Non-linear arrival time tomography

Constantinos B. Papazachos[1] **and Guust Nolet**[2]

[1] *Institute of Engineering Seismology and Earthquake Engineering (ITSAK), Foinikas, Thessaloniki, Greece*

[2] *Department of Geological and Geophysical Sciences, Princeton University, Princeton, USA*

Abstract

The use of 1D or pseudo- 3D ray tracing techniques in linearized tomographic problems leads to solutions for which it is difficult to assess the true resolution and error distribution. For this reason, we employ a revised 3D bending algorithm (Moser *et al.*, 1992) and show that it can be used efficiently for a non-linear inversion in a stepwise scheme. Initial paths are determined from graph theory in order to avoid local minima in bending. The importance of 3D ray tracing in inversion studies and the limitations of the standard 1D approach are demonstrated through synthetic examples. The speed of the ray tracing and the simple scaling scheme allow for an implementation in large-scale tomographic problems.

Key words *non-linear inversion – arrival times – graph theory – bending – 3D ray tracing*

1. Introduction

Since the first tomographic results were presented (Aki and Lee, 1976; Aki *et al.*, 1977; Dziewonski *et al.*, 1977), various researchers have studied different aspects of the tomographic problem. These studies have focused on the three main stages of the tomographic algorithm, namely the parameterization of the model, the computation of the matrix elements (the ray tracing algorithm in the case of arrival time tomography) and the solving the final system of equations together with obtaining estimates for the errors and the resolution of the final solution. The first stage significantly affects the latter two, since both the ray tracing technique employed and the inversion algo-

rithm used depend on the way the model is represented. Usually, after linearization of the problem, we need to solve a linear system of the form:

$$r = As \qquad (1.1)$$

For the arrival time tomography, r is the travel time residual vector, A is the appropriate Jacobian matrix and s is the slowness perturbation matrix. Even in the case of a local model representation the inverse problem (solution of eq. (1.1)) cannot be considered overdetermined. A unique solution can only be obtained by assuming additional constraints on our model. If we accept the Bayessian point of view, we do this by imposing an *a priori* covariance matrix for our solution (Tarantola, 1987). This results in defining a «least-norm» solution (Franklin, 1970; Tarantola and Nercessian, 1984; Constable *et al.*, 1987) by minimizing not only the data misfit but also additional constraints like the model norm or the model second derivative norm.

An important aspect of the tomographic problem is the way matrix A is constructed. This is controlled by the ray tracing technique

Mailing address: Dr. Constantinos B. Papazachos, Institute of Engineering Seismology and Earthquake Engineering (ITSAK), Georgikis Sxolis 46, P.O. Box 53, 55102, Foinikas, Thessaloniki, Greece; e-mail: costas@itsak.gr

used in arrival time tomography. Usually this step represents the heaviest computational effort. The traditional approach is either to rely on Fermat's principle and ignore the effect of 3D structure on rays (Aki and Lee, 1976) or employ a fast pseudo-3D ray tracing technique (Thurber and Ellsworth, 1980; Thurber, 1983). In very few cases, a 3D ray tracing technique has been applied in arrival time tomography (Sambridge, 1990). The usual 3D ray tracing approach is to the two-point boundary value problem with shooting or bending (Julian and Gubbins, 1977). Both techniques have drawbacks; shooting techniques as well as other algorithms (*e.g.* Sambridge and Kenett, 1989) fail to converge or converge too slowly in shadow zones where ray theory breaks down. On the other hand, bending frequently fails to find the global travel time minimum, unless an initial ray close to the global minimum is provided as a first guess. However, bending methods have no problem finding rays in shadow zones or even identifying a diffracted wave as the first arrival and therefore annealing the Wielandt effect (Wielandt, 1987) of biasing towards positive velocity anomalies.

In this paper we are concerned with the suppression of errors introduced in the elements of matrix A by inaccurate ray geometries. We address this problem using an efficient but accurate 3D ray tracing technique proposed by Moser *et al.* (1992), which has been shown to have very good convergence and accuracy properties. Global minimization is virtually assured by incorporating initial ray paths derived from graph theory (Moser, 1991). Moreover, we explore the significance of 3D ray tracing for tomographic purposes.

2. Non-linear arrival time inversion using 3D ray tracing

In order to trace rays in a three-dimensional medium we employed a bending algorithm as revisited by Moser *et al.* (1992). In this algorithm, rays are represented by β-splines which are defined by a series of support points. Our model was parameterized by a 3D cartesian grid, where velocities are determined at

each node and the velocity at each point is estimated by trilinear interpolation. Since we use a bending method, shadow zones or the Wielandt effect represent no problem. Details about the efficiency and accuracy of the method and comparisons with other techniques can be found in Moser *et al.* (1992).

As mentioned earlier, the most important problem in bending is that, if our starting ray is far from the global minimum, we may find a secondary arrival. Usually initial paths are calculated in a 1D model which is an average of our 3D structure. A better alternative is to use rays for a 1D model determined by locally averaging our 3D model in a limited area around the source-receiver cross-section (Thurber and Ellsworth, 1980) or use a ray derived by a pseudo-3D ray tracing technique. However, this does not necessarily solve our problem, especially in areas of strong heterogeneities. For this reason we employed initial paths as derived in a 3D medium from graph theory using a recent modification by Moser (1991). Graph theory systematically searches in a network of nodes for the shortest path to a specific point. If we measure the «lengh» of a path by its travel time, we can use graph theory to find minimum travel times from every node of our model to a given point, and therefore estimate seismic ray paths in accordance with Fermat's principle. Each resulting ray consists of line segments connecting different nodes. A limitation of this approach is that the accuracy of graph theory depends on the node density. Since our model consists of nodes between which velocity is estimated by trilinear interpolation, we need a much denser grid for the graph theory in order to simulate a bending ray. This is further illustrated in fig. 1a, where a possible ray geometry (dashed line) and the ray determined from graph theory (solid line) are shown, when the original velocity grid and the grid used in graph theory coincide. In this case, the minimum travel time path is constructed by examining the minimum paths from each point to its eight closest neighbors (which is referred to as forward star in graph theory). When we use paths only to such a small number of neighbors, we impose a specific angle discretization on the directions that
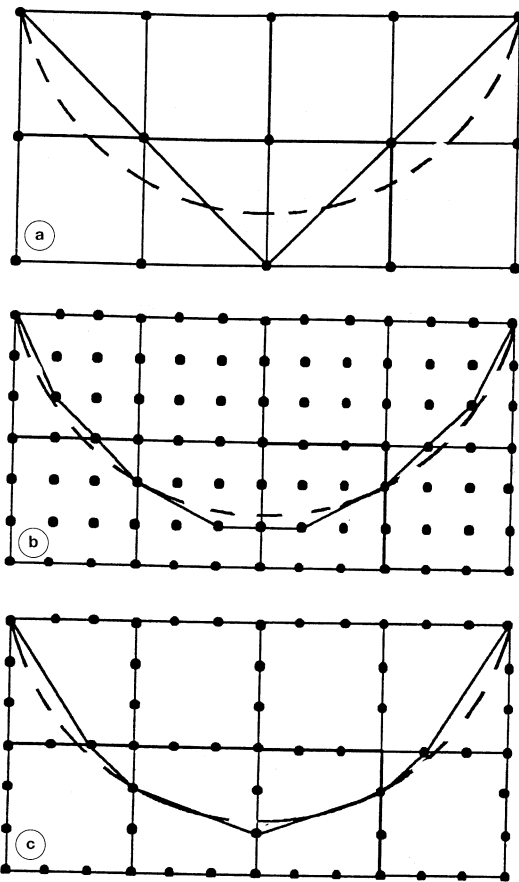
Fig. 1a-c. An example ray (dashed line) and three approximations (solid lines) by graph theory for different velocity-graph theory grid configurations. a) Ray approximation when graph theory is applied using the original velocity grid. The ray derived from graph theory cannot follow the true ray path. b) Same when graph theory uses a grid with a more dense spacing. The graph theory ray is improved, but the number of nodes incorporated increases drastically. c) Same when only nodes lying on the velocity cells are kept. We retain the same accuracy but the number of nodes used is significantly reduced.

a factor r, as shown in fig. 1b, and increasing the total number of nodes by a factor $\simeq r^3$. However, the number of operations involved is proportional to $n \log n$ (Moser, 1991), where n is the total number of nodes, therefore this change would result in $\simeq O(r^3)$ increase of the total CPU time. To reduce this inefficiency, we only keep the nodes that lie on the surface of the velocity-node cells (fig. 1c). This configuration allows the same angle discretization and increases the CPU time by a smaller factor of $3\,r\,(r-1)$, resulting in an almost 80% CPU reduction for a typical value $r = 4$. Unfortunately, because of the intrinsic limitations of graph theory, we cannot completely rely on its results. Further bending of the initial ray provided from graph theory allows an increased accuracy, which could not be achieved by graph theory without a very heavy computational effort.

The combination of initial paths from graph theory and further ray improvement by bending give excellent results. In a series of 2D tests in velocity structures where an analytic solution could be determined we found a very good agreement of the ray paths and an accuracy of 1 in 10^3-10^4 for the travel times. The highest accuracy can be achieved by a careful study of the parameters involved in the two algorithms. For models which involved a number of nodes of the order of a few thousands, an additional improvement of the order of 0.5-4% was achieved by bending, when graph theory with $r = 4$ was used. For bending, there is an optimum number of support points, corresponding to approximately one support point for each cell that the ray penetrates. Below this number, the error starts to increase because we do not have enough points to maintain a flexible ray, whereas above this number errors gradually accumulate because of round-off errors and the increasing difficulty of travel time minimization in a larger model space when we are already close to the minimum. In addition, the CPU time increases with the number of support points involved. The minimization also depends on how close to the minimum the starting ray is, as estimated from the graph theory. This is especially important for an efficient application in 3D arrival time inversion

the ray can take (45° in our example). This is the most serious limitation for the accuracy of graph theory. One possible alternative is to establish a denser grid only for the graph theory calculations, by decreasing the node distance by

studies as the two algorithms are used in a different way: since graph theory provides initial paths from every point in our model to a given location, we can choose to use the receivers as these locations, as we usually have many more sources than stations. Then the initial paths that correspond to events recorded at the receivers can be updated by bending. Hence, we can determine an optimum CPU time (for the same travel time accuracy), by «tuning» the two algorithms, depending on the specific configuration of our inverse problem.

3. Synthetic example for 3D ray tracing

In order to test the significance and efficiency of the 3D ray tracing proposed we show a second synthetic example. The model used was a nine-layered model with 200 virtual seismic events randomly distributed throughout the model. The horizontal and vertical node spacing were fixed at 40 km and 10 km, respectively, and 20 virtual stations were randomly distributed on the top side of the model (sur-

face). The final model consists of 2178 P and S nodes. The 1D background P velocity model is shown in table I and S velocities were calculated using a constant V_P/V_S ratio of 1.76.

The velocity model we used was identical to the 1D model except for layers 3 and 4 where a low velocity square block was put in the center of our model, as shown in fig. 2. The size

Table I. One dimensional model used in the synthetic tests.

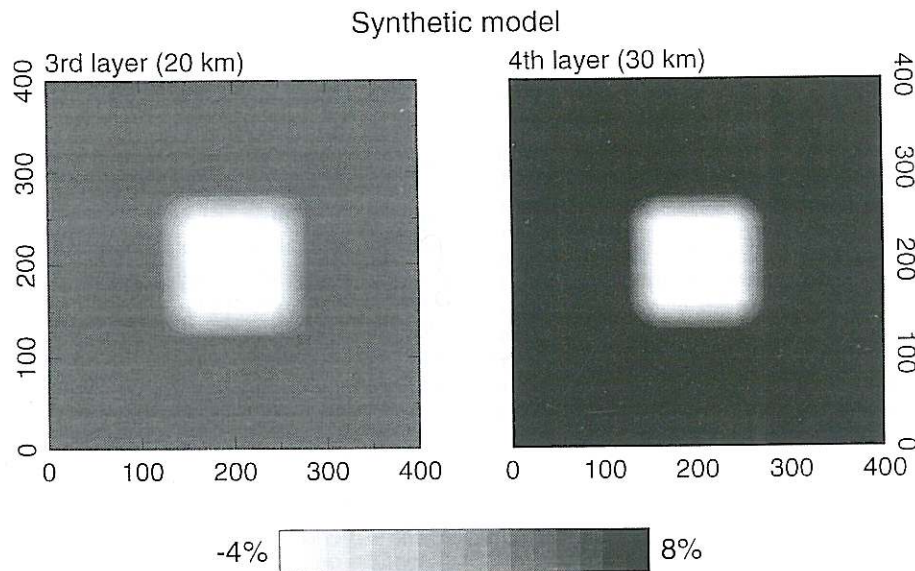| Layer	Depth (km)	P velocity (km/s)
1	0	5.5
2	10	6.0
3	20	6.5
4	30	7.2
5	40	7.9
6	50	7.95
7	60	8.0
8	70	8.05
9	80	8.1

Fig. 2. Velocity model used in the synthetic test to demonstrate the effect of 3D ray tracing. Velocity anomalies are confined in the 3rd (20 km) and 4th (30 km) layers.

of the block is 3×3 grid nodes (80 km $\times$ 80 km) and the corresponding P velocities are 6.1 km/s and 6.8 km/s for layers 3 and 4, respectively. At the same time the velocities outside this block were increased to 6.8 km/s and 7.9 km/s. This velocity structure is trying to simulate a crust which is locally thickening from 30 to 40 km. In total, 7400 synthetic travel times were calculated by 3D ray tracing in the true model. Since we only wanted to study the effect of 3D ray tracing, we added relatively small random errors to the P (0.1 s) and S (0.2 s) travel times. In order to test the effect of differential travel times we used S-P arrivals for three of the stations. Therefore, the P and S problems are not decoupled. Including event perturbations in the inversion will further increase the coupling between the P and the S solutions, although some of the velocity anomalies could creep into the event relocations.

We used the background model (table I) as the starting model and initially inverted eq. (1.1) only for velocity perturbations, using LSQR (Paige and Saunders, 1982). Figures 3 and 4 show the retrieved velocity anomalies for the top six layers (0-50 km) when 1D tracing (one iteration) and 3D ray tracing (after four iterations) are employed, respectively. We limit our result presentation only in the top six layers since practically no anomaly is retrieved for the three bottom layers. The contrast observed between figs. 3 and 4 is quite large: in the area of interest (layers 3 and 4) both the amplitude and the shape of the velocity anomalies are significantly improved by the non-linear inversion. Moreover, the noise introduced in the shallower and deeper layers almost disappears in the 3D ray tracing case. The artifacts in fig. 3 are quite interesting: some of the anomaly of the 4th layer creeps into the 5th layer and the pattern is flipped in the 6th layer to compensate for waves travelling deeper in the model.

The upper two layers show a distribution of mainly positive anomalies which matches perfectly with the station distribution, also shown in fig. 3 with solid circles. The three areas that exhibit negative anomalies correspond to the three stations for which only S-P differential travel times were used. The positive anomalies are mainly due to the fact that 1D rays are slower than real 3D rays. Therefore, 1D rays systematically exhibit negative residuals that are mapped as high velocity anomalies beneath the stations, in a form of station corrections. The initial r.m.s. misfit is 0.767 s and is reduced to 0.145 s after one iteration (1D ray tracing) and to 0.03 s (after 3D ray tracing).

The misfit after one iteration is quite comparable to the misfit expected from the random errors added in the data ($\simeq$ 0.13 s). However, this is not accurate: this misfit is estimated from the linear approximation (eq. (1.1)) and not from the real non-linear problem. This is clearly seen in fig. 5a where we plot the misfit (arbitrary scale) as a function of the number of steps performed in LSQR (circles-solid line). LSQR minimizes the misfit of the linear system by updating the model at each step. Between two iterations we recalculate the derivative matrix A by performing 3D ray tracing and recalculating the misfit for the solution derived from the previous iteration. Hence at these points we can compare the linear and the non-linear misfits. It is quite clear (see the blow-up in fig. 5a) that at the end of iteration 1 we are underestimating the misfit, which is at least 30% higher. This is due to the fact that we are using a derivative matrix, A, which was calculated for the 1D model. Hence, as the model is updated and deviates from the starting solution, the search directions determined by A are inadequate and the linear approximation residuals give us a false image of convergence of the non-linear problem. This behaviour verifies the suggestion (Tarantola, 1987) that, when solving a non-linear problem with a local linear approximation, all resolution and error estimates should incorporate the non-linear residuals. As we get closer to the true model, A is more accurate and this phenomenon practically disappears between iterations 2 and 3. We should also notice that the slope of the misfit curve changes between successive iterations, indicating that the new matrix A provides better search directions for the updated model. If instead of recalculating A we allow LSQR to
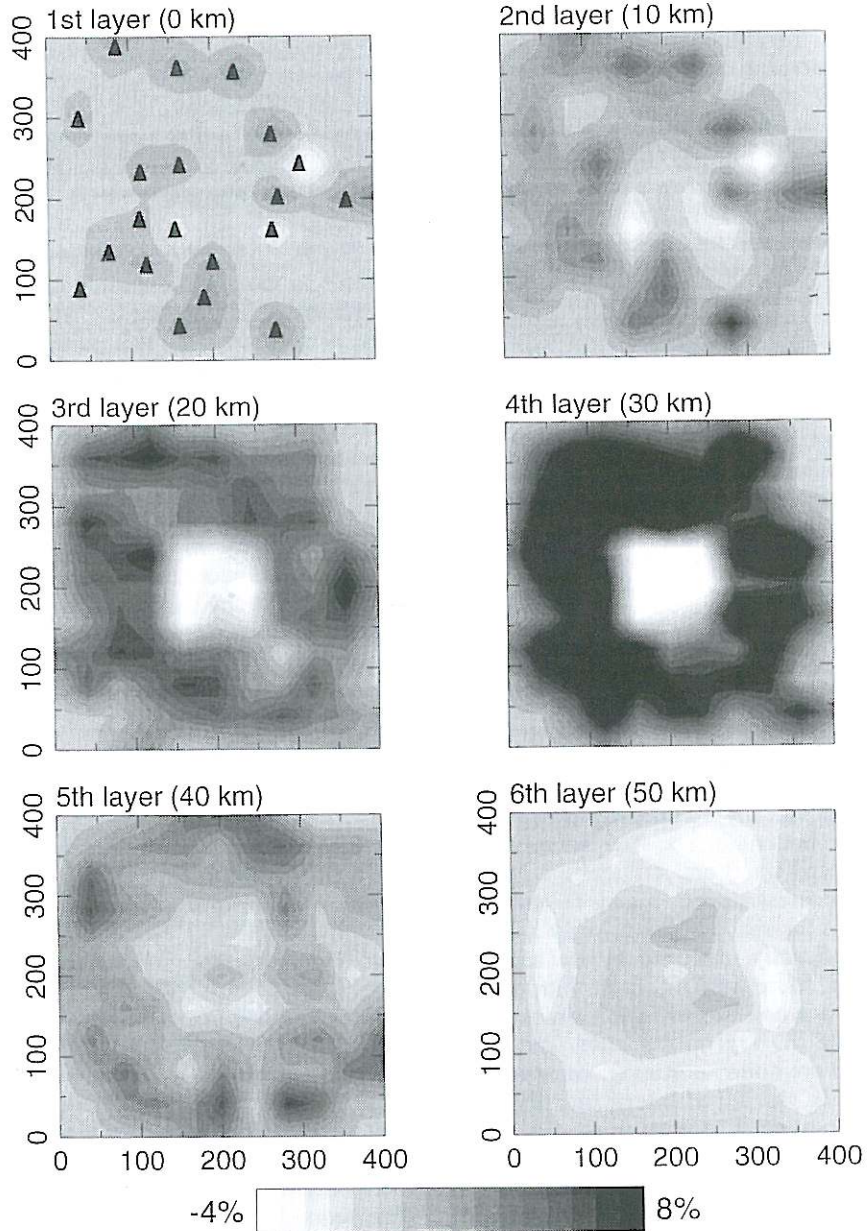
Fig. 3. Results of the inversion when 1D ray tracing is used. Hypocentral relocations were not included. The velocity anomaly is relatively well resolved in the 4th layer and partly in the 3rd layer, although a lot of noise is present. Notice the strong artifacts in the 1st and 2nd layers where no anomaly exists in the true model (see text for explanation). Stations are denoted by solid triangles in the first layer.
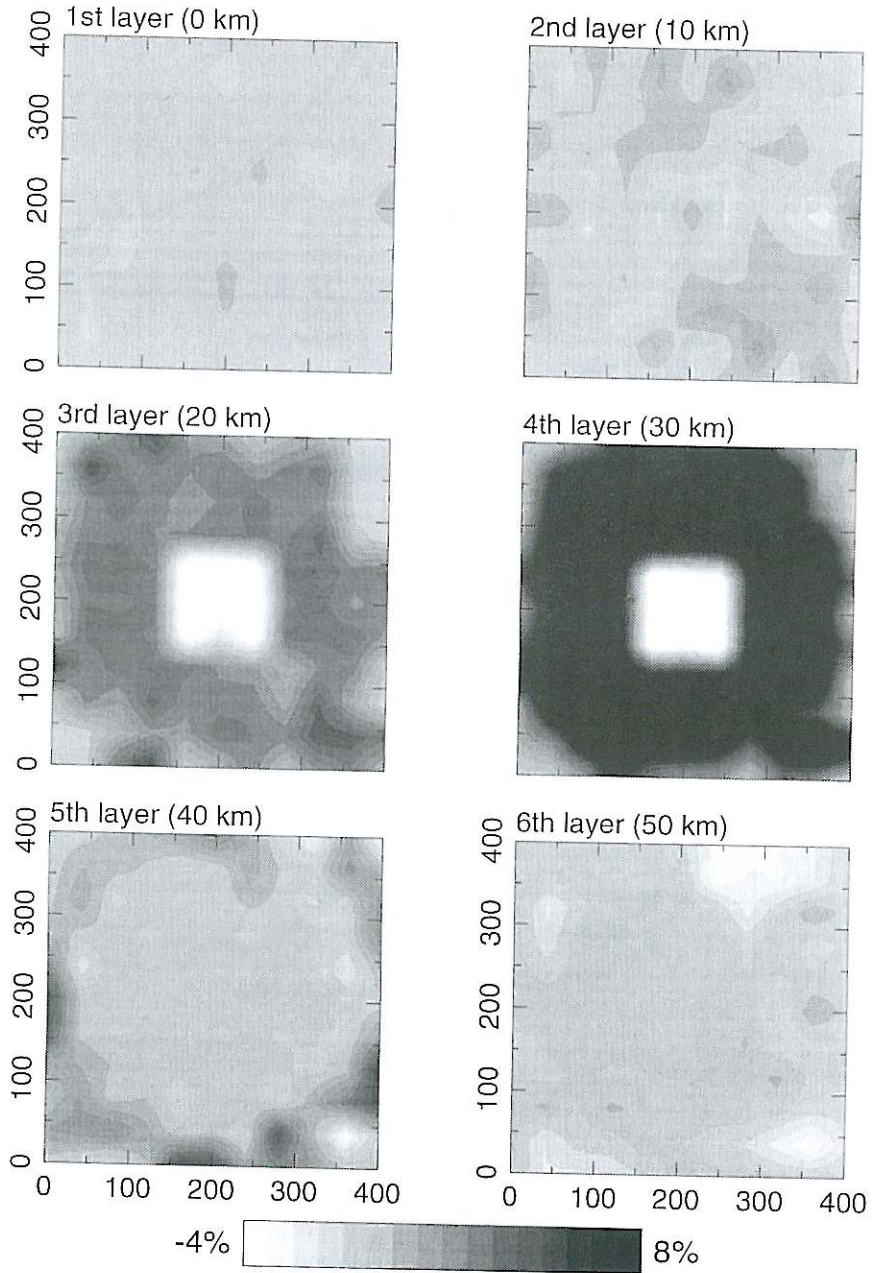
3-D ray tracing

Fig. 4. Same as fig. 3 when 3D ray tracing is used. The velocity anomalies in layers 3 and 4 almost match the corresponding images of fig. 2. Moreover, the artifacts in the rest of the model almost disappear.
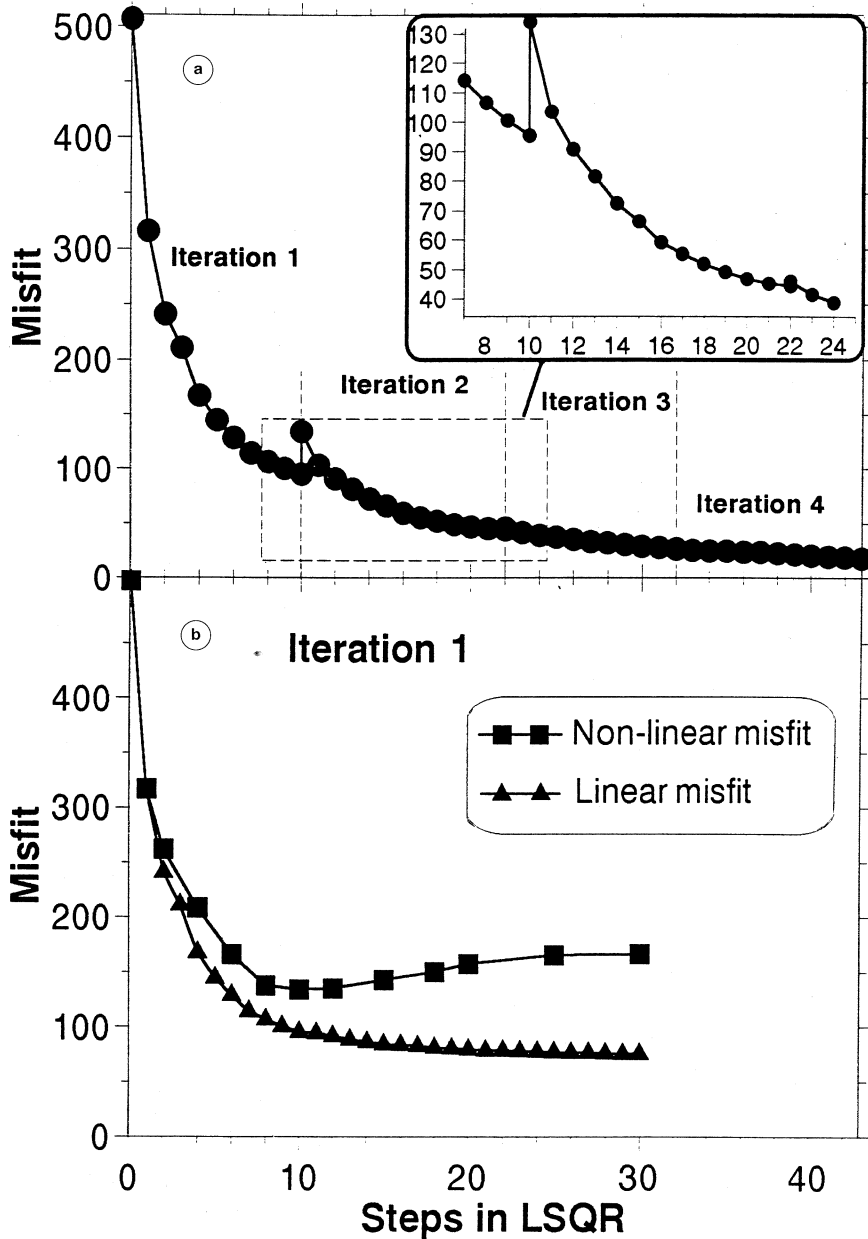
Fig. 5a,b. Breakdown of linearity in travel-time tomography. a) Plot of the travel-time misfit against LSQR steps in different iterations. The transition between the first, second and third iteration is seen in detail in the blow up. At the end of the first iteration the linear misfit underestimates the true misfit by ≃ 30%. The slope of the curve changes between successive iterations as the updated *A* matrix provides new search directions. b) Comparison between the linear and the non-linear (real) misfit for the first iteration. The real misfit decreases less rapidly than its linear approximation and after some LSQR step it gradually increases.

continue up to thirty steps in the first iteration (dashed line), we have a significantly smaller misfit reduction, as we are approaching the least-squares solution.

Although we would prefer to calculate the non-linear misfit at each step, non-linear estimation of the residuals is computationally very inefficient, since it requires full 3D ray tracing for the updated solution. Driven by the discrepancy between the linear and non-linear misfit, we performed these calculations for some solutions of the first iteration. The resulting curve (fig. 5b) verifies the breakdown of linearity: the real misfit (solid squares) decreases less rapidly than the linear approximation (solid triangles) and after a specific step in LSQR (= 10) it gradually increases. After 25 steps we reach an almost constant misfit value, since the model norm does not change significantly. The final misfit (30 LSQR steps) is ≃ 120% larger than our linear estimation, and 60% larger than its minimum value at the tenth iteration. This indicates that for similar non-linear problems, iterative solving of the linear approximation should stop well in advance before convergence. It is not clear how to identify the exact point at which we should stop iterating, since in our case we might consider performing all thirty LSQR steps in the first iteration and obtain an additional 20% misfit reduction. However, it is noteworthy that we made these calculations after we had obtained the final non-linear solution, shown in fig. 4. At the first iteration, we chose to stop LSQR at the specific step (= 10) because the misfit curve shows a prominent kink at this point due to a sudden decrease of the rate of the misfit reduction with LSQR steps. A similar change was observed in the P and S velocity model norms and this was attributed to the fact that matrix A had no more «power» to provide useful search directions to LSQR.

The inversion became more complicated when we simultaneously allowed event relocations and velocity perturbations in our model vector. The corresponding results are shown in figs. 6 and 7. Since a part of the velocity anomaly has crept in the event relocations, both figures look more «noisy» when com-

pared with the corresponding results without event relocations (figs. 3 and 4). However, in the case where 3D ray tracing is used (fig. 7) we clearly have much better resolved anomalies, especially in the area of interest (layers 3 and 4) compared with the 1D results (fig. 6). This is further demonstrated in fig. 8a-c where cross-sections of different models are presented along a specific ray path. Figure 8a shows our initial model and fig. 8b,c presents the recovered model when 1D and 3D ray tracing is employed, respectively. Again our final model is very close to the true one. The solid line in fig. 8c represents the final ray in the 3D model and the dashed line represents the original ray in the 1D model. Since in the 3D case the surface projection of the ray deviates more than 5 km from the great circle plane, we present its projection on the cross-section. The two rays have quite different paths and the 3D ray is faster than the 1D one. Therefore, when 1D ray tracing is used a systematic bias is introduced towards higher velocities. Since, in the 1D case, different rays will try to map these positive residuals in different places, the most convenient way for LSQR to explain them in a minimum norm way seems to be by putting higher velocities in the cells right beneath the stations, as observed in figs. 3 and 6.

Incorporation of 3D ray tracing in tomography is traditionally prohibited by computational inefficiency. However, the proposed method is quite efficient: on a SunSparc 10-Model 20, a regional model (dimension ≃ 1000×1000×200 km^3) with more than 12 000 P and S nodes required an average CPU time of 1.5 s per ray. An additional CPU time of 10 min is necessary for initial ray calculations (both P and S) for each station from graph theory. By distributing the computations on eight CPU's, a regional tomographic problem including 10^5 arrivals could be solved (Papazachos and Nolet, 1997) in ≃ ten hours. The use of supercomputing and efficient programming could make the application of 3D ray tracing and non-linear arrival time inversion in global problems with $O(10^6)$ arrivals a feasible target.
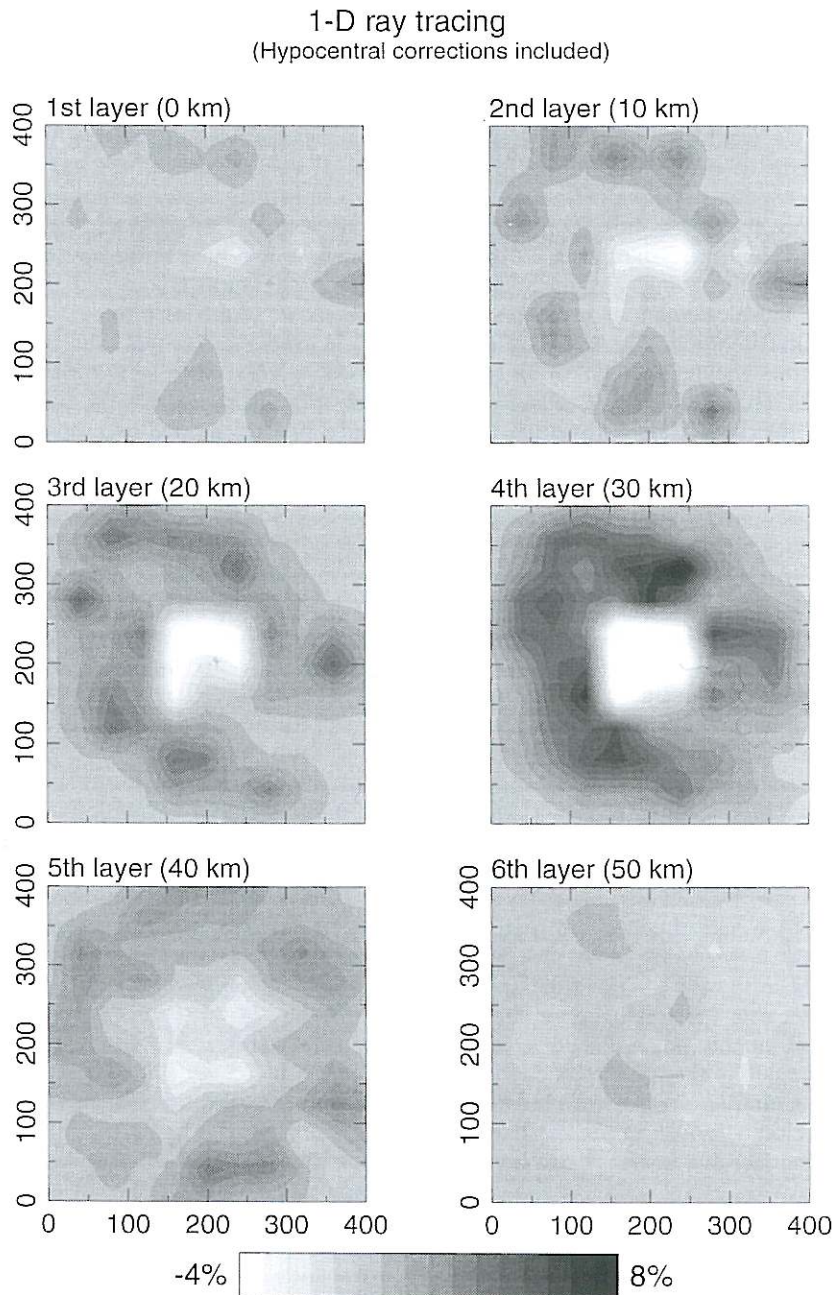
1-D ray tracing
(Hypocentral corrections included)

Fig. 6. Same as fig. 3 when hypocentral relocations are included in the inversion. Notice the amplification of the artifacts and the smaller amplitudes of the recovered anomalies since some «energy» has crept into hypocentral relocations.
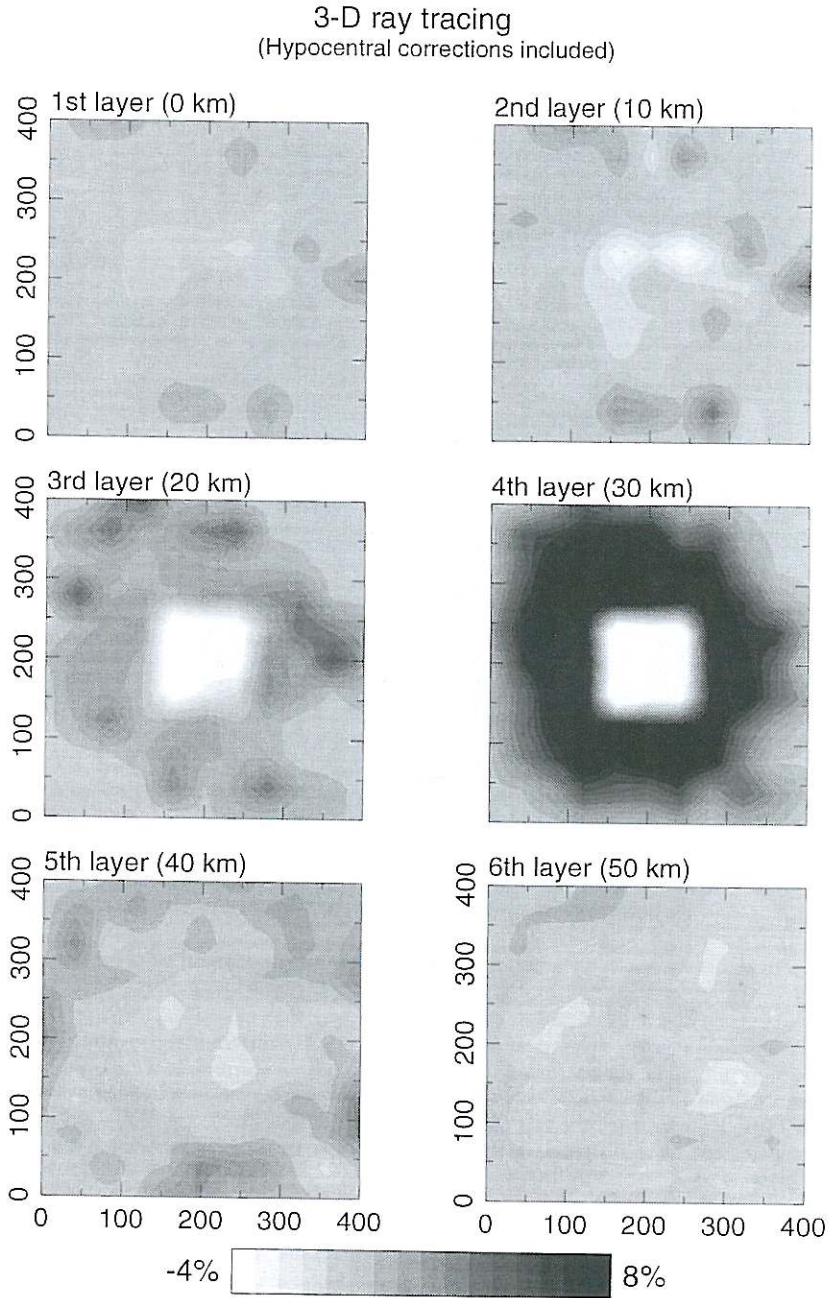
Fig. 7. Same as fig. 4 when 3D ray tracing is used. Although the introduction of hypocentral corrections reduces the amplitude of the anomalies observed, we still have a much better resolution in the area of interest (3rd and 4th layer).
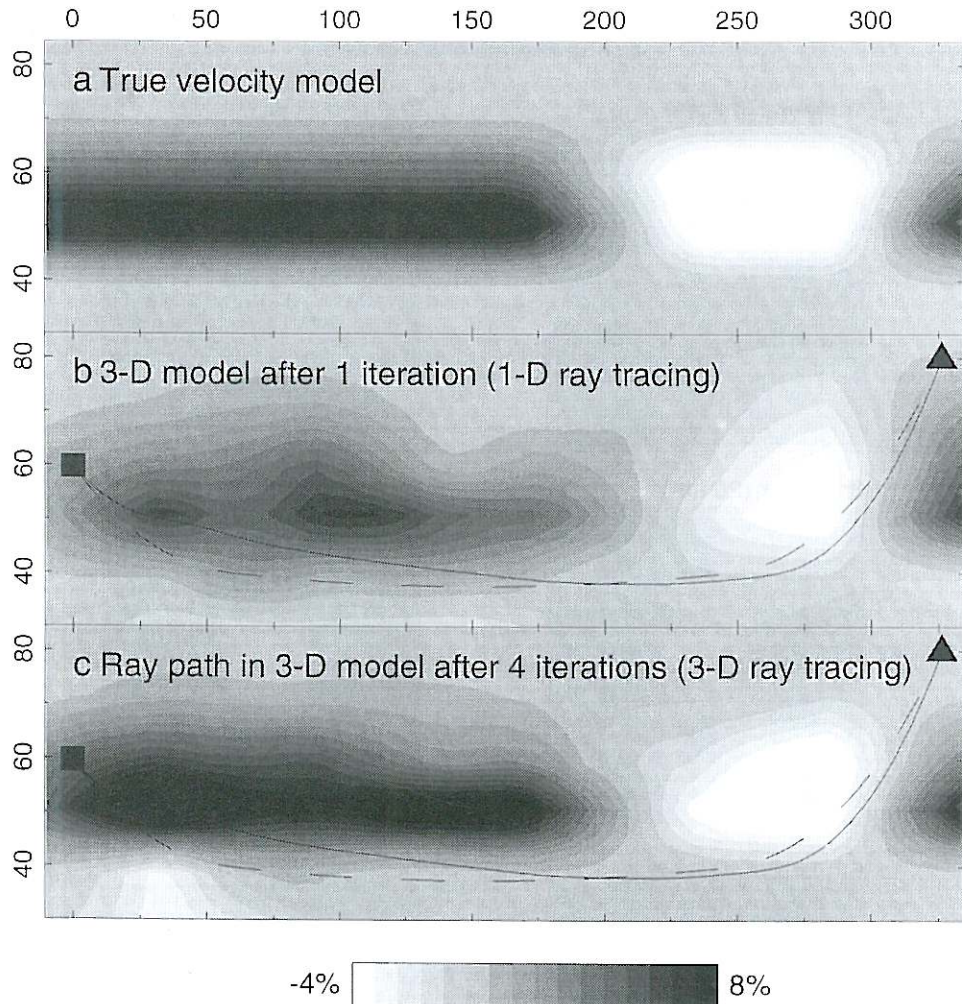
Fig. 8a-c. Cross-sections along a specific source-receiver path for different reconstructed models. a) Input model. b) Inversion result when 1D ray tracing is used. No additional constraints were used in the inversion of the linear system. c) Same as (b) when 3D ray tracing is used. A significant improvement of the recovered image is observed. In both (b) and (c) the ray paths for the 1D (dashed line) and 3D (solid line) case are also plotted. The two paths are quite different. Moreover, the 3D ray deviates more than 5 km from the vertical source-receiver plane.

4. Conclusions

The results presented in this paper focus on the effects of non-linearity on the inversion of arrival times. The problem is addressed by introducing three-dimensional ray tracing in the inversion and a significant improvement of to-mographic images is demonstrated through a synthetic example. The estimation of a global minimum in our ray path calculations is assured by incorporating initial paths from graph theory. The performed tests suggest that 3D ray tracing cannot be considered a small second-order effect in tomography. The limita-

tions of linearity as exhibited in fig. 6 imply that the proposed approach is necessary if we want to obtain reliable quantitative results. Moreover, as the effects of the starting solution are important and because iterative methods rely on the misfit of the linear approximation, we should probably perform a large number of iterations and update our solution very slowly. This approach avoids introducing large solution perturbations based on linear approximations far from the global minimum, which might introduce «phantom» anomalies in our final solution.

Acknowledgements

REFERENCES

AKI, K. and W.H.K. LEE (1976): Determination of three-dimensional velocity anomalies under a seismis array using first *P* arrival times from local earthquakes. 1. A homogeneous initial model, *J. Geophys. Res.*, **81**, 4381-4399.

AKI, K., A. CHRISTOFFERSON and E.S. HUSEBYE (1977): Determination of the three-dimensional seismic structure of the lithosphere, *J. Geophys. Res.*, **82**, 277-296.

CONSTABLE, S.C., R.L. PARKER and C. CONSTABLE (1987): Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, **52**, 289-300.

DZIEWONSKY, A.M., B.H. HAGER and R.J. O'CONNELL (1977): Large scale heterogeneities in the lower mantle, *J. Geophys. Res.*, **82**, 239-255.

FRANKLIN, J.N. (1970): Well-posed stochastic extensions of ill-posed linear problems, *J. Math. Anal. Appl.*, **31**, 682-716.

JULIAN, B.R. and D. GUBBINS (1977): Three-dimensional seismic ray tracing, *J. Geophys. Res.*, **43**, 95-113.

MOSER, T.J. (1991): Shortest path calculation of seismic rays, *Geophysics*, **56**, 59-67.

MOSER, T.J., G. NOLET and G. SNIEDER (1992): Ray bending revisited, *Bull. Seismol. Soc. Am.*, **82**, 259-289.

PAIGE, C.C. and M.A. SAUNDERS (1982): LSQR: an algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw.*, **8**, 43-71.

PAPAZACHOS, C.B. and G. NOLET (1997): *P* and *S* deep velocity structure of the Hellenic area obtained by robust non-linear inversion of travel times, *J. Geophys. Res.*, **102** (in press).

SAMBRIDGE, M.S. (1990): Non-linear arrival time inversion: constraining velocity anomalies by seeking smooth models in 3D, *Geophys. J. Int.*, **102**, 653-677.

SAMBRIDGE, M.S. and B.L.N. KENNETT (1989): Boundary value ray tracing in heterogeneous media: a simple and versatile algorithm, *Geophys. J. Int.*, **101**, 157-168.

TARANTOLA, A. (1987): *Inverse Problem Theory* (Elsevier, Amsterdam), pp. 613.

TARANTOLA, A. and A. NERCESSIAN (1984): Three-dimensional inversion without blocks, *Geophys. J. R. Astr. Soc.*, **79**, 299-306.

THURBER, C.H. (1983): Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, Central California, *J. Geophys. Res.*, **88**, 8226-8236.

THURBER, C.H. and W.L. ELLSWORTH (1980): Rapid solution of ray tracing problems in heterogeneous media, *Bull. Seismol. Soc. Am.*, **70**, 1137-1148.

WIELANDT, E. (1987): On the validity of the ray approximation for interpreting delay times, in *Seismic Tomography*, edited by G. NOLET (Reidel, Dordrecht), 85-98.