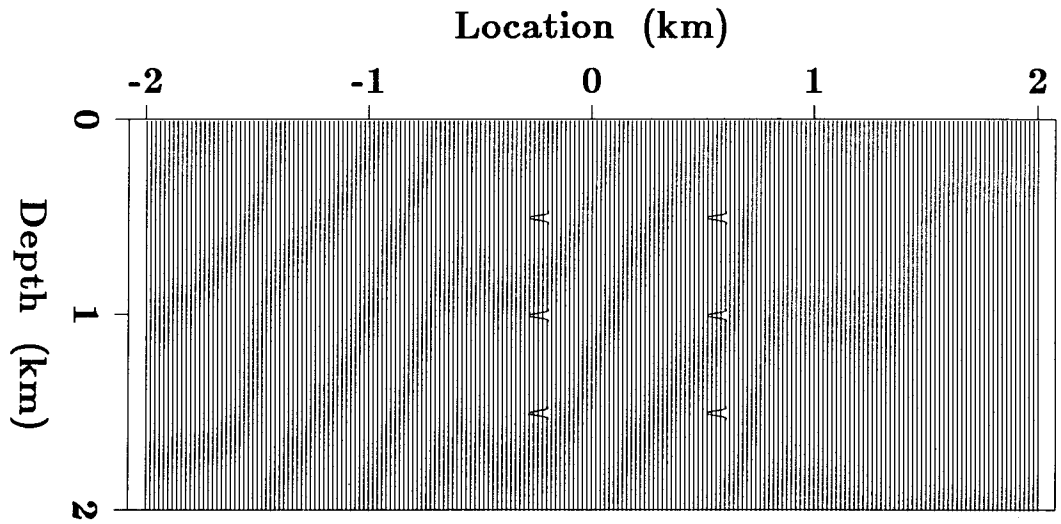
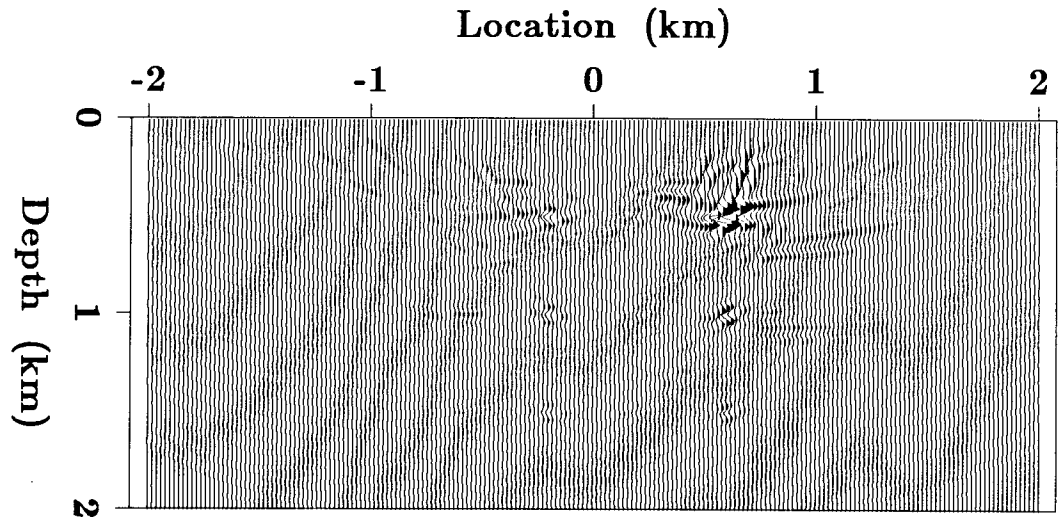


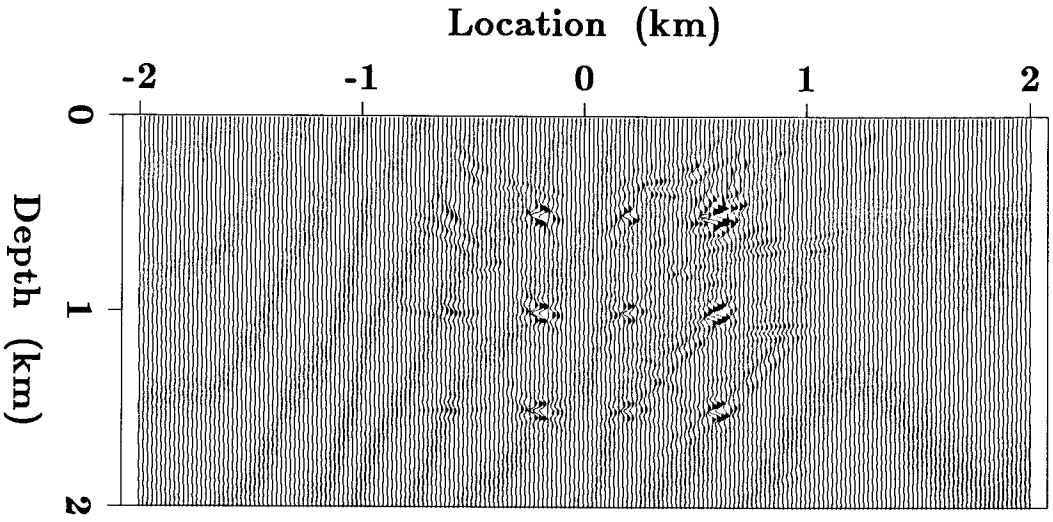
Figure 5: Residual after 5 iterations for the vertical source 12 diffractor example. (a) Vertical component, (b) horizontal component.


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

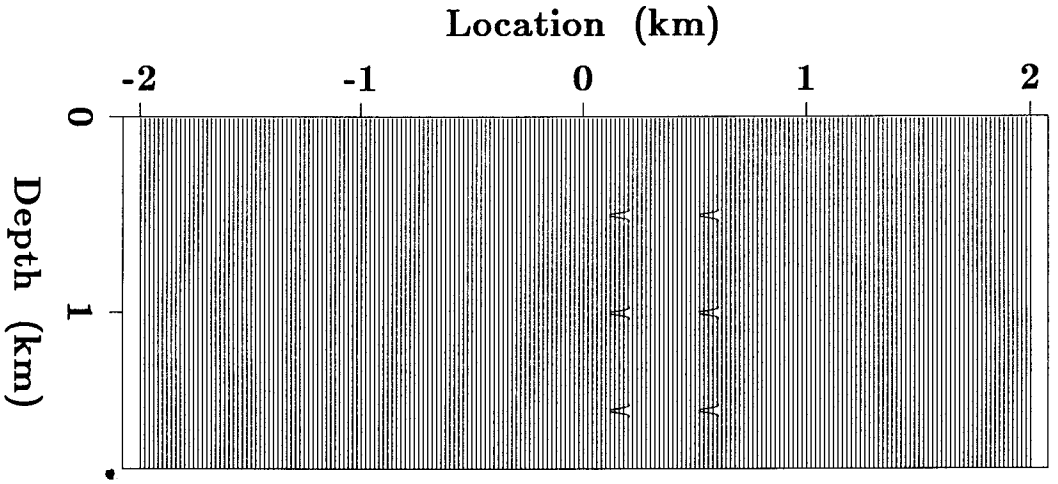
6 (d)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

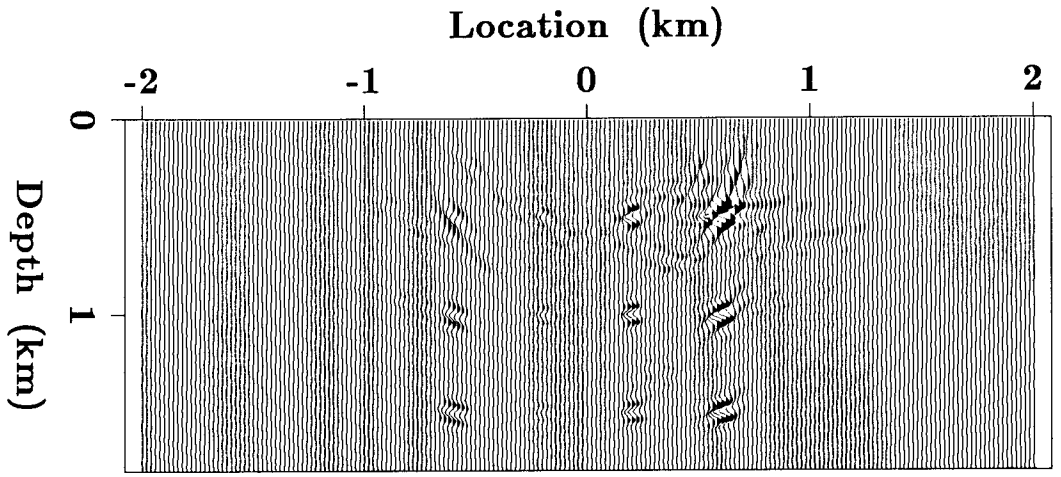
6 (e)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

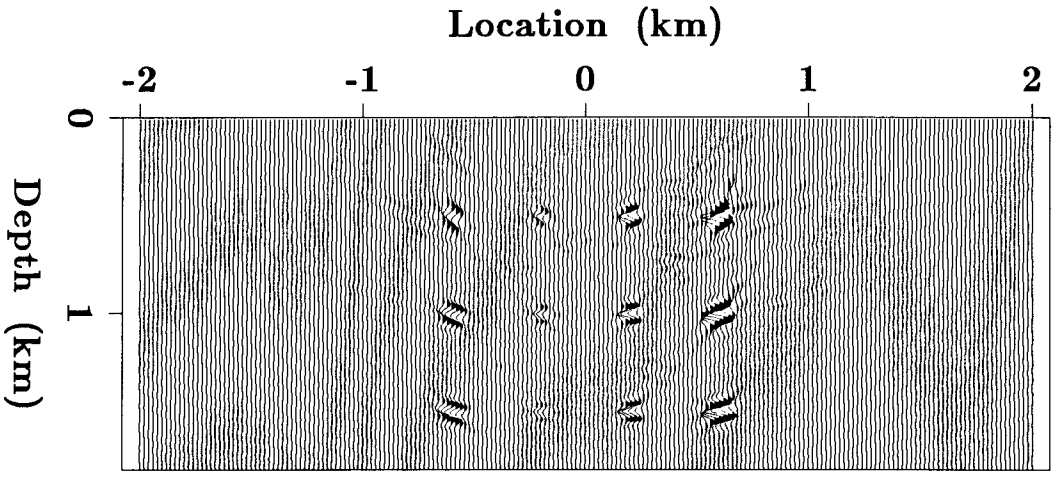
6 (f)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1

6(g)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1

6(h)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1

6(i)

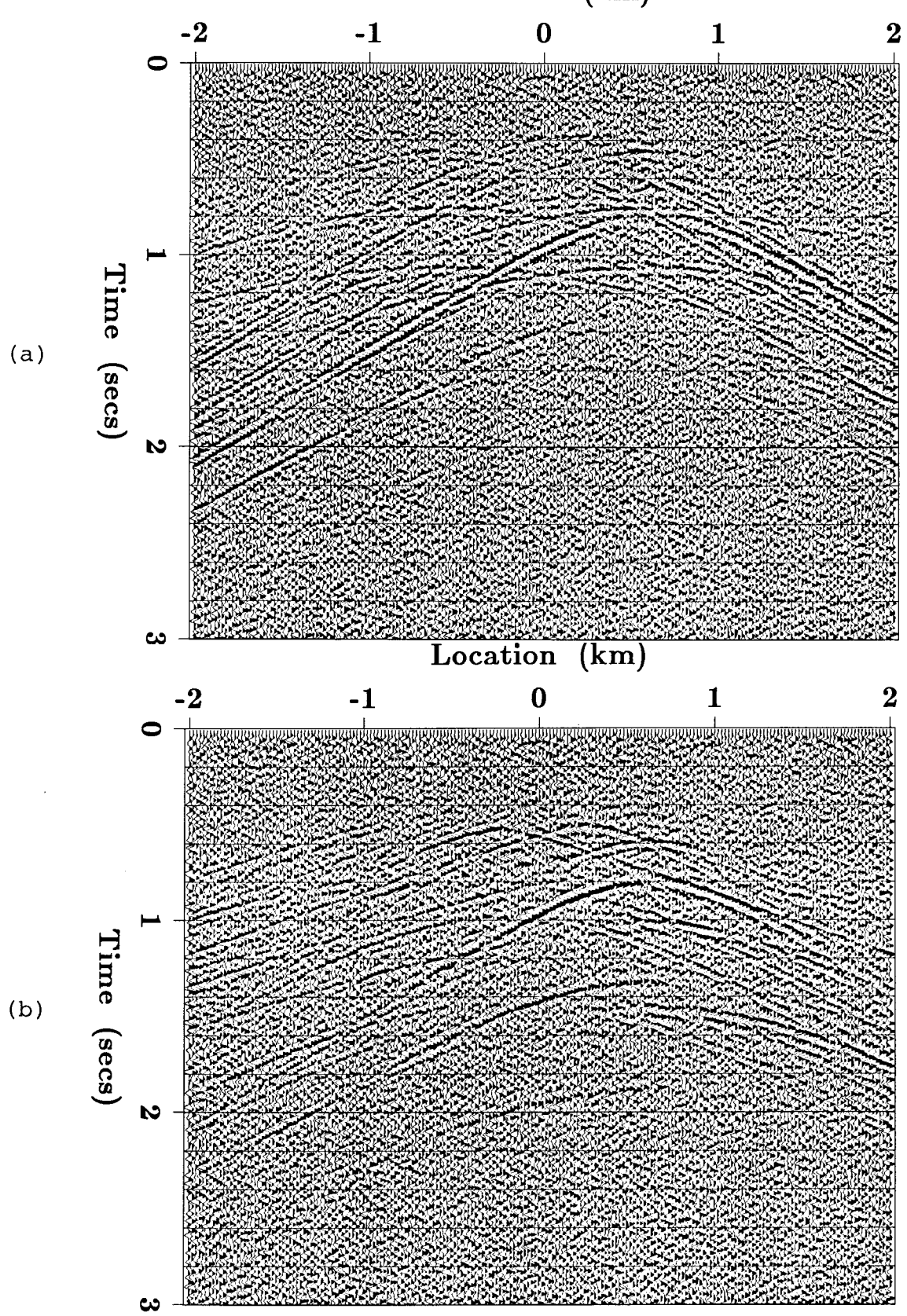


Figure 7: Noisy vertical source synthetic shot gather for the 12 diffractor model (the direct wave has been removed). (a) Vertical component, (b) horizontal component.

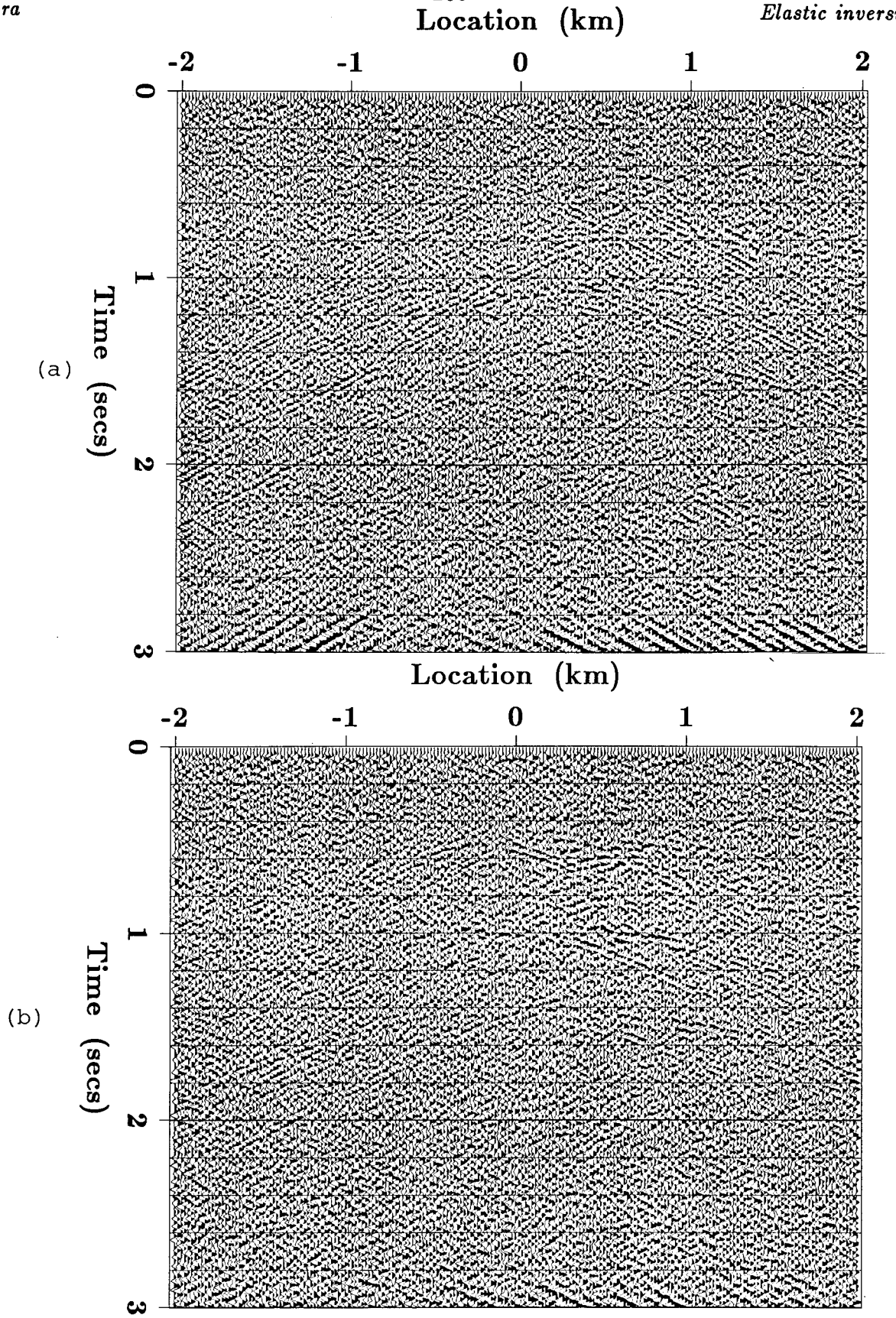


Figure 8: Residual after 5 iterations for the noisy vertical source 12 diffractor example. (a) Vertical component, (b) horizontal component.

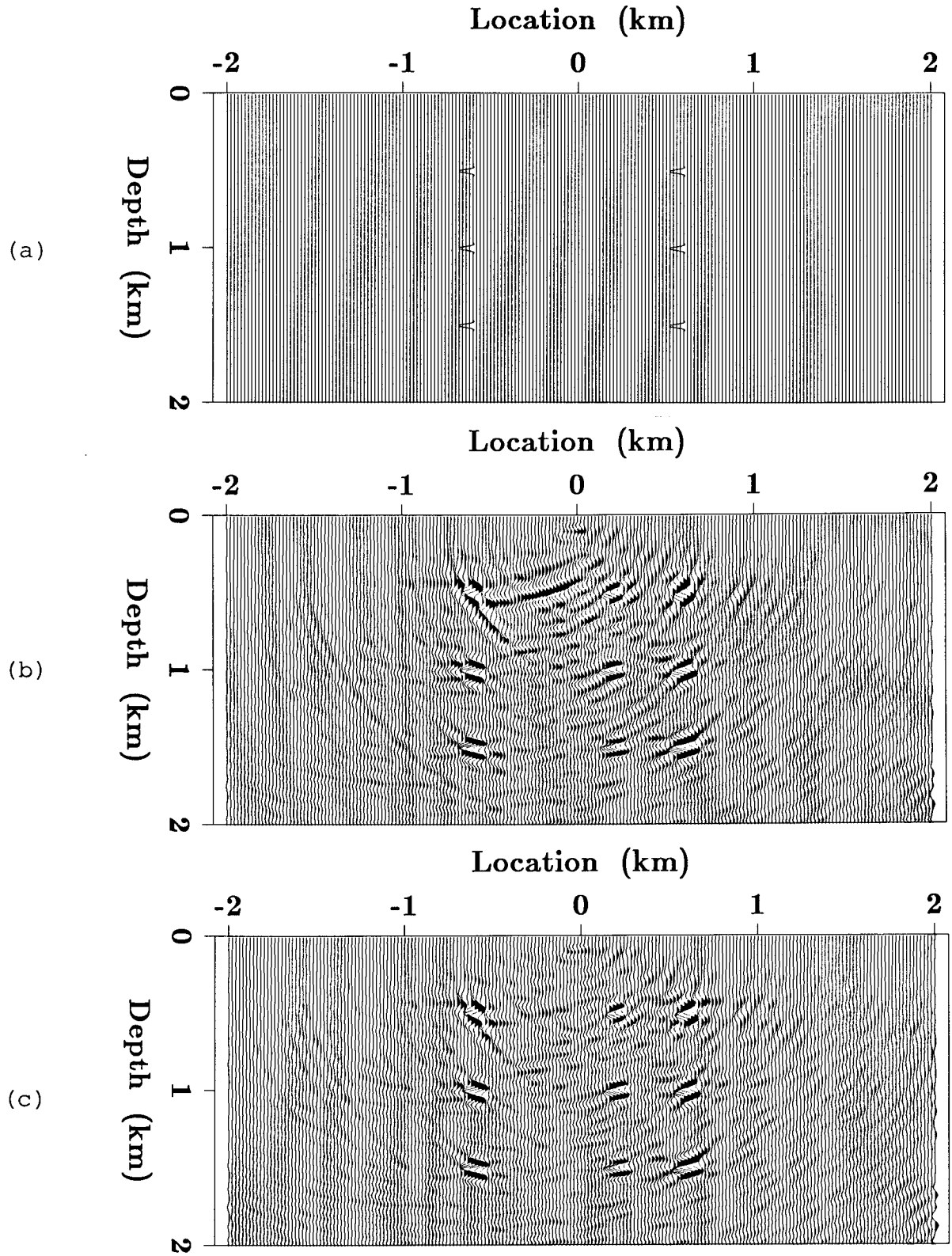
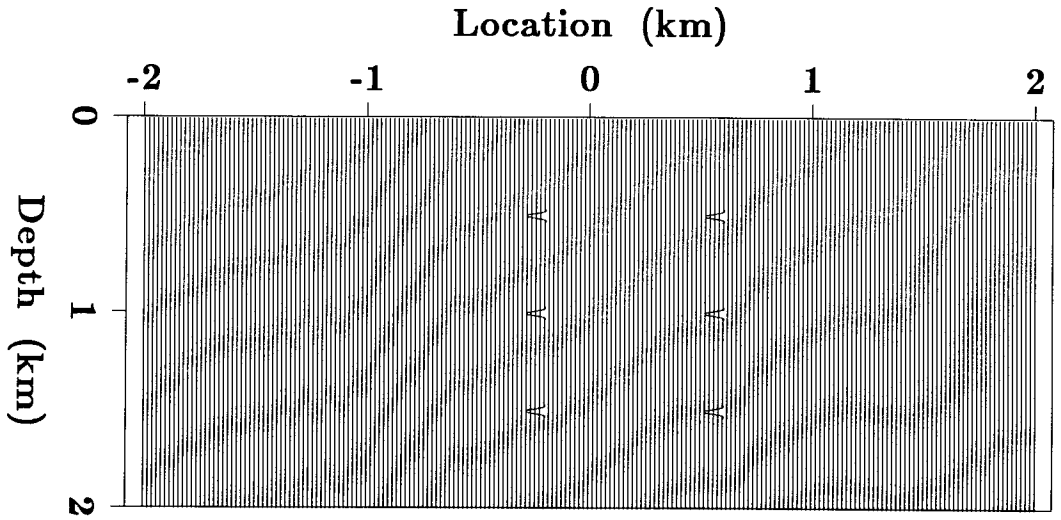
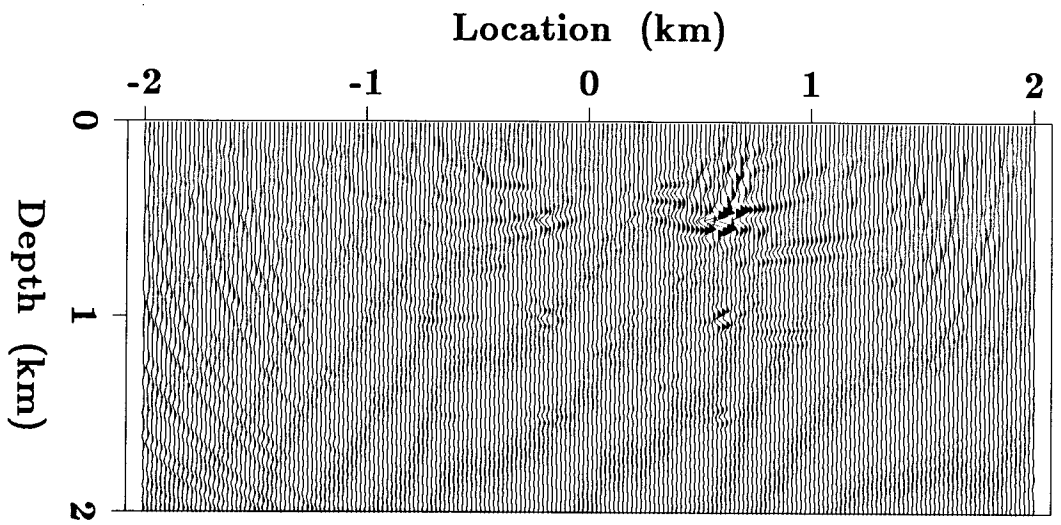


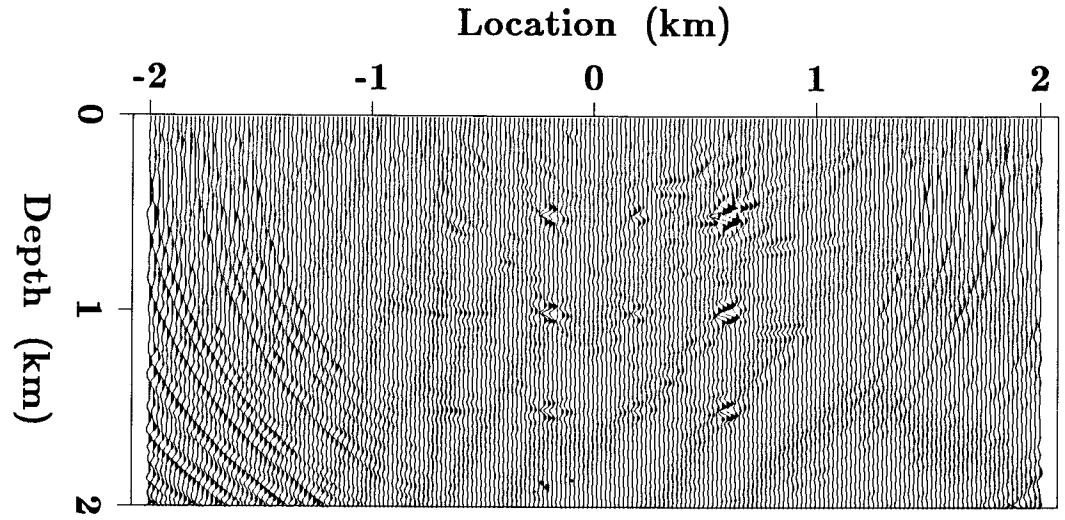
Figure 9: True model and inversion results for the noisy vertical source 12 diffractor example. (a) True P-wave velocity model, (b) P-wave velocity result after one iteration, (c) P-wave velocity result after five iterations, (d) true S-wave velocity model, (e) S-wave velocity result after one iteration, (f) S-wave velocity result after five iterations, (g) true density model, (h) density result after one iteration, and (i) density result after five iterations.


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

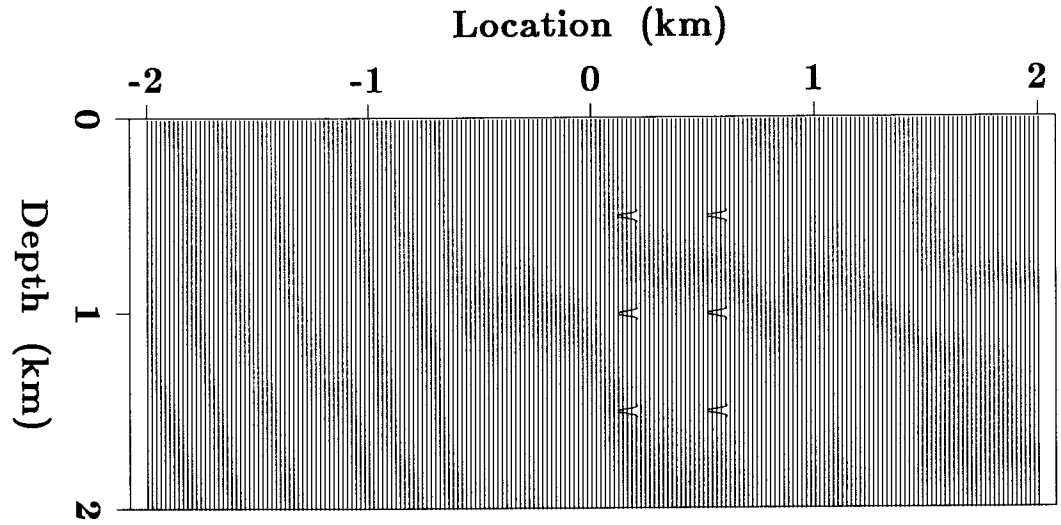
9(d)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

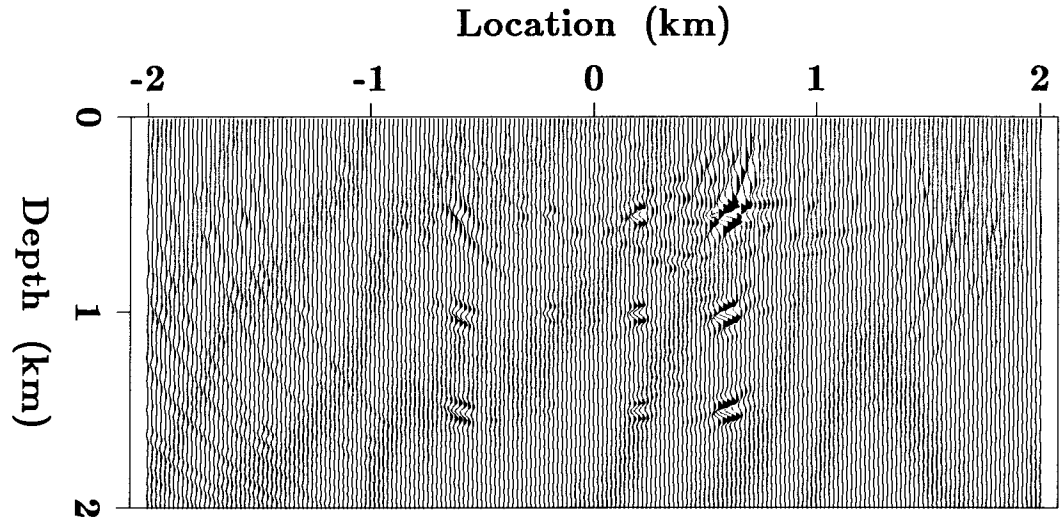
9(e)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

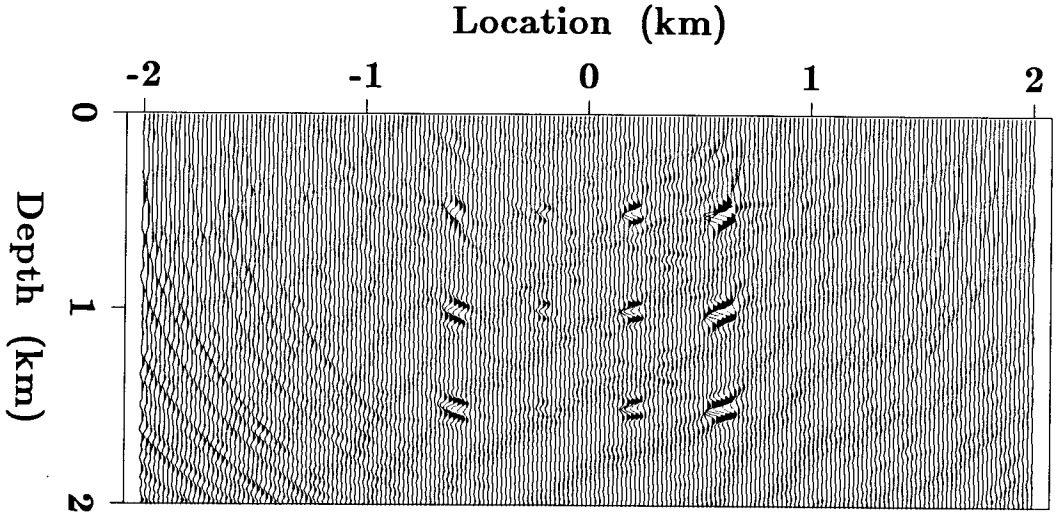
9(f)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

9(g)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

9(h)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

9(i)

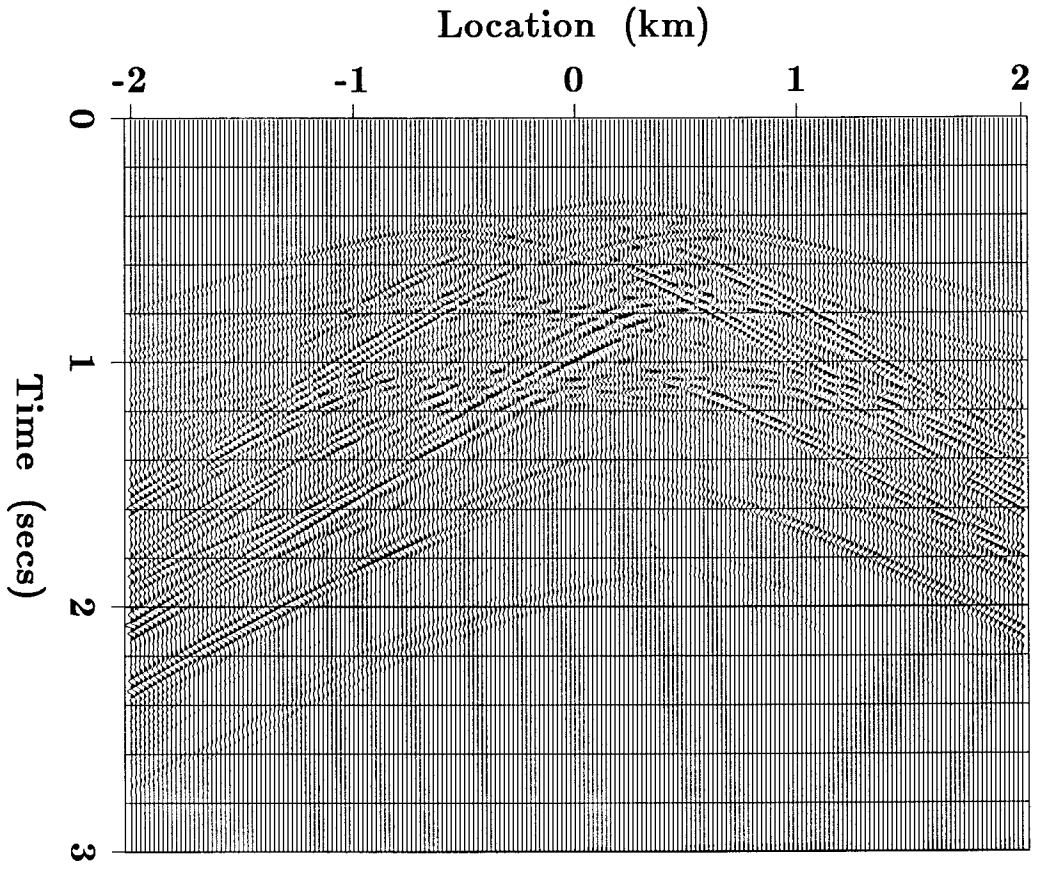


Figure 10: Vertical component residual after 5 iterations for the vertical source 12 diffractor example. Only the vertical component data of Fig. 4a was used in the inversion.

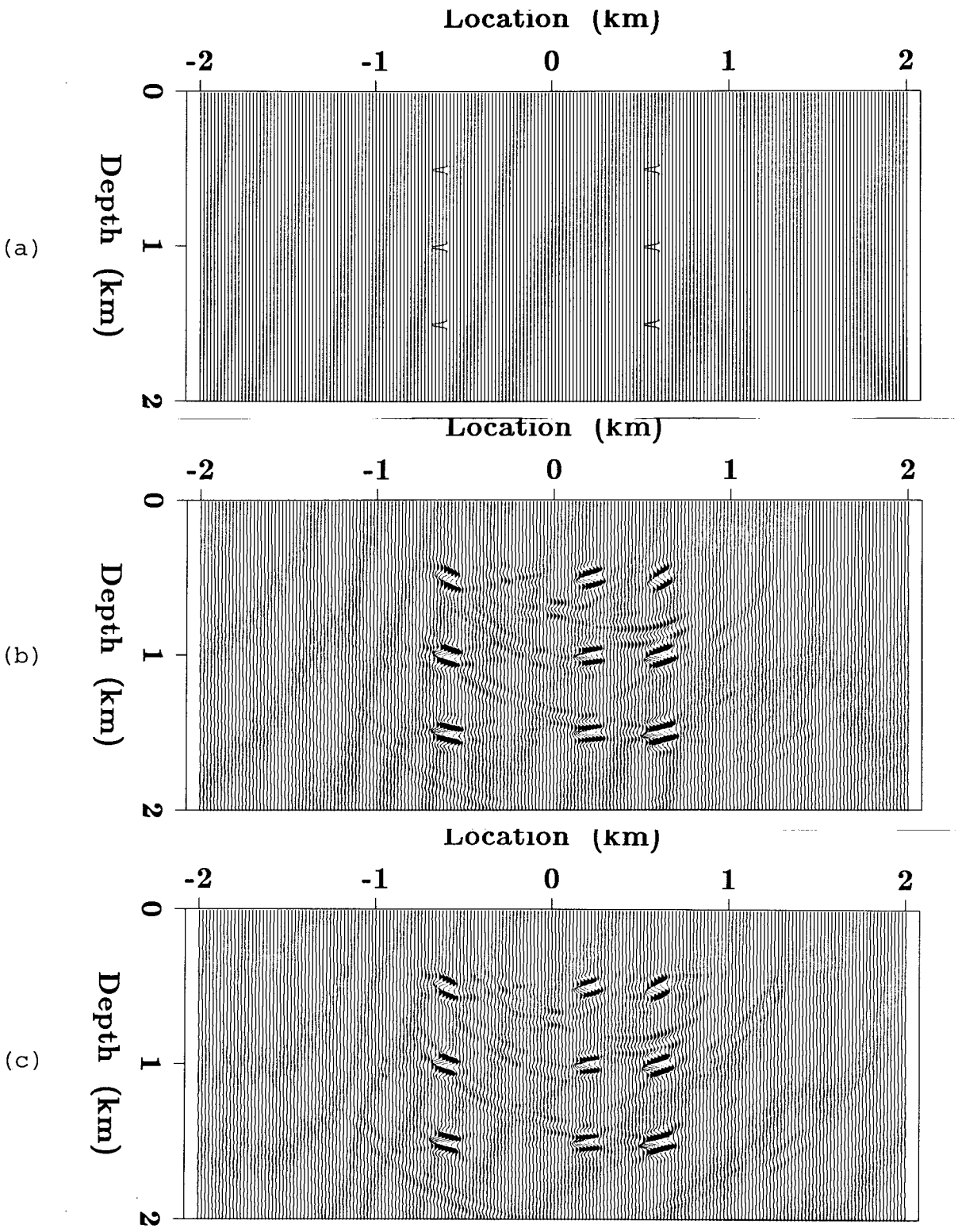
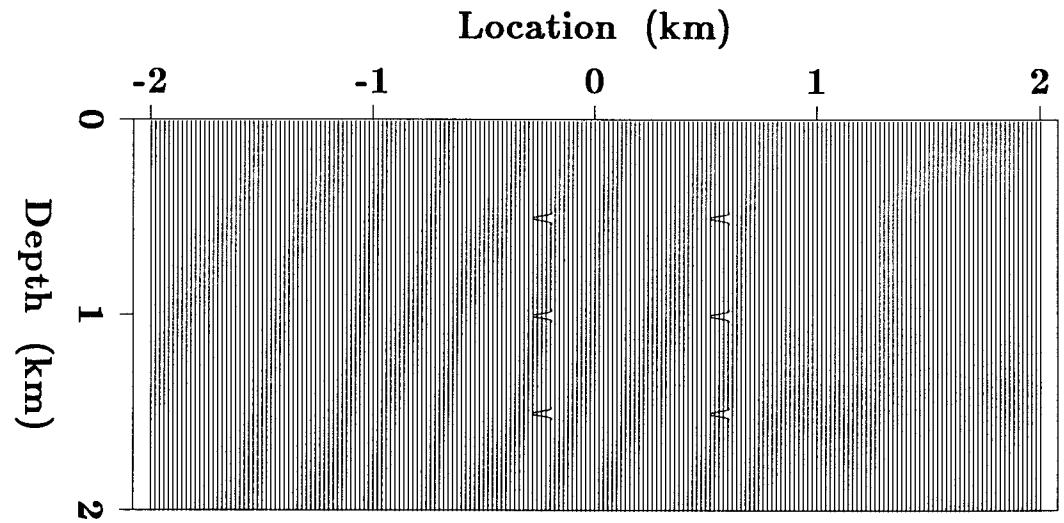
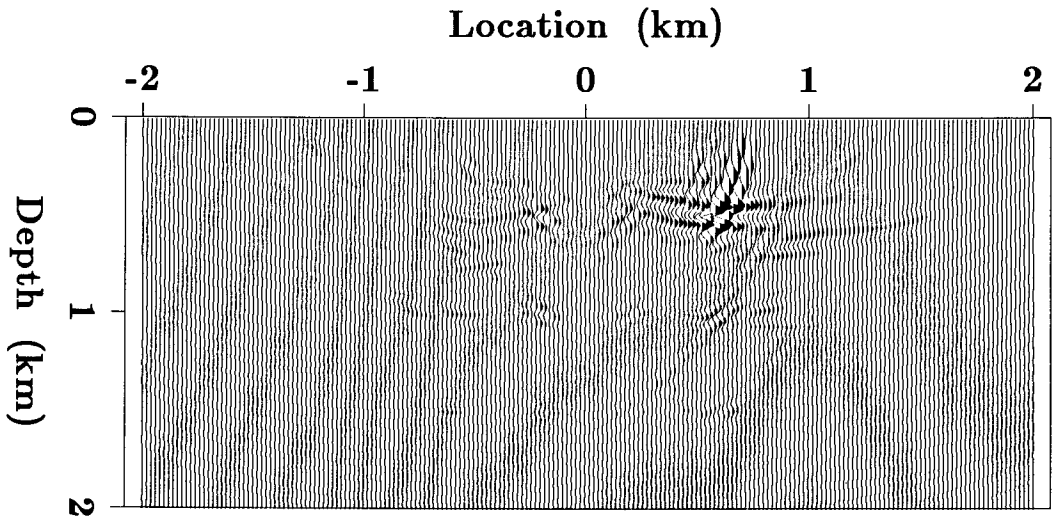


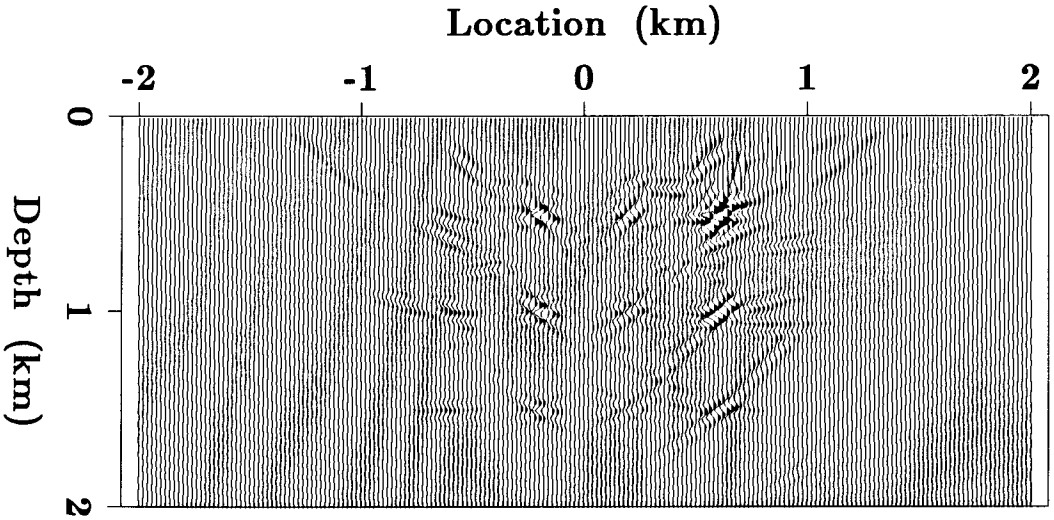
Figure 11: True model and inversion results for the vertical source 12 diffractor example when using only the vertical component data of Fig. 4a in the inversion. (a) True P-wave velocity model, (b) P-wave velocity after one iteration, (c) P-wave velocity after five iterations, (d) true S-wave velocity model, (e) S-wave velocity after one iteration, (f) S-wave velocity after five iterations, (g) true density model, (h) density after one iteration, and (i) density after five iterations.


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

ll(d)

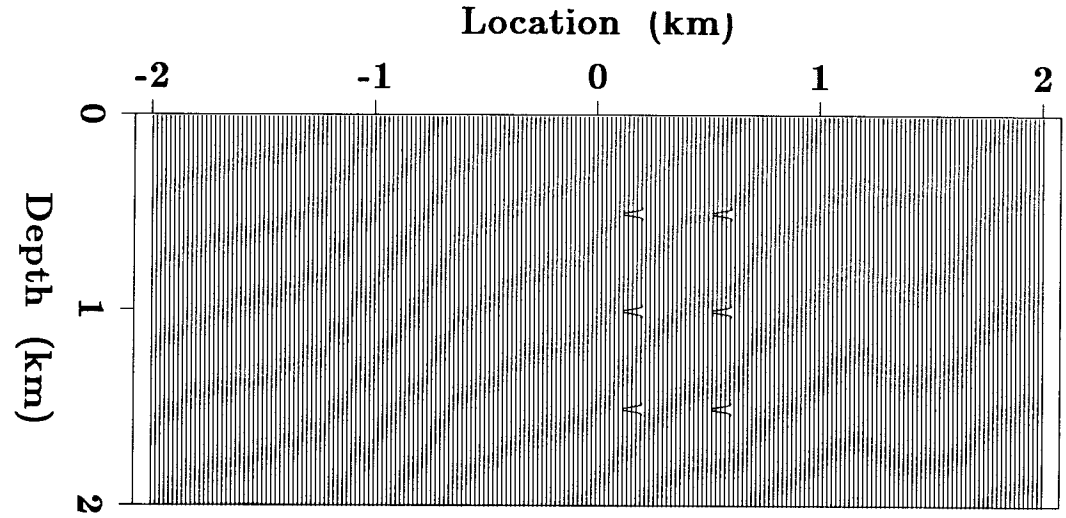

Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

ll(e)

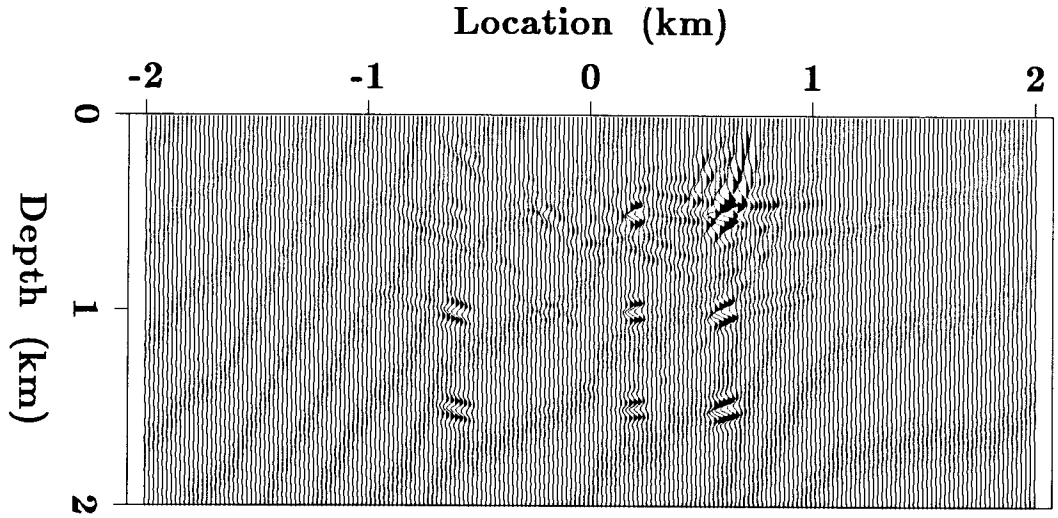

Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

ll(f)

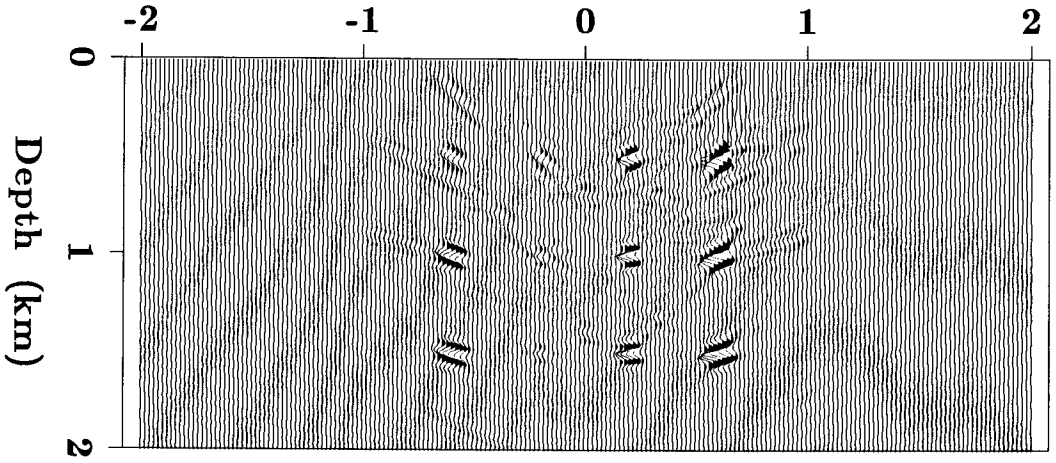
ll(g)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

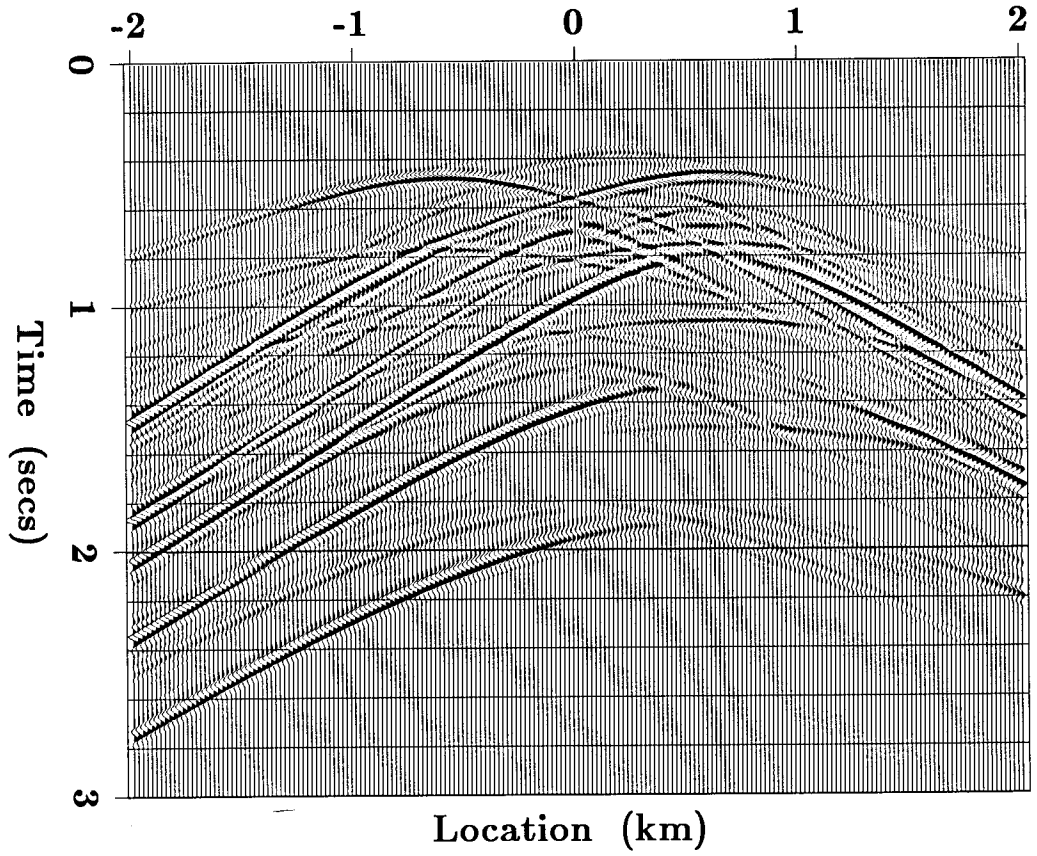
ll(h)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

ll(i)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

(a)



(b)

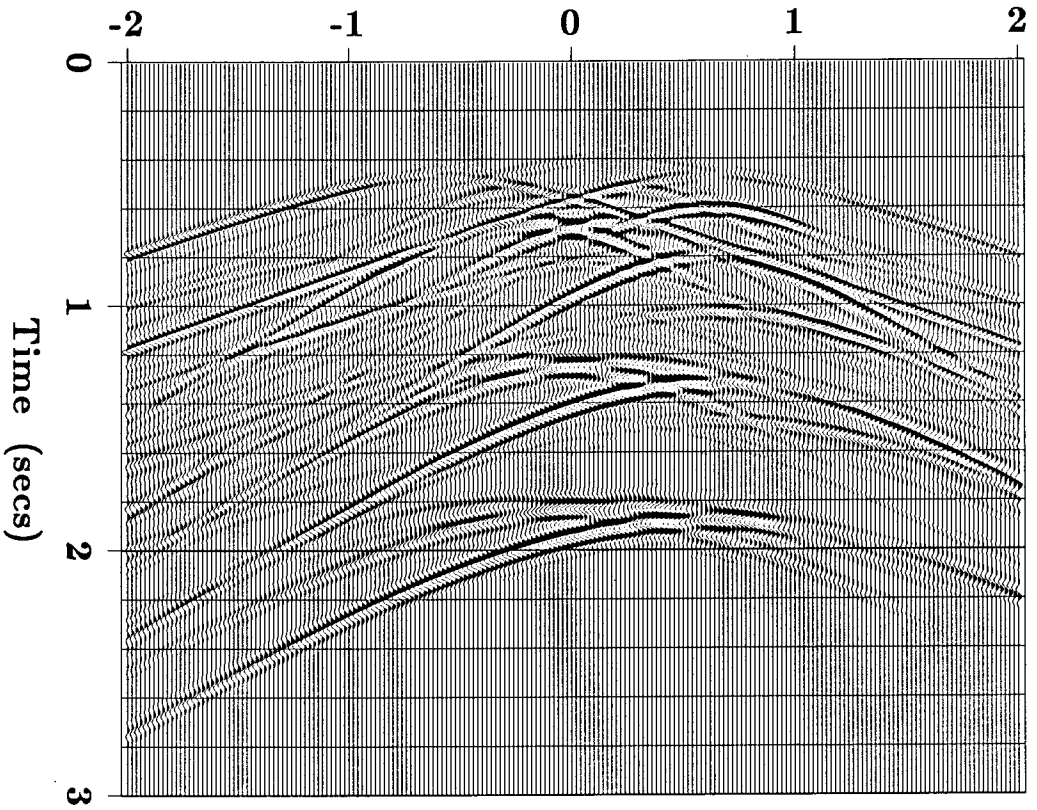
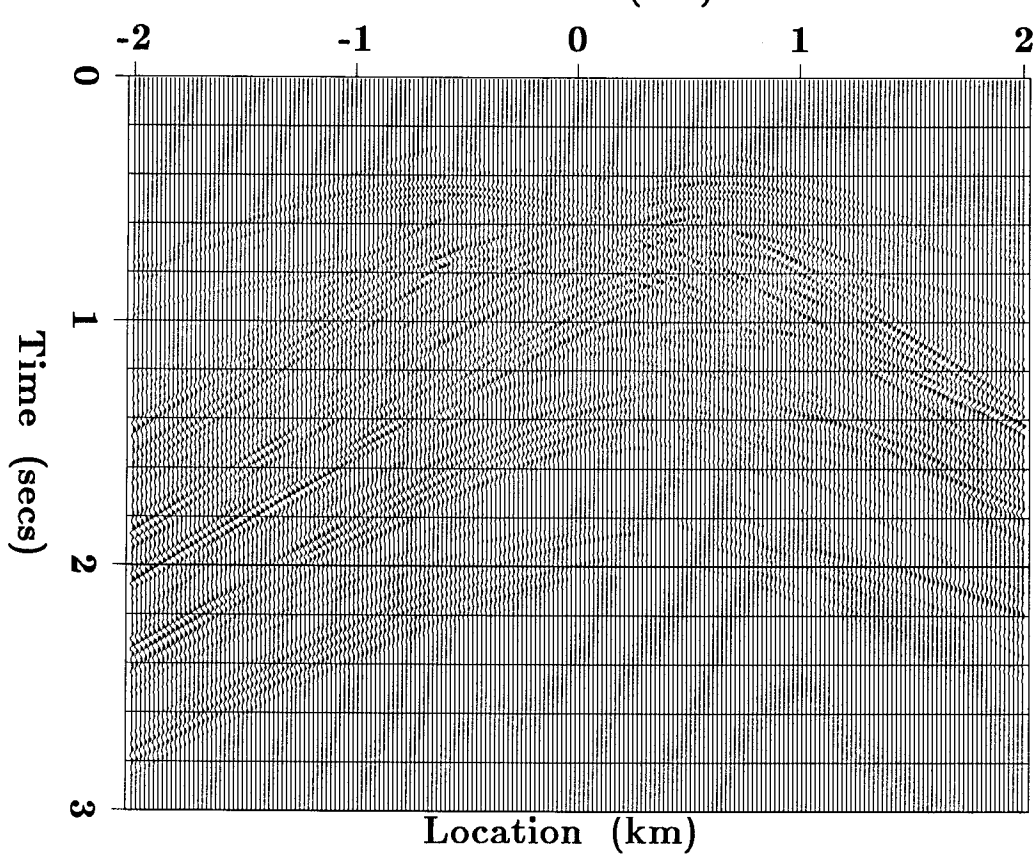


Figure 12: Horizontal source synthetic shot gather for the 12 diffractor model (the direct wave has been removed). (a) Vertical component, (b) horizontal component.

(a)



(b)

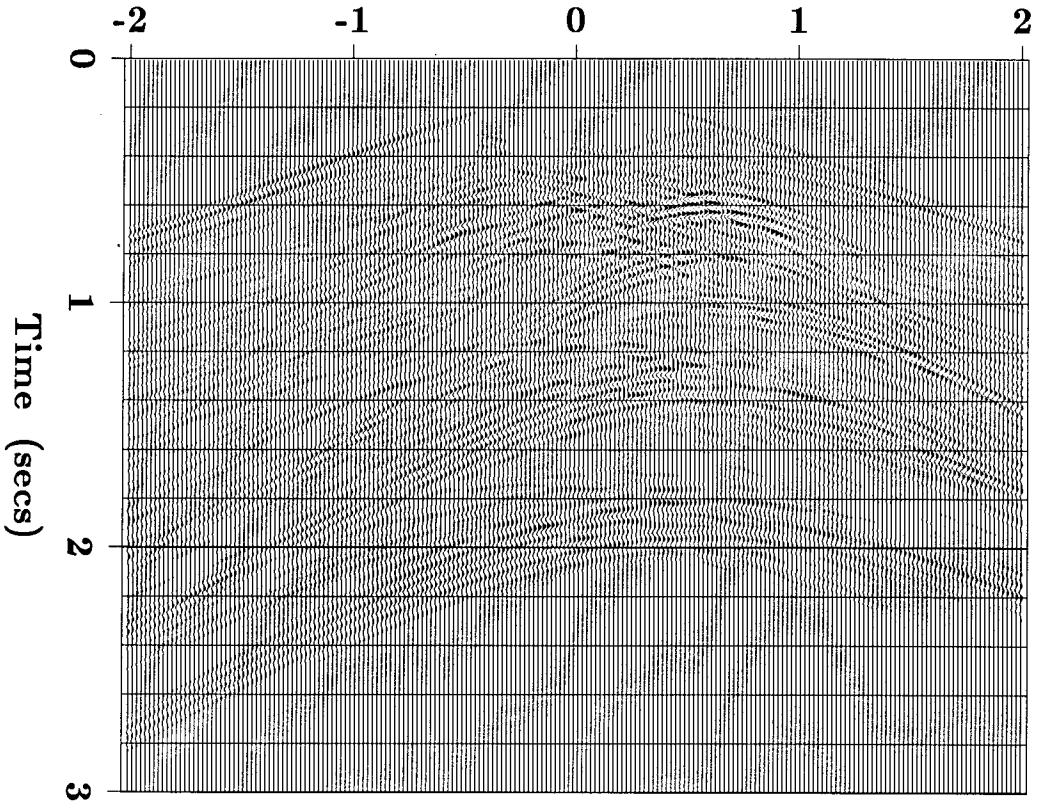


Figure 13: Residual after 5 iterations for the horizontal source 12 diffractor example. The vertical source data (Fig. 4) was also used in the inversion. (a) Vertical component, (b) horizontal component.

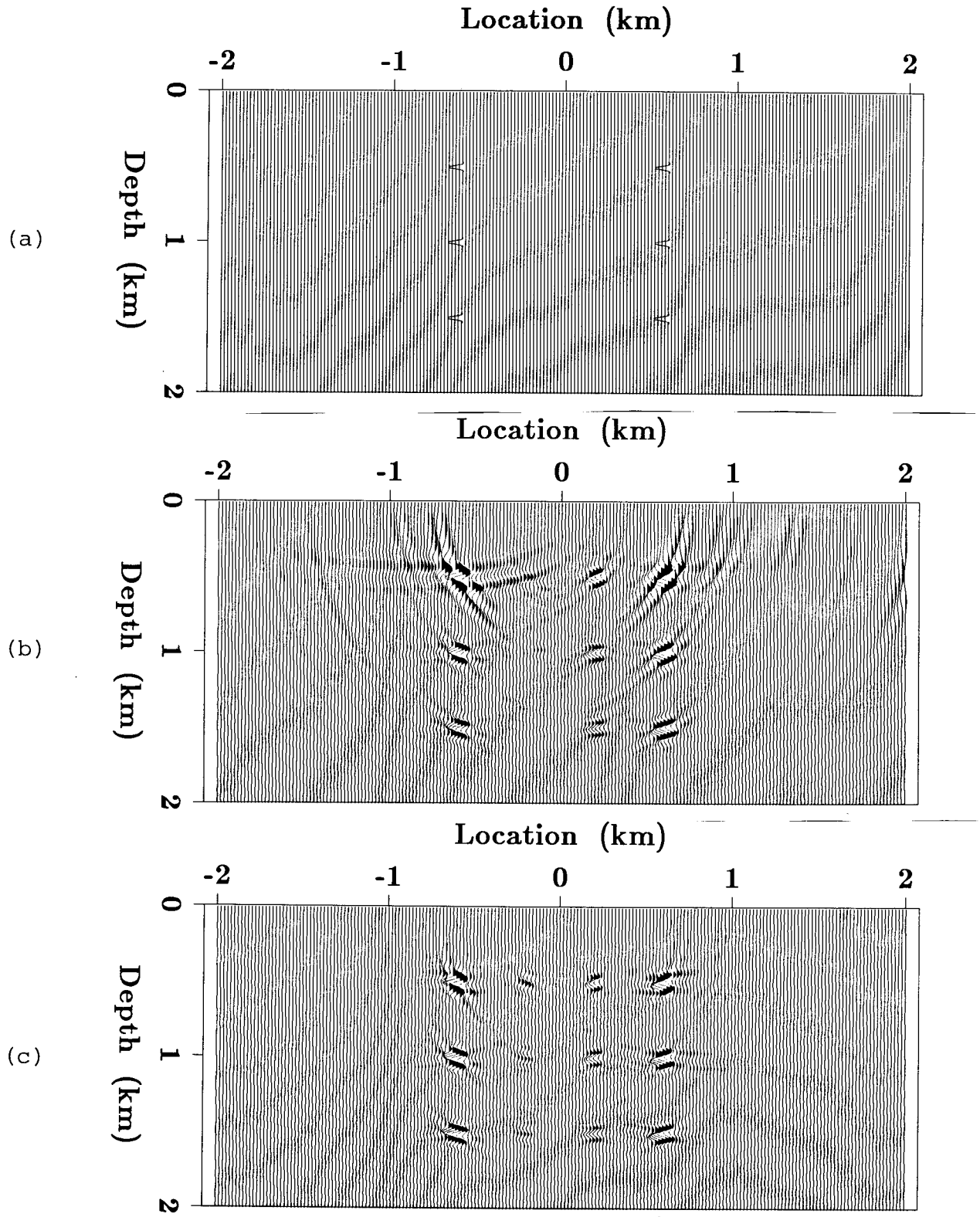
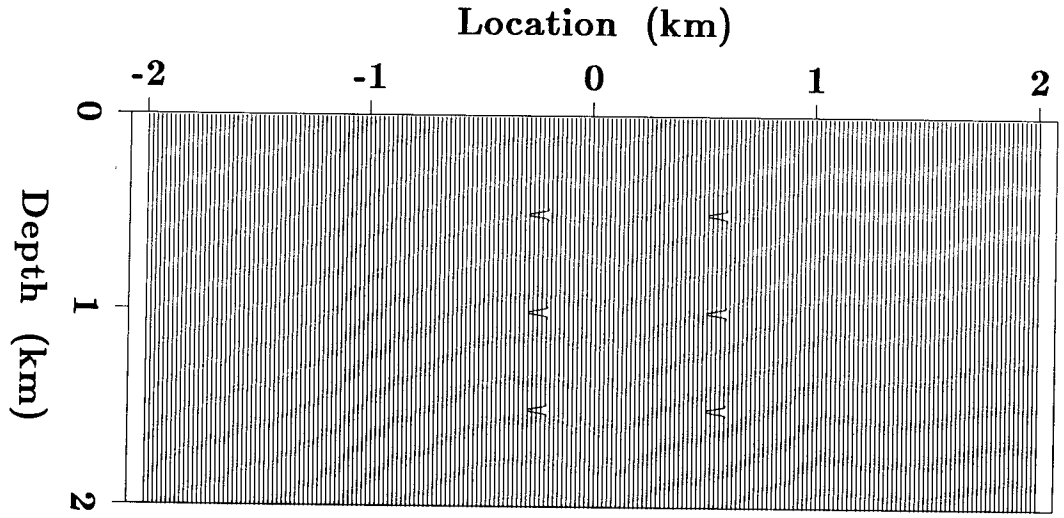
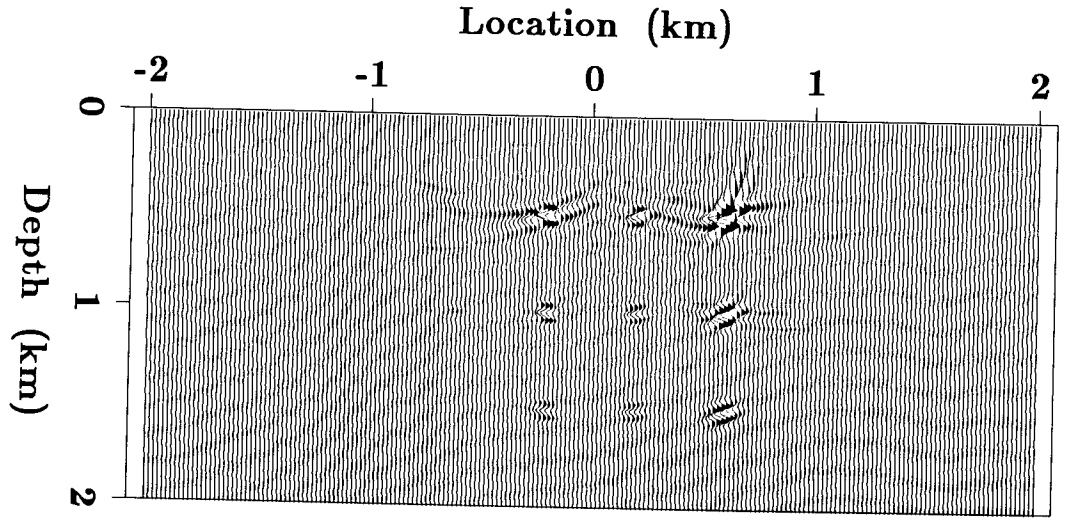


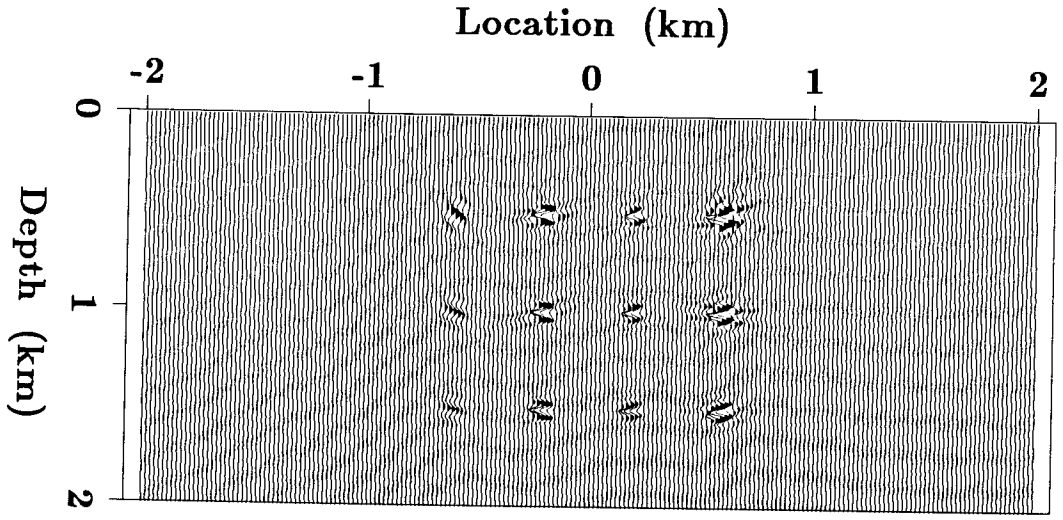
Figure 14: True model and results of a simultaneous inversion of the vertical and horizontal source data for the 12 diffractor example. (a) True P-wave velocity model, (b) P-wave velocity after one iteration, (c) P-wave velocity after five iterations, (d) true S-wave velocity model, (e) S-wave velocity after one iteration, (f) S-wave velocity after five iterations, (g) true density model, (h) density after one iteration, and (i) density after five iterations.


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

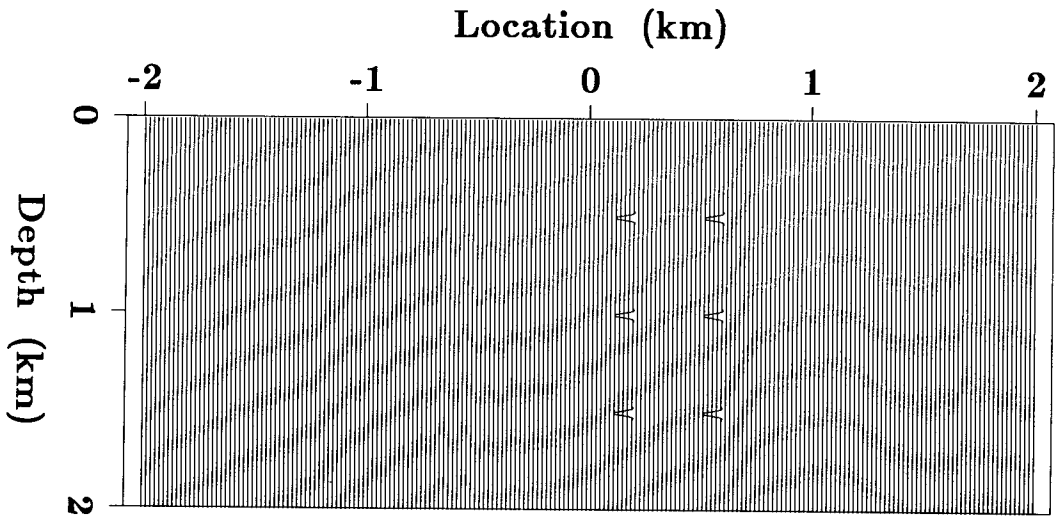
14(d)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

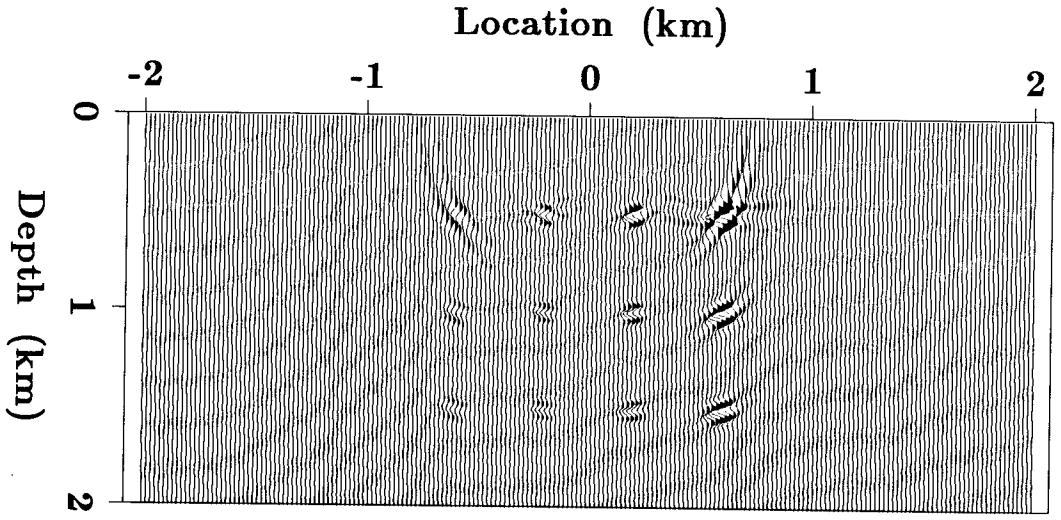
14(e)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

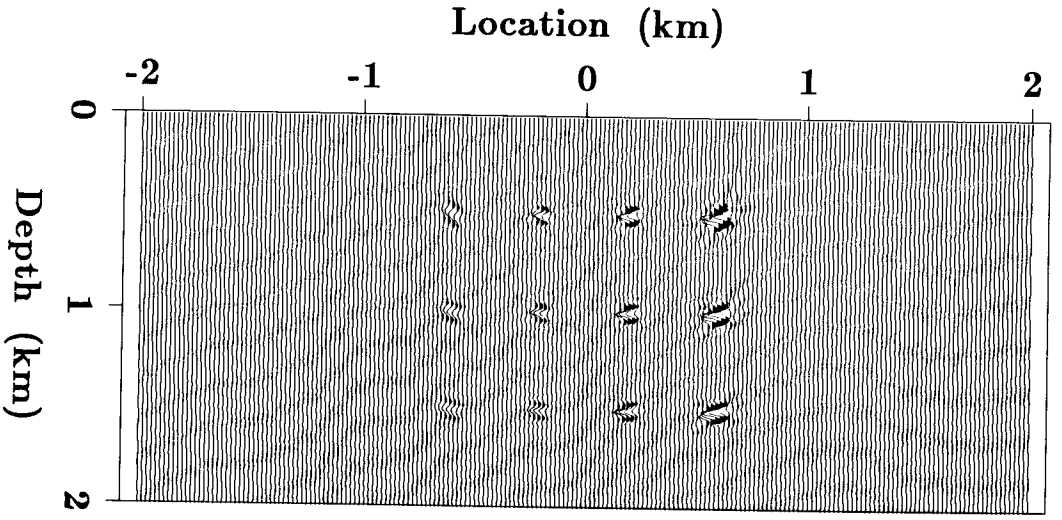
14(f)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

14(g)

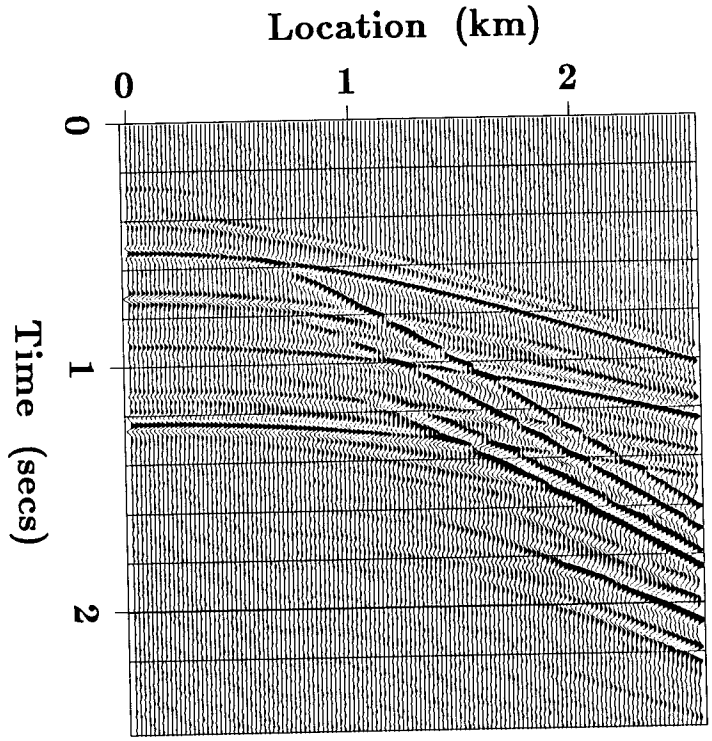

Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

14(h)


Location (km)
-2
-1
0
1
2
Depth (km)
0
1
2

14(i)

(a)



(b)

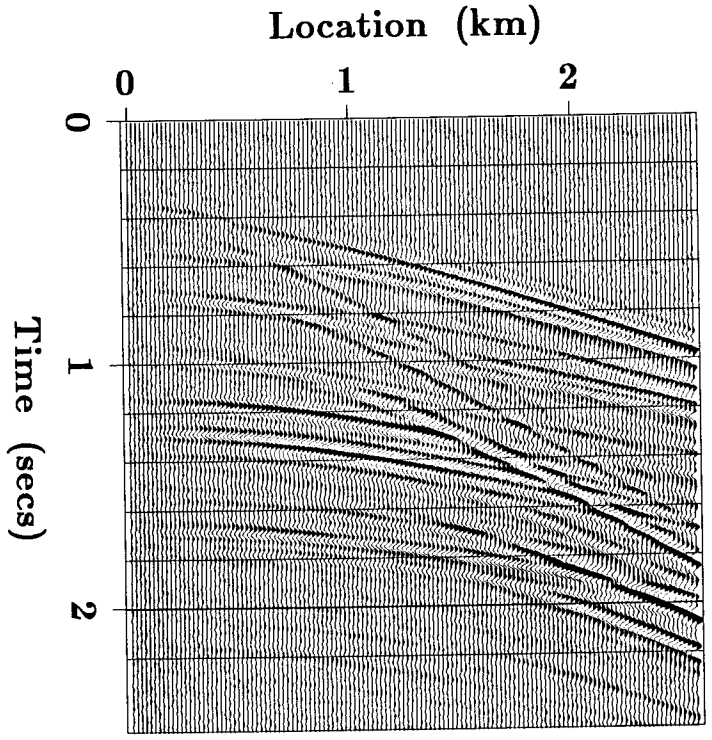
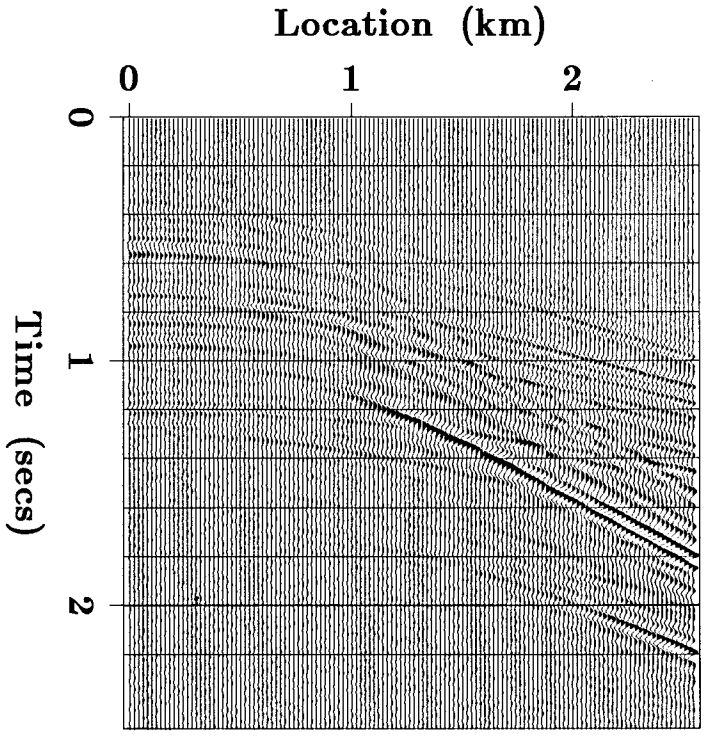


Figure 15: Synthetic shot gather produced from a vertical source for the layered model shown in the solid lines of Fig. 17 (the direct wave has been removed). (a) Vertical component, (b) horizontal component.

(a)



(b)

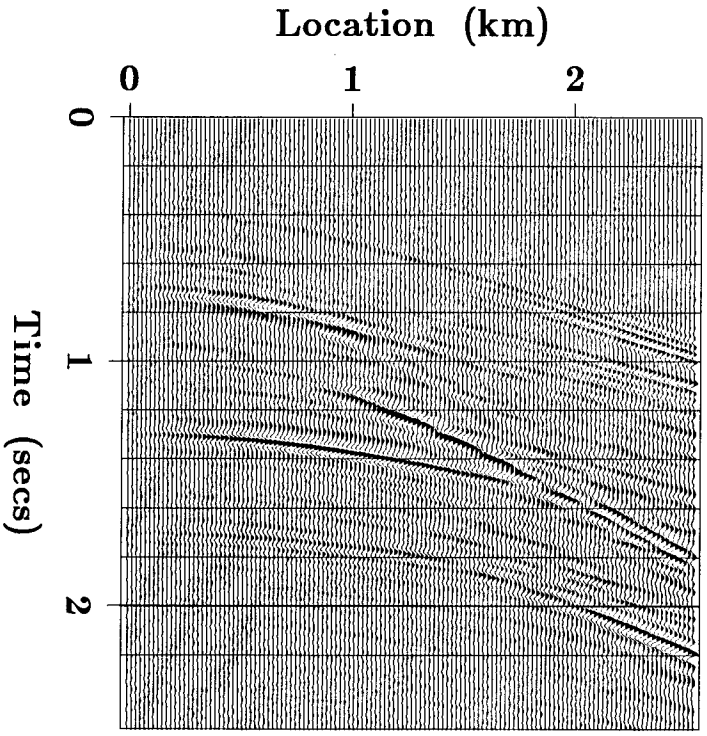


Figure 16: Residual after 15 iterations for the layered example. (a) Vertical component, (b) horizontal component.

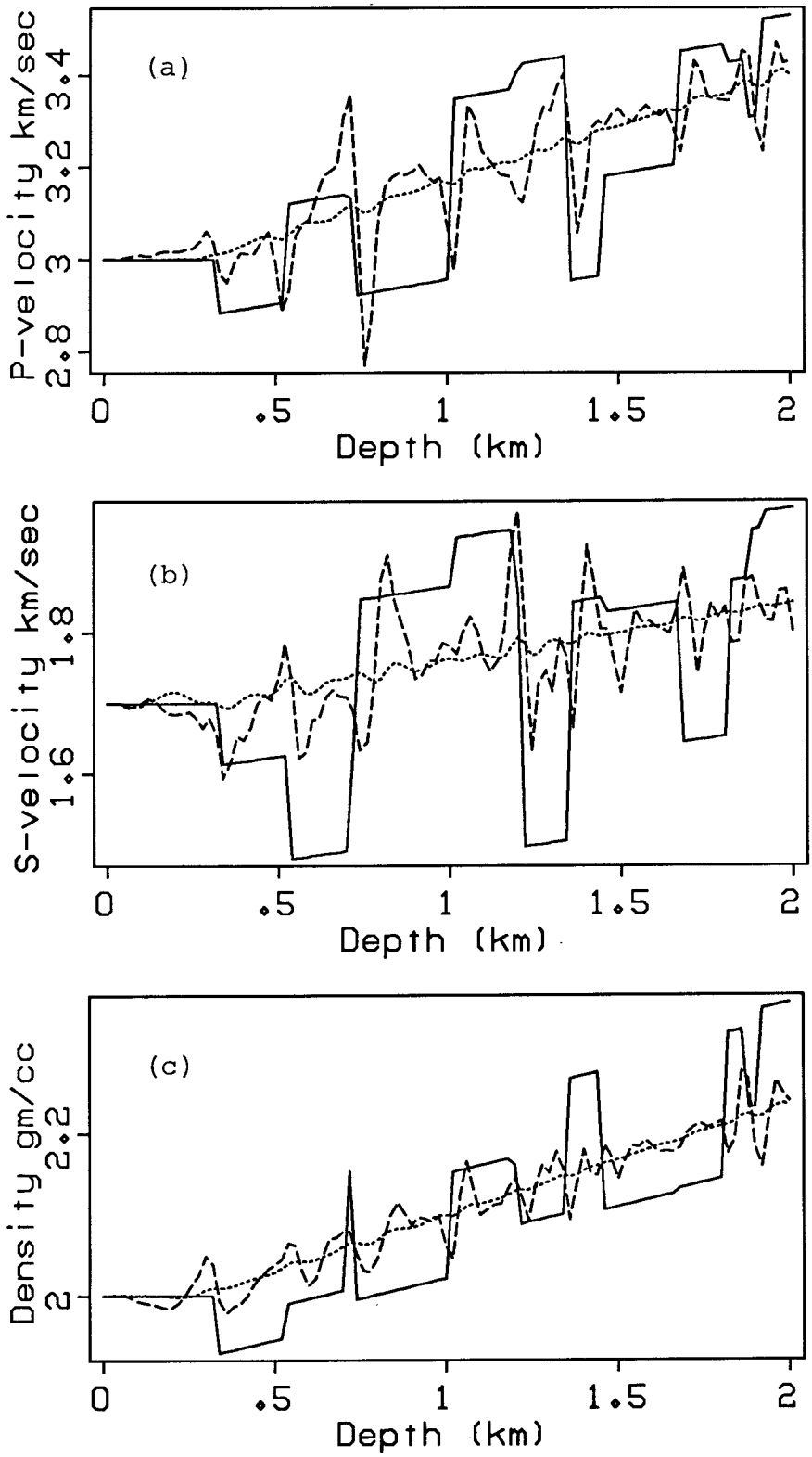


Figure 17: True model and inversion results for the layered example. (a) True P-wave velocity model (solid line) and the P-wave velocity result after 1 iteration (dotted line) and 15 iterations (broken line). (b) True S-wave velocity model (solid line) and the S-wave velocity result after 1 iteration (dotted line) and 15 iterations (broken line). (c) True density model (solid line) and the density result after 1 iteration (dotted line) and 15 iterations (broken line).

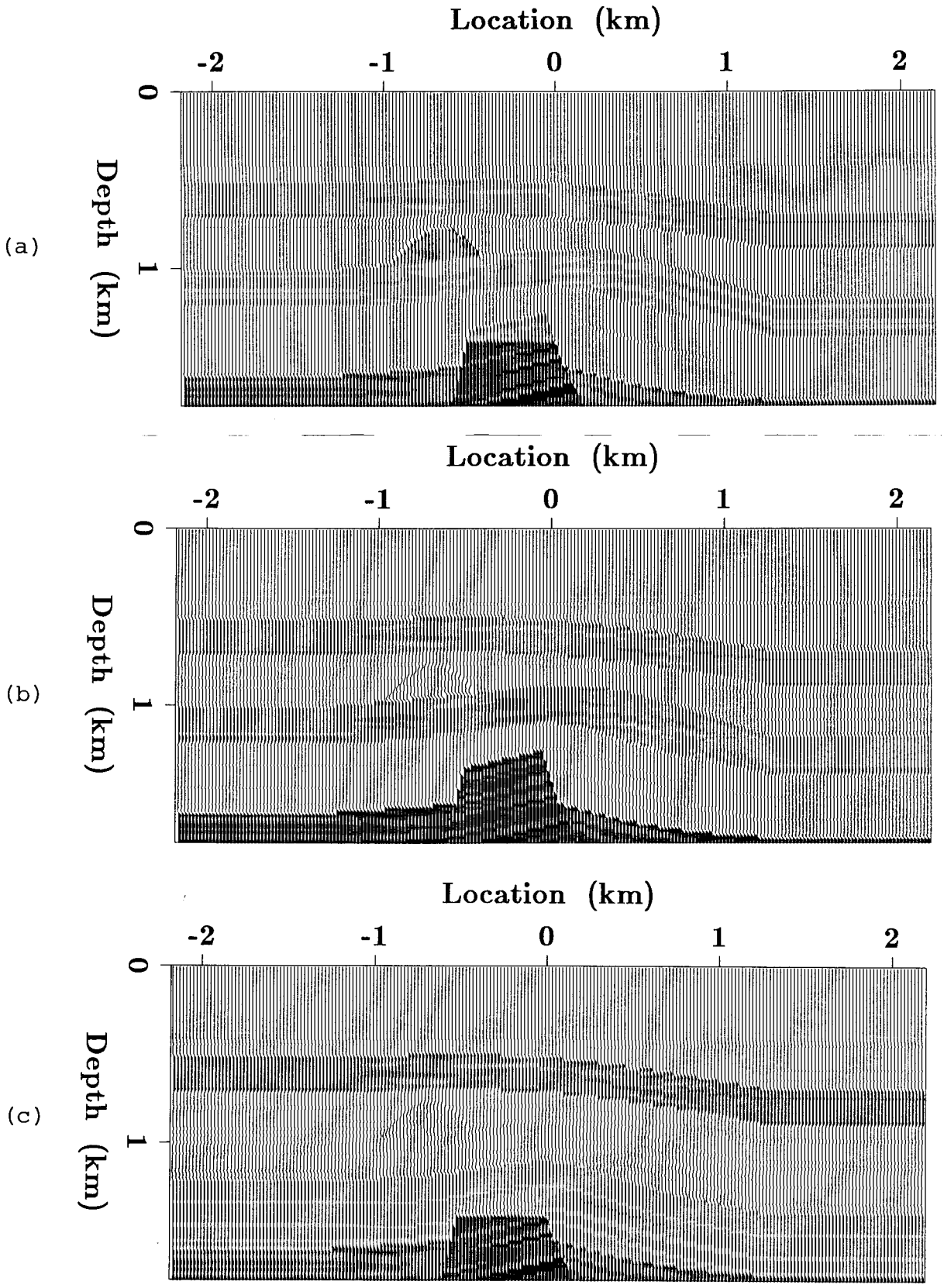


Figure 18: Horst model used to generate the synthetic data shown in Fig. 19. (a) P-wave velocity model, (b) S-wave velocity model, and (c) density model.

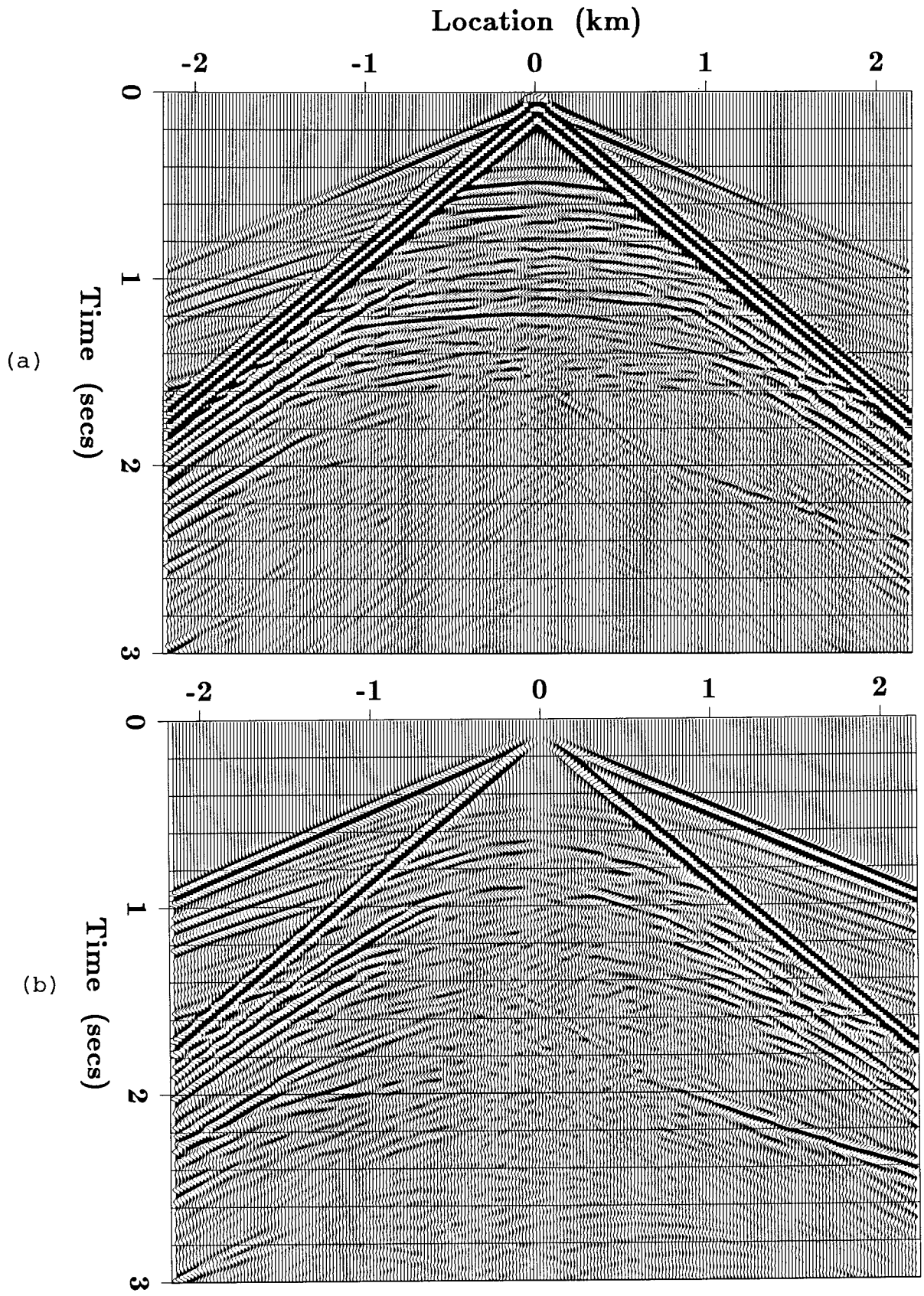


Figure 19: Synthetic shot gather for the horst model shown in Fig. 18. (a) vertical component, (b) horizontal component.

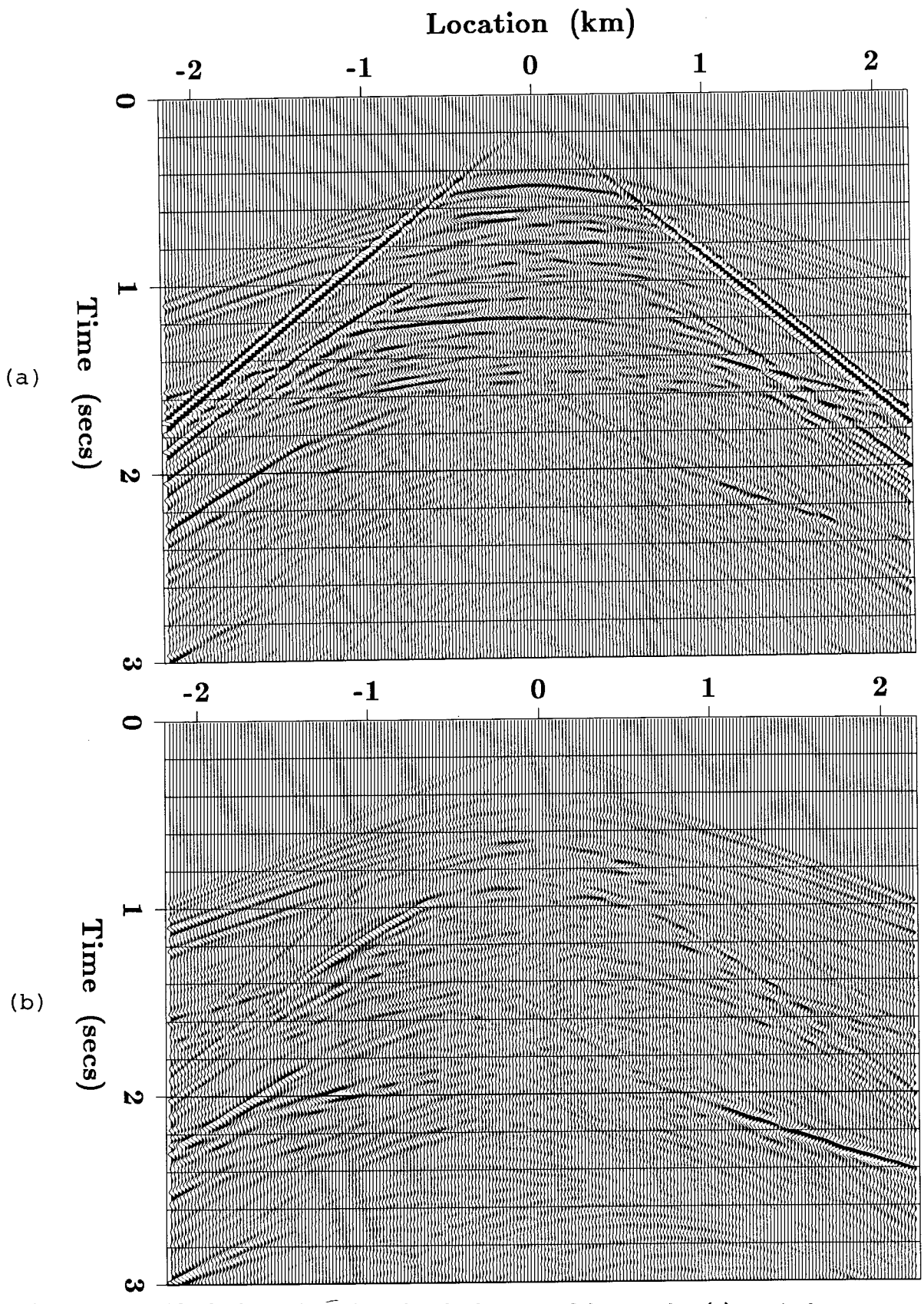


Figure 20: Residual after 3 iterations for the horst model example. (a) vertical component, (b) horizontal component.

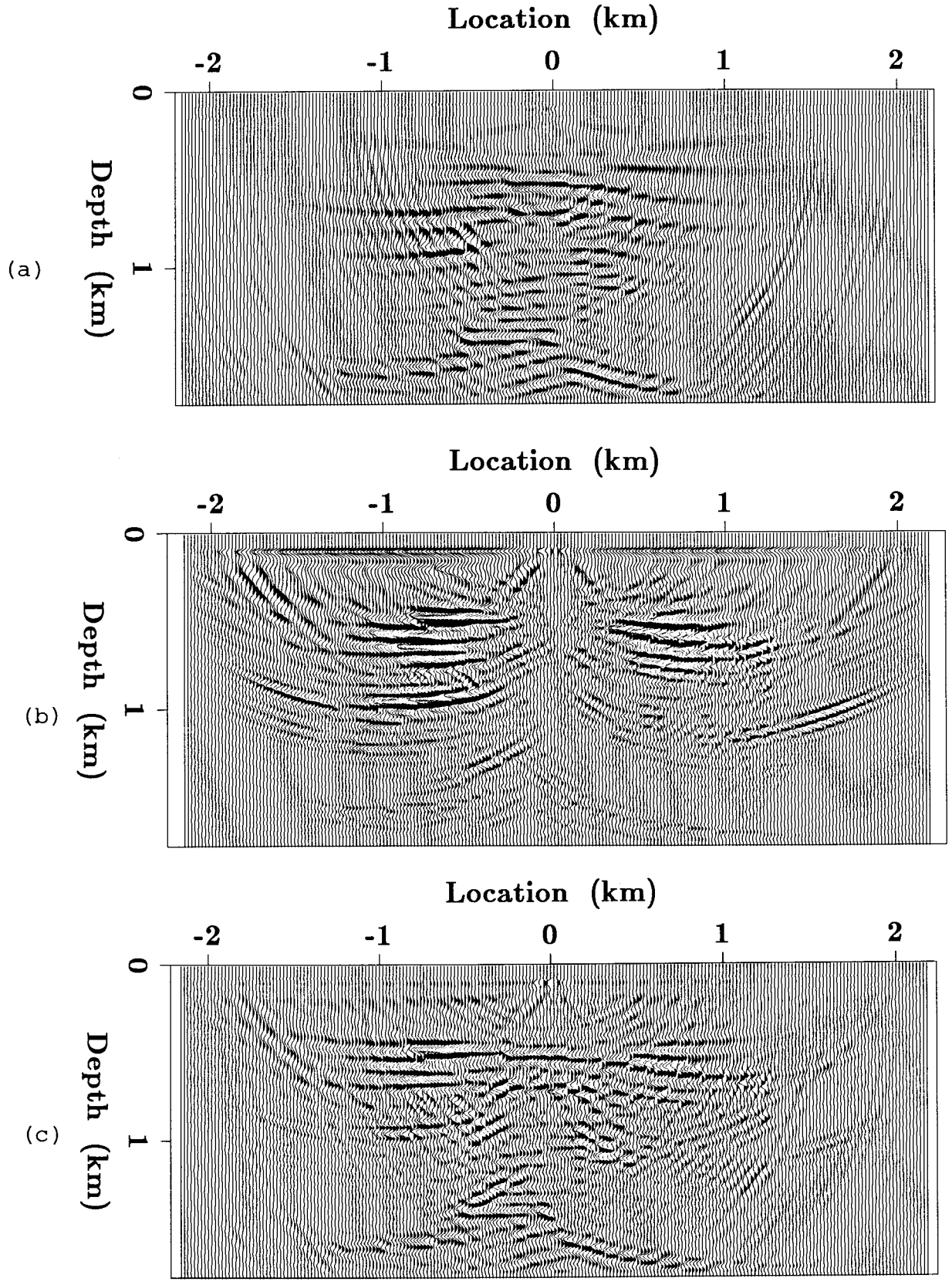


Figure 21: The high frequency component of the inversion result after 3 iterations. (a) P-wave velocity, (b) S-wave velocity, and (c) density.

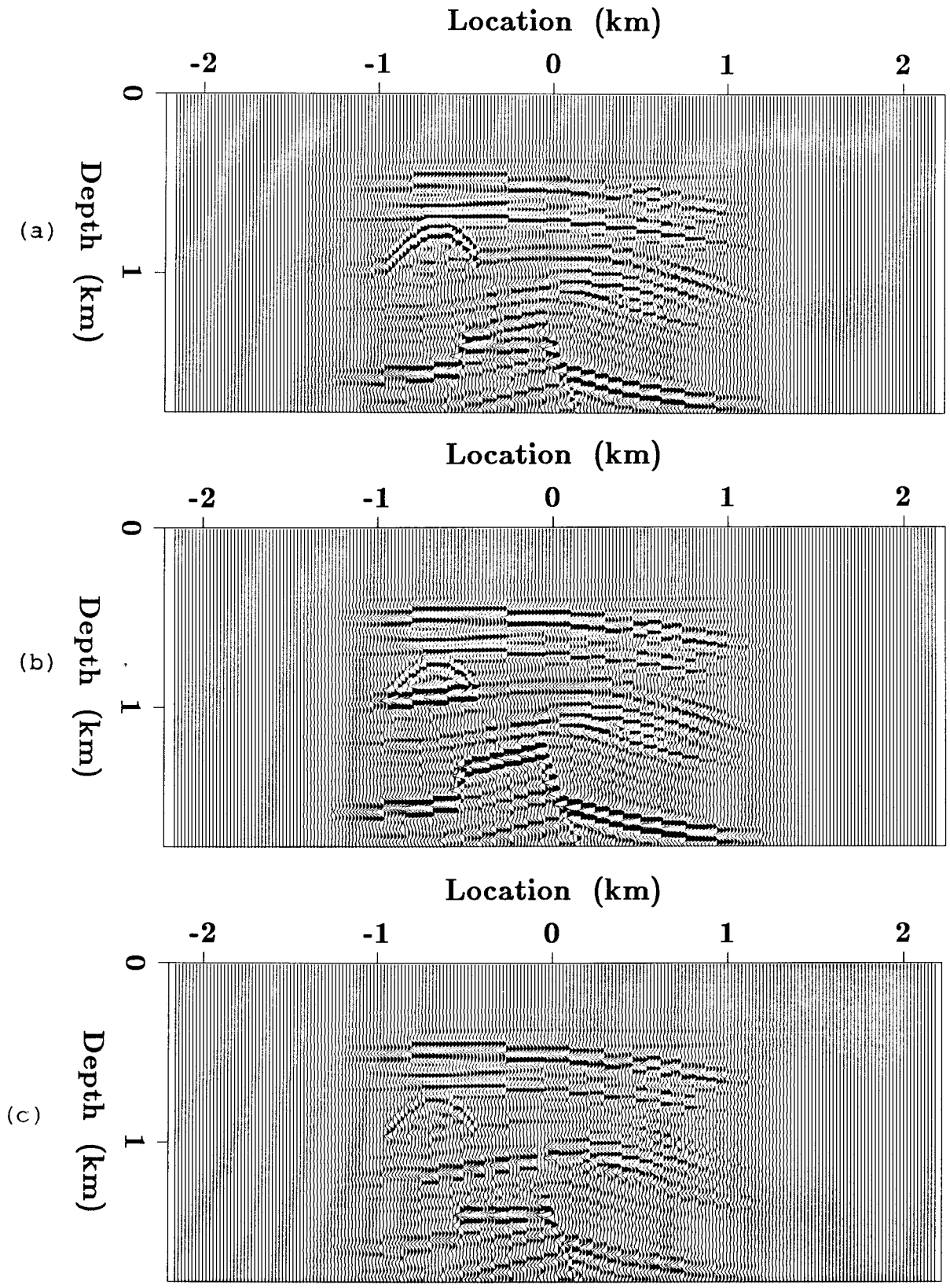


Figure 22: A high pass filtered version of the true model of Fig. 18 that has been zeroed in the regions where few reflections could occur to make it more comparable with the inversion result. (a) P-wave velocity, (b) S-wave velocity, and (c) density.