*Geophysical Journal* (1988) **95**, 135–147

# Non-linear reflection tomography

V. Farra*
*Institut Français du Pétrole, 1–4 Avenue du Bois Préau, 92506 Rueil Malmaison, France*

R. Madariaga
*Laboratoire de Sismologie, Institut de Physique du Globe, 4 Place Jussieu, tour 14, 75252 Paris Cedex 05, France*

Accepted 1988 April 21. Received 1988 April 21; in original form 1987 October 30

**SUMMARY**

Reflection tomography, the determination of velocity distribution and reflector position from reflection travel-time data, is a very non-linear inverse problem. Unlike in transmission tomography, ray paths have to be iteratively updated, because travel-time variations cannot be computed by integration of slowness along the original unperturbed ray paths. From a study of parameter sensitivity we conclude that in seismic reflection experiments the vertical variation of slowness inside layers is poorly resolved from travel-time data. For this reason, in each layer the slowness was modelled with functions varying only in the horizontal direction. A $B$-spline representation is adopted for lateral velocity heterogeneity and interfaces. These splines are local and well adapted for tomography because the spline parameters have a geometrical interpretation and they may be explicitly used as unknowns in the inverse problem. For each iteration, and for every source–receiver pair, two-point ray tracing was performed by paraxial ray tracing, and the inverse problem was solved by iterative least-squares. *A priori* data, necessary to stabilize the inverse problem, were introduced by a penalty function approach. This method is equivalent to using *a priori* convariance matrices, but it has a simpler physical interpretation and is faster to use. Damping was used to improve the convergence. The method was first tested in the inversion of synthetic data. These synthetic examples illustrate the limitations of reflection tomography: non-linearity effects, poor vertical resolution of the velocity, and decrease of the resolution with the ratio of maximum offset to interface depth. Finally, we inverted a data set from the Paris Basin. The inversion method reduces the residual norm to 6 ms, which is less than the expected error on the data.

**Key words:** Inverse theory, non-linearity, parameterization, resolution, travel-time inversion.

## 1 INTRODUCTION

Kinematical transmission tomography or travel-time inversion is commonly used by seismologists to study the 3-D heterogeneity of the lithosphere. The original method for modelling lateral velocity variations was proposed by Aki, Christofferson & Husebye (1977) who divided a horizontally layered medium into rectangular blocks, and then estimated a constant slowness perturbation for each block by inverting teleseismic travel times. This method has been refined and applied to many different data sets by a variety of authors (e.g. Thomson & Gubbins 1982; Nercessian, Hirn & Tarantola 1984). Another related seismological technique is the simultaneous inversion of earthquake location and crustal velocity structure from arrival-time data (Aki & Lee 1976; Pavlis & Booker 1980; Spencer & Gubbins 1980; Thurber 1983).

Recently, the kinematical tomography method of Aki *et al.* (1977) was adapted to reflection data by Bishop *et al.* (1985). The basic idea of reflection tomography is to use a large number of travel times for different reflected waves and source–receiver locations in order to determine the sub-surface velocity structure and reflector locations. Several authors have studied linearized reflection tomography even in 3-D. Most of the work done thus far in 3-D media assumes, however, constant velocity between interfaces (Chiu *et al.* 1986; Chiu & Stewart 1987), a hypothesis that simplifies ray tracing substantially. The assumption of lateral homogeneity of model layers is not very appropriate in areas with complex geological structures. In order to obtain these variations, Stork & Clayton (1985) and Bording *et al.* (1987) proposed to use tomography to obtain the velocity model and depth-migration in order to position the reflectors. This method can resolve small-scale velocity variations, but large-scale variations are strongly coupled with reflector position (Stork & Clayton 1986). The alternative approach is, of course, to use tomography to invert reflector positions as well as velocity distribution simultaneously (Bishop *et al.* 1985; Stork & Clayton 1986).

There are three major difficulties with reflection tomo-

* Present address: Laboratoire de Sismologie, Institut de Physique du Globe, 4 Place Jussieu, Tour 14, 75252 Paris Cedex 05, France.

graphy. First, inversion is very non-linear because small changes on the interface position produce large changes in ray trajectories. Thus, all source–receiver rays have to be retraced at each iteration of the inverse problem. Second, the model parameterization is crucial in order to obtain analytical expressions for the Fréchet differentials needed at each step in the inversion. Using boxes, for instance, requires smoothing the model between iterations in order to perform ray tracing in the updated model. It is better to describe velocity perturbations using splines that are sufficiently smooth and continuous so that ray tracing may be updated without smoothing. The last critical problem is the introduction of *a priori* information in the inverse problem. This additional information is needed because reflection tomography is intrinsically unstable due to poor resolution of velocity and interfaces near the borders of the model. Any tomographic inversion method has to make sure that error propagation from the borders of the model into the well-resolved areas is kept at a minimum level.

In this article we propose a method that takes into account the difficulties mentioned above and yet requires a reasonable computer time. A synthetic example is presented showing some of the problems that can be encountered in tomographic inversion: poor resolution, edge effects, non-linearity, etc. In order to stabilize the inverse problem, we include *a priori* information by a penalty function approach, and use a damping factor to accelerate convergence.

## 2 TRAVEL TIMES

The travel time along a ray path is expressed as the line integral:

$$T = \int_{\text{ray}} u(\mathbf{x})\, \mathrm{ds}, \tag{1}$$

where $s$ is the arc length along the ray and $u$ is the slowness (reciprocal of velocity). This relation is non-linear because the ray path $\mathbf{x}(s)$ depends on the slowness. To obtain an improved model of the medium we have to perturb the reference medium in order to minimize the difference between measured and computed travel times. The reference model is specified by giving the slowness $u_0$ within each layer and the reflector depths $z_0^i$. Let us perturb this reference medium,

$$\begin{aligned} u(\mathbf{x}) &= u_0(\mathbf{x}) + \delta u(\mathbf{x}) \\ z^i(x) &= z_0^i(x) + \delta z^i(x), \end{aligned} \tag{2}$$

where $\delta u(\mathbf{x})$ is the slowness perturbation and $\delta z^i(x)$ is the perturbation of the interface i. The corresponding first-order perturbation of the travel time is:

$$\Delta T = \int_{\text{ray}} \delta u(\mathbf{x})\, ds + \sum_{\text{interfaces}} \Delta p_z^i\, \delta z^i(x_i), \tag{3}$$

where $x_i$ is the horizontal coordinate of the incidence point of the ray on the interface i, and $\Delta p_z^i = (p_z^i - \hat{p}_z^i)$, where $\mathbf{p}^i$ and $\hat{\mathbf{p}}^i$ are the slowness vectors of the incident and the reflected/transmitted ray, respectively.

## 3 RAY TRACING

In order to use relations (1) and (3), the rays between every source–receiver pair have to be traced before computing the travel time and its partial derivatives. Paraxial ray theory (Červený 1985; Farra & Madariaga 1987) can be used to solve this problem. However, this theory supposes that the wavefront has continuous second-order derivatives, so that it is necessary to find all the ray branches before doing two-point ray tracing.

Let us assume that we have already traced a ray in the medium and solved the paraxial ray equations in the ray-centered coordinates system introduced by Popov (1969) and described by Červený (1985). The correction $\Delta\phi$ to the initial ray angle $\phi$ (Fig. 1) is estimated using the element

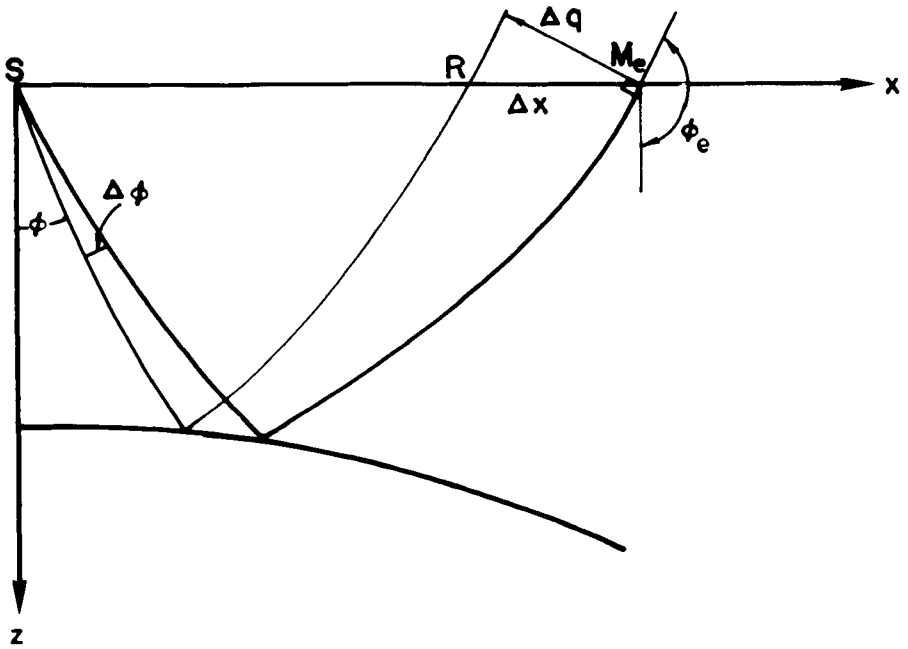


**Figure 1.** Geometry of the two-point ray-tracing problem. A ray leaving the source $S$ with the angle $\phi$ is traced in the medium. The correction of the take-off angle $\Delta\phi$ can be estimated to first order from the distance $\Delta x$ from the receptor $R$ to the endpoint of the ray $M_e$.

$Q_2$ of the dynamic ray-tracing matrix (Červený 1985). This matrix element is defined as:

$$Q_2(s) = \frac{1}{u_0}\frac{\partial q}{\partial \phi}(s),$$

where $u_0$ is the slowness at the source location and $q$ is one of the ray-centred coordinates $(s, q)$.

Let us denote $\Delta x$, the horizontal distance between the receiver and the endpoint of the ray $M_e, s_e$ the arc length between the source and $M_e$, and $\phi_e$ the emergence angle of the ray (Fig. 1). To first order,

$$\Delta x = \frac{\Delta q(s_e)}{\cos \phi_e}, \tag{4}$$

where $\Delta q(s_e)$ is the perturbation in position between the reference ray and the paraxial ray connecting the source and the receiver (Fig. 1). Moreover, from the definition of $Q_2$

$$\Delta q(s_e) = Q_2(s_e) u_0 \, \Delta \phi. \tag{5}$$

These two relations allow us to estimate $\Delta \phi$:

$$\Delta \phi = \frac{\cos \phi_e}{u_0 Q_2(s_e)} \Delta x, \tag{6}$$

which can be used to trace a new ray. Sophisticated numerical techniques (Fletcher 1980) may be used to improve the convergence towards the solution. For typical seismic reflection distances (2 km), fewer than five iterations are necessary to reach a receiver with an accuracy of 1 m.

## 4 PARAMETERIZATION OF THE MODEL

The slowness and reflector depth model must have continuous first- and second-order derivatives for the ray-tracing procedure to work. For this reason, cubic $B$-splines (de Boor 1978) were used to expand slowness and reflector depths. The use of $B$-splines accelerates the computation of the travel-time partial derivatives, because most of the splines are equal to zero except for those that have knots in the neighbourhood of the ray.

In order to obtain the $B$-spline interpolation of a function $f(x)$, we first choose a sequence of $m$ appropriate knot points $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_m)$. The density of knot points in a given segment of $x$ should reflect the desired amount of smoothness of the function $f(x)$ (de Boor 1978). The knots should be closer where the function varies rapidly, and more sparse where it is smoother.

Let us introduce the interpolated function:

$$g(x) = \sum_{j=1}^{n} P_j B_j(x) \qquad \text{with } n = m - 4, \tag{7}$$

where $P_j$ are the spline coefficients. If we take a point $x$ in some interval $[\tau_i, \tau_{i+1}]$, the sum in (7) contains only four terms $(i-3) \leq j \leq i$ because of the locality properties of $B$-splines.

Once the knot sequence $\boldsymbol{\tau}$ is defined, the coefficients $P_i$ have to be calculated so that $g(x)$ is a 'good' interpolation of the function $f(x)$. In our computations we adopted the following values for $P_i$ (de Boor 1978, Chap. IX):

$$P_i = f(x_i) \tag{8}$$

with $x_i = (\tau_{i+1} + \tau_{i+2} + \tau_{i+3})/3$. With the coefficients $P_i$ chosen in this form, the interpolating function $g(x)$ does not agree with the function $f(x)$ at the abscissa $x_i$, but is it is very close to it if the sampling is adequate. This interpolation does not produce the oscillations, which appear when one tries to force the interpolating function to fit the model function at the data points.

### *B*-spline representation of the slowness

We use a 2-$D$ $B$-spline expansion of the slowness within each layer.

$$u(x, z) = \sum P_{ij} B_i(x) B_j(z) \tag{9}$$

The slowness function is assumed to be sampled at regularly spaced points. The corresponding $B$-splines, which are called cardinal splines (de Boor 1978), are identical under translation. They have the interesting property that only four cubic polynomials, whose coefficients are computed once the ray is in the layer, are required to interpolate velocity and its derivatives.

### *B*-spline representation of the interfaces

Paraxial ray theory requires the explicit representation of all the interfaces where the velocity or its first derivative are discontinuous. Consequently, we use a $B$-spline representation for the reflector depths:

$$z^i(x) = \sum_{j=1}^{n} P_j B_j(x) \tag{10}$$

Because we want to model complicated structures, general non-cardinal $B$-splines with unevenly distributed knots on the interfaces were used. One of the most important properties of $B$-splines is that

$$\min P_j \leq z^i(x) \leq \max P_j, \qquad j \in \{i-3, i\}, \qquad \text{for } x \in [\tau_i, \tau_{i+1}].$$

That is, the interface depth is bounded by the value of the four nearmost coefficients $P_j$. Thus, the interfaces are bounded by the spline coefficients. This property suggests a very useful technique for the computation of the intersection of rays with interfaces: around every interface segment we can construct a box whose vertical edges are along the straight lines $x = \tau_i$, $x = \tau_{i+1}$, and whose horizontal edges are along the straight lines $z = \min P_j$ and $z = \max P_j$, with $j \in \{i-3, i\}$. It is also possible to construct a hierarchy of boxes which includes several basis boxes and even a global box which includes the whole interface and whose edges are $z = \min P_j$, $z = \max P_j$, $j \in \{1, n\}$, and the vertical sides of the model. Testing the intersection of the ray with these edges is much faster than testing the intersection with the true interface, because it does not require computation of $B$-spline values at each step of integration of the ray equations.

## 5 INVERSE PROBLEM

### Classical inverse approach

The problem to be solved is to find the set of parameters $\mathbf{m}$ that minimizes the function:

$$S(\mathbf{m}) = \| T_{\text{obs}} - T(\mathbf{m}) \|^2, \tag{11}$$

where $T_{obs}$ are the observed travel times and $\mathbf{m}$ is the set of spline coefficients $P_{ij}$ of the velocity distribution (9), and $P_i$ of the interfaces (10). This non-linear least-squares problem is solved iteratively by the Gauss–Newton method, which linearizes expression (11) about the current model $\mathbf{m}_c$ to obtain the linear least-squares problem:

$$E(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2, \tag{12}$$

where $\mathbf{y}$ is an $n$-vector containing the time residuals $\Delta T = T_{obs} - T(\mathbf{m}_c)$, $\mathbf{x}$ is an $m$-vector containing the parameters perturbations $\mathbf{m} - \mathbf{m}_c$, and $\mathbf{A}$ is an $n \times m$ matrix containing the partial derivatives of travel time with respect to parameters.

The least-squares solution of the linearized problem (12) is:

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}. \tag{13}$$

Because of the poor resolution of the velocity distribution with depth in the tomographic method, the matrix $\mathbf{A}^T\mathbf{A}$ is ill-conditioned and the inversion is unstable. In order to study the properties of the linearized inverse problem (12), we use the singular-value decomposition approach (Jackson 1972). Following Lanczos (1961), an $(n \times m)$ matrix $\mathbf{A}$ with rank $p < \min(n, m)$ can be factored as:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T, \tag{14}$$

where $\mathbf{U}$ contains the $n$ eigenvectors of $(\mathbf{A}\mathbf{A}^T)$ and $\mathbf{V}$ contains the $m$ eigenvectors of $(\mathbf{A}^T\mathbf{A})$. $\boldsymbol{\Lambda}$ is an $(n \times m)$ semi-diagonal matrix partitioned as:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_p & 0 \\ 0 & 0 \end{bmatrix}$$

where $\boldsymbol{\Lambda}_p$ is a $(p \times p)$ diagonal matrix that contains the non-zero singular values $\lambda_i$ of $\mathbf{A}$ arranged in order of decreasing size. The decomposition of $\mathbf{A}$ is then:

$$\mathbf{A} = \mathbf{U}_p\boldsymbol{\Lambda}_p\mathbf{V}_p^T \tag{15}$$

where $\mathbf{U}_p$ and $\mathbf{V}_p$ consist of the $p$ columns of $\mathbf{U}$ and $\mathbf{V}$, respectively. Introducing the partition $\mathbf{V} = [\mathbf{V}_p, \mathbf{V}_0]$, the general solution of the linearized inverse problem can be written (Menke 1984):

$$\mathbf{x} = \mathbf{A}^{-g}\mathbf{y} + \mathbf{V}_0\boldsymbol{\alpha}_0, \tag{16}$$

where $\mathbf{A}^{-g} = \mathbf{V}_p\boldsymbol{\Lambda}_p^{-1}\mathbf{U}_p^T$ is the generalized inverse of $\mathbf{A}$. Vector $\boldsymbol{\alpha}_0$ can be chosen arbitrarily to fit some preconceived properties of the solution.

In order to use $\mathbf{A}^{-g}$, one must be able to identify the number $p$ of non-zero singular values. If some of the eigenvalues are small, errors in the data could cause strong fluctuations in the solution. One way of suppressing these undesirable effects is to use the damped least-squares approach (Levenberg 1944), which consists in adding a positive constant $\theta_d^2$ to the main diagonal of the matrix $\mathbf{A}^T\mathbf{A}$, so that the solution (13) is modified to:

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A} + \theta_d^2\mathbf{I})^{-1}\mathbf{A}^T\mathbf{y}. \tag{17}$$

Writing $\mathbf{A}$ in terms of $\mathbf{U}$, $\boldsymbol{\Lambda}$ and $\mathbf{V}$, we find:

$$\mathbf{x} = \mathbf{V}(\boldsymbol{\Lambda}^2 + \theta_d^2\mathbf{I})^{-1}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{y}.$$

Thus, the damping factor restricts the solution to the vector space generated by the singular values greater than $\theta_d$. This projection can be unphysical if the 'null space' corresponds to parameter combinations without clear physical meaning. Tarantola & Valette (1982) proposed a stochastic inverse that explicitly includes *a priori* information in the inverse problem through covariance matrices $\mathbf{C}_D$ and $\mathbf{C}_M$ of the data and parameters. The corresponding least-squares solution is:

$$\mathbf{x} = (\mathbf{A}^T\mathbf{C}_D^{-1}\mathbf{A} + \mathbf{C}_M^{-1})^{-1}(\mathbf{A}^T\mathbf{C}_D^{-1}\mathbf{y} + \mathbf{C}_M^{-1}\mathbf{x}_0), \tag{18}$$

where $\mathbf{x}_0$ is the *a priori* solution of the problem. The damped least-squares solution (17) corresponds to choosing $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{C}_D = \mathbf{I}$ and $\mathbf{C}_M^{-1} = \theta_d^2\mathbf{I}$.

### Penalty function approach

*A priori* covariance matrices are not always easy to define; for this reason we preferred to use another technique to introduce *a priori* information. Generally, one wants the model parameters to approximately verify a set of non-linear functional relations of the form $\mathbf{f}_0(\mathbf{m}) \simeq \mathbf{d}_0$. For example, $\mathbf{f}_0$ may represent the slope of an interface, and $\mathbf{d}_0$ the value that we would like this slope to take. Let us redefine the cost function (11) as:

$$S(\mathbf{m}) = \sigma_D^{-2}\|T_{obs} - T(\mathbf{m})\|^2 + \theta^2\|\mathbf{d}_0 - \mathbf{f}_0(\mathbf{m})\|^2, \tag{19}$$

where $\sigma_D$ is the estimated uncertainty of the data and $\theta$ is the weight of the *a priori* information compared with that brought by the data. In finite-element practice, $\theta$ is called a penalty parameter. Linearizing the function $T(\mathbf{m})$ and $\mathbf{f}_0(\mathbf{m})$ around the current model $\mathbf{m}_c$, we obtain:

$$E(\mathbf{x}) = \sigma_D^{-2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \theta^2\|\mathbf{y}_0 - \mathbf{A}_0\mathbf{x}\|^2, \tag{20}$$

where

$$\mathbf{y}_0 = \mathbf{d}_0 - \mathbf{f}_0(\mathbf{m}_c)$$

and $\mathbf{A}_0$ is a matrix containing the partial derivatives of the function $\mathbf{f}_0$ with respect to the model parameters. $E(x)$ may be written as:

$$E(\mathbf{x}) = \|\mathbf{y}_t - \mathbf{A}_t\mathbf{x}\|^2, \tag{21}$$

where

$$\mathbf{A}_t = \begin{pmatrix} \sigma_D^{-1}\mathbf{A} \\ \theta\mathbf{A}_0 \end{pmatrix}, \qquad \mathbf{y}_t = \begin{pmatrix} \sigma_D^{-1}y \\ \theta y_0 \end{pmatrix}. \tag{22}$$

The solution of the corresponding least-squares problem is

$$\mathbf{x} = (\mathbf{A}_t^T\mathbf{A}_t)^{-1}\mathbf{A}_t^T\mathbf{y}_t,$$

which can be written explicitly as:

$$\mathbf{x} = (\sigma_D^{-2}\mathbf{A}^T\mathbf{A} + \theta^2\mathbf{A}_0^T\mathbf{A}_0)^{-1}(\sigma_D^{-2}\mathbf{A}^T\mathbf{y} + \theta^2\mathbf{A}_0^T\mathbf{y}_0). \tag{23}$$

Comparing this expression with (18), it is easy to see that the penalty-function approach is equivalent to using an *a priori* model covariance matrix given by its inverse $\mathbf{C}_M^{-1} = \theta^2\mathbf{A}_0^T\mathbf{A}_0$. The *a priori* solution is replaced by the vector $\mathbf{x}_0 = (\mathbf{A}_0^T\mathbf{A}_0)^{-g}\mathbf{A}_0^T\mathbf{y}_0$ which acts as an attractor for the solution $\mathbf{x}$. A further refinement of the penalty-function method would be to assign separate weights $\theta_i$ to each constraint; this may be done replacing $\theta$ in (22) by a diagonal matrix. In (19) we also assumed that the standard deviations of the observations were uniform; this restriction may be relaxed replacing $\sigma_D$ in (22) by a diagonal matrix.

The penalty-function approach is very economical because the vector $\mathbf{x}_0$ and the matrix $\mathbf{C}_M$ do not have to be explicitly computed. All that is needed is to construct matrix $\mathbf{A}_t$ as shown in (21) by adding rows to the Jacobian matrix $\mathbf{A}$, and to solve the corresponding least-squares problem by (23).

### Iterations

The solution $\mathbf{x}$ in (23) is obtained by linearization of the problem. Because of the strong non-linearity of reflection tomography, this procedure has to be used iteratively. First, rays are traced through the current model $\mathbf{m}_c$; then, the mismatch between the computed travel times and the data is used to update the model. Between two iterations, the cost function $S$ should decrease. If this is not the case, a damping factor can be used to improve the convergence towards the solution. The introduction of a damping coefficient $\theta_d$ changes (23) to

$$\mathbf{x} = (\sigma_D^{-2}\mathbf{A}^T\mathbf{A} + \theta^2\mathbf{A}_0^T\mathbf{A}_0 + \theta_d^2\mathbf{I})^{-1}(\sigma_D^{-2}\mathbf{A}^T\mathbf{y} + \theta^2\mathbf{A}_0^T\mathbf{y}_0). \quad (24)$$

In the limit when $\theta_d$ is very large, (24) reduces to

$$\mathbf{x} \simeq -\frac{1}{\theta_d^2}\gamma,$$

where $\gamma = -(\sigma_D^{-2}\mathbf{A}^T\mathbf{y} + \theta^2\mathbf{A}_0^T\mathbf{y}_0)$ is the gradient of the cost function calculated at $\mathbf{m}_c$. By increasing the damping factor, one forces the search direction to be closer to the cost-function gradient so that the cost function decreases. Iteration is stopped when the cost function does not decrease significantly any more.

### A *posteriori* covariance matrix

The statistical formalism implied in the least-squares approach (Tarantola & Valette 1982) yields an estimate of the model confidence interval. This estimate may be obtained by calculating the covariance $\mathbf{C}_m$ of $(m - \hat{m})$, where $m$ and $\hat{m}$ are the unknown solution and its estimation, respectively. Following Tarantola (1987),

$$\mathbf{C}_m = (\mathbf{A}^T\mathbf{C}_D^{-1}\mathbf{A} + \mathbf{C}_M^{-1})^{-1}. \quad (25)$$

The corresponding expression in terms of matrix $\mathbf{A}_t$ is

$$\mathbf{C}_m = (\mathbf{A}_t^T\mathbf{A}_t)^{-1}, \quad (26)$$

or in terms of its singular-value decomposition:

$$\mathbf{C}_m = \mathbf{V}_t^T\Lambda_t^{-2}\mathbf{V}_t.$$

## 6 ANALYSIS OF THE RESOLUTION OF THE INVERSE PROBLEM

Because of the complexity of geological media, we need a large number of parameters to obtain a good representation of the seismic velocity distribution. Some of these parameters may have little or no influence on the given set of observed travel-time data; keeping them as unknowns in the inverse problem destabilizes the solution. The knowledge of the number and the nature of the degrees of freedom which can be attributed to a geological medium is very important if one does not want to be confronted with an intractable numerical problem. Before presenting inversion results, we will study the sensitivity of travel times to parameter perturbation by analysing the singular values and the associated eigenvectors of matrix $\mathbf{A}$ for a simple model.

The velocity model used as the reference medium to invert travel times of reflection data consists of two homogeneous layers with velocities of 2 and 3 km s$^{-1}$, respectively. The reflector is horizontal and is at 2 km depth. The experimental set up is as follows. Over a distance of 5 km, 25 sources were regularly spaced at 200 m intervals. For each source, 10 receivers were regularly distributed every 200 m, the first one being located 200 m to the right of the source.

The slowness perturbation is defined by the 2-D $B$-spline series:

$$\delta u(x, z) = \sum \alpha_{ij} B_i(x) B_j(z)$$

with nodes distributed evenly at every 1 km in $x$ and every 0.67 km along the vertical axis. Thus, we have 48 parameters to represent the slowness perturbation.

The reflector depth perturbation is defined by the $B$-spline interpolation:

$$\delta z^i(x) = \sum P_j B_j^i(x)$$

with knot points distributed every kilometre. Thus, we have eight parameters to represent the perturbation of the reflector position.

For each source–receiver pair, we computed the partial derivatives of the travel time with respect to the parameters $\alpha_{ij}$ and $P_j$, whose units are s km$^{-1}$ and km, respectively. Because the parameter perturbations are expected to be of the same order of magnitude, it was not useful to do a normalization (normalization would have been necessary had we used other units, such as m and s m$^{-1}$). The singular values $\lambda_j$ of the Jacobian matrix and the corresponding eigenvectors $v_j$ of the parameter space define the parameter combinations which are well determined or not by the data. A perturbation proportional to $v_j$ will be well determined if the corresponding $\lambda_j$ is large. The computed singular values spread over a large interval between 20 and the minimum resolution of the computer. The computational errors on travel times and partial derivatives are of the order of $10^{-3}$. For this reason, the eigenvectors associated with singular values whose ratio with the largest singular value is less than $10^{-3}$ will be poorly determined. Four eigenvectors, which are representative of this poorly determined space, are shown in Fig. 2. For each singular value, the upper figure represents the perturbed interface, and the lower figure represents the level lines of the slowness perturbation. The deviation $\Delta u$ between two level lines is indicated in the figure in units of s km$^{-1}$. These parameter combinations, which cannot be obtained by inversion, correspond to slowness perturbation varying with depth. When the reference medium has homogeneous layers, these vectors have no component along the interface parameters and the vertical average of the slowness perturbation is zero.

The parameter combinations which are well determined are the lateral variations of slowness and the shape of the interfaces. The largest singular values correspond to

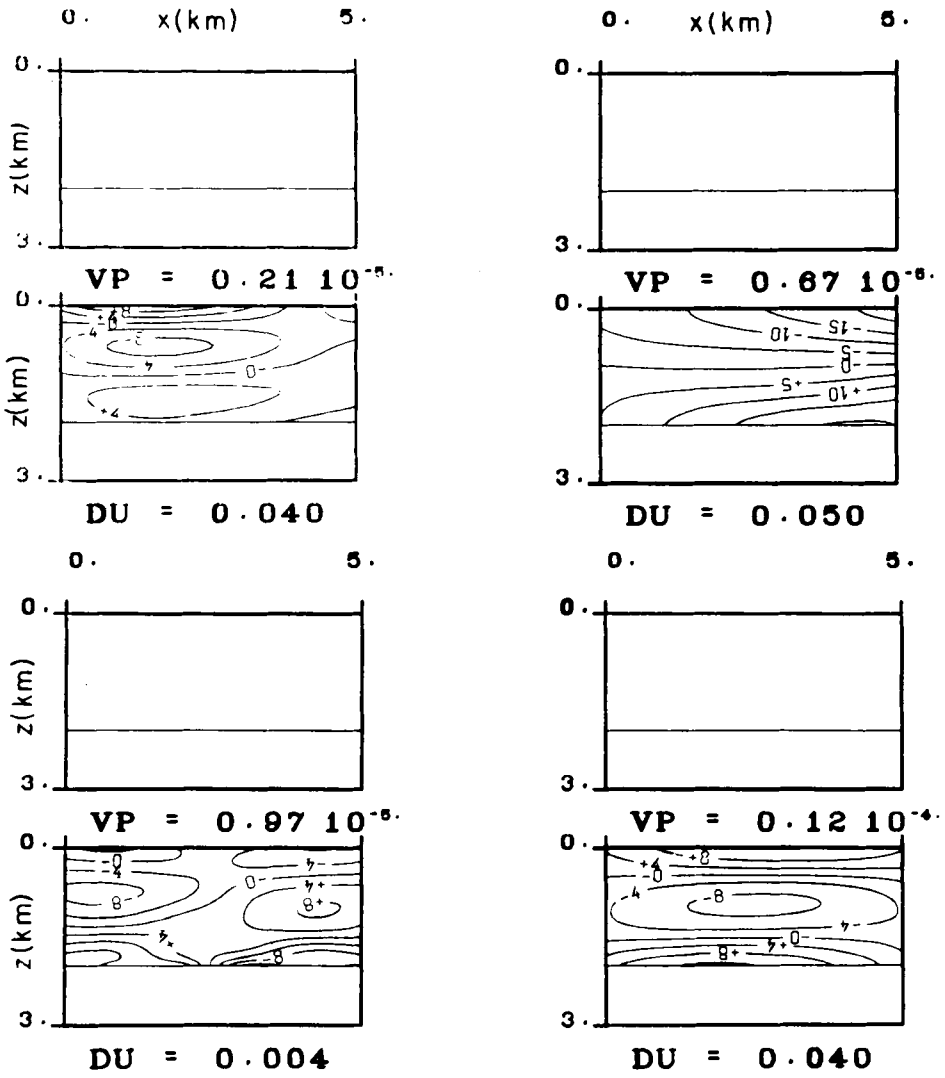


**Figure 2.** Study of resolution on a test example: geometry of eigenvectors corresponding to the four small singular values of the Jacobian matrix of linearized tomography. For each eigenvalue we present two figures: the upper one represents the shape of the interface perturbation of the associated eigenvector; the lower one shows the level lines of the slowness perturbation. DU gives the difference between two level lines. These poorly resolved eigenvectors consists mainly of vertical oscillations of velocity perturbation in the layer.

parameter combinations (Fig. 3) which produce constructive effects on travel-time perturbations, as Stork & Clayton (1986) observed. The parameter combinations whose travel-time effects interfere destructively correspond to smaller singular values (Fig. 4). These remarks lead us to replace the slowness parameterization by another one that depends only on $x$:

$$\delta u(x) = \sum \alpha_j P_j(x). \tag{27}$$

In this way, we introduce the *a priori* information that vertical velocity variations within a layer cannot be resolved from travel times of a reflection profile.

## 7 A SYNTHETIC EXAMPLE

From a large number of synthetic examples we chose the three-layer model shown in Fig. 5(a), because it illustrates most of the problems that appear in non-linear tomography. In the first layer, the velocity has the following form:

$$v(x, z) = 1.8 + 0.7z - 0.2 \exp -[(x-2)^2/2].$$

The two other layers have constant velocities of 4.5 and 5 km s$^{-1}$, respectively. The second layer is pinched out to the right.

The distance between successive sources is 200 m. For each source, 10 receivers were regularly distributed along a 2 km line, the source being located at the left end of the cable. The arrival times of the three primary reflections are shown in Fig. 6. Each line connects travel times computed for one source. The lateral velocity variation of the first layer and the presence of the pinch out in layer 2 are clearly seen in the data.

We inverted these synthetic data, starting from the plane layered model shown in Fig. 5(b). In each layer, the slowness and the interface depth perturbations were

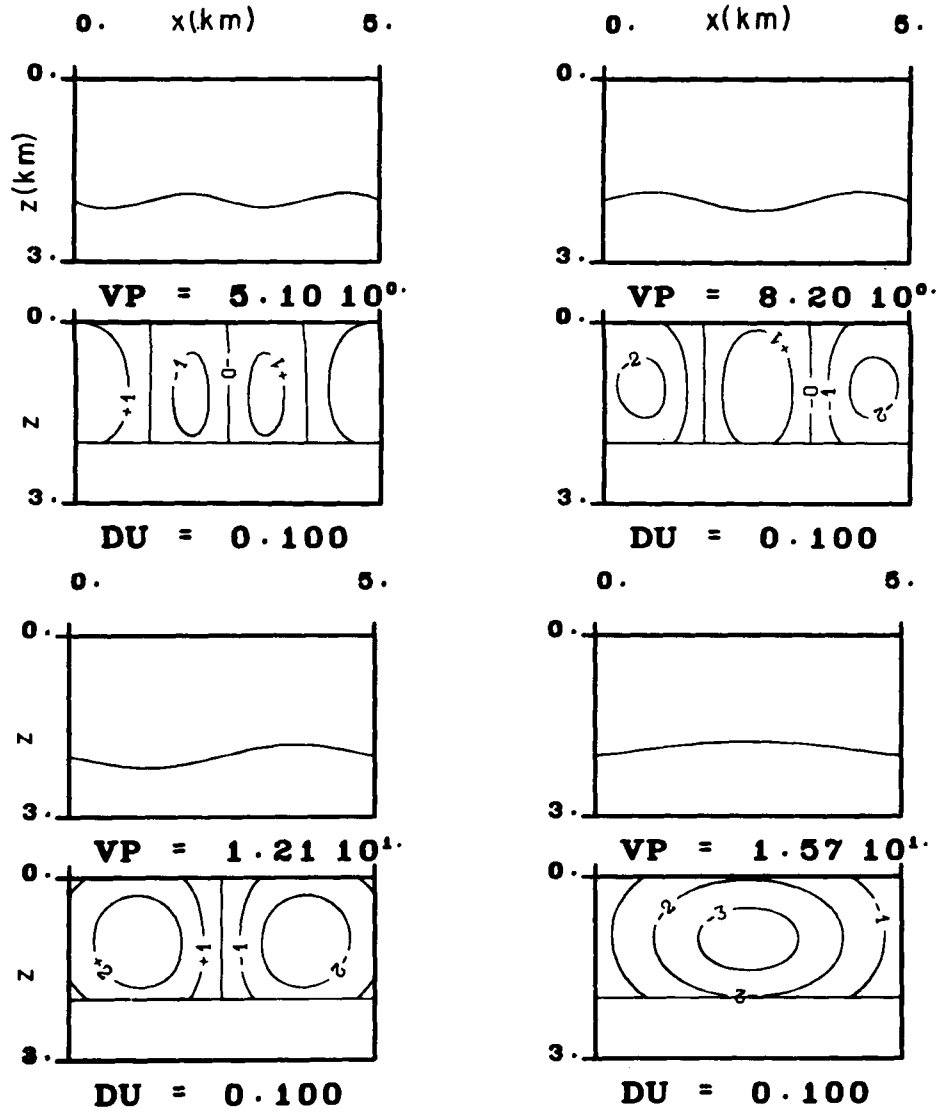


**Figure 3.** Study of resolution on a test example: geometry of eigenvectors corresponding to the four largest singular values of the Jacobian matrix. The representation is the same as that in Fig. 2. These eigenvectors correspond to lateral variations of velocity and interface geometry which have constructive effects on the travel times. These are the best resolved features in the inversion.

interpolated by $B$-splines with nodes regularly distributed every kilometre. Two sorts of *a priori* information were added to stabilize the inverse problem. Because travel-time data are scarce near the edges of the model, we limited the size of second derivatives of slowness and reflector depth by means of the penalty-function approach proposed in Section 5. The lack of data on the right-hand side of the model requires that we force the first and second interfaces to be sealed along a certain distance:

$$z^2(x) + \delta z^2(x) \simeq z^1(x) + \delta z^1(x),$$

where $z^1(x)$, $z^2(x)$, $\delta z^1(x)$ and $\delta z^2(x)$ are the interface depths and their perturbations, respectively. This constraint can be written in the linearized form:

$$\mathbf{A}_0\mathbf{x} \simeq \mathbf{y}_0,$$

where $\mathbf{y}_0$ contains the terms $(z^2(x) - z^1(x))$ calculated at several abscissae and $\mathbf{A}_0$ contains the corresponding $B$-spline values. The matrix $\mathbf{A}_0$ is sparse because of the local $B$-spline representation and is very easily computed. The penalty coefficient $\theta$ was chosen so that the condition number of the matrix that needs to be inverted in (23) was smaller than $10^{-3}$.

The following procedure was used for the inversion. We began by inverting the first layer with the first reflection data set. When the corresponding norm of the residues was smaller than the expected data error, we began to invert the second layer and so on, so that the inversion gradually became global. At each iteration, the parameters of the upper layers were reconsidered in the inversion, as well as the corresponding travel-time data set. During the inversion of the third layer, coupled oscillations appeared because the offsets were not large enough to separate slowness and interface effects. In order to decrease the condition number, we used a smaller set of nodes to represent the last interface. Four intermediate stages of the iterative inversion

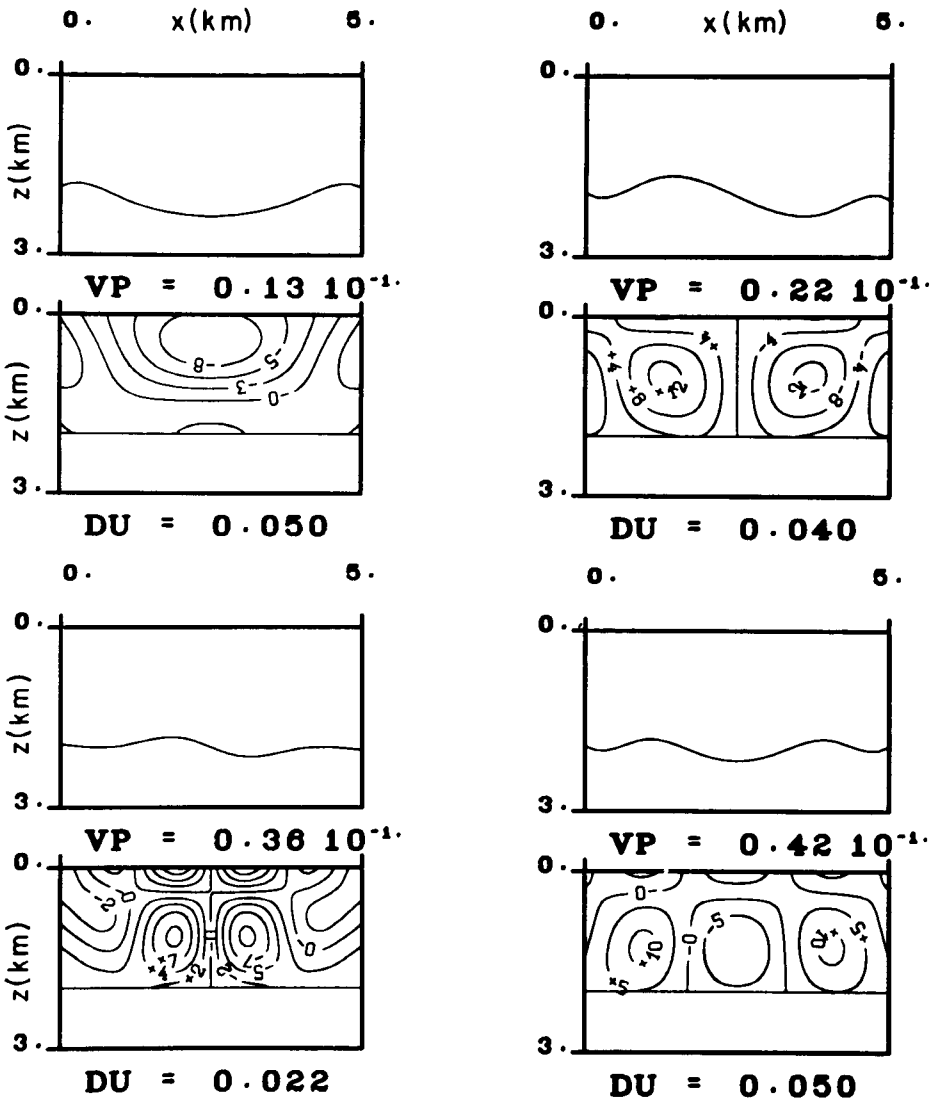


**Figure 4.** Study of resolution on a test example: geometry of eigenvectors corresponding to lateral variations with destructive travel-time effects. The corresponding singular values are much smaller than those in Fig. 3, and the features are less well resolved.

are shown in Fig. 7. In the last iteration the norm of the residues was reduced to 0.45 ms, which is smaller than the numerical error in the computation of travel times. The reflector depths obtained by inversion are systematically deeper than in the true model because of the neglect of the vertical-velocity gradient. The travel-time residuals are smaller in the right-hand side of the model because the rays are closer to the vertical axis, so that they are less sensitive to the vertical gradient. For the same reason, we recovered the correct value of the velocity in the right-hand side of the medium. On the left-hand side, the neglect of the velocity gradient leads to poor velocity resolution in the deeper layers.

## 8 A FIELD EXAMPLE FROM THE PARIS BASIN

The section under study contains Mesozoic terrains and mostly Cretaceous chalk above 700 m depth. The arrival times of the reflections on four interfaces were picked on partially stacked sections for offsets 160, 480, 880, 1280 and 1680 m. For the first two interfaces, only the first two offsets were available. Sources were spaced at 160 m intervals. The uncertainties on the arrival times were considered to be uncorrelated with a 10 ms standard deviation. The travel-time data set is shown in Fig. 8. Each line connects the travel times corresponding to a common shot. Lateral variations are evident in the first layer and their effects are clearly seen on all the reflections. These velocity anomalies come from the shallower tertiary cover or from local dolomitization of the chalk. The exploration target is the fourth interface.

We performed a travel-time inversion using a reference model consisting of four horizontal layers with constant velocities, equal to 2.8, 2, 4 and 5 km $s^{-1}$, respectively. The interfaces have corresponding depths of 1, 1.2, 1.5 and 1.8 km. In each layer, the slowness and the interface depth are functions only of $x$ and were interpolated by $B$-splines.

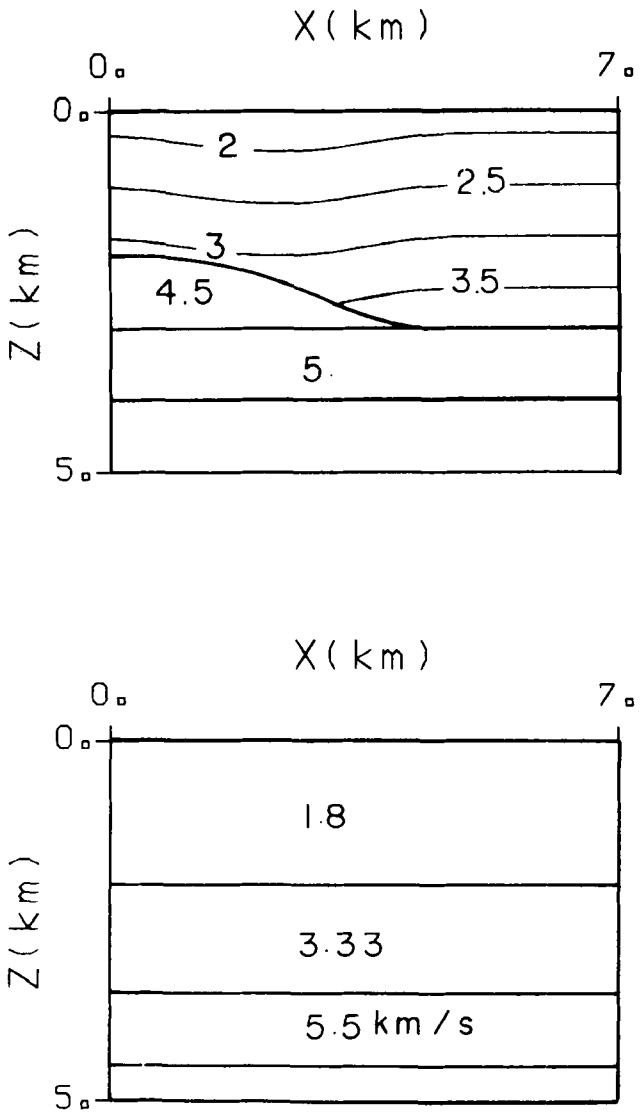


**Figure 5.** A synthetic example of tomography in a three-layered model. In the first layer, level lines of velocity are plotted. In layers 2 and 3, velocities are constant and equal to 4.5 and 5 km $s^{-1}$, respectively. The lower figure shows the starting medium that was used to invert the synthetic data set. The three layers have constant velocities of 1.80, 3.33 and 5.5 km $s^{-1}$, respectively.

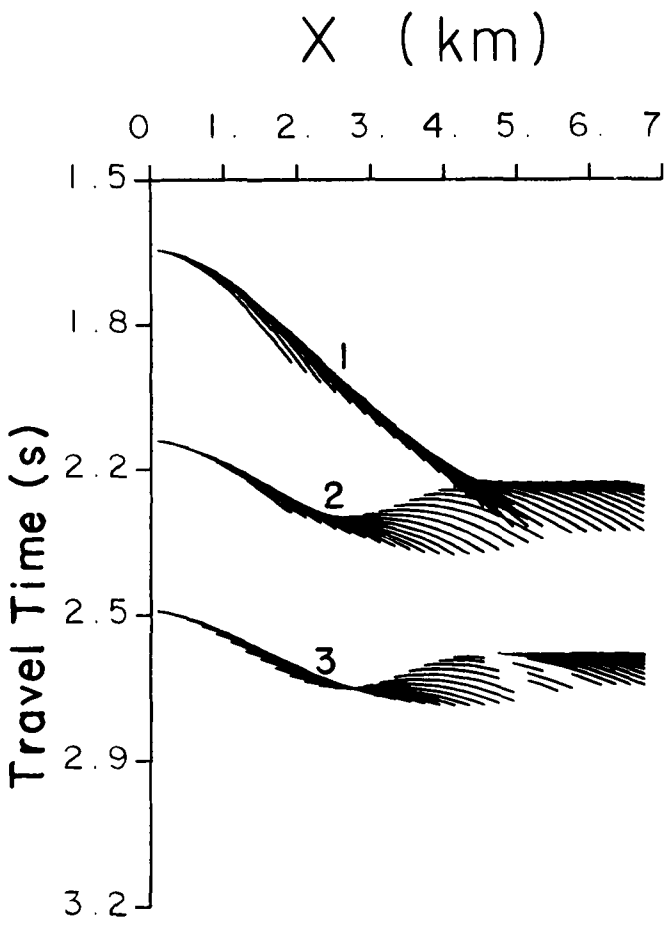


**Figure 6.** Synthetic travel time-data computed for the three-layer model shown on Fig. 5(a). Each line connects the travel times corresponding to a common source. Sources are spaced at 200 m intervals. For each source, 10 receivers are regularly distributed every 200 m, the first one being located 200 m to the right of the source.

The parameterization of the medium is displayed in Table 1.

Constraints were added on the parameters near the sides of the model, where there was not enough information available from the seismic profile. We imposed the condition that the first derivative of the slowness and the reflector depth be small. In addition to these, we imposed constraints on the second derivative of the slowness of the second and the fourth layer in order to avoid coupled oscillations. This phenomenon appears when the data cannot resolve simultaneously the lateral variations of the slowness and the interfaces.

The inverse problem was solved iteratively as in previous sections. The final inversion result obtained after 10 iterations is displayed in Fig. 9. At the top of this figure we show the interfaces inverted from the travel-time data. One can observe that there are practically no lateral variations of the interface depths, which is in agreement with the known structure of the Paris Basin. At the bottom, we show the velocities as a function of $x$ obtained for the last iteration. These velocities show much larger lateral variations. The residual norm is 6 ms, which is smaller than the estimated error of the travel-time data. The comparison of the inversion result with well data available in the neighbourhood shows a systematic error in the depth of the fourth interface. The seismic depths are larger than the well-log depths. This observed depth anomaly is probably caused by velocity anisotropy of sub-surface layers.

The *a posteriori* covariance function of the depth of the fourth interface (Fig. 10) shows a moderate standard deviation (30 m) and a short correlation length (6 km), which is in fact the length of each $B$-spline. This function is easily obtained from the covariance matrix of the $B$-spline parameters because of the linear relation connecting them with the interface depth. The local property of $B$-spline limits the correlation length of errors. This could be interesting if one wanted to estimate the uncertainty of the

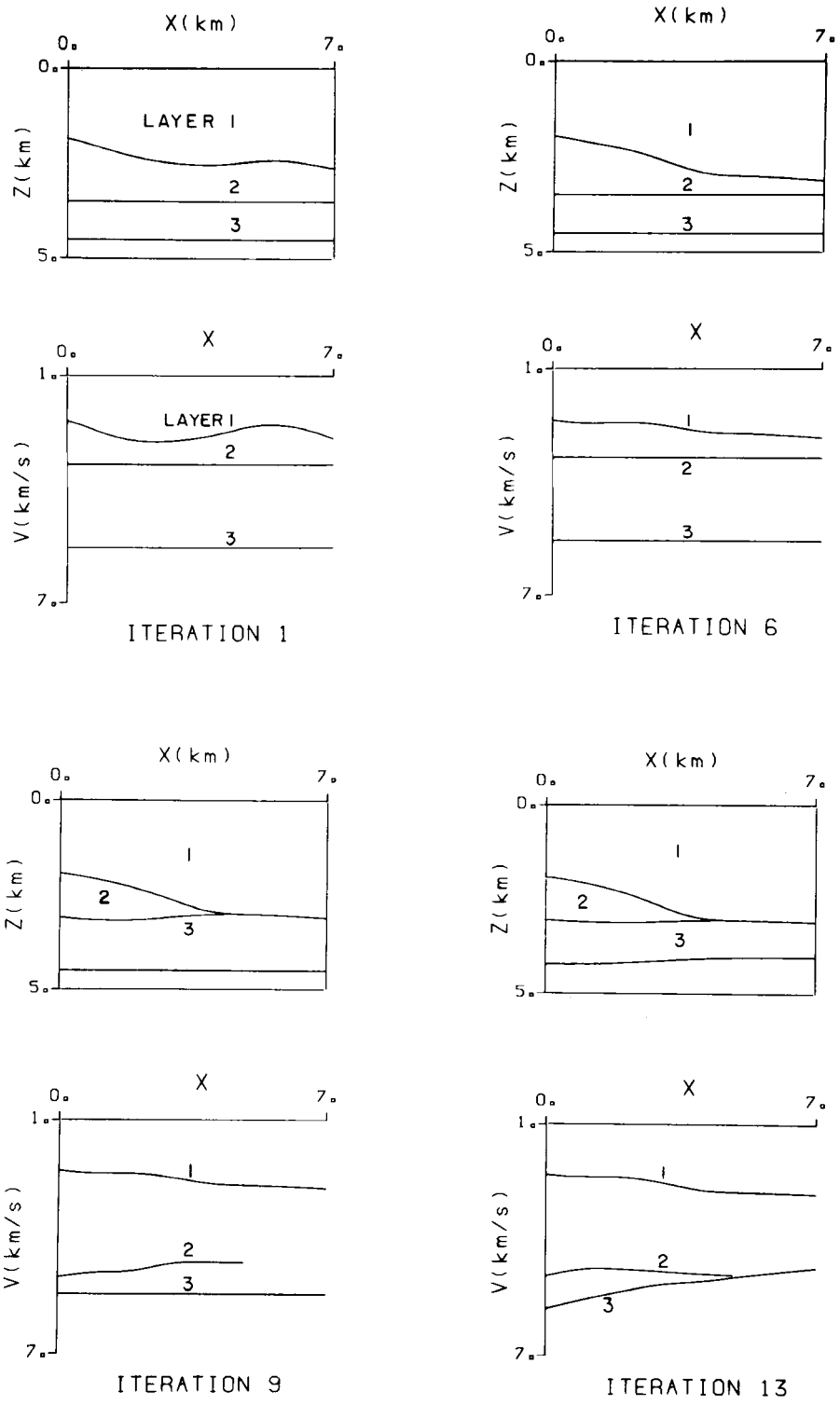


**Figure 7.** Result of the inversion of the model shown in Fig. 6. For each iteration, the upper figure shows the interface depths and the lower figure shows the velocity as a function of distance $x$. The corresponding iterations are numbers 1, 6, 9 and 13, respectively.

closure of an expected oil trap. Standard deviations are important within a 3 km range from the vertical borders of the model, especially for the velocity. The covariance function between the reflector depths and the slowness in the fourth layer shows a short correlation length (6 km) and a strong negative correlation between the slowness and the reflector-depth determination.

The total CPU time for the 10 iterations necessary for the inversion of Paris Basin data set was 20 minutes on a Cray computer. The most time-consuming element of our

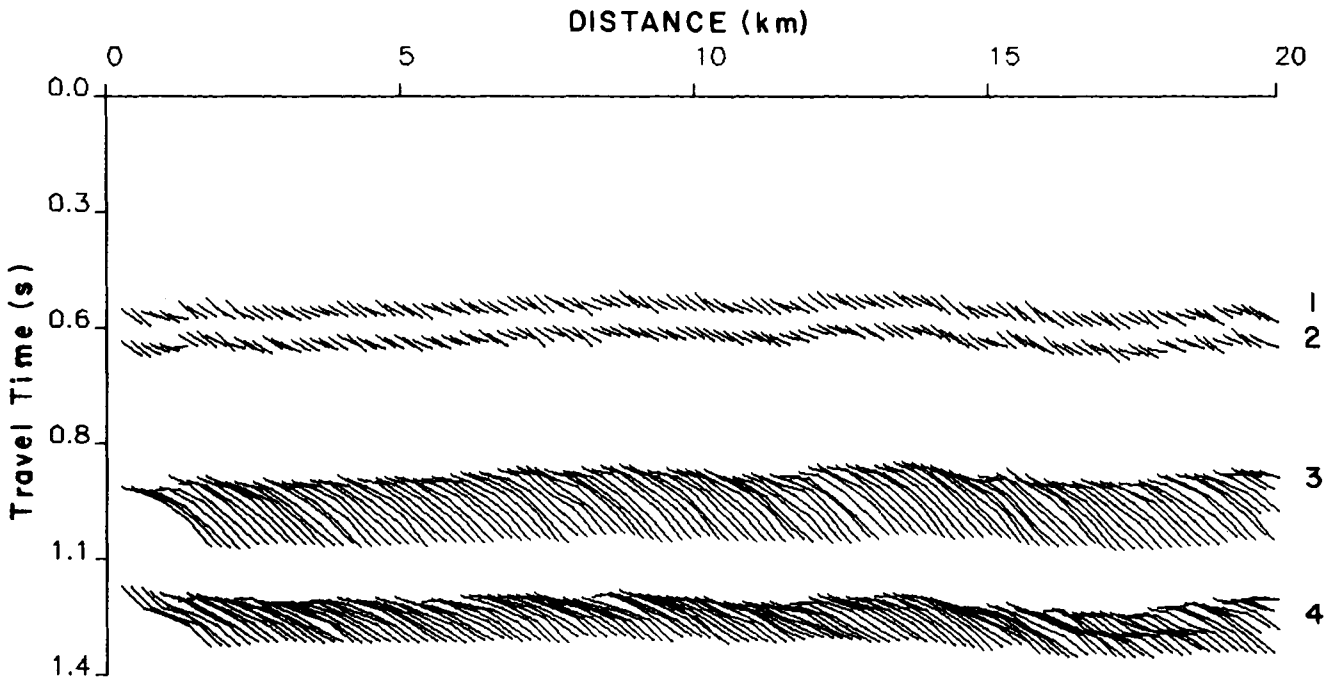


**Figure 8.** Travel-time data set for the tomographic inversion of the Paris Basin Structure. Travel times corresponding to four reflections are shown. Representation is the same as in Fig. 6.

**Table 1.** Model parameterization: number of data points and distance between knots.

| Layer | Slowness | | Interface | |
|---|---|---|---|---|
| 1 | 23 | 1. | 8 | 4. |
| 2 | 8 | 4. | 8 | 4. |
| 3 | 13 | 2. | 13 | 2. |
| 4 | 13 | 2. | 13 | 2. |

program is ray tracing: about 80 per cent of the computing time was spent in solving the two-point ray-tracing problems. Our program uses the Runge–Kutta method for integrating the ray equations. This process is too time-consuming for routine applications. In order to accelerate ray tracing we are presently considering the use of finite-element methods (Langan, Lerche & Cutler 1985;

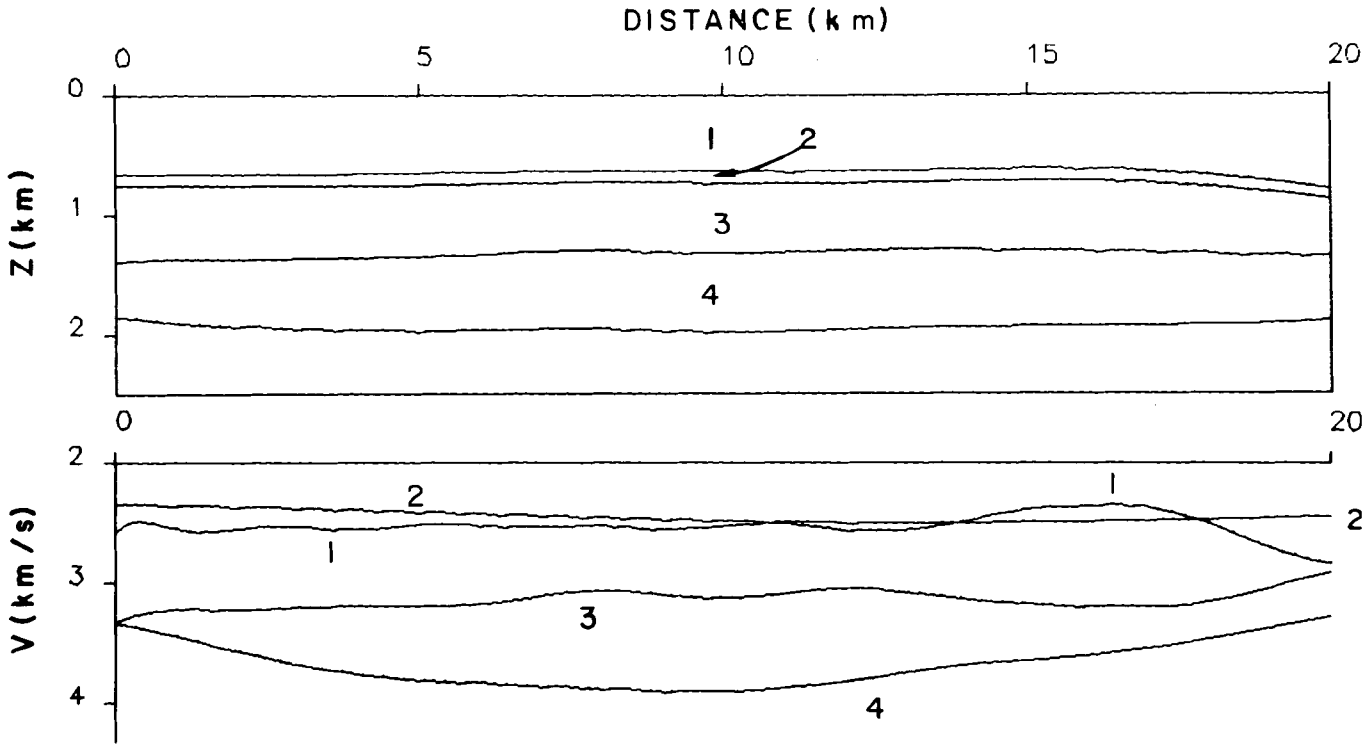


**Figure 9.** Paris Basin model obtained after 10 iterations from the inversion of Paris Basin travel-time data shown in Fig. 8. In each layer, the velocity is modelled as a function of distance $x$ only (lower figure). The upper figure shows the depths of the four reflectors.

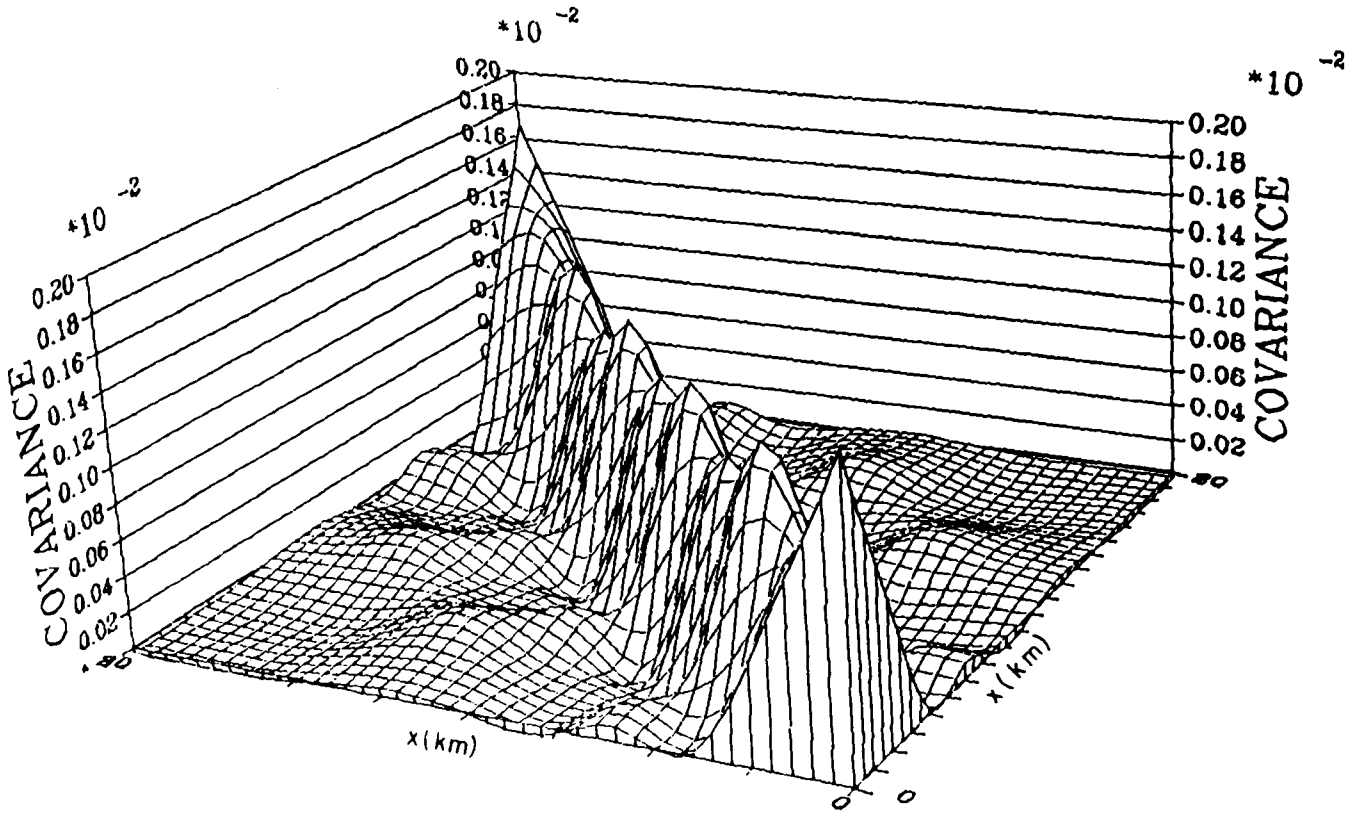


**Figure 10.** Paris Basin model: two point covariances of the fourth interface depth as a function of the abscissae $x_1$ and $x_2$ of the points under consideration.

Chapman 1985; Červený 1987; Virieux *et al.* 1988). Another way to improve the efficiency of the program is to use the information previously obtained by ray tracing from neighbouring sources. One can use the propagator matrix not only to perturb the receiver position, as we do in Section 3, but also to perturb the source position. This technique, sometimes called continuity, will be implemented in future studies.

The linear inversion consumes little time and any method that solves the least-squares problem (21) may be used. Using the singular-value decomposition method (SVD) was not found intractable for this application. For larger data sets, alternative methods, like SIRT (simultaneous iterative reconstruction techniques) or projection techniques (van der Sluis & van der Vorst 1987), could be used to accelerate the procedure.

## 9 CONCLUSION

We proposed an iterative non-linear inversion technique for reflection tomography. A sensitivity study led us to propose a suitable parameterization for this problem: in each layer, the slowness and the interface depth were modelled with functions that vary only in the horizontal direction. This parameterization is adopted because seismic reflection travel-time data have little information on the vertical variations of slowness. For an adequate ratio of the maximum offset to the interface depth, it is possible to separate lateral variations of the slowness and the interface depth, but the poor vertical resolution of slowness may produce a systematic error on the location of interfaces.

The $B$-spline representation of the slowness and the interface depth is well adapted for ray tracing and tomography because of the local property of the basis functions. $B$-splines provide also a fast computation method for travel-time derivatives. Finally, the local property limits the correlation length of the errors.

Using this parameterization, the inverse problem was iteratively solved by a least-squares approach. Because of the inherent poor resolution of the travel-time data near the edges of the model and in order to avoid interlayer oscillations we introduced *a priori* information in the form of penalty functions. This technique that is frequently used in non-linear finite element modelling is very practical and easy to use. The functional relations that we impose on the model are multiplied by a penalty coefficient and added directly to the cost function. When the problem is linearized, these penalty functions appear as additional lines in the Jacobian matrix of the inverse problem. This method, like any other linear inverse, can be cast in the form of a stochastic inverse; in fact, we can identify the corresponding *a priori* covariance matrix and *a priori* solution of the stochastic inverse. For the tomographic problem, where *a priori* information is naturally introduced in the form of non-linear functions of the parameters, the penalty-function approach is easier and faster to implement, and it has a simple geometrical interpretation. *A priori* data may not be sufficient to accelerate convergence of the non-linear inverse problem towards a minimum. A damping factor that forces the search direction to be close to the direction of maximum descent turns out to be a very efficient means of making the cost function diminish.

A synthetic example presented in this paper shows the limitations of reflection tomography. The most important limitations are poor vertical resolution of the velocity and decrease of resolution with the ratio of maximum offset to interface depth. In addition, poor vertical resolution associated with the non-linearity of the inverse problem can introduce errors that are greater than the computed errors obtained in the linear approximation.

As an example of the use of real data, the travel times of four reflections picked on seismic profiles run on the Paris Basin, were inverted using the techniques proposed in this paper. The minimum norm of the residues obtained after several iterations was about 6 ms, less than the predicted error of the reading of travel data.

## ACKNOWLEDGMENTS

Numerous discussions with Michel Leger (IFP) and Jean Virieux (IPG) are gratefully acknowledged. A careful review of the original version of this paper by an anonymous reviewer was extremely useful in producing a manuscript that we hope is more readable and better focused than the original one. The data from the Paris Basin were kindly provided by Compagnie Française des Pétroles. I.P.G. Publication N°1008.

## REFERENCES

Aki, K. & Lee, W. H. K., 1976. Determination of three-dimensional velocity anomalies under a seismic array using first *P* arrival times from local earthquakes, 1, A homogeneous initial model, *J. geophys. Res.*, **81,** 4381–4399.

Aki, K., Christofferson, A. & Husebye, E. S., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *J. geophys. Res.*, **82,** 277–296.

Bishop, T. N., Bube, K. P., Cutler, R. T., Langan, R. T., Love, P. L., Resnick, J. R., Shuey, R. T., Spindler, D. A. & Wyld, H. W., 1985. Tomographic determination of velocity and depth in laterally varying media, *Geophysics,* **50,** 903–923.

Bording, R. P., Gersztenkorn, A., Lines, L. R., Scales, J. A. & Treitel, S., 1987. Applications of seismic travel-time tomography, *Geophys. J. R. astr. Soc.*, **90,** 285–303.

Cěrvený, V., 1985. The application of ray tracing to the numerical modeling of seismic wavefields in complex structures in *Handbook of Geophysical Exploration,* section 1, pp. 1–119, *Seismic Exploration,* Vol. 15A, Geophysical Press, London.

Cěrvený, V., 1987. Ray tracing algorithms in three-dimensional laterally varying layered structures, in *Tomography in Seismology and Exploration Seismics,* ed. Nolet, G., Reidel, Dordrecht.

Chapman, C. H., 1985. Ray theory and its extensions: WKBJ and Maslov seismograms, *J. Geophys.*, **58,** 27–43.

Chiu, S. K. L., Kanasewich, E. R. & Phadke, S., 1986. Three-dimensional determination of structure and velocity by seismic tomography, *Geophysics,* **51,** 1559–1571.

Chiu, S. K. L. & Stewart, R. R., 1987. Tomographic determination of three-dimensional seismic velocity structure using well logs, vertical seismic profiles and surface seismic data, *Geophysics,* **52,** 1085–1098.

de Boor, K., 1978. *A Practical Guide to Splines.* Springer, New York.

Farra, V. & Madariaga, R., 1987. Seismic waveform modeling in heterogeneous media by ray perturbation theory, *J. geophys. Res.*, **92,** 2697–2712.

Fletcher, R., 1980. *Practical Methods of Optimization,* Vol. 1, John Wiley, New York.

Jackson, D. D., 1972. Interpretation of inaccurate, insufficient and inconsistent data, *Geophys. J. R. astr. Soc.*, **28,** 97–110.

Lanczos, C., 1961. *Linear Differential Operators,* Van Nostrand-Reinhold, Princeton, New Jersey.

Langan, R. T., Lerche, I. & Cutler, R. T., 1985. Tracing of rays through heterogeneous media: an accurate and efficient procedure, *Geophysics,* **50,** 1456–1465.

Levenberg, K., 1944. A method for the solution of certain non-linear problems in least-squares, *Q. appl. math.*, **2,** 162–168.

Menke, W., 1984. *Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press, Orlando, Florida.

Nercessian, A., Hirn, A. & Tarantola, A., 1984. Three-dimensional seismic transmission prospecting of the Mont Dore volcano, France, *Geophys. J. R. astr. Soc.*, **76,** 307–315.

Pavlis, G. L. & Booker, J. R., 1980. The mixed discrete-continuous inverse problem: Application to the simultaneous determination of earthquake hypocenters and velocity structure, *J. geophys. Res.*, **85,** 4801–4810.

Popov, M. M., 1969. Free oscillations of a multimirrored oscillator, *Vestnik Leningrad University,* **22,** 42–54.

Spencer, C. & Gubbins, D., 1980. Travel time inversion for simultaneous earthquake location and velocity structure determination in laterally varying media, *Geophys. J. R. astr. Soc.*, **63,** 95–116.

Stork, C. & Clayton, R. W., 1985. *The iterative tomographic and migration reconstruction of seismic images,* Paper presented at the 55th annual SEG Meeting, Washington, D.C.

Stork, C. & Clayton, R. W., 1986. Analysis of the resolution between ambiguous velocity and reflector position for travel time tomography, Paper presented at the 56th annual SEG Meeting, Houston.

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

Tarantola, A., 1987. *Inverse problem theory: Methods for Data Fitting and Model Parameter Estimation,* Elsevier, Amsterdam.

Thomson, C. J. & Gubbins, D., 1982. Three-dimensional lithospheric modelling at Norsar: linearity of the method and amplitude variations from the anomalies, *Geophys. J. R. astr. Soc.*, **71,** 1–36.

Thurber, C. H., 1983. Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, Central California, *J. geophys. Res.*, **88,** 8226–8236.

Van der Sluis, A. & van der Vorst, H. A., 1987. Numerical solution of large, sparse linear algebraic systems arising from tomographic problems, in *Tomography in Seismology and Exploration Seismics,* ed. Nolet, G., Reidel, Dordrecht.

Virieux, J., Farra, V. & Madariaga, R., 1988. Ray tracing for earthquake location in laterally heterogeneous media, Submitted to *J. geophys. Res.*