ANNALI DI GEOFISICA, VOL. XXXVII, N. 1, April 1994

Comparison of inversion methods in seismic tomography: application to tectonic structures in Northwestern Italy

Fabio Maggio([1]), **Fabrizio Malfanti**([2]), **Mario Bertero**([3]), **Marco Cattaneo**([2]) **and Claudio Eva**([2])
([1]) *CRS4, Cagliari, Italy*
([2]) *Dipartimento di Scienze della Terra, Università di Genova, Italy*
([3]) *Dipartimento di Fisica, Università di Genova and INFN, Genova, Italy*

Abstract
In this paper we apply various inversion methods to a set of teleseismic data collected by a network operating along the Ligurian Belt in the transition region between Alps and Apennines. In particular, we consider the regularization method, the truncated singular value decomposition, the Landweber method (with the Related Simultaneous Iterative Reconstruction Technique) and the conjugate gradient method. All the methods provide rather similar velocity models which are well approximated by that provided by back-projection (used with an appropriate normalization constant). A drawback of these models seems to be the large discrepancy (of the order of 40%) between the observed time residuals and those computed from the model itself. However, for each station of the network, the azimuth dependence of the computed time residuals reproduces rather well the observed one so that it is believable that the most significant information contained in the data has been exploited. The computed velocity models indicate strong heterogeneities in the first 200 km below the Apennines.

Key words *seismic tomography – tectonic structures – Northwestern Italy*

1. Introduction

Northwestern Italy, in particular the region extending from the western side of the Alpine arc to the Northern Apennines, is characterized by strong lateral heterogeneities in the lithosphere and the upper mantle, as revealed by gravimetric and seismic-wave propagation anomalies. From a seismological point of view, structural deep heterogeneities have been suggested by many authors (Baer 1980; Panza *et al.*, 1980; Babuska *et al.*, 1984; Spakman, 1986; Granet and Cara, 1988), using different approaches and data sets.

In previous papers (Cattaneo *et al.*, 1986; Cattaneo and Eva, 1990; Cattaneo and Spallarossa, 1991), we showed the presence of strong variations in the time residuals of teleseismic waves in the area, depending on both azimuth and distance. The size of the azimuthal variation of residuals has been interpreted in terms of high velocity contrast due to the presence of the «Ivrea Body», in the crust, and of Alpine and Apenninic «roots», reaching at least a depth of 200 km, in the upper mantle. In this paper we investigate the complex structure associated with the contact between the Alps and the Apennines, using teleseis-

mic data recorded by stations distributed along the Ligurian coast.

Since the geographic region involved has a rather small size and the number of seismic stations and events is not too large (roughly 10^3 arrival-time residuals), the tomographic problem we have to solve is of reasonable dimension. For this reason it is possible to apply to this problem several inversion methods, including those which are more expensive from the computational point of view. Therefore the scope of the paper is twofold: to derive information on the lateral heterogeneities of the region considered and to compare the ability of various inversion methods in the interpretation of seismic data.

Concerning the second point, different inversion methods could provide rather different solutions, because they are different methods for solving approximately the same basic equation which is severely ill-conditioned. A rather satisfactory result we have obtained is that all the inversion methods considered provide quite similar velocity models. A possible difficulty is that the discrepancy between the real and computed data using these models is rather large, of the order of 40%. Even the discrepancy corresponding to the generalized solution is of the same order of magnitude in spite of the fact that this solution, which is not physically reasonable because it is strongly influenced by experimental errors, is that giving the smallest discrepancy.

The previous result may indicate a large inconsistency between the models considered and the seismic data because this discrepancy is much larger than the estimated experimental errors ($\sim$ 10%). However, the computed time residuals have the same azimuth dependence of the observed ones for all the stations of the network. As a result, for each station the difference between the observed and the computed data is a fluctuating vector with approximately zero mean. This random behaviour may suggest that the experimental errors have been underestimated and that the most significant information contained in the time residuals has been exploited. Anyway further investigation of this phenomenon is required. It should also be interesting to verify if it is a common feature of all the models used in the interpretation of teleseismic data.

In section 2 we briefly describe the data we have used and, in section 3, the models we have considered. In section 4 we discuss the inversion methods we have applied to our problem. In section 5 we give the numerical results obtained, while their geophysical implications are discussed in section 6.

2. Data sources and analysis

The main source of teleseismic traveltime data used in this paper is the regional network operating along the Ligurian Belt and the South-western Alps, based at the University of Genova (Eva *et al.*, 1985; Cattaneo and Augliera, 1990). Also teleseismic data recorded by two temporary networks installed in the North-western Apennines (Augliera *et al.*, 1990) have been considered. Thus data from 23 stations (see fig 1a) have been analyzed all together. Although not all the stations were operating simultaneously all the time, it was possible to have a consistent set of data using the centralized net as a reference. The areal coverage of the network was not optimal: stations are mainly installed around the Ligurian Sea, leaving the Po Plain incompletely covered, because of the presence of very strong cultural noise (see fig. 1b).

After careful tests, a set of 166 teleseismic events ($40^o < \Delta < 90^o$), with high quality recorded signals, was selected, considering not only the signal-to-noise ratio, but also the azimuthal and distance coverage. Figure 2 shows the azimuthal distribution of the considered events. Most of the foci were concentrated on the borders of the Pacific Ocean and the Andes region. The absolute minimum of coverage in the

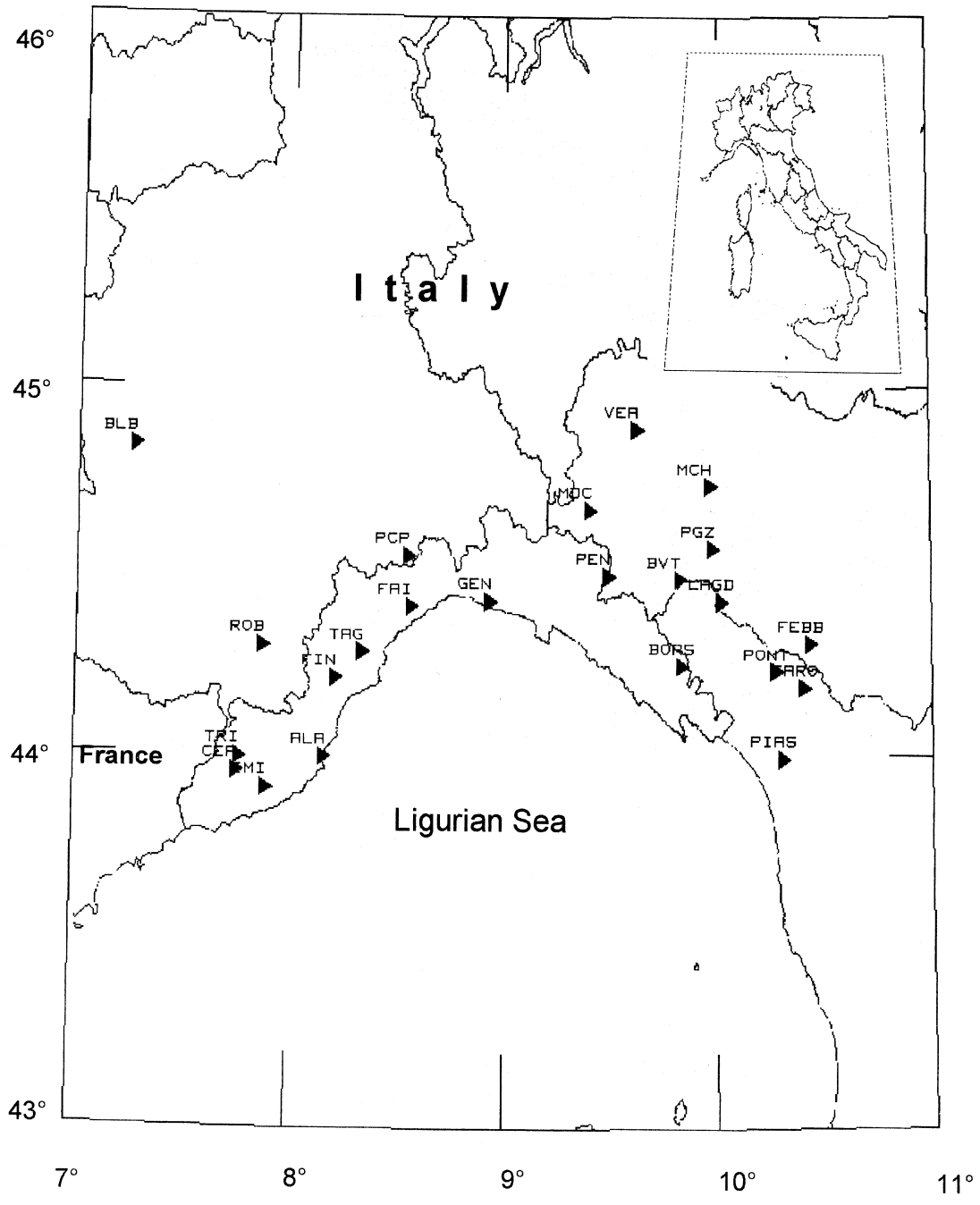


Fig. 1a. Map of the seismic stations used for this work.

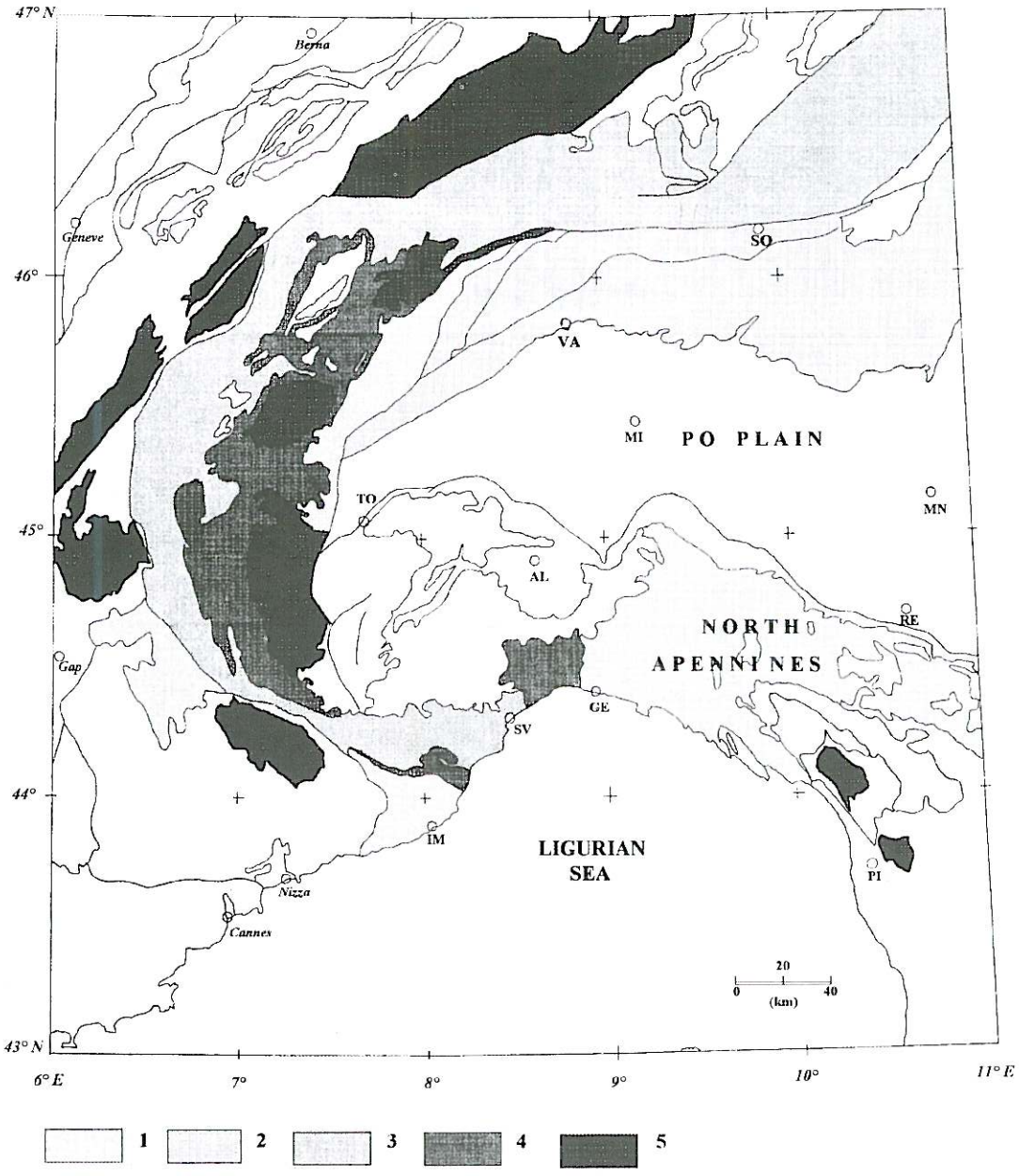


Fig. 1b. Structural sketch map of the Western Alps and their correlation with the North-western Apennines and the Ligurian sea: 1) Alpine flysh and Ligurides apenninic formations; 2) Southern alpine structures; 3) Pennidic units; 4) Brianconnais units; 5) Cristalline massifs.

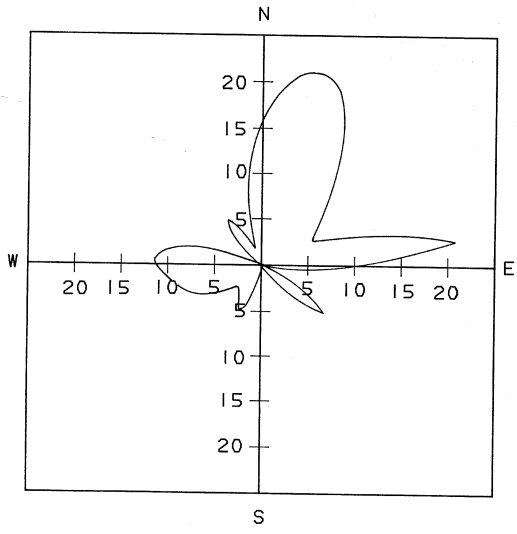


Fig. 2 Azimuthal distribution of the events considered in this paper.

southern sector is due to presence of the nearly aseismic African craton.

Since the analysis method requires the knowledge of the relative time residuals between the stations, all digital signals have been processed using a cross correlation operator applied on a small time-window around the first arrivals of pairs of seismic records (Cattaneo *et al.*, 1986; Cattaneo and Eva, 1990). The application of this operator allows to identify coherent phases and to determine the relative time shift needed to obtain the maximum likeness between them.

The procedure (see fig. 3a-c) can be synthesized in the following in steps:

a) selection of a set of seismic traces relative to the teleseismic event recorded by the network;

b) consideration of a suitable time window centered around the P-phase (less than 5 s of record);

c) evaluation of the phase shift (τ_i) between station pairs (generally assuming as reference the master station of the net), applying the cross-correlation operator;

d) realignment of all traces, performing a time shift (τ_i) between the traces, to check the obtained result (in this way, all arrival-time differences are related to the «master» station).

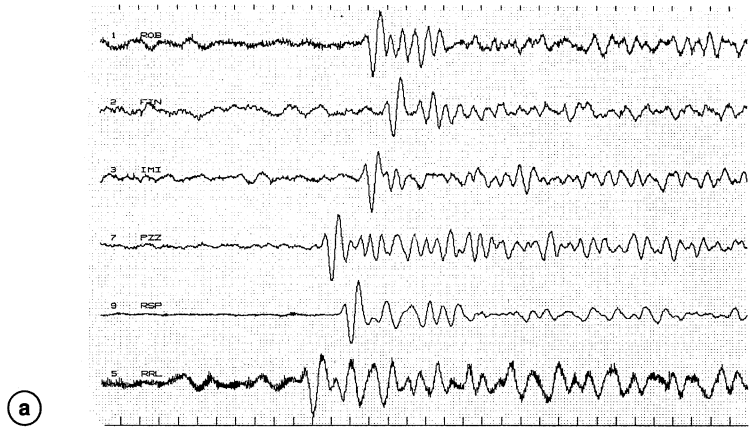
This method is more effective than a direct reading of arrival times which, in presence of noise, can be uncertain. In fact, in this way, the evaluation of relative residuals becomes nearly independent of the determination of the first motion arrival time, which is strongly influenced by the signal-to-noise ratio. The procedure also increases the time-difference resolution that, in some cases, can reach some hundredths of a second. The transformation of arrival-time differences into relative residuals was easily obtained once the earthquake focus was determined and a mean reference velocity model was adopted.

As is known, the main advantage of working with relative residuals is that errors in event location and in long-range velocity models only marginally affect the result.

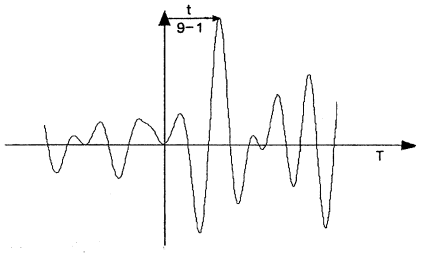
3. The models: singular system analysis

Since we consider teleseismic data and we use the classical ACH method (Aki *et al.*, 1977), only the last part of the ray path, immediately beneath the network, is modelled while the Earth's structure outside this volume is assumed to be known. Clearly this assumption is justified only for network dimensions small compared with the mean length of the ray paths, but this condition is satisfied in the case of the data considered in this paper. Velocity perturbations are parametrized starting with homogeneous layers of constant average velocity and dividing each layer into blocks of constant dimensions.

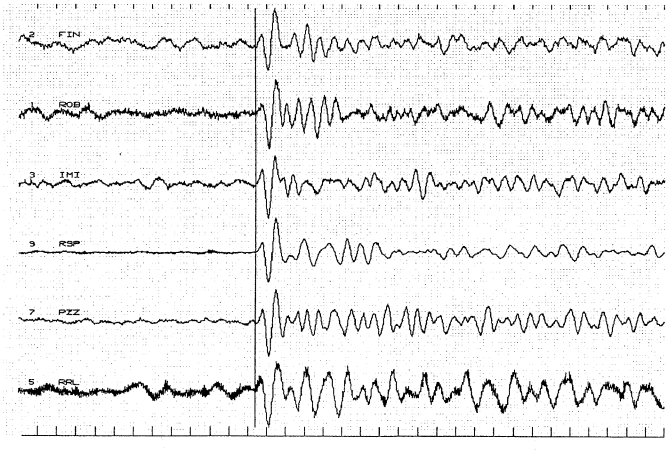
The volume considered corresponds to the geographical area of fig. 1a and its dimensions are 280 km in the E-W direction,



a



b



c

Fig. 3a-c. a) Seismic traces of the same teleseismic event recorded at some different stations of the network; b) cross-correlation function between traces 1-9; c) realignment of the traces of the plot (a).

Table I. Values of the parameters characterizing the models considered in this section. In the last column we also give the values of the sampling index which is defined as the ratio between the number of sampled blocks and the total number of blocks.

Model	Layers	Lateral sizes (km)	Number of blocks	Number of sampled blocks	Sampling index
1	4	35×35	192	139	72.4%
2	6	35×35	288	211	73.3%
3	4	23×23	432	270	62.5%
4	6	23×23	648	422	65.1%
5	4	17×17	768	433	56.4%
6	6	17×17	1152	660	57.3%

210 km in the N-S direction and 200 km in depth. Concerning the choice of the layers we have considered two cases: four layers (layer depths: 40 km for the first one, 50 km for the second and third, 60 km for the last one) and six layers (layer depths: 30 km for the first four layers and 40 km for the last two layers). For each one of these two cases we have considered three different choices of lateral dimensions of the blocks, as summarized in table I.

The initial 1-D velocity distribution, used for computing the travel-times corresponding to the six layer models, is shown in fig. 4 (full line). This is an approximation of the preliminary reference Earth model given in (Dziewonski and Anderson, 1981), which is also reported in fig. 4 (dotted line).

In section 5 we will present our numerical results mainly for the case of model 4. For this model the size of the blocks is still sufficiently large with respect to the wavelength of the seismic waves but the resolution in the lateral directions is better than that of the models 1 and 2 while the resolution in depth is better than that of model 3. Moreover, as follows from fig. 5, the blocks of the last three layers are sufficiently sampled and the sampling is rather uniform. This is not true for models 5 and 6.

According to the ACH method, for each model we must solve the linear algebraic system:

$$Gs = t \tag{3.1}$$

where the unknown s is the vector whose components are the slowness anomalies of the blocks and the data t is the vector whose components are the travel-time residuals minus their average over all the station for each event. The matrix G can be computed from the travel-times derived from the velocity distribution of fig. 4. The unit of both G and t is the second.

In order to reduce the dimension of the problem, it is convenient to remove the unsampled blocks since they correspond to columns of zeros in the matrix G. This is a rectangular matrix with m rows (number of data) and n columns (number of sampled blocks). As follows from table I, we have $m > n$ for all the models we have considered. The rank p of G is smaller than n. In fact, as proved in (Aki *et al.*, 1977) we have $p \leq n - N_L$, where N_L is the number of layers.

The matrix G is also ill-conditioned and its ill-conditioning can be investigated by computing its singular system. This task is possible in our case since the dimensions of G are not too large (in the case of model 6 it is 1077 × 660).

The singular system of G, however, depends on the definition of the norms of the

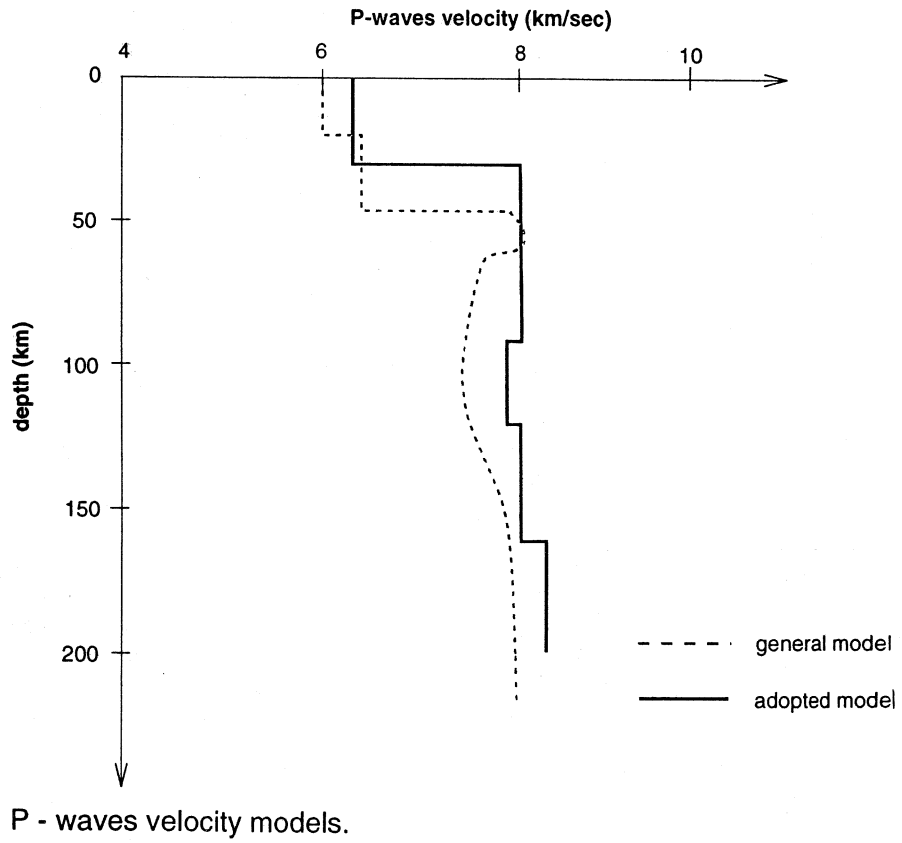


Fig. 4. Velocity distribution of the preliminary reference Earth model given in (Dziewonski and Anderson, 1981) (dotted line) compared with the approximation used as initial 1-D velocity distribution for the six layer models.

solution and data vectors. In general we can consider weighted Euclidean norms of the following form:

$$\|\boldsymbol{s}\|^2 = \sum_{j=1}^{n} \gamma_j \mid s_j \mid^2, \quad \|\boldsymbol{t}\|^2 = \sum_{i=1}^{m} \frac{1}{\varrho_i} \mid t_i \mid^2 \tag{3.2}$$

The simplest choice is, of course, $\gamma_j = 1$ and $\varrho_i = 1$, while the choice corresponding to the Simultaneous Iterative Reconstruction Technique (SIRT) is (van der Sluis and van der Vorst, 1987)

$$\gamma_j = \sum_{i=1}^{m} \mid G_{ij} \mid^{\alpha}, \quad \varrho_i = \sum_{j=1}^{n} \mid G_{ij} \mid^{2-\alpha} \tag{3.3}$$

where α is a given parameter satisfying the condition $0 \leq \alpha \leq 2$. In the following these will be called the SIRT weights.

If we introduce the matrices

$$C = \text{diag}\ \{\gamma_j\}, \quad R = \text{diag}\ \{\varrho_i\} \tag{3.4}$$

then the adjoint matrix G^* is given by

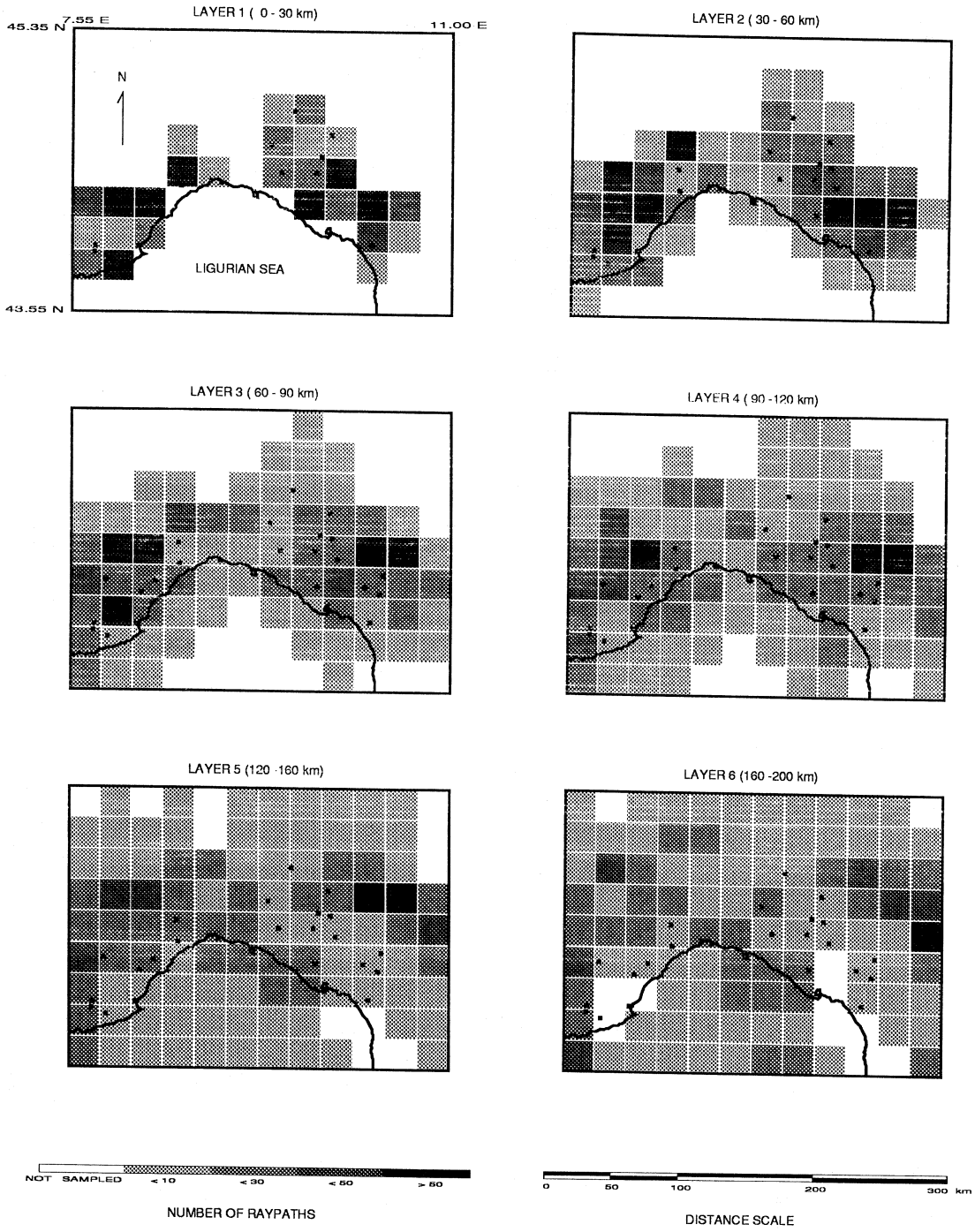


Fig. 5. Distribution of the sampling of the blocks of model 4, table I.

Table II. Parameters characterizing the matrices G corresponding to the models of table I. In the column of the values of the rank p we also give (in brackets) the values of $n - N_L$. In the third and fourth column we report the values of the condition numbers σ_1/σ_p both for unit weights and for SIRT weights. In the fifth and sixth column we give the values of the effective ranks corresponding to 10% data errors.

Model	p	(σ_1/σ_p)	$(\sigma_1/\sigma_p)^{\mathrm{SIRT}}$	p_{eff}	$p_{\mathrm{eff}}^{\mathrm{SIRT}}$
1	134 (135)	65.8	45.5	92	126
2	203 (205)	1.85×10^{4}	4.36×10^{3}	110	167
3	260 (266)	2.36×10^{5}	4.54×10^{4}	164	231
4	390 (416)	2.69×10^{10}	5.14×10^{9}	190	295
5	400 (429)	2.93×10^{8}	8.53×10^{7}	248	332
6	559 (654)	1.41×10^{13}	4.90×10^{13}	278	403

$$G^* = C^{-1}G^TR^{-1} \tag{3.5}$$

and the singular system of G is defined as the set of the solutions $\{\sigma_k; \boldsymbol{v}_k, \boldsymbol{u}_k\}$ of the homogeneous equations

$$G\boldsymbol{v}_k = \sigma_k\boldsymbol{u}_k, \quad G^*\boldsymbol{u}_k = \sigma_k\boldsymbol{v}_k; \quad k = 0, 1, \ldots, p \tag{3.6}$$

When the weights are not identically one, this singular system can be computed by means of standard routines (which apply to the case of the usual Euclidean norm) if one rescales the singular vectors and the matrix as follows (van der Sluis and van der Vorst, 1987)

$$\boldsymbol{x}_k = C^{1/2}\boldsymbol{v}_k, \quad \boldsymbol{y}_k = R^{-1/2}\boldsymbol{u}_k, \quad L = R^{-1/2}GC^{-1/2} \tag{3.7}$$

so that the eq. (3.6) become

$$L\boldsymbol{x}_k = \sigma_k\boldsymbol{y}_k, \quad L^T\boldsymbol{y}_k = \sigma_k\boldsymbol{x}_k; \quad k = 0, 1, \ldots, p \tag{3.8}$$

We point out that in the case of the SIRT weights the singular values (as well as the matrix L) are dimensionless while in the case of unit weights (*i.e.* $L = G$), the unit of the σ_k is the second.

We have computed the singular system for all the models of table I both in the case of unit weights and in the case of the SIRT weights. The singular value spectra in the case of unit weights are plotted in fig. 6a,b. In the case of the SIRT weights we have found that the singular values do not strongly depend on α. For example, the first singular value of the matrix G corresponding to model 4 is in the range $0.45 \div 0.58$ for α between 0 and 2. The largest value of σ_1 corresponds approximately to $\alpha = 1$ and the behaviour of σ_1 as a function of α is roughly symmetric with respect to this point. Also the behaviour of the singular value spectrum is roughly independent of the value of α. For this reason the results reported in table II correspond to the case $\alpha = 1$.

We already know that the matrix G is rank deficient since $p \leq n - N_L$. Therefore we have tried to determine numerically its rank because, in this way, it is also possible to determine the true condition number of G, *i.e.* the ratio σ_1/σ_p between the largest and the smallest singular value. In fact, due to the finite precision of the computer, the number of singular values of G different from zero can be just n so that the spurious singular values can be detected only by changing the precision of the computation.

For all the models we have found several

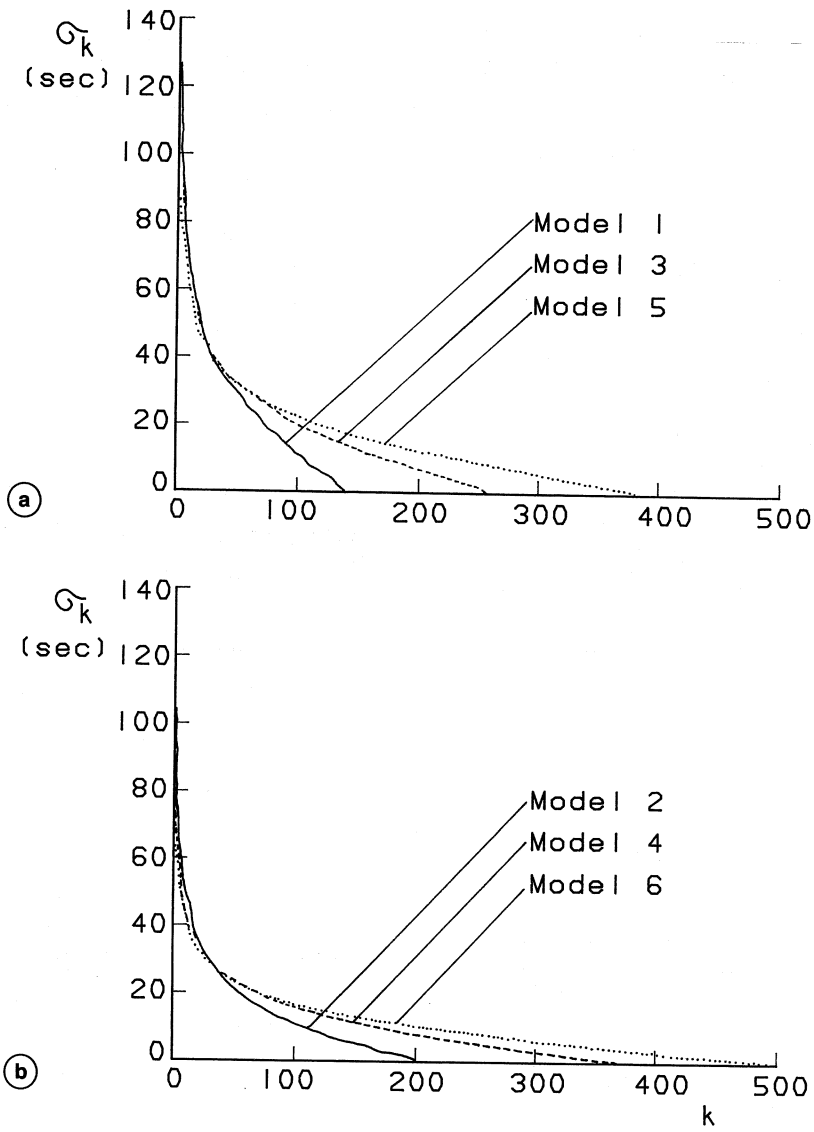


Fig. 6a,b. Singular value spectra for the models of table I: a) model 1 (full line), 3 (dashed line), 5 (dotted line); b) model 2 (full line), 4 (dashed line), 6 (dotted line).

singular values of the order of the computer precision ($\delta = 10^{-15}$). For example, in the case of model 4, we have $n - N^L = 416$ but only 390 singular values are $> 10^{-13}$ in the case of unit weights and $> 10^{-16}$ in the case of SIRT weights. We have recomputed the singular values with a greater precision and we have found that the first 390 singular values are stable while the others decrease and are still comparable with

the new computer precision (for instance, $\delta = 10^{-30}$). Therefore we conclude that, in such a case, $p = 390$.

Once the rank of the matrix has been determined, one can compute its condition number σ_1/σ_p, *i.e.* the parameter controlling error propagation from the data to the solution. We find that the ill-conditioning of the matrix G increases with the number of blocks of the model. Moreover, for a given model, the ill-conditioning corresponding to SIRT weights is smaller than that corresponding to unit weights.

Another interesting parameter is the «effective rank» of the matrix, which can be defined as the number of singular values such that σ_1 / σ_k is smaller than the inverse of the relative error on the data. This effective rank is, in general, much smaller than the rank of the matrix. The values reported in table II correspond to 10% data errors, *i.e.,* the number of singular values σ_k such that $\sigma_1/\sigma_k \leq 10$. As follows from this table the effective rank corresponding to SIRT weights is larger than the effective rank corresponding to unit weights.

4. The regularized inversion methods

In this section we give a short account of the methods we have used. Even if they are well documented both in the mathematical and in the geophysical literature (see, for instance, Bertero, 1989; van der Sluis and van der Vorst, 1987), where some of them are reviewed), we wish to stress the relationship between these methods since each of them can be considered a filtering of the generalized solution. This is the solution of minimal norm of the least squares equation

$$\tilde{G}\boldsymbol{s} = G^*\boldsymbol{t} \tag{4.1}$$

where $\tilde{G}$ is the self-adjoint $m \times m$ matrix

$$\tilde{G} = G^*G \tag{4.2}$$

and G^* defined in eq. (3.5). In terms of the singular system of G it is given by

$$\boldsymbol{s}^+ = \sum_{k=1}^{p} \frac{1}{\sigma_k} (\boldsymbol{t}, \boldsymbol{u}_k)\, \boldsymbol{v}_k \tag{4.3}$$

where $(\boldsymbol{t}, \boldsymbol{u}_k)$ is the scalar product associated with the norm (3.2) and p is the rank of the matrix. This solution is numerically unstable since the condition number of G, cond $(G) = \sigma_1/\sigma_p$, is, in general, rather large, as follows from table II.

A blurred solution, however, can be easily obtained by means of the so-called back-projection method. We define the back-projected solution $\boldsymbol{s}_{BP}$ as follows:

$$\boldsymbol{s}_{BP} = \frac{1}{\sigma_1^2} G^*\boldsymbol{t} \tag{4.4}$$

For its computation we only need the largest singular value σ_1, which can eventually be computed (without computing the complete singular system) using the power method (Ralston, 1965). If we represent $\boldsymbol{t}$ as a linear combination of the singular vectors $\boldsymbol{u}_k$, we find that

$$\boldsymbol{s}_{BP} = \sum_{k=1}^{p} \left(\frac{\sigma_k}{\sigma_1}\right)^2 \frac{1}{\sigma_k} (\boldsymbol{t}, \boldsymbol{u}_k)\, \boldsymbol{v}_k \tag{4.5}$$

and therefore $\boldsymbol{s}_{BP}$ is a filtered version of the generalized solution $\boldsymbol{s}^+$

$$\boldsymbol{s}_{BP} = W_{BP}\boldsymbol{s}^+ \tag{4.6}$$

with a filter W_{BP} given by

$$W_{BP} = \text{diag}\left\{\left(\frac{\sigma_k}{\sigma_1}\right)^2\right\} \tag{4.7}$$

In the case of an ill-conditioned problem the back-projection provides a very rough approximation of the unknown solution. More refined approximations can be obtained using methods of the so-called *reg-*

ularization theory (Tikhonov and Arsenine, 1977).

4.1. *Tikhonov regularized solution*

This is given by

$$\boldsymbol{s}_\lambda = (\tilde{G} + \lambda I)^{-1} G^* \boldsymbol{t} \qquad (4.8)$$

with $\tilde{G}$ defined in eq. (4.2) and G^* defined in eq. (3.5). The parameter λ is called the regularization parameter (damping parameter in geophysical literature).

When $\gamma_j = 1$, $\varrho_i = 1$, the regularized solution (4.8) is just the well-known damped least squares (DLS) solution. When γ_j and ϱ_i are defined as in eq. (3.3) then the solution (4.8) is the limit of the DSIRT algorithm considered in (Trampert and Leveque, 1990).

If we write $\boldsymbol{s}_\lambda$ in terms of the singular system of G, then we easily obtain that $\boldsymbol{s}_\lambda$ is a filtered version of $\boldsymbol{s}^+$

$$\boldsymbol{s}_\lambda = W_\lambda \boldsymbol{s}^+ \qquad (4.9)$$

where

$$W_\lambda = \operatorname{diag}\left\{\frac{\sigma_k^2}{\sigma_k^2 + \lambda}\right\} \qquad (4.10)$$

It is obvious that the diagonal elements of W_λ are of the order of 1 if $\sigma_k^2 \gg \lambda$ and are negligible when $\sigma_k^2 \ll \lambda$.

The optimum value of λ is related to the inverse of the signal-to-noise ratio (see Bertero, 1989 for a review). When this quantity is not known, a procedure, which is used very often in practice, consists of computing the value of λ which minimizes the so called generalized cross-validation (GCV) function (Craven and Wahba, 1979), defined by

$$V(\lambda) = \frac{\|\boldsymbol{t} - G\boldsymbol{s}_\lambda\|^2}{[Tr[I - G(\lambda)]]^2} \qquad (4.11)$$

where

$$G(\lambda) = G(\tilde{G} + \lambda I)^{-1} G^* = \hat{G}(\hat{G} + \lambda I)^{-1} \qquad (4.12)$$

and $\hat{G} = GG^*$ (see also Bertero, 1989 for a review).

We have also considered the case of a regularized solution defined as the minimizer of the following functional:

$$\Phi_\lambda[\boldsymbol{s}] = \|G\boldsymbol{s} - \boldsymbol{t}\|^2 + \lambda(\|\boldsymbol{s}\|^2 + \|H_1\boldsymbol{s}\|^2 + \|H_2\boldsymbol{s}\|^2 + \|H_3\boldsymbol{s}\|^2) \qquad (4.13)$$

where H_1 is the matrix which associates to each block the variation of its slowness with respect to that of the preceding one in the x-direction; H_2 and H_3 have a similar designation relative to variations in the y- and z-directions. No variation is associated to a boundary block in the directions orthogonal to its boundary surfaces. In short, the functional (4.13) implies regularization with a constraint on the gradient of the slowness function.

By means of a suitable change of variables (Seatzu, 1986) it is possible to write the minimizer of the functional (4.13) in the form (4.8) with a different definition of the matrices G and G^*. Again the estimate of λ can be obtained by means of the GCV criterion.

4.2. *Truncated SVD*

The method of truncated Singular Value Decomposition (SVD) (Bertero, 1989) consists in looking for an approximate solution given by

$$\boldsymbol{s}_K = \sum_{k=1}^{K} \frac{1}{\sigma_k}(\boldsymbol{t}, \boldsymbol{u}_k)\boldsymbol{v}_k \qquad (4.14)$$

with $K < p$. The number of terms plays here a role similar to that of the regularization parameter. As concerns the choice of K, if the signal-to-noise ratio is known, one keeps in eq. (4.14) only the terms corre-

sponding to singular values σ_k greater than the inverse of the signal-to-noise ratio. This solution corresponds to replacing the filter (4.10) by a squared filter of the type 1 or 0.

4.3. *Landweber iterative method*

This method, introduced in (Landweber, 1951), refined in (Bialy, 1959) and known in Russian literature as the *method of successive approximations*, is a generalization of the method known as SIRT in geophysical literature. If $\boldsymbol{s}^{(N)}$ is the result of the N-th iteration and if $\boldsymbol{r}^{(N)}$ is the corresponding discrepancy vector

$$\boldsymbol{r}^{(N)} = \boldsymbol{t} - G\boldsymbol{s}^{(N)} \tag{4.15}$$

then the iterative scheme is

$$\boldsymbol{s}^{(N+1)} = \boldsymbol{s}^{(N)} + \omega G^* \boldsymbol{r}^{(N)} \tag{4.16}$$

where ω, the so called *relaxation parameter*, is a fixed number satisfying the condition

$$0 < \omega < \frac{2}{\sigma_1^2} \tag{4.17}$$

If this condition is satisfied and $\boldsymbol{s}^{(0)} = 0$, then $\boldsymbol{s}^{(N)}$ converges to $\boldsymbol{s}^+$ in the limit $N \to \infty$. For a finite N, the result of the N-th iteration can be expressed as a filtering of $\boldsymbol{s}^+$ as follows:

$$\boldsymbol{s}^{(N)} = W^{(N)} \boldsymbol{s}^+ \tag{4.18}$$

where

$$W^{(N)} = \text{diag}\,\{1 - (1 - \omega\sigma_k^2)^N\} \tag{4.19}$$

It follows that the number of iterations plays the role of a regularization parameter. In such a case one needs a «stopping rule» but unfortunately no practical criterion seems to be available (except the obvious one of stopping the iterations when one finds a satisfactory solution).

When G^* is given by eqs. (3.3) and (3.5), then the iteration scheme (4.16) is identical to that of SIRT and condition (4.17) can be replaced by $0 < \omega < 2$ since $\sigma_1 \leq 1$ (van der Sluis and van der Vorst, 1987).

We also note that, if $\boldsymbol{s}^{(0)} = 0$ and $\omega = 1/\sigma_1^2$, then $\boldsymbol{s}^{(1)} = \boldsymbol{s}_{BP}$, the back-projected solution.

4.4. *Conjugate gradient*

This is a projection method that can be regarded as an iterative method for the determination of the solution of minimal norm of eq. (4.1). The important feature of this method is that it has a regularizing effect similar to that of the Landweber method, *i.e.* the number of iterations plays the role of a regularization parameter. Moreover, this method can be more practical than the Landweber method because, in general, it requires a smaller number of iterations for obtaining a reasonable approximate solution. The iterative scheme we have used is that given in (van der Sluis and van der Vorst, 1987).

Again the result of the N-th iteration can be described as a filtering of the generalized solution. This filter is data-dependent and it can be computed as follows: if $\theta_1^{(N)}$, $\theta_2^{(N)}$, ..., $\theta_N^{(N)}$ are the Ritz values associated with the N-th iteration (see van der Sluis and van der Vorst, 1987) and if $R_N(t)$ is the Ritz polynomial defined by

$$R_N(t) = \left(\prod_{j=1}^{N} \theta_j^{(N)}\right)^{-1} \prod_{j=1}^{N} (\theta_j^{(N)} - t) \tag{4.20}$$

then the result $\boldsymbol{s}^{(N)}$ of the N-th iteration can be written as follows:

$$\boldsymbol{s}^{(N)} = W^{(N)} \boldsymbol{s}^+ \tag{4.21}$$

where

$$W^{(N)} = \text{diag}\,\{1 - R_N(\sigma_k^2)\} \tag{4.22}$$

Finally, if $\tilde{s}$ is some approximate solution obtained by one of the previous methods, we define as *relative discrepancy* δ, associated with $\tilde{s}$, the quantity

$$\delta = \frac{\|G\tilde{s} - t\|}{\|t\|} \tag{4.23}$$

i.e. the ratio between the norm of the discrepancy vector and the norm of the data vector. It is obvious that the generalized solution is that corresponding to minimal discrepancy.

5. Numerical results

Since we have determined the rank of the matrix G and we have computed its singular system, for all the models of section 2 we can compute the generalized solution (4.3). As follows from table II, except in the case of model 1, this solution is highly ill-conditioned and the ill-conditioning increases by increasing the number of blocks. Therefore these solutions are more and more affected by numerical instability: the corresponding velocity models are not physically reasonable due to the dramatic effect of the propagation of data errors.

It is well known, however, that among all the approximate solutions of eq. (3.1), s^+ is that providing the *minimum* value of the relative discrepancy (4.23). Now, in the case of model 6, *i.e.* the model with the highest resolution we find $\delta = 0.31$ while in the case of model 4 we find $\delta = 0.45$.

Since we believe that the relative RMS error on the data is of the order of 0.1 (*i.e.* 10%) or less, the previous result seems to imply that models consisting of blocks of constant slowness are not able to reproduce the data within experimental errors. Moreover, it also implies that it is not possible to use in such a case a rather well-known method for the determination of the regularization parameter λ, eq. (4.8), known as *discrepancy principle* (Morozov, 1968), which consists of looking for a value of this parameter such that the discrepancy is of the order of the experimental errors. In fact since the relative discrepancy corresponding to the generalized solution is > 0.1, it follows that no regularized solution exists satisfying the previous criterion.

It is possible, however, that experimental errors have been underestimated. For each station the observed time residuals have a rather typical azimuth dependence. An example is given in fig. 7 where the time residuals of the Apenninic station (SARO) are represented in the central plot. The azimuthal dependence is clear as well as a systematic variation with the incidence angle (events with different incident angles are represented with different symbols, the decreasing number of edges of the symbols indicating an increase of the distance of the event). In the same figure the bottom plot represents the time residuals computed by means of the propagation model which will be discussed in the following. It is quite evident that most of the significant trends of the experimental data are correctly reproduced. Finally the top plot represents the difference between the computed and the experimental time residuals, *i.e.* the components of the discrepancy vector corresponding to this station. They show a nearly random behaviour, as expected for a vector representing experimental errors, with an average value which is approximately zero. Quite similar results are obtained for the other stations, indicating that the computed velocity model, in spite of the large discrepancy, is able to reproduce the most significant information contained in the data.

We have applied all the methods described in section 4 to all the models discussed in section 2. For each one of these models all the methods have provided quite similar results with comparable discrepancies. Therefore a detailed description of these results is not necessary. Moreover among all the models, model 4 is that having a reasonable resolution both in lateral directions and in depth and we will report our results mainly in this case.

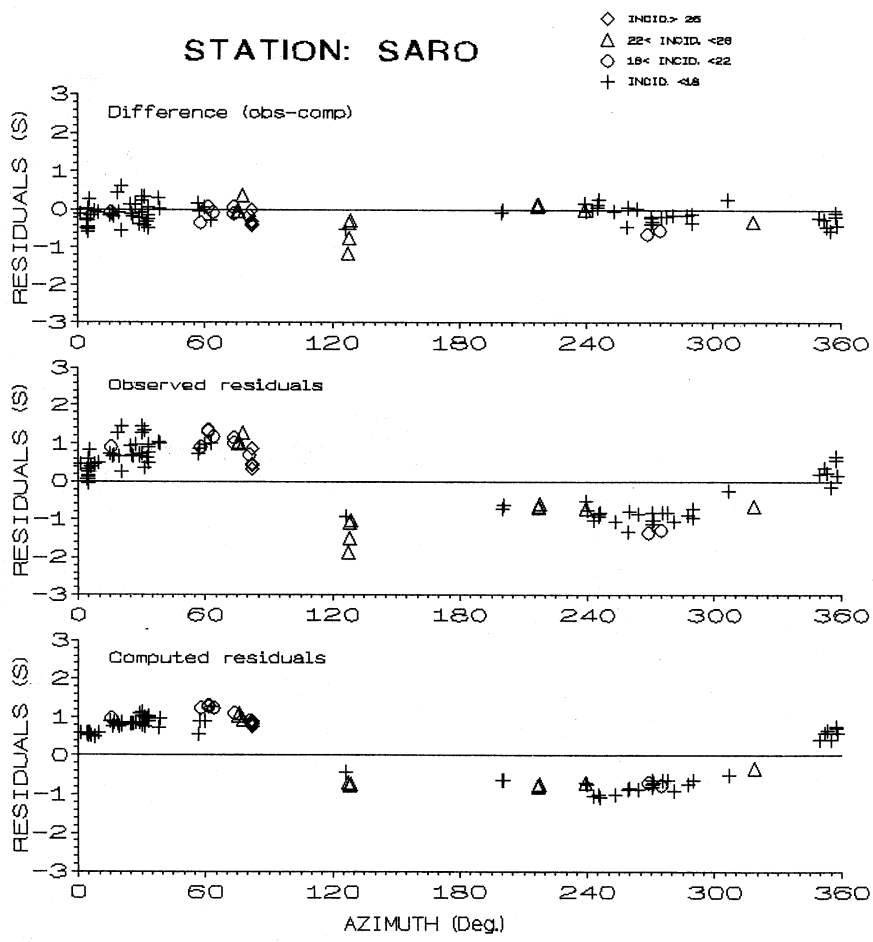


Fig. 7. Dependence of the time residuals on the azimuth for one of the stations of the network. The dependence on the incidence angle is also indicated by representing events with different incidence angles by means of different symbols (the number of edges of the symbol decreases if the distance of the event increases). The observed residuals are represented in the middle plot, the computed residuals are represented in the lower plot while the difference between observed and computed residuals is represented in the upper plot.

A first remark is that a reasonable solution is already provided by the back-projection method as defined by eq. (4.4). This can be obtained very easily since it only requires a matrix-vector multiplication once the largest singular value σ_1 is known. The corresponding relative discrepancy is $\delta = 0.63$. This value is rather high but comparable with the value corresponding to $\boldsymbol{s}^+$. The back-projected solution is given is fig. 8.

As follows from eq. (4.5) $\boldsymbol{s}_{BP}$ is a strongly filtered version of $\boldsymbol{s}^+$ and therefore it is a very «blurred» solution. In spite of this fact, it clearly indicates a structure which will be discussed in the next section. We do not give the back-projected solution

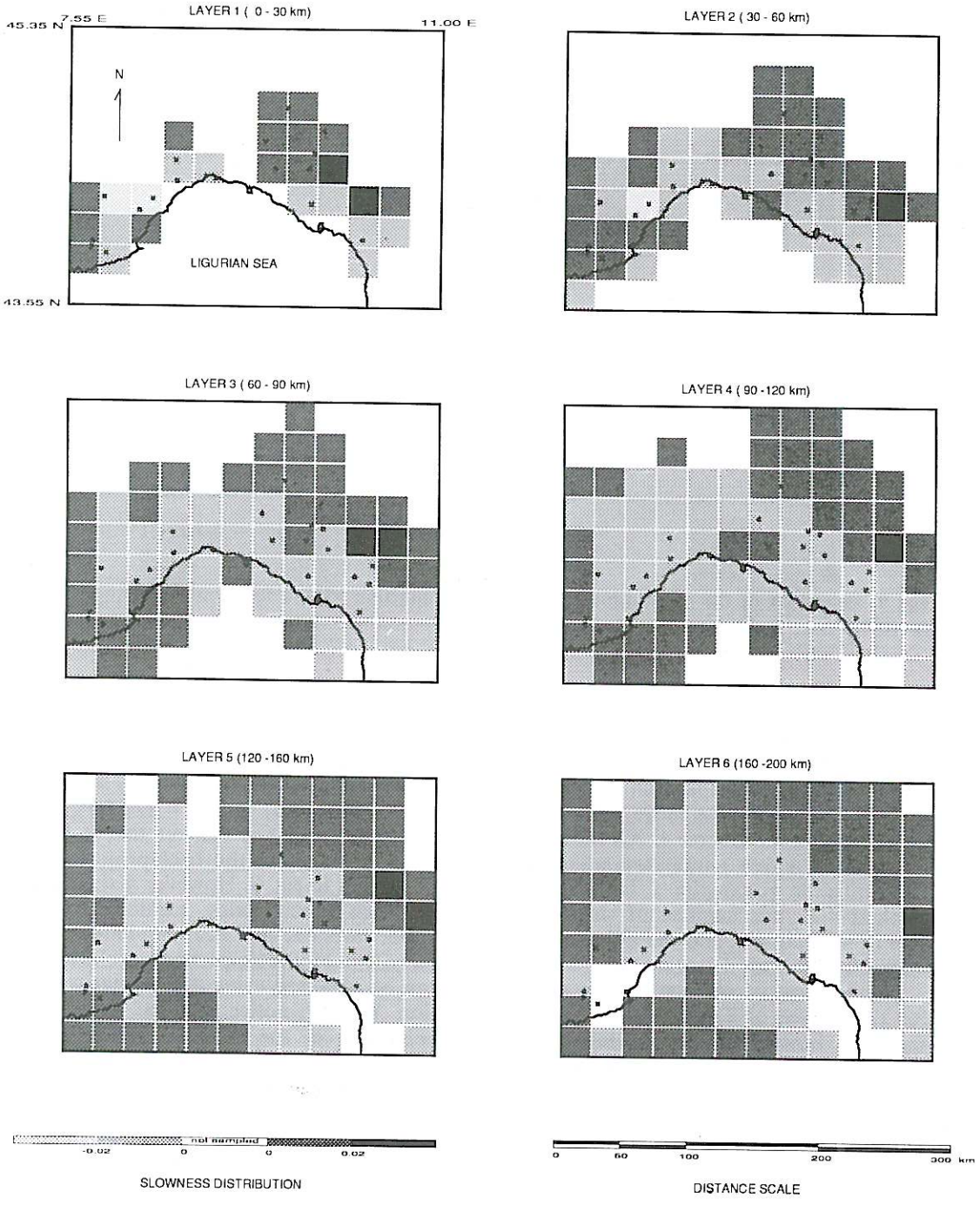


Fig. 8. Slowness distribution provided by the back-projection solution.

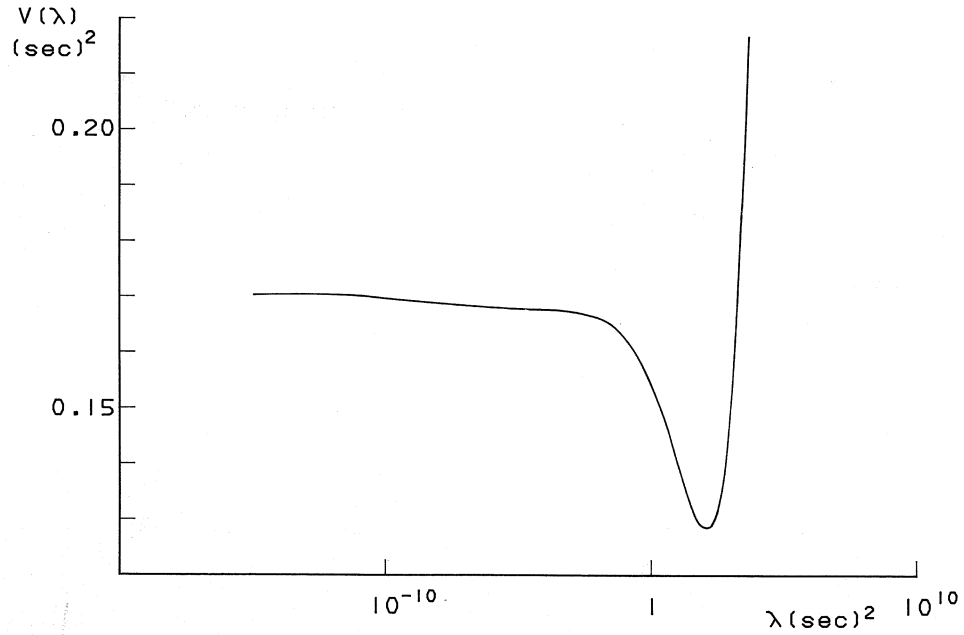


Fig. 9. Behaviour of the generalized cross validation function $V(\lambda)$, a function of the regularization parameter λ.

obtained using the SIRT weights because this is very similar to that of fig. 8.

Since the back-projected solution is quite blurred, we expect an improvement of resolution using the more refined methods described in section 4. All these methods contain a free parameter, which is the regularization parameter λ in the case of method 4.1., the number of singular vectors in method 4.2., and the number of iterations in methods 4.3. and 4.4.

In the case of method 4.1., a criterion for the choice of λ consists of minimizing the GCV function $V(\lambda)$ given in eq. (4.11). This can be easily computed once the singular system of the matrix G is known. In general $V(\lambda)$ is very flat and therefore very difficult to minimize. Moreover it can have several local minima. In our case we have verified the flatness of $V(\lambda)$ but we have found only one minimum at $\lambda = 122\ (\mathrm{s})^2$. In fig. 9 we plot the behaviour of $V(\lambda)$ as a function of log λ. As we see, in the wide range considered we obtain variations only on the second digit of the values of $V(\lambda)$.

The regularized (or damped) solution, as given by eq. (4.8) with $\lambda = 122\ (\mathrm{s})^2$, is plotted in fig. 10. It is evident that it has the same structure of the back-projected solution given in fig. 8. However a shrinking of the high-velocity body, particularly in the upper layers is also evident. This is clearly an effect of the improved resolution that can be obtained using the regularized solution. We also notice that the more regular behaviour of the solution in the deepest layers with respect to that in the upper layers is not a result of the smoothing of the inversion algorithm, as demonstrated by numerical simulations described in (Cattaneo and Eva, 1990). In fig. 11 we present an attempt of 3D-representation of the high-velocity body indicated by the previous results.

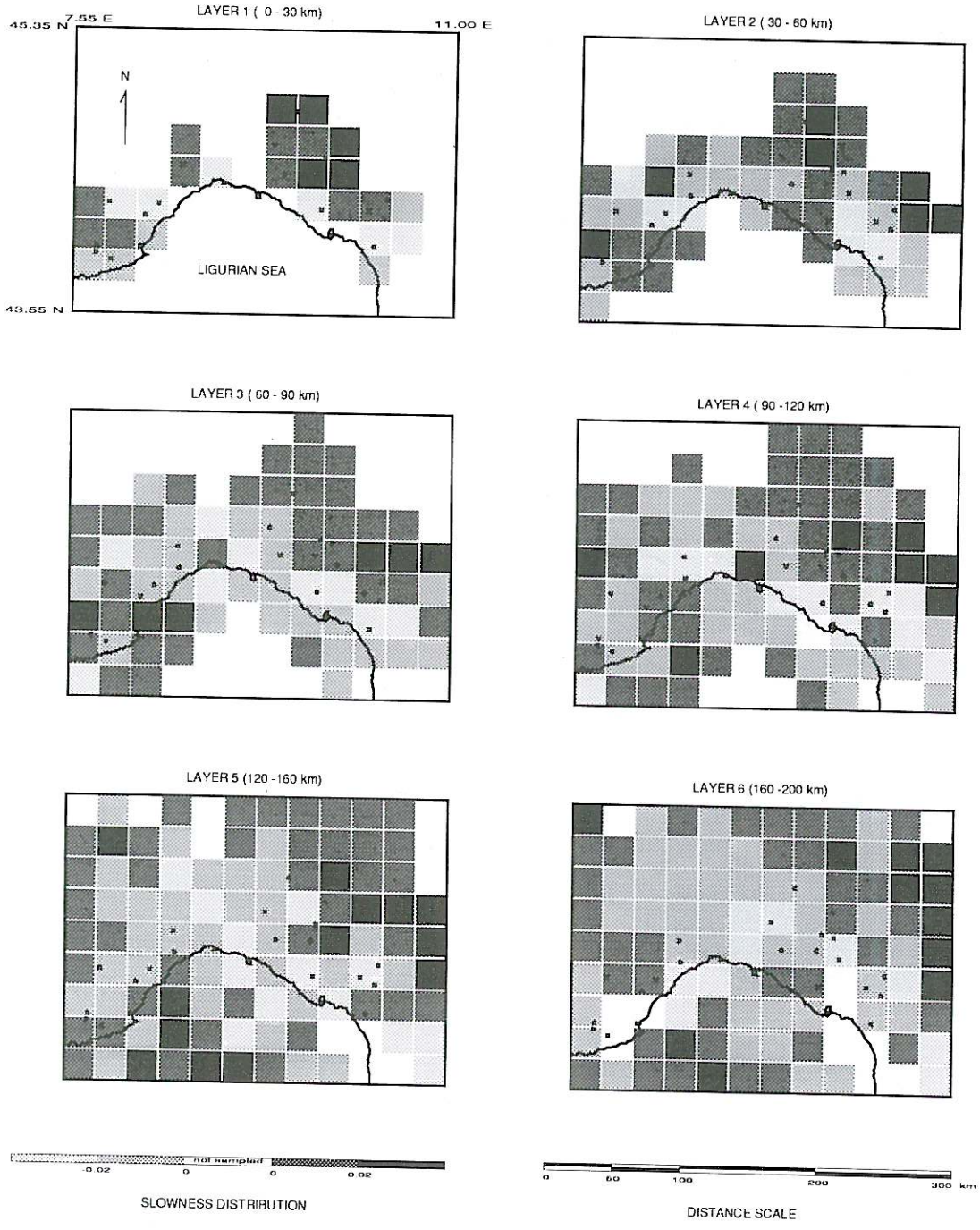


Fig. 10. Slowness distribution provided by the regularized solution with the value of λ minimizing $V(\lambda)$ (see fig. 9).

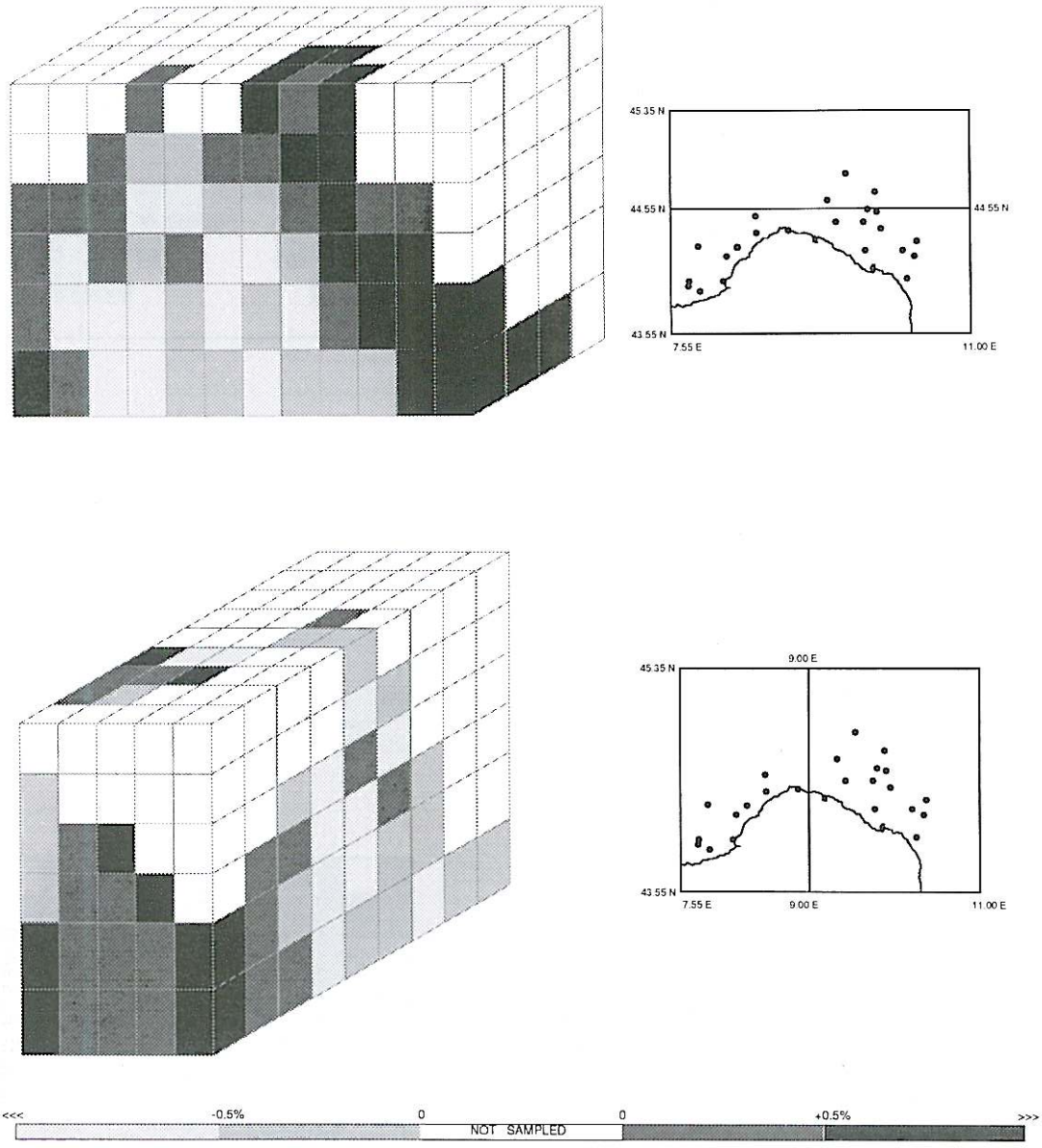


Fig. 11. 3D representation of the propagation model provided by the back-projection method.

A similar solution we have obtained by minimizing the functional (4.13). More precisely we have obtained a solution which is intermediate in terms of resolution between the back-projected solution of fig. 8 and the regularized solution of fig. 10. This is a reasonable result because, due to the constraints on the variations of s implied by the

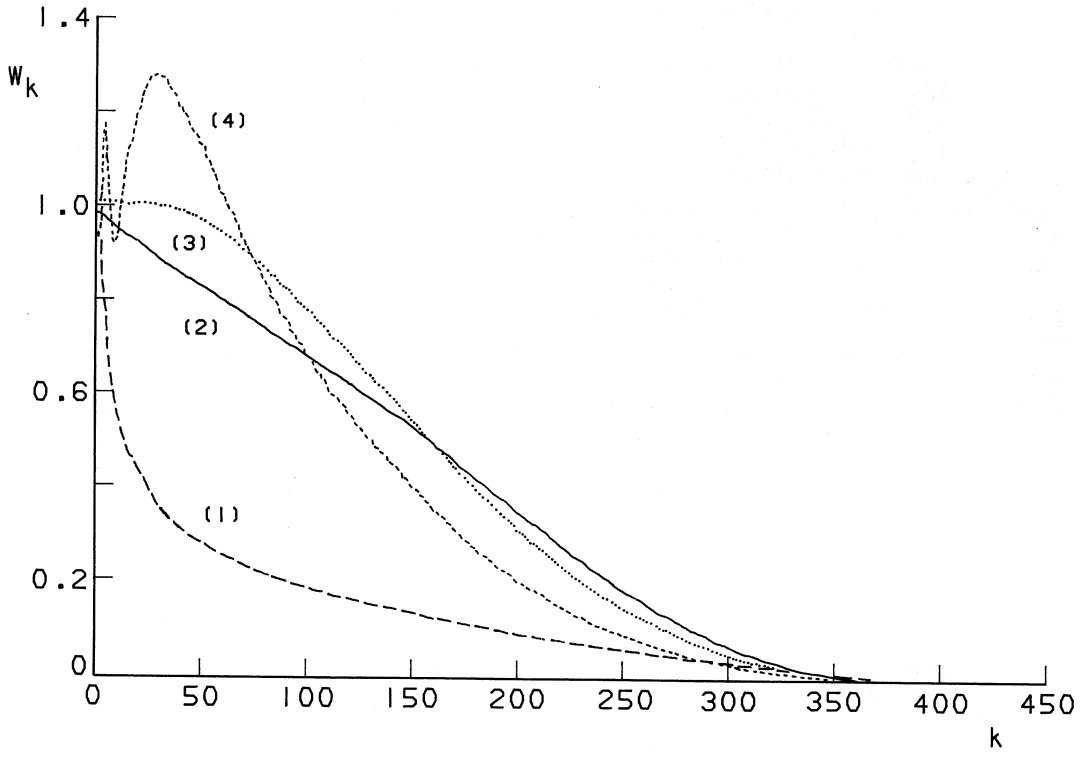


Fig. 12. Comparison of the filters corresponding to the various solutions considered in this paper: (1) = back-projection solution; (2) = Tikhonov regularized solution with $\lambda = 122$; (3) = Landweber solution with $N = 42$; (4) = conjugate gradient solution with $N = 5$.

functional (4.13), we expect that the minimum of this functional is smoother than the regularized solution (4.8).

As concerns the other methods described in section 4, by means of an appropriate choice of the number of singular vectors or of the number of iterations we have always obtained solutions quite similar to the regularized solution both in the case of unit weights and in the case of SIRT weights (this statement is only partially true in the case of method 4.2.). For this reason it is not interesting to plot all the solutions we have obtained. Since each method can be described by means of a filter we compare the various methods simply by comparing the corresponding filters. This is done in fig. 12. We only remark that the number of iterations required for obtaining the best solution (in the case of iterative methods) is not very large. In fact we need 42 iterations in the case of Landweber method and only 5 iterations in the case of the conjugate gradient method. Since the first iteration of the conjugate gradient method is essentially the back-projected solution (except for a factor of the order of 1) this method is highly recommended for a fast improvement of the resolution provided by the back-projection method.

This conclusion is confirmed by a comparison of the CPU times required for the various methods. For the computation of the singular system of the matrix G we

need approximately 1 min on a workstation IBM RISC/6000-560, using the SVD routine of the ESSL library. Then the computation of all the solutions based on the singular system ($\boldsymbol{s}^+$, $\boldsymbol{s}_\lambda$, $\boldsymbol{s}_K$ and also $\boldsymbol{s}^{(N)}$ through eq. (4.18)) takes approximately 1 s. This is also the computation time for computing $\boldsymbol{s}_{BP}$ through eq. (4.5). However, as remarked before, an approximation of $\boldsymbol{s}_{BP}$ is provided by the first iteration of the conjugate gradient method, which does not require the computation of the singular system of G. The CPU time for this first iteration is about 14 s (which is essentially the CPU time for computing G^TG). Each of the subsequent iterations requires approximately 1 s. Therefore in our case the conjugate gradient method is certainly the most convenient one from the computational point of view.

6. Discussion

The 3-D inversion, with the limitations due to the assumption of a block model, gives a sketch of the structural setting, expressed in term of velocity contrast, of the lithosphere and upper mantle in the transition region between the Alps and the Apennines. Basically all methods, when applied to our data set, show a pattern similar to that of fig. 10, *i.e.* scattered areas of high and low velocity zones in shallower layers and a more smoothed distribution in the deepest ones, with high velocity zones bounded by low velocity ones. As demonstrated by the numerical simulation described in (Cattaneo and Eva, 1990) the smoothing effect of the regularization (damping) parameter λ is roughly depth independent and therefore we believe that the above mentioned behaviour is not an artifact of the inversion procedure.

More in detail it is noteworthy that high velocities characterize, at all depth ranges, a band extending from the center of the Po Plain up to the inner part of the Northern Apennines. The lowest velocities appear mainly localized in the westernmost side of the Alps, the Ligurian Sea and the external Apennines. This might be correlated to the presence of Alpine and Apenninic roots (Mueller and Panza, 1986; Spakman, 1990a; Cattaneo and Eva, 1990), which are cold lithospheric bodies transported downwards by subduction processes. Thus the increased velocity and density in the root region is mainly due to the cooling of the surrounding relatively hot asthenosphere. This hypothesis is strongly supported by the close correlations with the heat flow map of the region (Cermak *et al.*, 1992).

The high heat flow values determined in the Ligurian Sea, due to the existence of a young lithosphere and a well developped asthenosphere, are in agreement with the low velocity zone depicted by our models. In addition the high velocity structure under the westernmost sector of the Po valley corresponds to an absolute heat flow minimum. Therefore, the deep extension of this zone should be attributed to a deep-reaching root, representing the remnant of an earlier subduction episode, probably broken and partly assimilated by the mantle during the course of time (Spakman, 1990a, b). More surprising is the extension of the roots in the Northern Apennines.

The presence of high velocity zones in the shallower layers ($h < 90$ km), along the Ligurian coast, could be correlated with a lithospheric indenter (Laubscher *et al.*, 1992) determined by the collision between the thin Ligurian Sea plate and the thicker Adria plate. The younger less deep penetrating Ligurian lithosphere beneath the Northern Apennines might produce a nearly vertical, deep Apenninic slab, as proposed by some authors for the Central Apennines (Mueller and Panza, 1986; Laubscher *et al.*, 1992). Some lithospheric slivers in the crust, as suggested by aeromagnetic surveys (AGIP, 1986; Bolis *et al.*, 1981) might be also responsible of the increase of the mean crustal velocity.

Even if a possible geometrical connection between the two structures appears in these figures, in the light of the present

data it is impossible to hypothesize any correlation between them.

Finally we point out that the general features of the velocity distribution can be inferred from the back-projection solution (fig. 8), while the more refined solution provided by the other methods (fig. 10) only gives more precise information about the extension of the high velocity zones. Therefore, back-projection can be extremely useful for an inversion of the seismic data since it is very fast and very significant, exhibiting already the principal features of the velocity distribution.

Acknowledgements

The work of F. Maggio was supported by the Sardinian Regional Authorities. The work of M. Bertero and F. Malfanti was also partially supported by the Istituto Nazionale di Fisica della Materia (INFM).

REFERENCES

AGIP (1986): *Carta delle anomalie aeromagnetiche d'Italia.*.

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

AUGLIERA, P., S. PASTORE and A. TOMASELLI (1990): Sismicità della Lunigiana-Garfagnana: primi risultati da una rete mobile, in *Atti IX Convegno GNGTS, 13-15 November 1990, Roma* (in press).

BABUSKA, V., J. PLOMEROVA, J. SILENI and M. BAER (1984): Large scale oriented structures in the subcrustal lithosphere of Central Europe, *Ann. Geophys.*, **B2**, 649-662.

BAER, M. (1980): Relative travel time residuals for teleseismic events at the new Swiss seismic station network, *Ann. Geophys.*, **38**, 119-126.

BERTERO, M. (1989): Linear inverse and ill-posed problems, in *Adv. in Electr. and Electron Physics*, edited by P.W. HAWKES (Academic Press, New York), Vol. 75, 1-120.

BIALY, H. (1959): Iterative behandlung linearer funktionalgleichungen, *Arch. Rat. Mech. Anal.*, **4**, 166-176.

BOLIS, G., V. CAPPELLI and M. MARINELLI (1981): Aeromagnetic data of the Italian area: instrumental to a better comprehension of the basement main characteristics in Italy, in *Proceedings of the 43rd Meeting Eu. Ass. Expl. Geophys, Venice 26-29 May 1981*.

CATTANEO, M., M. COVIELLO, C. EVA and M. PASTA (1986): P-waves time residuals in Northwestern Italy (in Italian with English abstract), in *Atti V Convegno GNGTS, 17-19 November 1986, Roma,* 439-449.

CATTANEO, M. and P. AUGLIERA (1990): The automatic phase picking and event location system of the IGG Network (NW Italy), in *Proceedings Workshop on «Seismic Networks and rapid digital data transmission and exchange», 2-4 October 1989, Walferdange* (GDL).

CATTANEO, M. and C. EVA (1990): Propagation anomalies in Northwestern Italy by inversion of teleseismic residuals, *Terra Nova*, **2**, 577-584.

CATTANEO, M. and D. SPALLAROSSA (1990): Inversione in 3D di residui di onde *P* telesismiche: modelli a differente scala, in *Atti IX Convegno GNGTS, 13-15 November 1990, Roma* (in press).

CERMAK, B., B. DELLA VEDOVA, F. LUCAZEAU, V. PASQUALE, G. PELLIS, R. SCHULZ and M. VERDOYA (1992): *A Continent revealed. The European Geotraverse,* edited by D. BLUNDELL, R. FREEMAN and ST. MUELLER (Cambridge Univ. Press).

CRAVEN, P. and G. WAHBA (1979): Smoothing noisy data with spline functions, *Numer. Math.*, **31**, 377-403.

DZIEWONSKI, A.M. and D.L. ANDERSON (1981): Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, **25**, 297-356.

EVA, C., M. CATTANEO and F. MERLANTI (1985): The seismic network in Liguria and Piemonte (in Italian with English abstract), in *Atti IV Convegno GNGTS, 29-31 October 1985, Roma,* 31-41.

GRANET, M. and M. CARA (1988): 3D-velocity structure beneath France in different frequency bands, *Phys. Earth Planet. Int.*, **51**, 133-152.

LANDWEBER, L. (1951): An iteration formula for Fredholm integral equations of the first kind, *Amer. J. Math.*, **73**, 615-624.

LAUBSCHER, H., G.C. BIELLA, R. CASSINIS, R. GELATI, A. LOZEJ, S. SCARASCIA and I. TABACCO, (1992): The collisional knot in Liguria, *Geol. Rundschau*, **81/2**, 275-289.

MOROZOV, V.A. (1968): The error principle in the solution of operational equations by the regularization method, *USSR Comput. Math. Math. Phys.*, **8**, 63-87.

MUELLER, ST. and G.F. PANZA (1986): Evidence of a deep-reaching lithospheric root under the Alpine Arc., in *The Origin of Arcs*, edited by F.C. WEZEL, Development in Geotectonics, **21**, 93-114.

PANZA, G.F., S. MULLER and G. CALCAGNILE (1980): The gross features of the lithosphere - asthenosphere system in Europe from seismic surface waves and body waves, *Pageoph*, **118**, 1209-1213.

RALSTON, A. (1965): *A First Course in Numerical Analysis* (Mc Graw-Hill, New York).

SEATZU, S. (1986): A remark on the numerical solution of linear inverse problems with discrete data, *Inverse Problems*, **2**, L27-L30.

SPAKMAN, W. (1986): The upper mantle structure in the Central European-Mediterranean Region, in *Proceedings 3rd ECT Workshop, 14-16 April 1986, Bad Honnef,* 215-221.

SPAKMAN, W. (1990a): The structure of the lithosphere and mantle beneath the Alps as mapped by delay time tomography, the european geotraverse: integrative studies, in *Results from the Fifth Study Centre,* edited by R. FREEMAN, P. GIESE AND ST. MUELLER, 213-220.

SPAKMAN, W. (1990b): Tomographic images of the upper mantle below Central Europe and the Mediterranean, *Terra Nova*, **2,** 542-553.

TIKHONOV, A.N. and V.Y. ARSENIN (1977): *Solutions of Ill-Posed Problems* (Winston/Wiley, Washington).

TRAMPERT, J. and J.J. LEVEQUE (1990): Simultaneous iterative reconstruction technique: physical interpretation based on the generalized least squares solution, *J. Geophys. Res.*, **95**, 12555-12559.

VAN DER SLUIS, A. and H.A. VAN DER VORST (1987): Numerical solution of large, sparse linear algebraic systems arising from tomographic problems, in *Seismic Tomography,* edited by G. NOLET (Reidel, Dordrecht), 49-83.

(received July 5, 1993;
accepted April 18, 1994)