

Pre-selection of seismic rays as a possible method to improve the inverse problem solution

S. Tinti and S. Ugolini

Università di Bologna, Dipartimento di Fisica, Settore di Geofisica, V. le Bertè Pichat 8, 40127 Bologna, Italy

Accepted 1989 December 19. Received 1989 December 12; in original form 1989 January 12

SUMMARY

The paper is a contribution to the study of the instabilities associated with traveltimes data inversion in tomographic seismology. It has been found that there are circumstances in which instabilities may be mitigated by reducing in a proper way the number of seismic paths taken into account for the inversion. In order to choose such a set of rays, it has been found useful to apply a pre-selection algorithm which, starting from the set of all available seismic paths, gives a new configuration of rays by minimizing the configuration score. The score is a scalar whose value depends upon some geometrical characteristics of the ray pattern, such as the local ray density and the azimuthal distribution. The method is presently applied to 2-D models where the region crossed by the seismic rays is partitioned into rectangular cells, which is a feature not unusual in tomographic investigations. Numerical experiments of inversion are carried out in order to show the 'goodness' of the new algorithm. Comparison with solutions computed without pre-selecting the seismic rays shows the advantage of applying our algorithm for the considered cases. As a measure of the performance, the magnitude as well as the space distribution of the discrepancy between the computed field and the true solution is used in addition to a number of other suitable indicators.

Key words: borehole inversion, inverse problem, ray selection, 2-D tomographic images.

INTRODUCTION

After the first systematic exploration of the mathematical structure of the inverse problem in geophysics made by Backus (1970a, b, c), seismic tomographic studies received a great impulse from the basic application by Aki, Christofferson & Husebye (1977). Now inversion of traveltimes data is one of the most powerful and most widely used tools to investigate the internal structure of the Earth. A major problem of the method, however, resides in the reliability evaluation of the tomographic maps obtained. Indeed, it is recognized that tomographic images of the Earth's interior are often distorted by disturbing effects whose influence is difficult to assess conveniently and to separate properly from the real slowness anomalies that are to be determined. Sources of such artificial heterogeneities are the insufficient illumination of the region by the seismic rays, the errors in the experimental data and the mathematical and numerical approximations involved in the computation procedures. Ray illumination is one of the most crucial points for the success of a tomographic

reconstruction: indeed, poor illumination may give rise to an underdetermined inversion problem even though we use many more data than unknowns. In practical cases slowness inversion is often performed in regions with uneven ray coverage or including sub-regions predominantly crossed by quasi-parallel ray paths, with the final consequence that the resulting tomographic mapping contains areas that are both well and poorly resolved by the available data. In this paper we emphasize the point of view that the image of the Earth's interior obtained by traveltimes inversion shows a significant dependence upon some relevant geometric properties of the ray distribution and that there are cases for which the solution may be improved by carrying out a ray selection prior to accomplishing the inversion. We will in fact show that in some circumstances a better reconstruction of the slowness field can even be computed by making use of a smaller number of data than those originally available, provided that an 'optimal' configuration of the seismic rays is used. It is worth stressing here that our pre-selection procedure has nothing to do with the selection of the seismic rays, based on the traveltimes data errors, which consists of

the elimination of outliers, i.e. rays somehow found to be affected by gross experimental errors: indeed, in the following we will always assume that such an elimination has already been performed. The theoretical criteria on which our selection is based aim mainly at overcoming the problems arising from an unsatisfactory space distribution of the seismic rays. Indeed, we think that a correct approach to the geophysical inverse problem should primarily take into account the peculiar non-uniformity of the data. In other words, it is more appropriate to operate chiefly on the data space rather than on the model space, since the effects can be more easily estimated and interpreted.

THE SCOPE OF SEISMIC RAYS PRE-SELECTION

We consider the problem of determining the slowness field $s(\mathbf{r})$ of the seismic waves, \mathbf{r} being the position vector within a 2-D region of the Earth, starting from a data set consisting of N traveltimes observations T_i ($i = 1, \dots, N$). The traveltimes T_i associated with the ray i assumes the simple form

$$T_i = \int_{C_i} s(\mathbf{r}) da, \quad (1)$$

where C_i is the ray path i through the Earth's region and da is the infinitesimal element of the ray path. The above expression is intrinsically non-linear in the unknown function s , since both the integrand and the integration path depend on the slowness; however, it can be linearized by introducing the reference model $s^0(\mathbf{r})$, which allows the computation of the traveltimes T_i^0 along the ray path C_i^0 . In terms of the delay time vector $d_i = T_i - T_i^0$ and of slowness perturbation $m(\mathbf{r}) = s(\mathbf{r}) - s^0(\mathbf{r})$, it is then usual to derive the linear system

$$d_i = \int_{C_i^0} m(\mathbf{r}) da. \quad (2)$$

In equation (2) the integration path may be taken as C_i^0 , since this introduces only a second-order influence on the delay time vector in virtue of Fermat's principle (see Nolet 1984; Tarantola 1987). The discretization of the system may be achieved by dividing the region into a set of M regular cells, in each of which the perturbation vector is approximated by a constant value. It follows that

$$d_i = \sum_{j=1}^M G_{ij} m_j \quad (i = 1, \dots, N), \quad (3)$$

where the matrix element G_{ij} represents the path length of the ray i intersecting the cell j . The objective of the analysis is therefore that of finding a suitable vector \mathbf{m} satisfying the linear system (3), on the base of which the tomographic image of the region crossed by the seismic rays can be drawn. In practical applications, however, the system may be ill-conditioned and problems of solution instability frequently arise (see Tarantola 1987). One of the most important sources of such instabilities, after data errors obviously, is the poor illumination of the region under investigation by the seismic rays. Of course, the cell illumination depends upon the combined geometry of the ray pattern and of the cell discretization. Dividing the region

into a smaller number of larger cells and allowing cells unequal in size and in shape may improve the illumination (see e.g. Hirahara & Mikumo 1980), but the former certainly degrades the resolution, while the latter implies some additional complexities in the ray computation, especially when the problem involves many blocks and many data. Therefore, the common strategy is to use a grid consisting of small cells of uniform size, to accept that some of the cells are poorly illuminated by the available data set and to adopt inversion algorithms able to control the instabilities arising, with the consequence, however, that the reliability of the resulting tomographic mapping is difficult to ascertain and the true slowness anomalies may be masked by numerical artifacts. The problem of choosing a cell partition, optimizing cell illumination versus field resolution, is still quite open and is mainly dealt with on an empirical basis *a posteriori* after computing the solution of the linear system (3). One of the main difficulties in evaluating how good is the illumination of a given cell is that illumination seems to depend not only upon local properties, but also on geometrical characteristics of the global configuration of the seismic rays. Let us now illustrate a case that may serve to elucidate better this point.

The configuration depicted in Fig. 1(a) is deliberately of extreme simplicity, consisting of 13 seismic rays spanning a region sub-divided into nine square cells. The distribution of rays is rather unequal, since the paths originating uniformly from sources equally spaced on the left boundary of the field all terminate on the central and right cells of the top layer. General ray orientation is from bottom left to top right. Let us assume the rigid constraint that no more rays can be added to improve the illumination of the whole region. The only practical alternative is therefore to try to search for a sub-region that is better illuminated by a subset of the

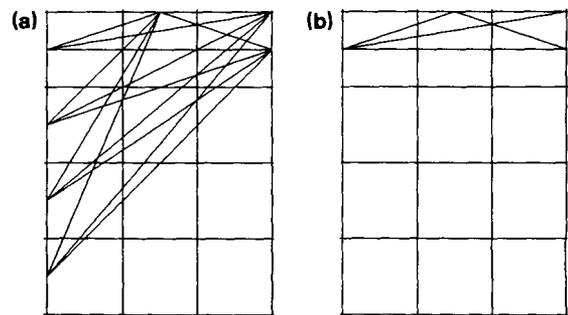


Figure 1. (a) Example of 2-D configuration of seismic rays useful for a preliminary illustration of the ray pre-selection; 13 seismic paths, crossing nine equal square cells, are mostly concentrated in the upper layer and particularly in the cell in the top corner on the right, the only one with two receivers. Observe that the inversion results are quite sensitive to the position of the upper receiver, which is placed quite close to, but not exactly in, the corner. The configuration score (see the following section) is $S = 0.80$. (b) Ray configuration obtained after removal of nine rays and six cells. The remaining three cells may be interpreted as the core of the original configuration. It exhibits a more balanced ray illumination, though the ray density in all cells has been drastically reduced. In spite of ray rarefaction, however, the configuration score is now much smaller, being $S = 0.71$. This configuration is shown to give a better tomographic image than the initial ray pattern (see results in Table 1).

seismic paths. In the present case it can be shown that such a sub-region exists; it is the three-cell set covered by the ray configuration illustrated in Fig. 1(b), formed by only four of the original rays. Though surprising and paradoxical at first sight, especially because the global ray density is considerably lower, nonetheless this distribution ensures a better illumination of the top layer and a more reliable slowness solution there. This point has been proved by means of many numerical experiments, one of which is given in Table 1. The true slowness s_t is assumed to be uniform throughout the region (0.2 s km^{-1}) as well as the reference slowness s_0 which is taken to be in excess of the true field by 20 per cent. No noise is added to the data so that the delay time vector d_i is essentially determined by the discrepancy between the true and the reference fields. Inversion computed for the whole region by means of the Singular Value Decomposition (SVD) algorithm is compared to the inversion restricted to the upper layer. The indicators chosen to facilitate the comparison are the normalized misfit μ between the true and the calculated slowness and the mean resolution width W , which is a measure of the average independence of the parameters in the model space (both indicators will be given a rigorous

definition later on). It may be observed that the global inversion results are less satisfactory, having a normalized misfit value almost double that computed by inverting the first arrivals in the sole upper sub-region. Remarkably, the global inversion cannot attain the same level of accuracy even if the slowness irregularities are smoothed out by cutting off the smallest eigenvalues. The best global solution is obtained by discarding one eigenvalue, but it is manifestly worse with regard to both indicators (larger normalized misfit and larger mean resolution width). It is also interesting to observe that the slowness discrepancies computed for the original configuration are not negligible at all depths; even in the upper layer where the ray coverage is substantial, the calculated solution is less satisfactory than that computed by means of only the four selected rays illustrated in Fig. 1(b). This is shown in the central column of Table 1, where the global inversion is used to compute both normalized misfit and mean resolution width for the sole upper layer. Equivalent results were obtained by taking different reference slowness field and by using different inversion algorithms such as the LSQR method, which is a variant of the conjugate gradient method (Paige & Saunders 1982) used extensively in seismic tomographic problems by Nolet and his group (see Nolet 1985; Spakman & Nolet 1988, Spakman 1988). Therefore, it may be concluded that in the present case the initial configuration includes a core, consisting of the three upper cells, whose tomographic image is reconstructed better by using the four paths entirely comprised within this region. All other rays do not improve the illumination of the core, but contrarily produce a degradation of the solution. This is an important consideration that suggests to us the usefulness of having a means whereby one could evaluate *a priori*, that is prior to entering the matrix inversion phase, which one of two different ray configurations is likely to provide a better tomographic map. If a method of comparison were available, it could be applied to solve a large number of problems. For example, it could be helpful to study the influence of the cell partition on the final solution, i.e. to decide the dimension and the position of the cells, or to optimize the disposition of the sources and receivers in a tomographic seismic exploration, or, as in the example given above, to find a subset of the available seismic rays providing a better illumination to a core region. In our paper we have investigated this possible method, restricting our attention to the last mentioned application. In order to compare the original configuration with the one consisting of the sole rays crossing the core, we attribute a score to each ray distribution. According to the definition given in the next section and in the Appendix, the smaller the score, the better the ray pattern, and the better the results likely to be expected from the inversion. As to the case illustrated, the score value for the whole configuration is $S = 0.80$ versus the value $S = 0.71$ of the core ray distribution. From the above example we learned that the existence of a core for the initial configuration of rays is advantageous, since limiting the inversion to this sub-region enables us both to compute a more reliable tomographic image of the core and to avoid its contamination by instabilities related to the poorer illumination of the remainder of the region. We should point out, however, that (i) the core does not always exist, since there are cases where the initial configuration is not

Table 1. Results for the configurations displayed in Figs 1(a) and (b). Normalized misfit μ , defined in equation (19), and mean resolution width W , defined with the aid of equation (23), are given in each column of the table as a function of the number n of the largest singular values used to compute the solution. SVD method is applied for the inversion and all contributions associated with the smallest $p - n$ singular values are cut off, according to Wiggins (1972). The true field is homogeneous as well as the guess reference field, which exceeds the true values by 20 per cent. The data are noise-free. The second column reports indicators computed for the complete configuration illustrated in Fig. 1(a). The results of the same inversion are used to compute the indicator values given in the next column to the right. Here the values are obtained by averaging over the sole upper layer, rather than over the whole region. Differences between the two columns are remarkable only for the resolution width, which assumes reasonable values slightly larger than 1 in the upper cells. The rightmost column refers to the core configuration of Fig. 1(b), including three cells, and consequently it contemplates only three possible entries. Observe that the case with all singular values retained (the one on the top of the right column) shows a misfit of 0.42 per cent and a perfect resolution ($W = 0.0$). According to our indicators, it gives the best image of the true homogeneous field in the upper layer. The same accuracy cannot be achieved by inverting the complete set of data, for any level of regularization. This was also found for regularization based on damped least-squares and for LSQR inversion.

n	GLOBAL INVERSION		CORE SOLUTION FROM GLOBAL INVERSION		CORE INVERSION	
	μ (10^{-2})	W	μ (10^{-2})	W	μ (10^{-2})	W
9	0.74	0.0	0.70	0.0		
8	0.50	1.2	0.50	1.1		
7	0.58	1.3	0.50	1.1		
6	0.92	1.3	0.73	1.1		
5	1.75	1.4	0.66	1.2		
4	1.70	1.5	0.72	1.3		
3	3.00	1.5	2.64	1.4	0.42	0.0
2	3.00	1.6	2.64	1.5	1.88	0.9
1	16.0	1.8	12.1	1.8	1.88	1.1

improvable and that (ii) the core illumination, though better than the whole region, may be nonetheless insufficient to assure a successful inversion and a reliable tomographic reconstruction. If the extent of the core is slightly smaller than the originally considered region, then limiting the seismic tomographic inversion to the core is not a big sacrifice, being generally compensated by the improvement of the results. However, if the core is small, naturally the problem arises of computing the slowness in the whole field. We have not systematically investigated the numerical techniques most suitable for extending the solution from the core to the entire region, but this will certainly be a good subject for a further study. As for the example treated in this section, we have performed a two-step inversion. We first computed the slowness in the core, by solving a three-unknown problem, then we calculated the solution in the six off-core cells by taking into account the slowness values already computed in the core. From a mathematical viewpoint, the original 13×9 matrix \mathbf{G} in equation (3) results in being partitioned into a set of sub-matrices of which the one 4×3 concerning the core may be independently inverted. Then, applying tomography to the off-core region implies that a 9×6 sub-matrix be inverted on a local delay time vector whose components are adequately modified depending on the results of the previous inversion. The computed normalized misfit μ for the whole region is 0.57×10^{-2} , which shows that the solution obtained is more satisfactory than that calculated by the one-step SVD inversion on the complete set of rays (see upper row in Table 1). In the following sections we will focus on the algorithm enabling us to find the configuration core and on numerical experiments in which inversion in the core gives better results than the inversion on the whole region crossed by the seismic paths.

THE PRE-SELECTION ALGORITHM

The pre-selection algorithm is based on the idea that the 'goodness' of the solution is strongly dependent on some relevant properties of the global geometrical configuration of the seismic ray paths. For a given cell partition of the seismic region and for a given set of observations and paths, the algorithm essentially provides a way to determine a sub-configuration of the rays likely to produce a more accurate solution. Before giving the detailed description of the procedure, it is convenient to introduce the quantities that we use as descriptors of a given set of rays, i.e. the mean density D , the dispersion index σ and the anisotropy index α . Let us define the density δ_j of the cell as the total sum of the seismic paths intersecting the cell, properly normalized over the characteristic cell length $l_j = A_j^{1/2}$, A_j being the 2-D cell area. The set formed by the densities $\delta_j \neq 0$ of the p cells actually crossed by at least one ray ($p \leq M$, M being the total number of cells) may be viewed as the experimental sample of a density distribution for which we may compute the mean and the deviation around the mean; these quantities are taken as the mean density and the dispersion index of the ray configuration, respectively, and are calculated by means of the formulae

$$D = \frac{\sum_{j=1}^p \delta_j w_j}{\sum_{j=1}^p w_j}, \quad (4a)$$

$$\sigma = \frac{\sum_{j=1}^p |D - \delta_j| w_j}{\sum_{j=1}^p w_j} \quad (4b)$$

where the w_j s form a suitable weighting scheme discussed in the Appendix. The mean density D may be regarded as a measure of the amount of information that, on average, the data carry for any cell of the region under study, while the dispersion index σ points out how unequally such information is distributed among the cells. In order to introduce the anisotropy index of the ray configuration designated as α , we divide the cell j into q equal azimuthal sectors and consider the ray densities $\delta_{j,k}$ ($k = 1, \dots, q$) for each sector. If the mean density of the sectors in the cell j is denoted by $\bar{\delta}_j$, i.e. if

$$\bar{\delta}_j = \frac{1}{q} \sum_{k=1}^q \delta_{j,k} = \delta_j / q \quad (5a)$$

then the quantity α_j given by

$$\alpha_j = \sqrt{\frac{1}{q} \sum_{k=1}^q (\delta_{j,k} - \bar{\delta}_j)^2} \quad (5b)$$

may be taken as the anisotropy coefficient of the cell j ; indeed the smaller is α_j , the better will be considered the azimuthal ray coverage of this cell. The anisotropy coefficient reaches a maximum value when all rays crossing a cell belong to the same sector, in which case we have

$$\alpha_{j,\max} = q^{-1/2} \delta_j (1 - q^{-1}). \quad (6)$$

The above expression may be conveniently used to normalize the anisotropy coefficient of the cells and to define the anisotropy index α of the global ray configuration as

$$\alpha = \frac{\sum_{j=1}^p (\alpha_j w'_j / \alpha_{j,\max})}{\sum_{j=1}^p w'_j}. \quad (7)$$

As may be seen, α is a weighted mean of the coefficients α_j (for the definition of the weights w'_j , see the Appendix) and may assume values ranging from 0 to unity; at the lower extreme the ray coverage has a uniform azimuthal distribution over all the cells, whereas at the upper extreme all cells exhibit the maximum possible degree of anisotropy.

To better illustrate the meaning of the configuration descriptors introduced above, Fig. 2 shows a map of seismic rays selected by Basili, Mele & Valensise (1987) from the seismic bulletins published by the Istituto Nazionale di Geofisica (ING, Roma) and other seismic agencies; the configuration includes 1078 rays relative to 20 earthquakes received by almost 200 stations located in Italy and the surrounding areas. The region is partitioned into $1^\circ \times 1^\circ$ cells, of which $p = 74$ cells are crossed by seismic paths. For the configuration shown, the mean density D is 44.7 (see equation 4a), which means that on average there are nearly 45 rays crossing each cell. The density distribution is rather unbalanced, the dispersion index σ having the value of 42.0 (see equation 4b); the maximum density value is almost eight times greater than the mean value. The azimuthal coverage is visibly far from being uniform, since the rays are predominantly oriented NNW–SSE, due to geographical distribution of the seismic network. As a consequence, the anisotropy index is rather large, i.e. $\alpha = 0.52$, the number of sectors q being equal to 18. The goal of our algorithm is that

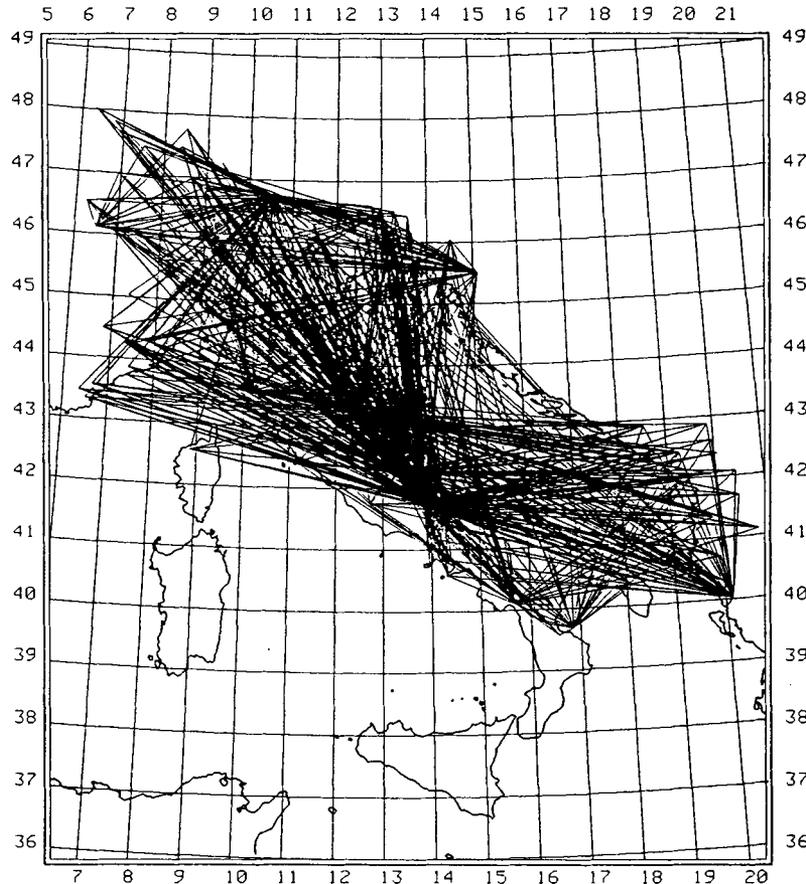


Figure 2. Seismic ray pattern for a set of 20 recent earthquakes. The map reproduces a subset of the data selected by Basili, Mele & Valensise (1987) from the microseismic bulletins published by the Istituto Nazionale di Geofisica and other seismic agencies. The ray configuration is shown only to illustrate clearly the meaning of the configuration descriptors introduced in our paper. There are 1078 rays crossing 74 ($1^\circ \times 1^\circ$) regular cells. The mean ray density D associated with the configuration is 44.7. According to the definition (see equation 4a), there are on average about 45 rays crossing each cell. The density value of the maximum density cell is 350.8. At a glance, the density distribution appears quite unequal. This is confirmed by the relatively high value of the dispersion index $\sigma = 42.0$, defined by equation (4b), which is very close to the mean value. The azimuthal ray distribution is not uniform. Most of the rays follow a NNW-SSE path because the location of the seismic stations is strongly dictated by the geographic form of the Italian peninsula. Therefore, the anisotropy index of the configuration defined in equation (7) is large, i.e. $\alpha = 0.52$. The number of sectors q used in the computations is equal to 18.

of improving the ray configuration resulting from the available observations. Now, given a model (the cell partition) and the data set (the rays), the parameters D , σ and α may be computed for the initial configuration consisting of all possible rays. A further parameter that can be calculated is the upper extreme of the density distribution, i.e. the maximum density found among the cells, here denoted by δ_{\max} . Let us define the best possible configuration as the one where all cells share the same maximum value of ray density δ_{\max} and where the azimuthal coverage is uniform. This configuration, that we will call the target, has the valuable properties of having $D = \delta_{\max}$, $\sigma = 0$ and $\alpha = 0$. In an appropriate 3-D vector space, we then measure the distance between the actual ray configuration and the target, and use this value to judge the goodness of the configuration: that is each configuration is associated to a proper scalar score S (for the computation details see the Appendix). Hence the goal of our procedure may be interpreted as that of approaching the target, i.e. of minimizing the distance (the scalar S) between the initial configuration and the best possible one. The algorithm

proceeds iteratively. The main steps may be illustrated as follows: (i) starting from a given ray configuration with score S , a cell is eliminated, that is all rays crossing the cell are dropped from the configuration; (ii) the score S' associated to the new configuration is calculated; (iii) the score S' is compared with the old configuration score S . If $S' < S$, then the new configuration is taken as a better configuration, the ray elimination is confirmed and the algorithm resumes from the first step (i). Otherwise, the old configuration is preserved and the sequence is restarted from the point (i), by trying to eliminate a different cell. The procedure stops when all attempts to reduce the configuration score fail. We observe that the procedure does not guarantee that the absolute minimum of the score function is reached, since the iteration path may lead to a local minimum, but it has the advantage of being very simple and in this paper we use it only as a means to support the principle that expressing the properties of a ray configuration with a scalar function may lead to useful applications. We observe further that the process whereby the final configuration is reached implies the removal of seismic rays and cells, which means that we

renounce part of the available data and information, reducing at the same time the number of unknowns because of the cell elimination. In the next section we will show examples for which the final configuration is such that the selected subset of data permits a better evaluation of the remaining unknowns.

NUMERICAL EXPERIMENTS

We carried out a lot of experiments to study the performance of our algorithm on different initial ray configurations, but we will show here only a small selection of such experiments taken from significant examples published in the literature. For the sake of simplicity, in our simulations we consider a homogeneous field where the true value of the seismic P -wave velocity is 5 km s^{-1} . The choice enables us to assume rectilinear ray paths, avoiding any superfluous ray tracing difficulty. Furthermore, the experiments involve only a small number of observations and of model parameters, offering the advantage that we are thus allowed to use the SVD algorithm for matrix inversion and to compute the variance-covariance matrix as well as the resolution matrix to accomplish a complete evaluation of the quality of the tomographic images we obtain. Simulations with thousands of unknowns and tens of thousands of first arrivals, although closer to large-scale tomographic problems, would have implied the use of iterative row-action inversion methods such as the LSQR and the SIRT family techniques which are notoriously less precise and do not permit an easy calculation of the solution accuracy indicators (see Clayton & Comer 1984; Humphreys & Clayton 1988; Spakman & Nolet 1988; Spakman 1988). The test of the performance of our pre-selection algorithm on a real large-scale set of data is therefore left for future work.

The procedure we used to compute the seismic slowness in the 2-D cells may be summarized as follows. According to least-squares theory, equation (3) is transformed into the set of normal equations

$$\mathbf{G}^T \mathbf{G} \mathbf{m} = \mathbf{G}^T \mathbf{d}, \quad (8)$$

where \mathbf{G}^T denotes the transpose of the matrix \mathbf{G} . If we denote the diagonal elements of the square $M \times M$ matrix $\mathbf{G}^T \mathbf{G}$ by γ_i and introduce the matrix $\mathbf{B} = \text{diag}(\gamma_i^{-1/2})$, then a convenient scaling leads to the new system

$$\mathbf{L} \mathbf{w} = \mathbf{z}, \quad (9)$$

where \mathbf{L} is an $M \times M$ matrix more suitable for the inversion than $\mathbf{G}^T \mathbf{G}$. The linear transformations involved in obtaining equations (9) are

$$\begin{aligned} \mathbf{L} &= \mathbf{B} \mathbf{G}^T \mathbf{G} \mathbf{B}, \\ \mathbf{w} &= \mathbf{B}^{-1} \mathbf{m}, \\ \mathbf{z} &= \mathbf{B} \mathbf{G}^T \mathbf{d}. \end{aligned} \quad (10)$$

In the framework of the SVD procedure (see Aki & Richards 1980), the matrix \mathbf{L} is suitably factorized as

$$\mathbf{L} = \mathbf{U}_l \mathbf{\Lambda}_l \mathbf{V}_l^T, \quad (11)$$

where \mathbf{U}_l and \mathbf{V}_l are real $M \times l$ matrices ($l \leq M$), satisfying the properties $\mathbf{U}_l^T \mathbf{U}_l = \mathbf{V}_l^T \mathbf{V}_l = \mathbf{I}$, and where $\mathbf{\Lambda}_l$ is a diagonal $l \times l$ matrix whose non-zero elements λ_i are known as singular values and, with no loss of generality, may be

assumed to be arranged in decreasing order: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_l > 0$. According to SVD theory, the generalized inverse of \mathbf{L} , \mathbf{L}_g^{-1} , is given by

$$\mathbf{L}_g^{-1} = \mathbf{V}_l \mathbf{\Lambda}_l^{-1} \mathbf{U}_l^T \quad (12)$$

and the solution of (9) is expressed in the form

$$\mathbf{w} = \mathbf{L}_g^{-1} \mathbf{z}. \quad (13)$$

Hence, in terms of the original model and data vectors, the solution of the system (10) may be written as follows:

$$\mathbf{m} = \mathbf{B} \mathbf{L}_g^{-1} \mathbf{B} \mathbf{G}^T \mathbf{d}. \quad (14)$$

Practically, numerical errors often prevent the existence of exactly null singular values λ , but ill-conditioning caused by poor ray illumination results in the presence of singular values that are very small and which may have a dramatic influence on the final solution. Regularization of the computed field is generally achieved by modifying the generalized inverse in one of the two following ways (see Van der Sluis & Van der Vorst 1987): (i) the smallest $r = p - n$ eigenvalues λ_i are set equal to zero and, as a consequence, a value of $l = n$ is used in (11) and (12), with the effect of cutting off their effects on the solution; (ii) all singular values are slightly removed from zero by the addition of a positive constant c , with the consequence that the contribution of all λ_i smaller than c is strongly damped. As we will see, both regularization methods are quite effective, but their performance is strongly dependent upon the sequence of the singular values and the values of the parameters n and c , whose choice, on the other hand, is commonly quite problematical and difficult to perform *a priori* (see Scales & Gersztenkorn 1988). The explicit computation of the generalized inverse \mathbf{L}_g^{-1} allows us to deduce the variance-covariance matrix \mathbf{C}_m and the resolution matrix \mathbf{R} . As to the former, if \mathbf{C}_d is the covariance matrix for the data, i.e. for the observed traveltimes, the covariance matrix for the model parameter vector \mathbf{m} may be easily obtained with the aid of the expression

$$\mathbf{C}_m = \mathbf{F} \mathbf{C}_d \mathbf{F}^T, \quad (15)$$

where the transformation operator \mathbf{F} is an $M \times N$ rectangular matrix given by

$$\mathbf{F} = \mathbf{B} \mathbf{L}_g^{-1} \mathbf{B} \mathbf{G}^T. \quad (16)$$

As regards the resolution matrix, according to the definition by Wiggins (1972), it may be simply deduced by substituting the data vector \mathbf{d} in (14) with the expression given by equation (3). It thus results in

$$\mathbf{R} = \mathbf{B} \mathbf{L}_g^{-1} \mathbf{B} \mathbf{G}^T \mathbf{G}, \quad (17)$$

which, in regularization methods based on singular values cut-off, reduces to

$$\mathbf{R} = \mathbf{V}_l \mathbf{V}_l^T. \quad (18)$$

In order to perform a thorough evaluation of the goodness of the inverse solution, it is quite useful to introduce a number of indicators. One of these is the normalized misfit μ between the true slowness s^t , which we assume to know in our simulations, and the computed field s , i.e.

$$\mu = \left[\sum_{j=1}^p (s_j - s_j^t)^2 \right]^{1/2} \left[\sum_{j=1}^p (s_j^t)^2 \right]^{-1/2} \quad (19)$$

which obviously measures how faithful a picture the resulting tomographic image is of the reality. Statistical errors are commonly used as a further means to estimate the solution. In terms of the model covariance matrix, the statistical error, say ε_j , that should be associated with the slowness determination for the j th cell, may be written as

$$\varepsilon_j = \sqrt{(\mathbf{C}_m)_{j,j}}, \quad (20)$$

where $(\mathbf{C}_m)_{j,j}$ is the diagonal element of the j th row for the matrix \mathbf{C}_m . The mean statistical error ε is simply obtained by averaging the ε_j over the p cells crossed by seismic rays. In the following the data covariance matrix is assumed to be the identity matrix, since, as we will see, it is not the absolute size of the statistical errors which is of major concern here, but rather their distribution in the cells and how much they change as we pass from the global configuration inversion to the core inversion or as we vary the regularization parameters. The resolving power of the inversion over the parameter space is deduced by analysing the resolution matrix \mathbf{R} , whose row j is commonly taken as a discrete smearing function of the slowness determination in the cell j , though caution is necessary in the interpretation since the row elements of the matrix are not all positive and do not sum to 1 when cut-off or damping is applied to \mathbf{L}_g^{-1} (see Nolet 1985). In this paper we prefer to consider the absolute value of an element of the matrix \mathbf{R} , say $|r_{j,k}|$, as a cross-action coefficient of the j th distortion function, i.e. of a function that in acting on the cell j has the double effect of smearing and of amplifying the solution in the cell. The function amplification factor ρ_j may be computed by summing up all coefficients in the row j :

$$\rho_j = \sum_{k=1}^p |r_{j,k}|, \quad (21)$$

while the mean amplification ρ may be simply taken as the average of the ρ_j over all the resolution matrix rows. If we normalize the distortion function for the cell j , by dividing each coefficient $|r_{j,k}|$ by the amplification factor ρ_j , we can single out the smearing action of the function. In order to associate a typical smearing length to the cell j , the so-called resolution width, benefit may be taken from the mathematical resemblance between the normalized distortion function and a discrete probability density function. In probability theory, the second moment of a probability distribution is commonly taken as a measure of the variate dispersion and, consistently, the second moment square root is taken as the typical dispersion length (see e.g. Kalbfleisch 1979). Since our normalized distortion function for the cell j is defined on a 2-D region, it may be regarded as a probability density for two discrete variables x and y , which may assume only the values corresponding to the central points of the cells in a suitable reference frame oriented parallel to the partition grid. Consequently for both variables, it results in the related marginal distributions being determined, and their variances may be computed as follows:

$$W_{x,j}^2 = \frac{1}{\rho_j} \sum_{k=1}^p (x_k - x_j)^2 |r_{j,k}|, \quad (22a)$$

$$W_{y,j}^2 = \frac{1}{\rho_j} \sum_{k=1}^p (y_k - y_j)^2 |r_{j,k}|, \quad (22b)$$

where the cell midpoints are involved. From the above expressions, our following definition of the resolution width for the cell j in cell units is naturally derived:

$$W_j = \frac{1}{l_j} \sqrt{W_{x,j}^2 + W_{y,j}^2}, \quad (23)$$

where $l_j = A_j^{1/2}$ is a typical length and the mean resolution width W may be taken as the average over all the cells. In cases where the resolution is perfect, the matrix \mathbf{R} reduces to the identity matrix with the consequence that all amplification factors ρ_j are unitary and all the resolution widths W_j vanish.

The first simulation concerns the very simple initial configuration displayed in Fig. 3(a), quite similar to that utilized in a theoretical paper by McMechan (1983) to study the resolving properties of the algebraic reconstruction technique (ART) in obtaining tomographic images. The set of 14 seismic paths crossing 19 cells gives rise to a strongly ill-conditioned problem, as may be seen from the ratio between the largest and the smallest singular values which is of the order of 10^{10} . The total score associated with this ray distribution by our algorithm is $S = 0.84$, whereas the contributions due to the score components related to ray density, density dispersion and anisotropy are 0.67, 0.20 and

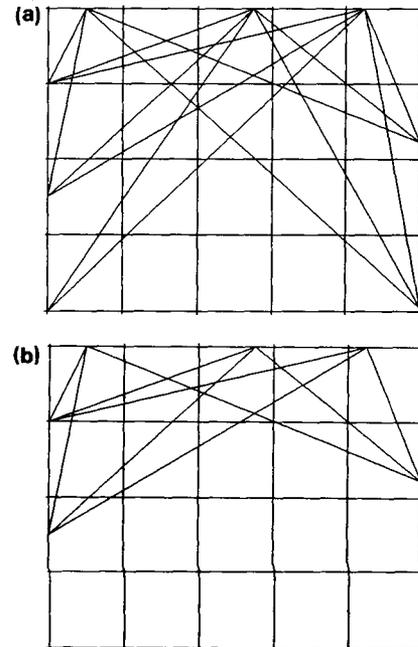


Figure 3. (a) McMechan initial ray configuration. The simple ray pattern is very similar to that utilized by McMechan (1983) to investigate the properties of algebraic reconstruction techniques in tomographic inversion. The configuration departs slightly from central symmetry and consists of 14 rays crossing 19 cells. The total configuration score is 0.84, while the components related to the ray density, the dispersion index and the anisotropy index are 0.67, 0.20 and 0.48, respectively. (b) McMechan final ray configuration. The pre-selection algorithm eliminates five rays, reducing the number of cells in which to compute the solution to 11. The configuration score is 0.80. The ray concentration per cell is much higher than for the initial data set and compensates for the worsening of the azimuthal distribution of the rays. The score components for the ray density, the dispersion index and the anisotropy coefficient are given by 0.50, 0.19 and 0.59, respectively.

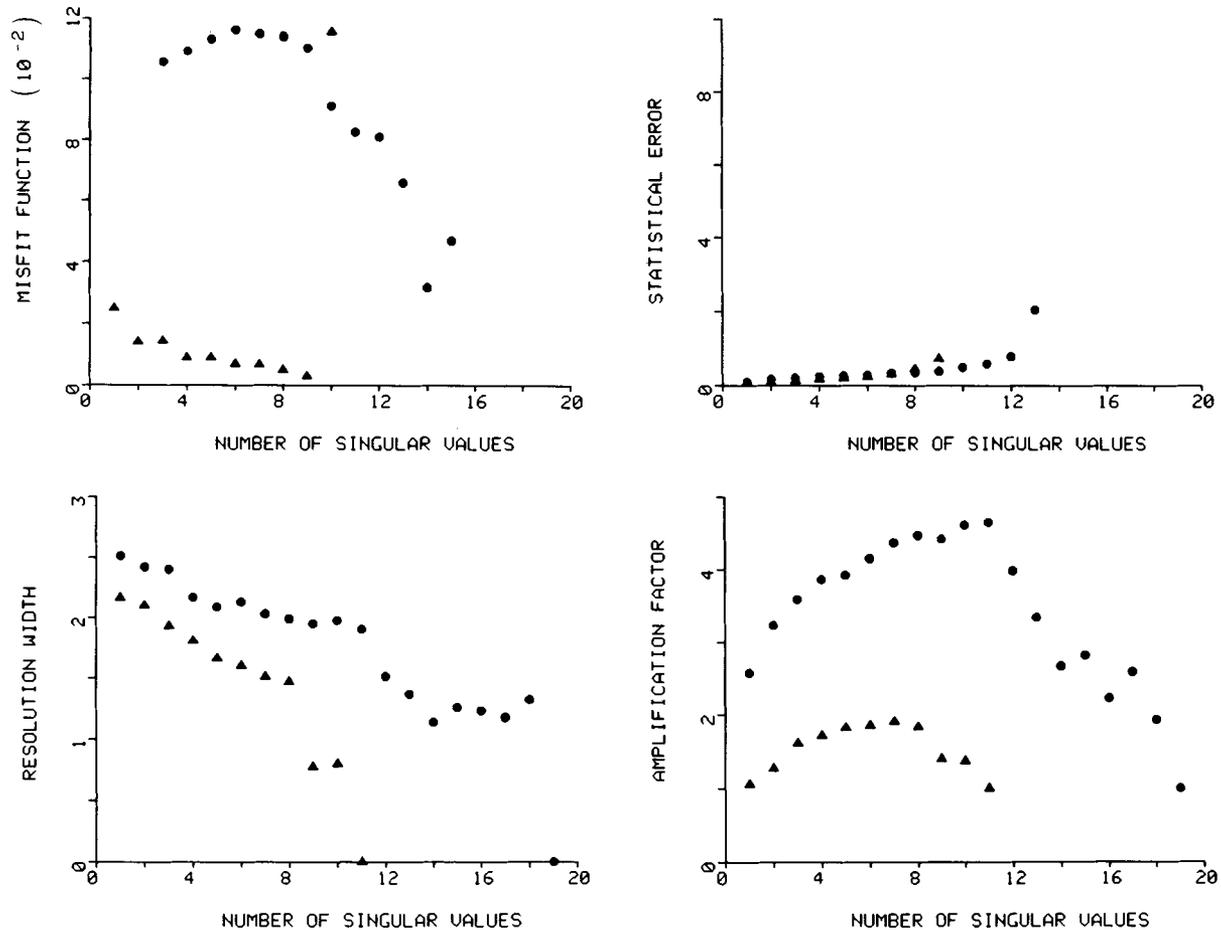


Figure 4. Goodness indicators of the solution computed for both configurations of Fig. 3; all indicators represent averaged quantities over the region covered by the rays (see definition in the text). They are plotted as a function of the number n of the largest singular values used to calculate the slowness, according to Wiggins' regularization method. Solid-circle curves, ranging from $n = 1$ to 19, refer to the original pattern with 19 cells, while those marked by solid triangles are confined within $n = 12$ and refer to the configuration core. Note that in the upper panels the values corresponding to the largest n are not plotted, being out of the chosen scale. The IMSL routine LSVDF was used to compute singular values and related eigenvectors. The reference slowness field was taken to be homogeneous and to exceed the true homogeneous field by 5 per cent. No noise was added to the true traveltimes.

0.48, respectively. The final configuration obtained by the application of the iterative procedure described in the previous section is given in Fig. 3(b); five rays have been eliminated and, consequently, eight cells. The total score has been lowered to $S = 0.80$ by the applied pre-selection; the main contribution comes from the reduction of the density component passed to a value of 0.50 (measuring the normalized difference between the mean ray density D and the maximum ray density of a cell δ_{\max} as explained in the Appendix), while the density dispersion component decreases slightly to 0.19 and the anisotropy index increases to 0.59. A large number of simulations have been performed with this case. Here we will show the results obtained starting from a reference homogeneous slowness 5 per cent in excess of the true field, without adding noise to the true traveltimes. In Fig. 4 the normalized misfit μ , the mean statistical error ϵ , the mean resolution width W and the mean amplification factor ρ are plotted versus the number, say n , of the largest singular values used to compute the solution for both the initial as well as the final configurations. Observe that, due to the strong ill-

conditioning, the misfit and the statistical errors take an unacceptably large value (out of scale in the graph), as soon as n is greater than 15 and 10, respectively, for the initial and final ray geometries. Observe, further, that the mean resolution width W has a general increasing trend as the number of the singular values n used decreases; starting from a null value when all singular values are retained, it approaches a value of the same order as the mean characteristic region length measured in cell units, which is expected according to our definition. This is the main point of difference with the resolution analysis initially formulated for continuous fields by Backus & Gilbert (1967) and successively adapted to a discrete partition by Nolet (1985), where the resolution width may assume values much larger than the grid dimension, making the original interpretation of typical mean length of influence of the distortion function quite difficult. A further interesting observation regards the amplification factor ρ which, especially for the original ray configuration, considerably exceeds the unit value. The general behaviour of the four functions is quite similar for the two cases, i.e. without and with our pre-selection

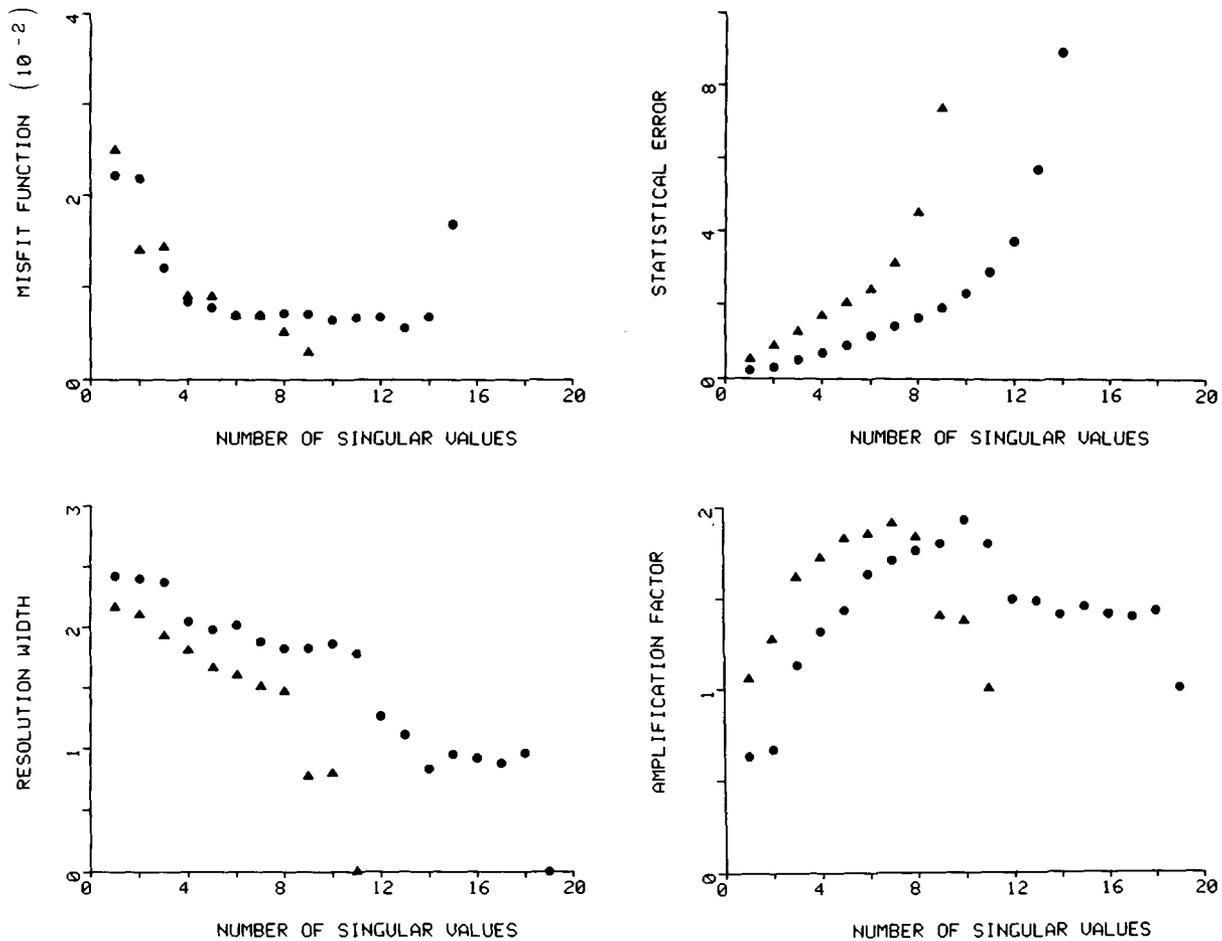


Figure 5. Goodness indicators to compare the solutions computed in the core. All quantities are averages extended over the cells forming the core, i.e. the cells covered by the set of selected seismic rays (see Fig. 3b). The solution calculated from the original configuration (solid circles) is compared to that derived from the final ray pattern (solid triangles). Note the improvement as regards the misfit as well as the resolution. For other details, see caption to Fig. 4.

applied, but it is remarkable that the indicators are almost constantly better in the second case, which should be taken as good evidence in favour of the usefulness of applying our method for the proposed numerical experiment. But we can add one more convincing consideration. One could argue that the above comparison is rather unfair, since the results obtained for the global ray geometry are too heavily affected by the very bad determination of the off-core cells in the bottom row; the inversions of the two data sets should be compared side by side only in the core region. In order to answer this possible objection, the indicators related to the global inversion and averaged over the cells forming the core are shown together with those computed for the final ray pattern in Fig. 5 as a function of n . The discrepancy between the curves is not as large as in Fig. 4, but it is still quite evident that the pre-selected ray configuration gives rise to better results. Hence, analogously with the case illustrated in the second section of this paper, it is shown that the elimination of a number of rays from the initial configuration leads to a more satisfactory tomographic image of the core. We emphasize further that the above important conclusion is apparently independent of the regularization method used. This was proven by applying the damping regularization method instead of the Wiggins'

cut-off. We performed a set of experiments where the chosen damping factors were $c = \lambda_n$ with the effect of attenuating the influence of all the singular values smaller than or equal to λ_n , n ranging from one to the number of unknowns p . We obtained curves with behaviour quite similar to those displayed in Figs 4 and 5, showing the convenience of using the reduced ray pattern. We observed also that, in general, damping is slightly more effective than cut-off in smoothing the artificial inversion irregularities, since meaningful slowness values can be obtained over a wider range of n and, moreover, all the used indicators exhibit a better and more regular behaviour as n changes. Taking advantage of the substantial equivalence of the two regularization methods as regards the main purpose of our paper, in the following discussion on this experiment only cut-off will be considered explicitly and it will be implicitly admitted that all the runs were also parallelly duplicated with damping. To complete the analysis of our example, it is convenient to compare the best solutions obtained for both the examined ray geometries. Looking at Figs 4 and 5, it appears that the best images are computed by using $n = 13$ and $n = 9$ singular values for the respective cases of the initial and the final data sets. In Figs 6 and 7 (the former being related to the global pattern, the latter to the selected

2000 0.4	2007 0.7	1994 0.4	1999 0.3	1998 0.6	0.4 1.0	1.3 1.7	1.3 1.5	1.1 1.3	0.7 1.1
2013 0.6	2016 0.8	2015 1.0	1996 0.5	2003 0.4	1.3 1.7	1.3 1.8	1.3 1.9	1.0 1.5	0.8 1.1
1975 0.6	2007 0.2	1898 2.4	2033 0.5	1982 0.5	1.8 1.7	1.5 1.3	1.7 5.0	1.7 1.7	1.3 1.7
1998 0.8	1953 24		1439 3.3	2017 0.2	1.7 1.4	2.0 18		2.1 17	1.7 1.0

Figure 6. Results of the inversion of the Fig. 3(a) ray pattern, computed through the Wiggins' cut-off method; the 13 largest singular values are used in the computations. The calculated slowness and the statistical error are shown for each cell one above the other in the left panel. Remember that the true slowness value is everywhere $2000 \times 10^{-4} \text{ s km}^{-1}$, while the reference field is 2100 in the same units. Resolution width and amplification factor are displayed in the right panel. Observe the anomalous irregularities close to the bottom boundary of the field with almost meaningless values of slowness and amplification factor.

2000 0.4	2008 0.7	1990 0.2	2003 0.5	2000 0.6	0.0 1.0	0.9 1.5	1.0 1.2	0.8 1.3	0.0 1.0
2002 0.7	2010 0.8	1992 2.1	2003 1.2	2000 0.7	0.9 1.5	1.1 1.8	1.3 2.6	0.9 1.3	0.0 1.0
1995 0.2					1.6 1.3				

Figure 7. Results of the inversion for the core region, illustrated in Fig. 3(b). Only the nine largest singular values are used in computing the slowness, for this seems to provide the best solution for the considered configuration if we look at Figs 4 and 5. For other details, see caption to Fig. 6. Note that the very small subset of nine seismic paths is sufficient to give a quite satisfactory image of the upper layer of the original region.

core), the slowness (in $10^{-4} \text{ s km}^{-1}$) and the statistical error (in arbitrary units) distributions are displayed in the left-hand panel, whereas the resolution width and the amplification factor are given together in the other panel. A first glance at the results of Fig. 6 clearly shows that the tomographic reconstruction of the blocks in the bottom region is largely unsatisfactory; slowness estimates in two cells are unacceptably too far from the true value and even worse than the initial guess values. This is not surprising though, for the bottom cells are very poorly illuminated. It may be observed that the solution evaluated in the top two layers is much better, but that the quality of the results degrades from the first to the second layer, where a set of contiguous cells with higher slowness determination gives rise to a spurious lateral heterogeneity. A possible explanation is that the bad illumination of the bottom half region has negative effects even in the confining cells of the second layer. The hypothesis is confirmed by the solution computed for the geometry with selected ray paths. The fields shown in Fig. 7 exhibit a much more regular aspect over the whole sub-region covered by the rays; the relative error does not exceed 0.5 per cent, and the resolution width has satisfactorily small values. The bottom part of the core (i.e. the cells in the second row together with the leftmost cell in the third row) is certainly more accurate in Fig. 7 than in Fig. 6. And the obvious conclusion that can be drawn from the example shown is that removing five rays from the initial configuration has a beneficial influence on the inversion; a suitable removal process, though lowering the density of the ray paths in the core cells, preserves the

core from the contamination of the instabilities possibly existing in the off-core region. In order to evaluate the effectiveness of our strategy in computing tomographic maps in a more realistic situation, we allowed the true traveltimes to be contaminated by normally distributed errors, with standard deviation as large as 5 per cent of the mean absolute value of the residual observation vector d_i (see equation 3). From a numerical point of view, noise addition practically consists of drawing an experimental sample of as many random Gaussian numbers as the total number of data. In our case a sample S_{14} of 14 random elements drawn from the normalized Gaussian distribution (with zero mean and unitary variance) is associated with the initial ray configuration, while a subset S_9 of nine elements is associated with the final ray pattern. Due to the very small sample sizes, the variability of the experimental samples is expected to be very large; in practice, both the mean and the variance of the randomly chosen numbers may depart greatly from the values of the theoretical probability distribution and may change considerably not only from one experimental sample to another, but also from one sample of 14 elements S_{14} to the related sub-sample of nine elements S_9 . Now, if the random sub-sample S_9 happens to be much 'better' than the sample S_{14} (this meaning that the noise is mostly concentrated in the removed rays), it should be obviously expected that the initial ray configuration leads to a worse image of the core than the final configuration. On the contrary, if the sample S_{14} is largely more Gaussian than its subset S_9 (this generally meaning that the five eliminated rays are affected by much smaller noise than the others), the

result of the comparison is obviously favourable to the initial configuration. The most interesting occurrences are therefore the cases where the two sets S_{14} and S_9 have approximately the same mean and variance, i.e. the cases where the added noise acts evenly on both the data sets. If we consider that in our simulation the Gaussian noise plays the double role of the experimental errors in the arrival time determination and of our *a priori* uncertainty in choosing the reference slowness model (reflecting our simulated ignorance of the earth structure), this condition corresponds to the reasonable assumptions that there is no reason to prefer the time readings of the core rays to the others (or vice versa) as well as there is no reason to consider that the initial reference value of the slowness is better in the core than outside (or vice versa). We performed, therefore, a number of experiments with data corrupted by noise addition, under the condition that the means m_{14} and m_9 and the variances v_{14} and v_9 of both random samples S_{14} and S_9 were sufficiently close to one another and, further, not too far from the theoretical Gaussian distribution; i.e. $|m_{14} - m_9| \leq 0.02$ and both means smaller than 0.1, together with $|v_{14}^{1/2} - v_9^{1/2}| \leq 0.1$ and both variances smaller than 1.0. Table 2 reports the results of 10 different experiments of

noise addition, satisfying the above conditions. Each experiment consists of a set of inversions where the cut-off method was applied over the whole range of admissible values of n , i.e. of the number of singular values used in the calculations; we computed, therefore, curves analogous to those shown in Figs 4 and 5. For each experiment the results of the best inversion are reported in the table, namely the normalized misfit μ and the resolution width W . The first consideration is that the addition of noise implies a less accurate solution everywhere in the core as well as outside. For instance, in inverting the initial ray configuration, we obtained a misfit μ around 7.8×10^{-2} for the global region and around 7.8×10^{-3} for the core (see first and second columns of Table 2), while the values computed for the noise-free case are, respectively, 3.2×10^{-2} and 5.6×10^{-3} . The application of tomographic inversion of the final ray geometry (see third column in Table 2) gives a mean misfit of 7.1×10^{-3} versus the noise-free value of 3.0×10^{-3} . The second consideration is that in all cases the reduced configuration leads to a solution that is, on the whole, largely more acceptable (with misfit values one order of magnitude smaller). Third, comparing the performances in the core itself, the mean values of the misfit and of the

Table 2. Inversion results are given for each of the 10 different experiments of noise addition for the configurations shown in Figs 3(a) and (b). Each experiment corresponds to a random Gaussian drawing having zero mean and unitary variance; the true traveltimes are contaminated by errors with standard deviation as large as 5 per cent of the mean absolute value of the residual observation vector d_i (see equation 3). For each of the 10 cases the best solution is selected, according to the minimum value of the normalized misfit function μ (equation 19), in the set of inversions we performed by applying the SVD method with the cut-off regularization technique over the whole range of singular values. The performance of each experiment is synthesized by the normalized misfit and by the mean resolution width, which tends to grow larger as the number of ignored singular values increases. The second column reports the inverse solution indicators relative to the original ray configuration of Fig. 3(a). The third column refers to the same ray configuration but the indicators are obtained by averaging only over the cells belonging to the so-called 'core region', which are the cells covered by the configuration of Fig. 3(b). Finally, the last column shows the inversion results computed for the final ray configuration of Fig. 3(b), singled out by our pre-selection algorithm. Observe that, according to our indicators, the inverse solution calculated using the final ray configuration is generally more accurate; in 10 experiments, seven runs give rise to a better value of the misfit function and eight runs to a smaller resolution width.

	GLOBAL INVERSION		CORE SOLUTION FROM GLOBAL INVERSION		CORE INVERSION	
	μ (10^{-2})	W	μ (10^{-3})	W	μ (10^{-3})	W
I	7.3	1.4	6.4	0.8	6.0	1.5
II	8.7	2.0	8.3	1.8	6.1	1.5
III	8.0	2.0	8.6	2.0	6.0	1.5
IV	7.0	1.5	7.8	1.1	5.4	1.5
V	7.3	1.9	8.4	1.9	7.4	1.5
VI	8.2	2.0	7.8	2.0	8.1	1.7
VII	8.2	2.0	7.8	2.0	7.4	1.6
VIII	7.9	2.0	7.9	2.0	8.7	1.6
IX	8.0	2.0	7.8	2.0	8.3	1.7
X	7.7	2.0	7.7	2.0	7.3	1.6

2022 2.5	1996 1.8	1985 1.7	2011 2.2	1989 2.0	2024 1.6	1992 1.2	1980 1.1	2003 1.3	2002 1.6
1990 2.4	1996 1.7	2019 1.7	1982 1.8	2007 2.3	1985 1.8	2013 1.4	1998 1.5	2005 1.3	2004 2.0
1963 2.1	2028 1.9	1832 1.8	2031 1.9	1981 2.4	1990 1.7				
2038 2.9	1309 2.4		1495 2.5	2016 2.5					

Figure 8. Images derived from data corrupted by Gaussian noise (case 3 of Table 2). The cut-off method is applied to data affected by random noise as large as 5 per cent of the mean absolute value of the time residuals. The left-hand panel illustrates the best field computed for the total region, obtained by neglecting the contribution of the 13 smallest singular values. The right panel shows the best solution computed from the final configuration, calculated by cutting off the three smallest singular values. In each square cell, the estimated slowness, in 10^{-4} s km $^{-1}$, and the resolution width are given one above the other. By comparing the result with the images shown in Figs 6 and 7, the perturbing effect of the noise is easily seen. Comparison between the solutions shown in the two panels makes it evident that the selected ray geometry in the core performs better than the original one.

2022 2.4	1993 1.8	1987 1.7	2015 2.2	1996 2.5	2001 1.8	1993 1.4	1993 1.3	2021 1.4	1978 1.6
2005 1.9	2007 1.6	2023 1.8	1973 1.9	2008 1.9	1976 1.8	2036 1.5	2008 1.6	2007 1.7	2016 1.9
1980 2.1	2034 2.0	1827 1.8	2026 1.9	1977 2.2	2002 1.7				
2043 2.8	1355 2.3		1456 2.4	2012 2.4					

Figure 9. Images derived from data corrupted by Gaussian noise (case 8 of Table 2). The best solution computed for the original ray configuration is compared with the best one derived from the final geometry using only nine seismic rays. The former image is the result of the inversion with the nine smallest singular values neglected, while the latter is derived from ignoring the five smallest ones. The images are displayed as in Fig. 8. The anomaly in the left cells of the intermediate rows compromises the performance of the selected ray geometry in the core, shown in the right-hand panel. Note that this is the least favourable case of Table 2. Observe, however, that the resolution width in the right panel is everywhere smaller than in the panel on the left.

resolution width show that the images computed from the final configuration are generally more accurate. In a set of 10 runs, seven cases give rise to a smaller normalized misfit and eight cases to a smaller resolution width. Figs 8 and 9 complete the comparison. Fig. 8 displays the solution for the most favourable run, i.e. the third one. The left-hand panel shows the slowness and the resolution width in each cell computed from the original geometry, whereas the panel on the right shows the same quantities for the geometry obtained after the pre-selection. The improvement in the solution is evident as well as the improvement in the resolution width, which is smaller in all the cells. Fig. 9 refers to the eighth case, which was the least favourable of the 10 experiments; an anomaly in the two leftmost cells in the second layer degrades the solution of the core inversion, but the resolution width is still smaller everywhere, reflecting the good resolution power of the reduced configuration.

The second ray distribution we present in this paper is depicted in Fig. 10(a); it gives a configuration similar to a field experiment conducted at Kråkemåla in southern Sweden, extensively studied by Ivansson (1984). Explosions detonated at the left-hand bore-hole were recorded by geophones placed on the earth surface and in the opposite bore-hole. The vertical cross-section is partitioned into 40 equal cells and a total number of $N = 325$ seismic rays is considered. Our algorithm eliminates 13 out of the original

seismic rays as well as six cells in the bottom row, reducing the total number of unknowns to 34, which implies only a small diminution of the original region extension. Fig. 10(b) illustrates the ray configuration obtained after the ray selection. The score 0.75 associated with the initial configuration has been minimized to 0.69, essentially as the effect of the reduction of the density component of the score, passing from 0.65 to 0.59, which means that the resulting final configuration has a larger concentration of rays per cell. As regards the anisotropy index (see its definition in the Appendix), it remains nearly unchanged around a value of 0.31, since the rigid arrangement of sources and receivers cannot allow a substantial improvement of the azimuthal distribution. Even for this case we have performed all the experiments carried out for the McMechan simple configuration. We will show the curves computed by utilizing damping on noise-free data and a table summarizing the results of experiments on data corrupted by Gaussian noise. Fig. 11 shows the four indicators previously used, i.e. the normalized misfit μ , the statistical error (in arbitrary units), the resolution width and the related amplification factor, all averaged over the whole domain, as a function of the index n of the damped singular value λ_n . The curves are quite regular and tend to approach each other as the number of distinct undamped singular values is made smaller. The largest differences may be seen in the upper part of the interval for large values of n , where

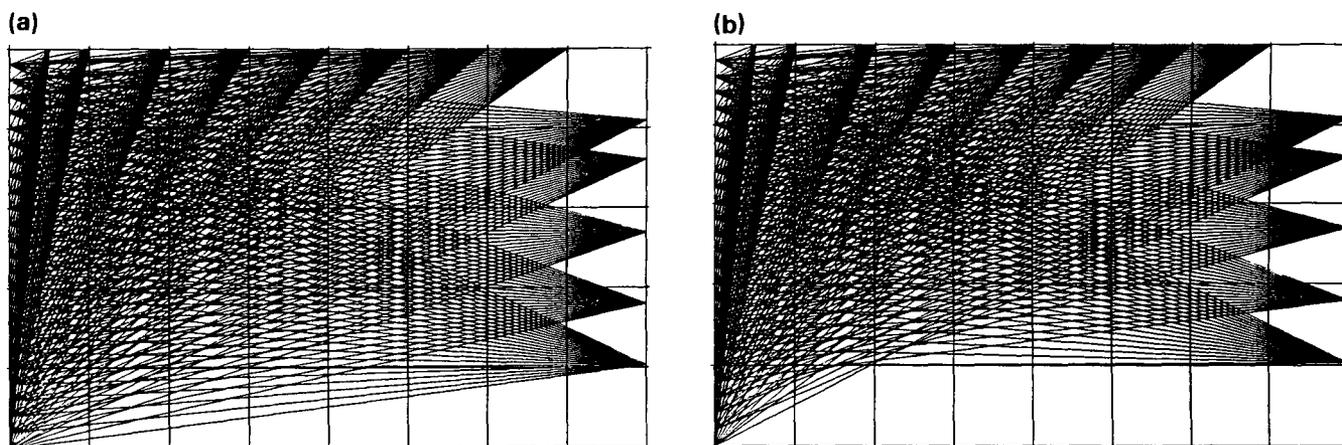


Figure 10. Kräkemåla initial ray configuration. The main geometrical features of the seismic ray configuration of the experiment performed in Kräkemåla, Sweden (see Ivansson 1984), are reproduced. In the vertical cross-section shown the explosion sources are located on the left side, while the receivers are placed on the opposite bore-hole and on the earth's surface. The seismic pattern consists of 325 rays crossing 40 cells. About 30 per cent of the rays are hole-to-hole paths according to the ray distribution in the real experiment. The score associated with the configuration is 0.75. The values of the score components related to the ray density, the dispersion index and the anisotropy index according to the metric explained in the Appendix are 0.65, 0.22 and 0.31, respectively. (b) Kräkemåla final ray configuration. Through the application of the pre-selection algorithm, 13 seismic rays of the original data set are eliminated. Consequently, six cells in the bottom row disappear from the parameter space. The total score S is reduced to 0.69, with score components related to the ray density, the dispersion index and the anisotropy index given by 0.59, 0.20 and 0.31. Comparison with the original configuration shows that the score improvement is mainly due to the diminishing of the density component, i.e. the final data set provides a higher concentration of seismic rays per cell.

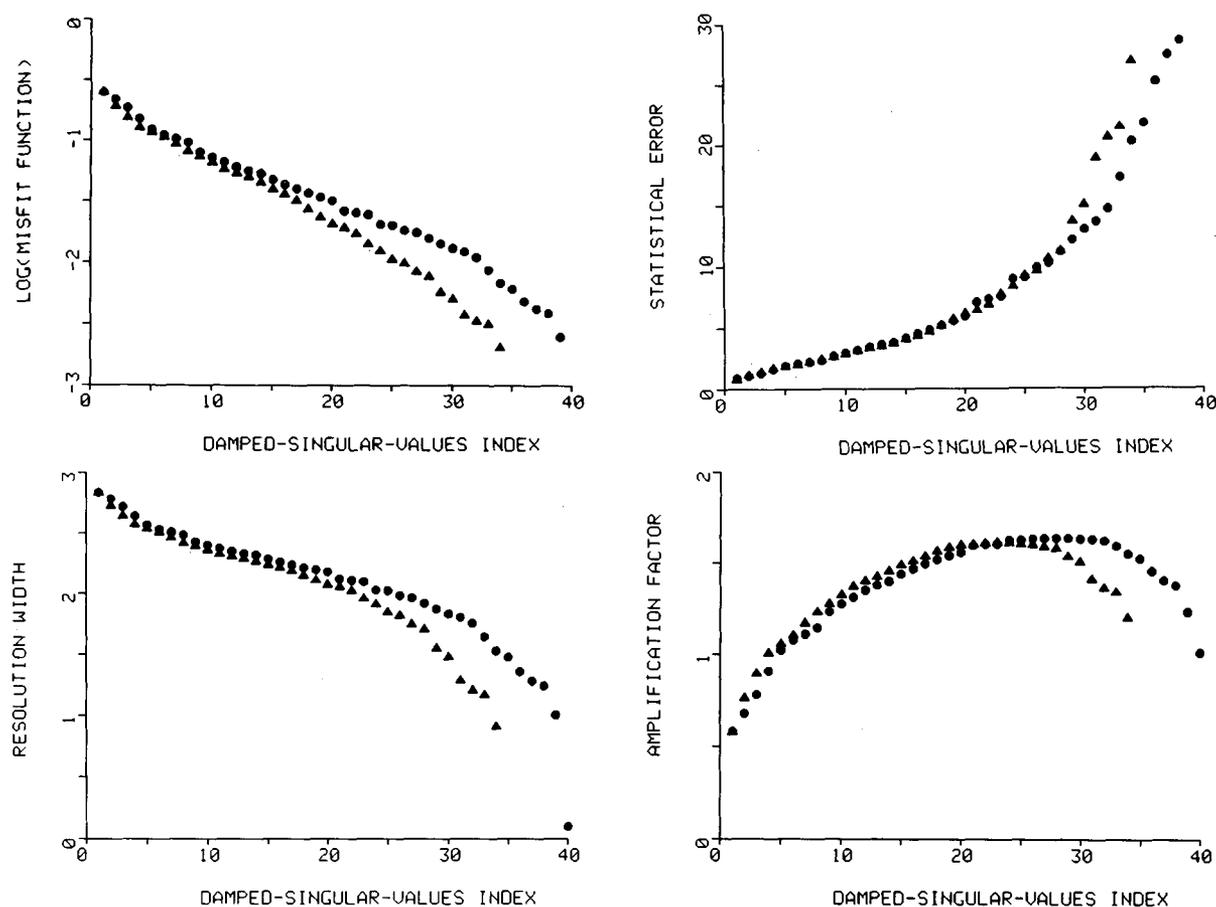


Figure 11. Goodness indicators computed for the original ray pattern of the Kräkemåla experiment (solid circles) and for the configuration obtained after pre-selection (solid triangles). Damping regularization method is applied with $c = \lambda_n$. The index n ranges from 1 to 40 and from 1 to 34, respectively, for the initial ray pattern and for the configuration core. The time data are noise-free. The reference slowness is $3000 \times 10^{-4} \text{ s km}^{-1}$, which is 50 per cent in excess of the true homogeneous field. Note that for $n = 40$ both the normalized misfit and the mean statistical error are out of scale.

2001 0.1	2001 0.1	2001 1.0	2000 1.2	2000 1.2	1999 1.3	2001 1.3	1997 2.1
2001 0.1	2001 1.2	2001 1.9	2000 2.3	2002 2.6	1992 3.9	2010 3.9	1995 3.2
2000 0.1	2001 1.8	2001 2.8	2001 3.2	2003 3.7	1991 4.2	2012 4.5	1995 3.7
2000 0.1	2000 2.7	2001 3.9	2001 4.2	2003 4.4	1991 5.0	2013 4.8	1993 4.1
2000 1.3	2000 4.3	2001 6.1	2002 7.1	1999 1.3	2006 6.4	2006 6.4	1986 9.5

0.6 1.0	0.5 1.0	0.7 1.1	0.7 1.1	0.7 1.1	0.7 1.1	0.7 1.1	1.0 1.1
0.6 1.0	0.6 1.1	0.8 1.2	0.8 1.2	0.9 1.2	1.1 1.4	1.2 1.4	1.1 1.2
0.6 1.0	0.8 1.1	0.9 1.3	1.0 1.3	1.0 1.3	1.0 1.4	1.2 1.5	1.3 1.3
0.7 1.0	1.1 1.2	1.1 1.3	1.1 1.3	1.1 1.3	1.3 1.5	1.3 1.5	1.5 1.3
0.9 1.1	1.5 1.3	1.5 1.5	1.4 1.5	1.4 1.4	1.4 1.3	1.5 1.3	1.7 1.5

Figure 12. Kråkemåla initial configuration results. The related ray configuration is shown in Fig. 10(a). In a way analogous to Figs 6 and 7, the left panel illustrates both cell slowness and statistical error, while the right panel shows resolution width and amplification factor. The example given here corresponds to the particular case $n = 39$ shown in the graphs of Fig. 11. Observe that the solution is reasonable everywhere, but shows some noticeable irregularities in the right part of the field where the receivers are few and distributed in an inadequate way.

2002 0.1	2001 0.1	2001 1.0	2000 1.2	2001 1.2	1999 1.2	2001 1.3	1998 2.1
2001 0.1	2001 1.2	2002 1.9	1999 2.3	2002 2.6	1992 3.9	2009 3.9	1996 3.2
2000 0.1	2000 1.8	2002 2.8	1999 3.3	2003 3.7	1992 4.2	2008 4.6	1997 3.8
2000 0.1	1999 2.7	2004 3.9	2000 4.4	2004 4.6	1991 5.2	2009 5.0	1997 4.3
2001 1.3	1997 4.9						

0.6 1.0	0.4 1.0	0.6 1.1	0.6 1.1	0.6 1.1	0.6 1.1	0.6 1.1	1.0 1.1
0.6 1.0	0.6 1.1	0.7 1.2	0.8 1.2	0.8 1.2	1.1 1.4	1.1 1.3	1.0 1.2
0.6 1.0	0.8 1.1	0.9 1.2	1.0 1.3	0.9 1.2	1.0 1.4	1.1 1.5	1.2 1.3
0.7 1.0	1.1 1.2	1.1 1.3	1.1 1.3	1.0 1.3	1.3 1.5	1.3 1.5	1.5 1.2
0.8 1.0	1.5 1.3						

Figure 13. Inversion results for Kråkemåla final configuration, displayed in Fig. 10(b). The case illustrated here uses a damping factor $c = \lambda_{34}$, which is the most favourable according to the curves in Fig. 11. Left and right panels show slowness together with statistical error, and resolution width together with amplification factor. It may be seen that the irregularities in the right portion are not completely eliminated, but are sensibly reduced.

the solution computed for the core proves to be better, especially as regards the normalized misfit and the resolution width. Note that the best solution for the reduced configuration is computed by damping the smallest singular value, i.e. for $n = 34$, which is an indication of good stability of the problem. On the other hand, damping the smallest singular value for the global configuration gives meaningless results (μ goes out of the logarithmic scale for $n = 40$). Figs 12 and 13 display the best fields calculated in the two experiments, corresponding respectively to $n = 39$ and $n = 34$ for the complete and the reduced set of rays. As expected, the tomographic inversion provides a rather faithful image of the field on the left side, which is well covered by the seismic paths radiated by a dense line of sources. Distortions become relevant in the cells of the three last columns on the right side, where ray illumination is less efficient. The removal of six cells in the bottom row has the effect of smoothing the distortions in the core, improving the global quality of the image. Addition of 0.2 per cent of the Gaussian noise to traveltimes data has also been taken into account. Table 3 shows a set of 10 experiments. The experimental samples S_{325} and S_{312} of random gaussian numbers used satisfy the same conditions regarding the mean m and the variance v that were assumed for the McMechan configurations treated previously. For each

experiment the complete set of inversions, with damping progressively increasing from the smallest to the largest singular value, has been carried out and the best performance is shown in the table. It may be observed that, in general, inversion on the pre-selected set of rays gives more satisfactory values of μ and W than inversion on the global configuration. This appears from comparing the goodness indexes computed for the solution over the whole covered region (shown in the first and the third columns) and also if attention is restricted to the core–core comparison (see second and third columns). From analysis of the results, it is evident that to forsake the determination of the slowness value in the six dropped cells permits better control of the numerical instabilities associated with the scarce ray illumination; it gives a general improvement in the global solution and a more satisfactory reconstruction of the right sector. The improvement in the core, however, is not as considerable as for the McMechan example previously discussed, since the core itself resulting from applying our pre-selection method is not exempt from instabilities. This shows that the idea of improving the solution of the inverse problem by means of a selection operated on the ray configuration (i.e. in the data space) may be effective, but that the pre-selection criteria deserve to be suitably refined in order to optimize the tomographic

Table 3. Inversion results for 10 experiments of noise perturbation on traveltimes data of the Kråkemåla ray configurations shown in Figs 10(a) and (b). The noise standard deviation is 0.2 per cent of the mean absolute value of the residual observation vector d_i (see equation 3). The reference slowness is $2100 \times 10^{-4} \text{ s km}^{-1}$. For other details, see description for the analogous Table 2. Core inversion generally provides a solution that is better than global inversion, both for the total region and for the core sub-region. Comparison between the second and the third columns shows that in none of the 10 experiments does the core inversion solution lead to a larger resolution width and that in seven cases it leads to a better normalized misfit.

	GLOBAL INVERSION		CORE SOLUTION FROM GLOBAL INVERSION		CORE INVERSION	
	μ (10^{-4})	W	μ (10^{-4})	W	μ (10^{-4})	W
<i>I</i>	7.1	1.2	6.3	0.9	7.1	0.9
<i>II</i>	11	1.0	8.5	1.2	6.9	1.2
<i>III</i>	13	1.0	9.5	1.3	9.3	1.2
<i>IV</i>	12	1.0	7.6	0.9	6.2	0.9
<i>V</i>	11	1.3	10	1.2	9.0	1.2
<i>VI</i>	9.1	1.0	8.0	0.9	7.1	0.9
<i>VII</i>	11	1.0	4.6	0.9	4.7	0.9
<i>VIII</i>	9.6	1.0	8.1	0.9	8.2	0.9
<i>IX</i>	9.3	1.0	6.7	0.9	6.5	0.9
<i>X</i>	13	1.3	11	1.4	10	1.3

image. We note further that improvements in the solution are obtained without adding any *a priori* information to the model; we do not impose any constraints on the solution in any sub-regions, such as lower and upper bounds on the slowness value or on the difference between slowness values in neighbouring cells. Of course, the joint utilization of such methods and a form of pre-selection such as that presented here is expected to fruitfully increase the effectiveness of the inversion procedure in treating singularities.

CONCLUSIONS

The most important point to emphasize as a conclusion of our paper is that in the inversion problems where the discretization is based on partitioning the region under investigation into cells, the goodness of the solution depends on some global properties of the configuration of the seismic rays. This result is certainly intuitive, but not trivial. The paper has pointed out the role of features such as the ray density, the density dispersion and the anisotropy of the ray configuration in influencing the accuracy of the inverse solution. It has been shown that there are circumstances where accuracy may be improved by dropping part of the available data following suitable criteria. That is to say, better results can be achieved even with a smaller amount of data (contrary to a very common idea), since the process of data removal leads to the determination of a core region which is better illuminated by the rays remaining after pre-selection and to which the inversion should be restricted. We point out that the algorithm described in the paper may be advantageously used before applying any inversion procedure to the data to enucleate a sub-region where the tomographic inversion may give better reconstructed slowness maps. In the procedure of score

minimization, computational simplicity is favoured over absolute correctness so that the minimum reached is not granted to be the absolute score minimum. Nonetheless, our algorithm is effective in reducing the original configuration score and in determining a sub-region with a more satisfactory ray pattern. In spite of the fact that our examples deal with a small number of rays and cells, since this permitted an easier and a more complete analysis of its performance, we are convinced that our procedure could be even more beneficial in problems with many more data and unknowns such as large-scale tomographic problems (see Spakman, 1988) where field illumination is generally fairly unequal and may critically affect the resulting slowness. Of course this extension, involving very large linear systems of equations, demands further studies and will be one of our future goals. It is, however, beyond the scope of the present paper, whose main purpose was to present the idea of evaluating a ray distribution by means of a single scalar, called the configuration score S . We remember that there are other important problems, such as that of optimizing the cell partition, given the set of seismic ray paths, to which the concept of a configuration score could be fruitfully applied. Among the problems deserving further analysis, we envisage the need to clarify better the complex relationship between the score decrease and the consequent potential improvement in the tomographic image. At the moment we are not able to predict quantitatively how significant a given decrease in the score value is for the expected solution improvement. What we can state is that the final configuration, obtained through score minimization, is likely to give a more accurate slowness map. A further point to investigate is the score definition. The examples shown indicate that the ray density generally plays the major role in reducing the score. This is essentially correct, but it could

be unnecessarily redundant if our method is applied in conjunction with regularization procedures proven to be effective in smoothing instabilities due to unequal density distribution of the seismic paths. Therefore, it could appear worthwhile to explore the possibility of favouring other aspects of the ray configuration characteristics, such as the azimuthal ray distribution.

ACKNOWLEDGMENTS

The authors wish to thank Professor M. Bonafede for valuable discussions. The work has been performed with the financial support of the Ministero della Pubblica Istruzione (MPI) and of the Consiglio Nazionale delle Ricerche (CNR).

REFERENCES

- Aki, K. & Richards, P., 1980. *Quantitative Seismology: Theory and Methods*, W. H. Freeman and Company, San Francisco.
- Aki, K., Christofferson, A. & Husebye, E. S., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *J. geophys. Res.*, **82**, 277–296.
- Backus, G., 1970a. Inference from inadequate and inaccurate data, I, *Proc. Nat. Acad. Sci. US*, **65**, 1–7.
- Backus, G., 1970b. Inference from inadequate and inaccurate data, II, *Proc. Nat. Acad. Sci. US*, **65**, 281–287.
- Backus, G., 1970c. Inference from inadequate and inaccurate data, III, *Proc. Nat. Acad. Sci. US*, **67**, 282–289.
- Backus, G. & Gilbert, F., 1967. Numerical applications of a formalism for geophysical inverse problem, *Geophys. J. R. astr. Soc.*, **13**, 247–276.
- Basili, A., Mele, F. & Valensise, G., 1987. Un modello crostale per la localizzazione di eventi sismici regionali rilevati dalla rete sismica nazionale centralizzata dell'Istituto Nazionale di Geofisica, *VI Convegno Nazionale GNGTS*, Roma.
- Clayton, R. W. & Comer, P., 1984. A tomographic analysis of mantle heterogeneities, *Terra Cognita*, **4**, 282–283.
- Hirahara, K. & Mikumo, T., 1980. Three-dimensional seismic structure of subducting lithospheric plates under the Japan Islands, *Phys. Earth planet. Inter.*, **21**, 109–119.
- Humphreys, E. & Clayton, R. W., 1988. Adaptation of back projection tomography to seismic travel time problems, *J. Geophys. Res.*, **93**, 1073–1085.
- Ivansson, S., 1984. Crosshole investigations—Tomography and its application to crosshole seismic measurements, *Thesis*, National Defence Research Institute, Stockholm.
- Kalbfleisch, J. K., 1979. *Probability and Statistical Inference*, Springer-Verlag, Berlin.
- McMechan, G. A., 1983. Seismic tomography in boreholes, *Geophys. J. R. astr. Soc.*, **74**, 601–612.
- Nolet, G., 1984. Global seismology and the investigation of deep continental structure, in *Problems and Methods for Lithospheric Exploration*, pp. 5–22, ed. Cassinis, R., Plenum Press, New York.
- Nolet, G., 1985. Solving or resolving inadequate and noisy tomographic systems, *J. Comp. Phys.*, **61**, 463–482.
- Paige, C. C. & Saunders, M. A., 1982. LSQR: an algorithm for sparse linear equations and sparse least-squares, *ACM Trans. Math. Soft.*, **8**, 43–71 and 195–209.
- Scales, J. & Gersztenkorn, A., 1988. Robust methods in inverse theory, *Inverse Problems*, **4**, 1071–1091.
- Spakman, W., 1988. Upper mantle delay time tomography, with an application to the collision zone of the Eurasian, African and Arabian plates, *Geologica Ultraiectina*, **53**, 299.
- Spakman, W. & Nolet, G., 1988. Imaging algorithms, accuracy and resolution in delay time tomography, in *Mathematical Geophysics: A Survey of Recent Developments in Seismology and Geodynamics*, pp. 155–187, eds Vlaar, N. J., Nolet, G., Wortel, M. J. R. & Cloetingh, S. A. P. L., Reidel, Dordrecht.
- Tarantola, A., 1987. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, Amsterdam.
- Van der Sluis, A. & Van der Vorst, H. A., 1987. Numerical solution of large, sparse linear systems arising from tomographic problems, in *Seismic Tomography*, pp. 53–87, ed. Nolet, G., Reidel, Dordrecht.
- Wiggins, R. A., 1972. The general inverse problem: implication of surface waves and free oscillation for Earth structure, *Rev. Geophys. Space Phys.*, **10**, 251–285.

APPENDIX

The three descriptors D , σ and α introduced in the text [see definitions (4a), (4b) and (7)] each refer to a particular property of the configuration of seismic rays. According to the definitions, all configuration descriptors are computed by means of a weighted mean over the partition cells, where the weighting schemes amplify the role of those cells that most affect the goodness of the inversion. As regards the mean ray density D and the dispersion index σ , the weights are given by

$$w_j = (1 + \beta)\delta_{\max} - \delta_j, \quad j = 1, \dots, p, \quad (\text{A1})$$

where β is a small quantity introduced to allow the cell with ray density equal to δ_{\max} to give a non-zero contribution to the total sum. In our computations β was assumed to be equal to 0.1. As will become apparent in the following, the above weighting distribution, which is a decreasing function of the ray density, favours the elimination of cells with the smallest ray concentration. The set of weights adopted in computing the anisotropy index is

$$w'_j = \delta_j^2, \quad j = 1, \dots, p; \quad (\text{A2})$$

they increase along the ray density of the cell. This means that it is much more important for the cells with the highest densities to have a good azimuthal coverage than for cells intersected by a smaller number of rays. In order to determine the optimal configuration it is more convenient to make use of a single quantity rather than a multiplicity of descriptors. This is obtained by first introducing the concept of the configuration target and then by taking the distance of the configuration from the target as a score with which to associate the actual ray configuration, where the distance is computed in an appropriate metric space. The above process may be formalized as follows. Let the actual configuration, say C , be represented by the three-component vector $C \equiv (D, \sigma, \alpha)$, and the target configuration, say C_t , by the vector $C_t \equiv (\delta_{\max}, \sigma = 0, \alpha = 0)$ where δ_{\max} is the upper end of the ray density distribution of C . Now, it is possible to introduce the distance between the configurations C and C_t , viewed as geometrical points in 3-D vector space. To this end, first we scale the ray density and the dispersion index with the aid of the scale factor δ_{\max} . In this metric space, the actual and the target configurations become $C \equiv (D/\delta_{\max}, \sigma/\delta_{\max}, \alpha)$ and $C_t \equiv (1, 0, 0)$, and their distance apart is assumed as the

actual configuration score, that is

$$\begin{aligned} S &= \sqrt{x_1^2 + x_2^2 + x_3^2}, \\ x_1 &= (D - \delta_{\max})/\delta_{\max}, \\ x_2 &= \sigma/\delta_{\max}, \\ x_3 &= \alpha. \end{aligned} \quad (\text{A3})$$

In the text, the quantities x_1 , x_2 , x_3 are often referred to as the components of the configuration score related to the ray density, the dispersion index and the anisotropy index in that order. The goal of the pre-selection algorithm is to search for the configuration with the minimum distance from the target \mathbf{C}_t . If \mathbf{C}^0 denotes the initial configuration, the algorithm finds a sequence of intermediate configurations \mathbf{C}_k approaching progressively \mathbf{C}_t and, in our terminology, the endpoint of the sequence is the final or the core

configuration. In general, all configurations \mathbf{C}_k have a mean density $D_k > D^0$, since the cells more prone to elimination are those with less rays, and there is a consequent stable increase of the density component of the score. However, as soon as $D_k < D^0$, it has been proven to be convenient to make use of a slightly different definition of the score for the k th configuration of the sequence, that is

$$\begin{aligned} S_k &= \sqrt{w_1 x_1^2 + w_2 x_2^2 + w_3 x_3^2}, \\ w_1 &= (D^0/D_k)^2, \\ w_2 &= w_3 = 3/(w_1 + 2). \end{aligned} \quad (\text{A4})$$

The above choice of weights practically adds a penalty to the configuration score, favouring the search for the seismic pattern with the highest concentration.