SEISMIC TOMOGRAPHY PROJECT

Department of Geophysics
Stanford University

Volume 1, No. 1
June 1990

TABLE OF CONTENTS

SEISMIC TOMOGRAPHY PROJECT

Volume 1, No. 1

Contributors:

Luis Canales
Mike Fitzpatrick
Jerry M. Harris
Caroline Lambert
Spyros Lazaratos
Gary Mavko
Reinaldo Michelena
Daniel Moos
Richard Nolen-Hoeksema
Pierre Samec

June, 1990

PAPER A

CROSS-WELL TOMOGRAPHIC IMAGING OF GEOLOGICAL STRUCTURES IN GULF COAST SEDIMENTS

Jerry M. Harris, Gary Mavko, Daniel Moos, and Richard Nolen-Hoeksema

Seismic Tomography Project

SUMMARY

In the following case history, cross-well seismic tomography was used to successfully image both structural and stratigraphic features in Gulf Coast Miocene sediments between wells 250 feet apart. A piezoelectric downhole seismic source and three hydrophones were used to acquire the data. In addition to the targeted fault delineation, the tomography successfully imaged a number of porous sandstone layers previously identified from type logs for the Miocene. Iterative inversion was applied to a set of approximately 5500 p-wave picks. Interpretation of the resulting velocity tomogram was consistent with the pre-survey description based on well log analysis. Borehole gravity meter (BHGM) data, obtained in both wells following the tomographic survey, were used to associate density variations with the velocity features imaged by the tomography. The BHGM derived low density zones were found to coincide with the seismically imaged low velocity zones, thus supporting the identification of porous sandstone layers.

INTRODUCTION

Surface-based seismic investigations of the subsurface are inherently limited by (1) the ability to propagate energy through attenuating Earth materials, (2) the resolving power of waves, which is essentially limited to identifying features not much shorter than the seismic wavelength, and (3) the fact that sources and receivers are remote from the volume of interest. Emerging methods known as cross-well tomography, and more generally cross-well seismology, place sources and receivers at depth, closer to the reservoir or geological targets of interest and below highly attenuating near-surface materials, and thus can employ much higher frequencies of investigation (kHz) than surface seismic methods. Cross-well seismology increases the resolving power, but because energetic sources pose risks to the wellbore, limits the distance between wells at which surveys can be safely run.

Cross-well seismic tomography has been successfully used to estimate velocities between wells (Dines and Lytle, 1979; Bregman et al., 1989; and Justice et al., 1989). Although much of the tomography has been for the purposes of enhanced oil recovery (EOR), the objective in this study was to demonstrate that cross-well tomography could be used to successfully image geological features such as faults, pinchouts, or truncations. With this objective in mind, we selected a site with a series of geologic targets of interest for reservoir characterization. These include structural features such as faults and stratigraphic features such as pinch-outs or facies changes. Other features such as gas/liquid interfaces, which might exist within the zone of interest, may be imageable also.

However, these other features were not of primary interest to this study. For the selected site, the primary targets were faults within a Miocene sand/shale section onshore Gulf Coast.

A piezoelectric downhole source and a three-hydrophone array were used to acquire the data. This system of hardware provided a flexible setup for investigating the transmission properties of Gulf Coast Miocene sediments. Signals to nearly 2000 Hz were recorded at distances over 400 feet in the poorly consolidated sediments. Nearly 7000 traces of data were recorded. Of those, over 5000 p-wave traveltimes were picked and processed.

The case study reported here summarizes the steps of selecting the site, designing and executing the data acquisition plan, then processing and interpreting the data. The results are evaluated by comparing the cross-well image to our pre-survey model. Of course, there is no certainty that the pre-survey model is accurate, and thus in a sense the cross-well data can also be said to provide a test for validating the pre-survey model.

SITE SELECTION

Criteria for site selection were developed and distributed to several oil companies. The most important of these were that the primary target zone must be in a sedimentary environment and contain an isolated feature which based on available data provided sufficient velocity contrast to be imageable. We also specified that enough prior information must exist to provide an unambiguous pre-survey geological model, even though the geophysical model may not be as well defined. Three companies responded with site proposals. Amoco's Gulf Coast area #1 was selected as our first site. (The field name and precise location are confidential.)

The proposed target was at 3000 to 4000 feet within a Miocene sand/shale sequence onshore in Southeast Texas. Four wells were situated at the site within a 1000-foot radius. The closest two wellheads were about 300 feet apart (Figure 1). Each of these wells had been logged within the target zone with conductivity and SP to provide stratigraphic ties. Analysis of these data in combination with log type sections from elsewhere in the field provided a model for the proposed site which included a series of faults penetrating the region of interest. These faults were previously mapped within deeper producing horizons and their fault planes were continued upwards into the region between the proposed wells. Thus, although the precise positions of the faults as shown in Figure 1 could only be estimated based on the logs, their existence was required by previous structural analyses and reservoir characterization studies. The fault offsets were determined based on intervals within the type logs which were missing in these wells and agreed with previously known offsets within the deeper horizons.

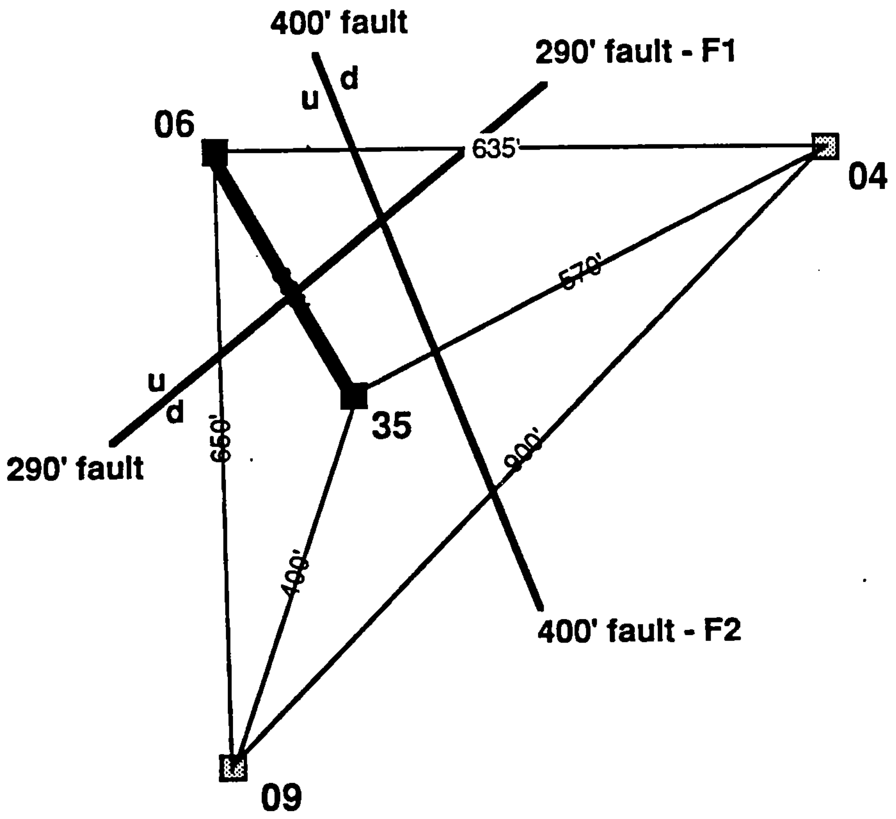


Figure 1. Plan view of study site. Faults F1 and F2 are shown at their approximate positions for a depth of 3300 feet.

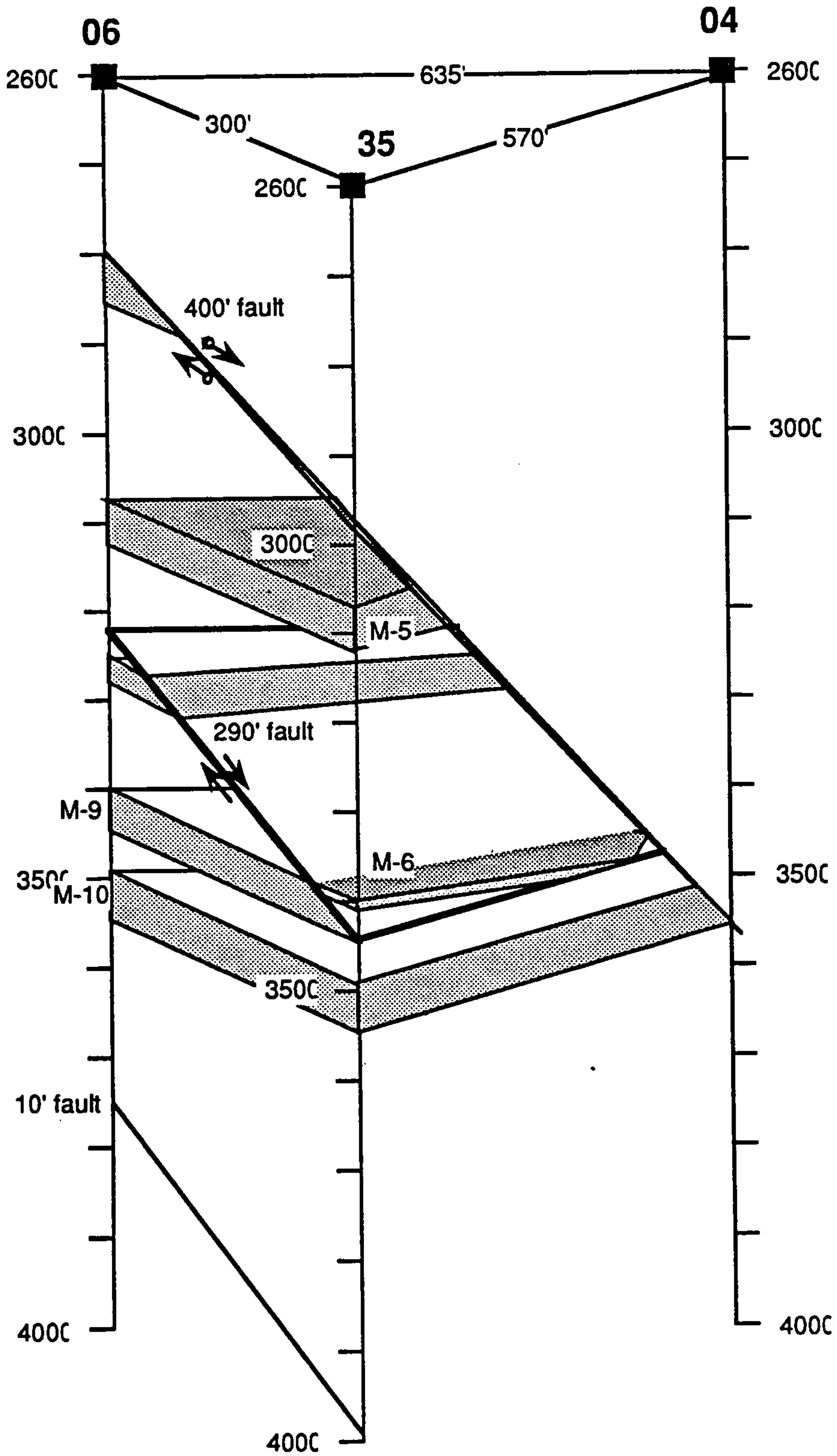


Figure 2. Interpreted pre-survey cross-section of area #1.

We chose the two closest wells (300 feet) for our first survey based on the desire to ensure that sufficient high frequency energy could be propagated between source and receivers. Four faults were identified in the section between these wells. Two of these are shown on the cross-section plotted in Figure 2. The deepest fault (not shown) has only 10' of missing section and is perpendicular to the cross-section. Proceeding up-section, the next fault (not shown) is inferred, since the M-10 to M-10A interval in both these wells is 35' too thin, relative to the type log. The existence of this fault was uncertain, as stratigraphic differences between the type well and these wells could explain the decreased thickness of these units. If the fault does exist it strikes sub-parallel to the section. The third fault (labelled F1) has about 290' of missing section, and strikes perpendicular to the section. This fault juxtaposes two sands (the M-8 and M-9) against a downthrown sequence of shales. The uppermost fault (labelled F2) removes 400' of section and is a major structural feature of the field. The fault F2 strikes sub-parallel to the section, and its location was poorly defined. The fault F1, at depths between about 3200 feet and 3450 feet, was chosen as the target of this experiment.

The physical properties of the sand/shale sequence are not well known. A density/-porosity well logging combination run in one of the wells yielded average sand densities of 2.00 to 2.33 gm/cc, and for the shales of 2.18 to 2.60 gm/cc. No velocity logs were available. Based on empirical relationships between density and velocity for these materials, a velocity model was constructed for the target fault, in which differences in velocity between the downthrown shales and the M-8 and M-9 sands provide the target.

EXPERIMENTAL DESIGN AND DATA ACQUISITION

The field survey used a piezoelectric bender bar downhole source constructed by Honeywell Marine Systems. Two bars were assembled and driven in phase to produce a source with a dominant monopole mode of vibration for the efficient generation of p-waves. The active section of the source structure is approximately 20 inches long and has a resonant frequency of about 800 Hz in free-field water at surface temperatures and pressures. The acoustic and electrical properties of the source in open water are well characterized. Unfortunately, operation in a borehole is not so well understood. The source couples acoustic energy to the formation through the borehole fluids and operates on a standard oil field wireline.

During setup and configuration prior to recording of the actual field survey, the source was driven with wavelets having different frequencies, durations, etc. These tests were instrumental in determining the frequency passband of the rocks, signal-to-noise ratio, and data acquisition parameters such as the acceptable stacking depth. Based primarily on energy considerations which affect the speed of data acquisition and is related to limitations of the recording hardware, a sweep signal was chosen for the tomographic survey. The chosen sweep was 400 ms long with start and stop frequencies of 400 Hz and 1600 Hz, respectively. Four sweeps were stacked to form the seismic trace.

Three OAS deep ocean hydrophones were used as detectors. These were separated by 10 feet of wireline. Signals detected by the hydrophones were preamplified and filtered downhole then transmitted analog up 7-conductor wireline to the surface for more filtering and digitizing. At the surface, the data were recorded at a sample interval of 0.1 ms using an Analogic waveform analyzer. The common receiver fan taken at a depth of 3366 feet and shown in Figure 4 is typical of the data recorded. The typical shooting pattern fixed the three hydrophones and scanned the source in an aperture of approximately +/- 250 feet or roughly +/- 45 degrees. Receiver points uniformly covered the depths of 2690 to 3980 feet at ten-foot intervals. The source point intervals ranged from 5-20 feet over depths between 2400 and 4000 feet - with the 5-ft intervals about the primary target zone of 3100 to 3500 feet. A few wide aperture scans were run to obtain high angle data for identification of possible reflections off the fault zone if present. This recording geometry provided about 6700 data traces for the entire survey. A diagram of the shooting pattern is illustrated in Figure 3. Notice that in addition to shooting the primary target zone about F1, the survey was extended well above F1 to cross F2 as well.

OTHER DATA COLLECTED

Prior to conducting the cross-well survey a series of geophysical well logs were run for the two wells through casing in the depth interval of interest (2000-4000 feet). These included natural gamma spectral, cement bond, and full waveform digital sonic logs. The cement bond log was necessary to verify that the casing was sufficiently well bonded to the formation to allow efficient transmission of energy from the wellbore source to the formation. The natural gamma log was run to provide depth ties to the earlier logs, as neither conductivity nor SP (which are routinely used in this field to locate marker sand beds) can be acquired through casing. Full waveform sonics were recorded in the hope that we could extract formation compressional wave velocities for comparison to the cross-well data.

Based on the cement bond log, the casing in each well was well-coupled to the formation. Consequently, slowness time coherence processing of the sonic waveforms in the field resulted in reasonable velocities. Comparison of the gamma log to SP and conductivity previously recorded in these wells revealed depth offsets of between 20 and 40 feet; the logs were shifted to correct for these offsets prior to interpretation. After completion of the cross-well acquisition and preliminary analysis of the data, Amoco Production Research recorded a borehole gravimetry (BHGM) survey which, when interpreted assuming flat-lying layers, results in an apparent density log.

TRAVELTIME PICKS AND DATA ANALYSIS

About 5500 first arrival times were picked for creating a velocity image from the data. A small subset (< 5%) of the traces where the arrivals could not be reliably identified were ignored.

We found that a useful way to display the traveltimes and survey geometry is to plot shot depth vs. receiver depth as shown in Figure 4. Each dot represents a recorded trace, exactly as in a "stacking chart" for surface reflection data. All of the traces along any vertical line correspond to a common source fan, and all of the traces along any horizontal

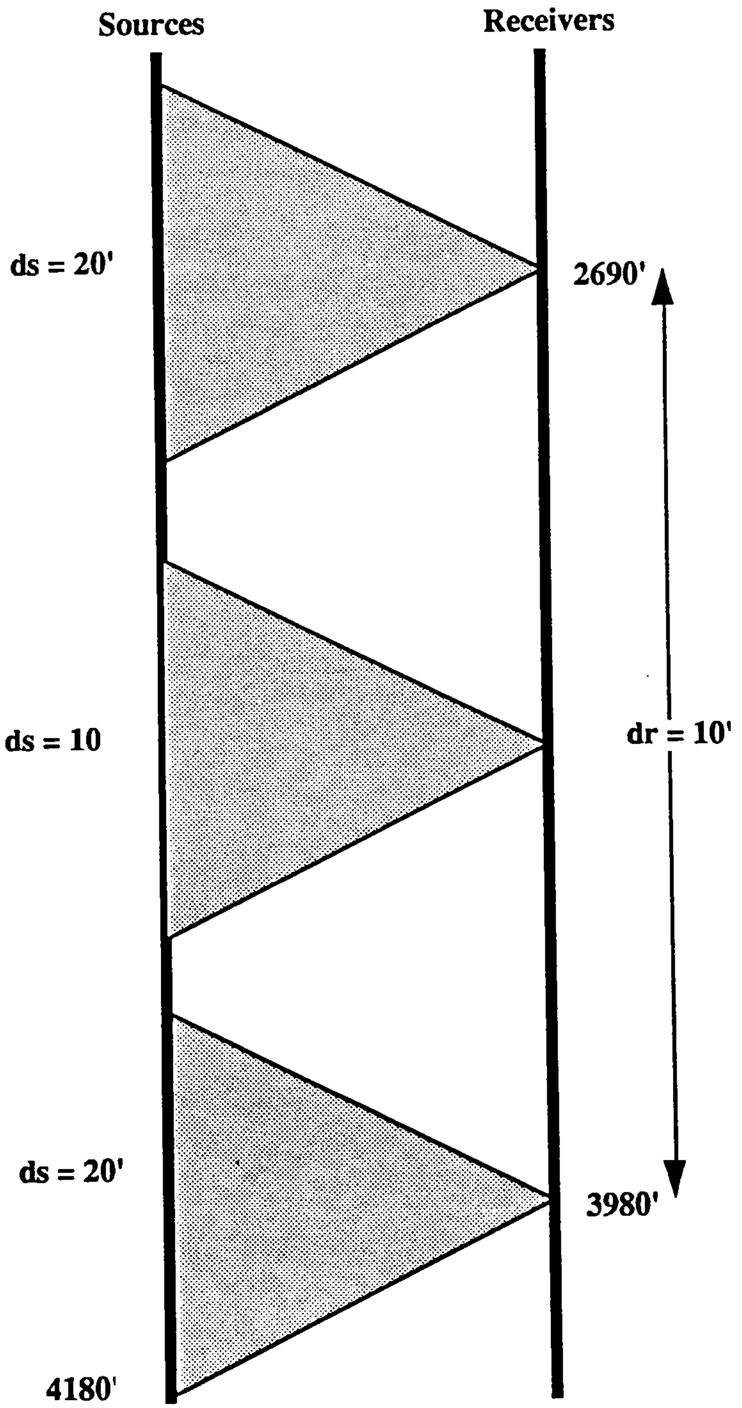


Figure 3. Overall survey geometry illustrating shooting pattern: Receivers are spaced at 10-foot intervals between 3980 feet and 2690 feet. Sources are spaced at 10-ft interval in primary target zone and 20-ft above and below target zone.

line correspond to a common receiver fan. When displayed this way, one can quickly see the fold, aperture dimensions, and variations of coverage with depth. The holes, where picks are missing, are also clearly evident.

The colors represent the picked arrival times. For more sensitivity, the times plotted here are actually the differences relative to traveltimes in a uniform velocity medium, $\Delta t = t_{pick} - r/V_0$, where t_{pick} is the picked arrival time, r is the distance between the source and receiver, and V_0 is the uniform reference velocity. Figure 5 shows the same picks interpolated to make the patterns more clear. The patterns directly indicate velocity variations in the formation between the wells. In general vertical streaks are related to near-source anomalies; horizontal streaks are related near-receiver anomalies, and streaks at other angles are related to anomalies between the wells. Source or receiver-consistent recording or geometry errors will also show up as vertical or horizontal streaks. Velocity variations with depth will show up as a color change along the center line, *source depth = receiver depth*; lateral variations will cause an asymmteric color variation with offset away from this center line. A consistent decrease of velocity with increased source-receiver offset may be an indicator of anisotropy. Please refer to the paper (Mavko and Lazaratos, 1990) in this volume for discussion on the uses of these "pick domain" plots for quality control and for interpreting velocity variations before performing an actual inversion. They also show that a first order, straight-ray tomography image is simply the "Radon transform" of the pick domain plot.

TRAVELTIME TOMOGRAPHY

The solution of the well-known non-linear inversion problem of traveltime tomography in strongly refracting media involves three distinct steps: first, we must pick observed traveltimes; second, we calculate traveltimes for an assumed slowness model; third, we invert a matrix equation where the data is given by the traveltime residuals (calculated minus observed) to obtain corrections to the assumed slowness model. Steps two and three are performed iteratively and constitute linearization of the original non-linear problem of finding both the slowness field and the raypaths. The iterations are usually stopped when an acceptable match between the calculated and the observed traveltimes is achieved. An outline of the process is shown in Figure 6.

A detailed description of the inversion procedure used on this data can be found in the paper by Harris, Lazaratos, and Michelena (1990). In brief summary, the method uses raytracing to model the traveltimes and a basis set of strings to parameterize the perturbations to the slowness field. Several general features of the inversion procedure which relate to the handling of the real data are summarized below:

Well deviation

To avoid the more complicated situation of inversion in three spatial dimensions, it is necessary to reduce the 3D deviation geometry to a plane of two dimensions. The common approach is to fix a plane intersecting the wellbores at some mean or median orientation and project all calculations of shot-receiver positions and traveltimes onto this plane. This can

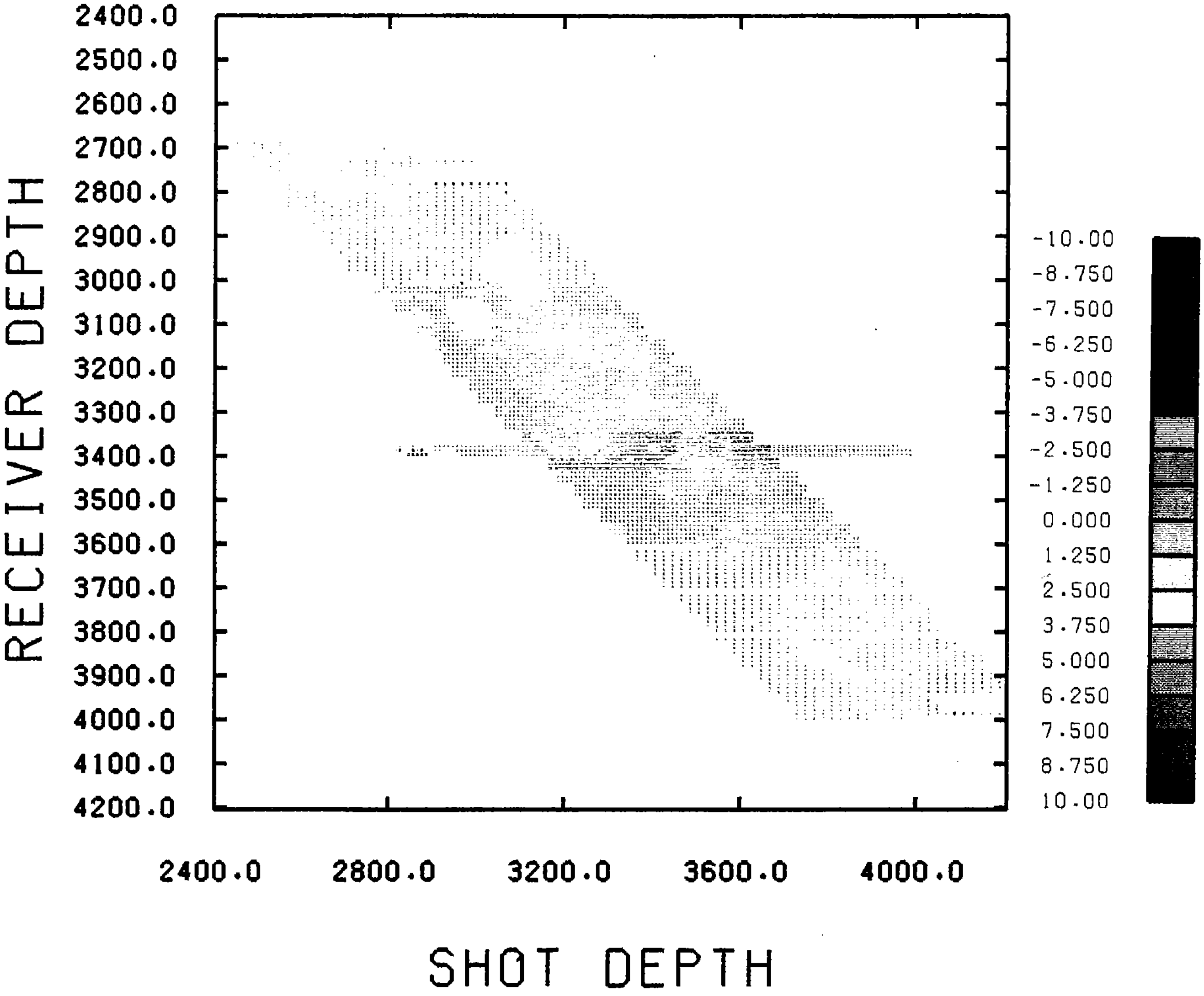


Fig. 4: Pick domain display of residual travel times (in milliseconds), showing survey geometry and rough travel time patterns.

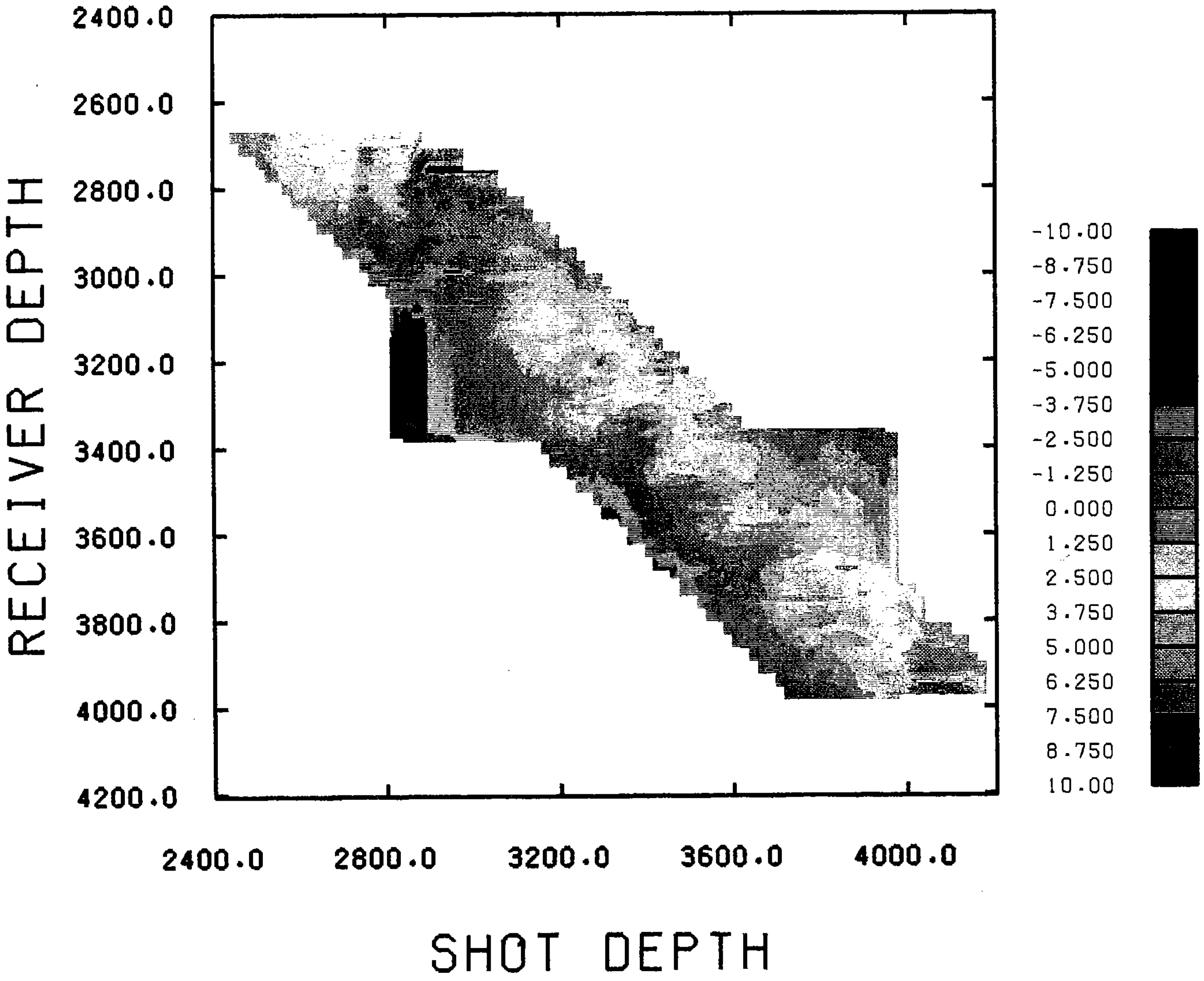


Fig. 5: Pick domain display of residual travel times, interpolated to better show the travel time patterns.

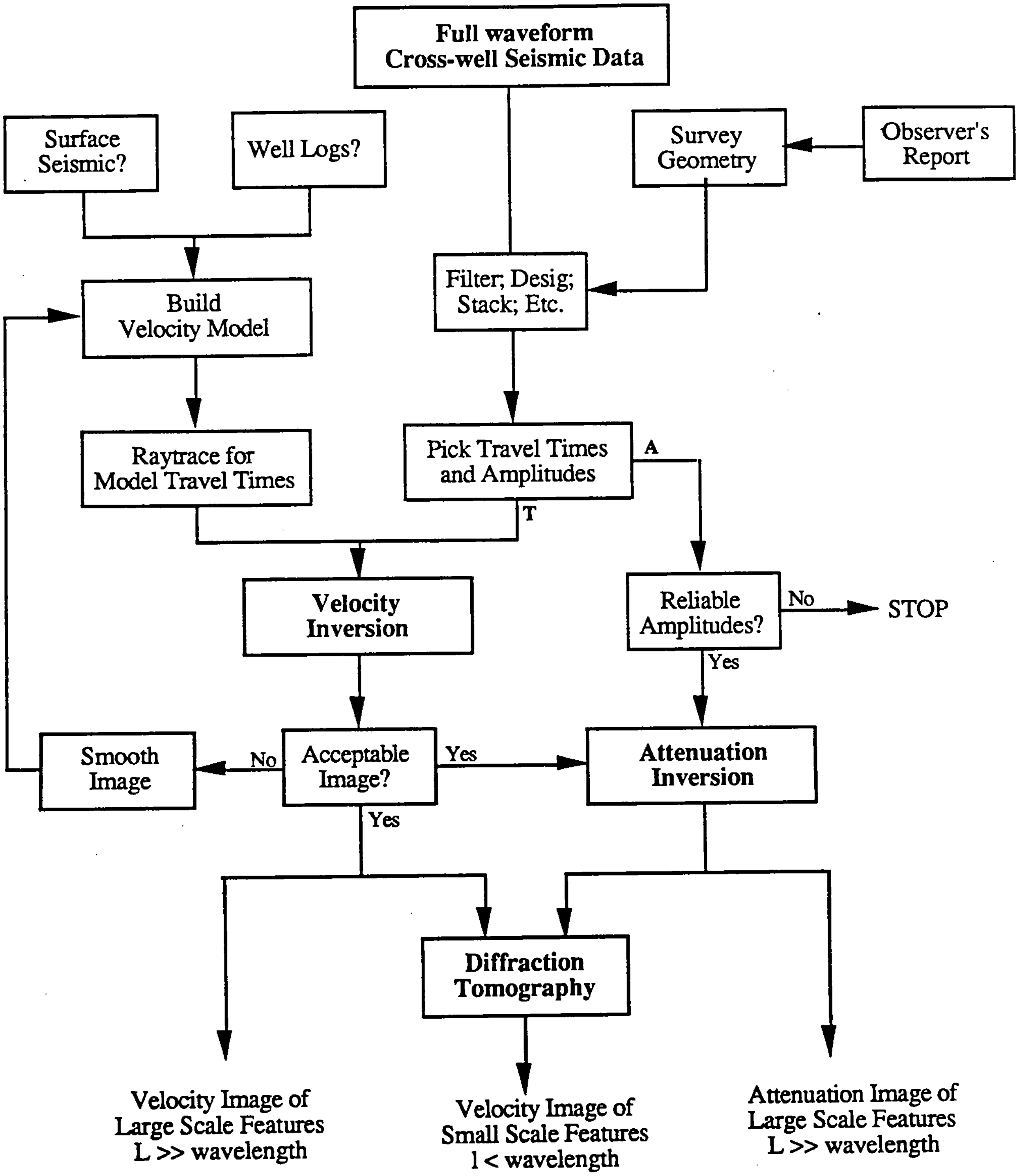


Figure 6. Schematic diagram of cross-well processing. Only traveltime processing was used on the field dataset.

result in large static corrections from the true positions to the image plane. A different approach is used here. First we note that the deviations from vertical are gradual or slowly varying; therefore, within a span of a few hundred feet, there is very little error associated with fitting a mean 2D plane to the actual deviations. Actually, we find the mean offset coordinates for each receiver point, that is, the mean orientation for the collection of shots associated with each receiver depth point. In effect, we calculate a pointing vector for each common-receiver-gather. This pointing vector then slowly rotates with depth, thus defining a local 2D plane for purposes of raytracing and inversion but varying with the actual deviation of the wells.

Raytracing

As mentioned above, raytracing is used to model albeit approximately the physical mechanism of generating the traveltimes. Calculated traveltimes are compared to observed traveltimes, thus producing a set of residuals which is inverted. This process requires calculating the traveltime for each source-detector pair or linking the source with the detector, i.e., solving the two-point boundary value problem which can be expensive and time consuming. To avoid this linking, we trace a very dense fan of rays, fit a curve through the observed common-receiver array of traveltimes, then use the curve to map observed traveltimes onto the calculated rays. This allows the raytracing to be vectorized on the computer and provides an important means of interpolation for the string-generated image.

Inversion

The inversion problem is now reduced to inverting a matrix equation relating the residual traveltimes (data) to perturbations of the assumed slowness field. For this purpose, a new procedure known as string tomography (Harris, Lazaratos, and Michelena, 1990) was used. Rather than parameterize the perturbations to the slowness field in square orthogonal pixels, this new procedure uses strings which are derived from the raypaths themselves. The approach avoids two costly operations in the inversion process, calculating elements of the projection matrix and inverting the projection matrix.

Results

String tomography was applied to the nearly 5500 traveltimes acquired during the imaging survey. The inversion procedure, outlined in Figure 6, involved several raytrace iterations. The results for two different starting models are substantially the same. See (Harris, et. al., 1990) for inversion details. The tomogram used for interpretation was generated after five iterations, starting with constant velocity slowness model. his result is shown in Figure 8. Recall the manner described above of handling the well deviations. In the tomogram display of Figure 10, well 06 is plotted straight along the vertical and all the deviation is allowed to appear in the offset well 35. Therefore, to register the coordinates to the real world, one should move down well 06 and out a radial distance toward well 35.

GEOLOGIC INTERPRETATION

Tomogram validation

Figure 7 illustrates the final velocity image overlain by the original geological model. Each of the faults predicted by the model have expressions in the final image. However, these expressions are different for the different faults. The original target of the experiment, the 290' fault labelled F1, truncates a series of beds including a very high-velocity shale below the M-8 against the overlying downfaulted shales. The position of this fault can be inferred from the bed truncations. Interestingly, the 400' fault labelled F2, which was oblique to the original section and poorly characterized in the geologic model, shows up very clearly at its predicted depth and orientation as a sharp velocity contrast, with higher velocities above than below the fault (trapped gas?). The two small offset faults below the primary target zone could also be interpreted, again based on bed truncations, but probably would not have been resolved based only on the image.

Comparisons of the tomogram to the other data

Looking in more detail at the relationship between the image and the well log data, there is a striking correlation between lithology and velocity. Figure 8a shows the image overlain by the SP logs, plotted at the locations of the wells on the image plane. Background shale readings are generally associated with intervals of higher velocity, whereas sands have distinctly lower velocities than the surrounding materials. For example, in well 06 the M-4 sand shows up as a pronounced low velocity (<7865 ft/s) interval. This sand continues across section and is similarly imaged near the 35 well. The M-9 sand, which is present in well 06 but faulted out by the 290' fault in 35, is also associated with a low velocity horizon, as are the M-10 and M-10A.

A similar comparison between the BHGM densities and the velocity image is shown in Figure 8b. Overall, low BHGM densities and low velocities are well correlated. This correlation between a deep-reading BHGM measurement and the cross-well image is a striking example of the ability of these two methods to provide valuable and complementary data. Furthermore, the congruence between low densities and low velocities generally confirms standard petrophysical assumptions. It should be noted, however, that the BHGM densities were obtained from 1-D inversion of the gravity station measurements. Differences between the resulting densities at the two wells, and the lateral discontinuities in the velocity image, suggest that the 1-D assumption is at best a first-order approximation.

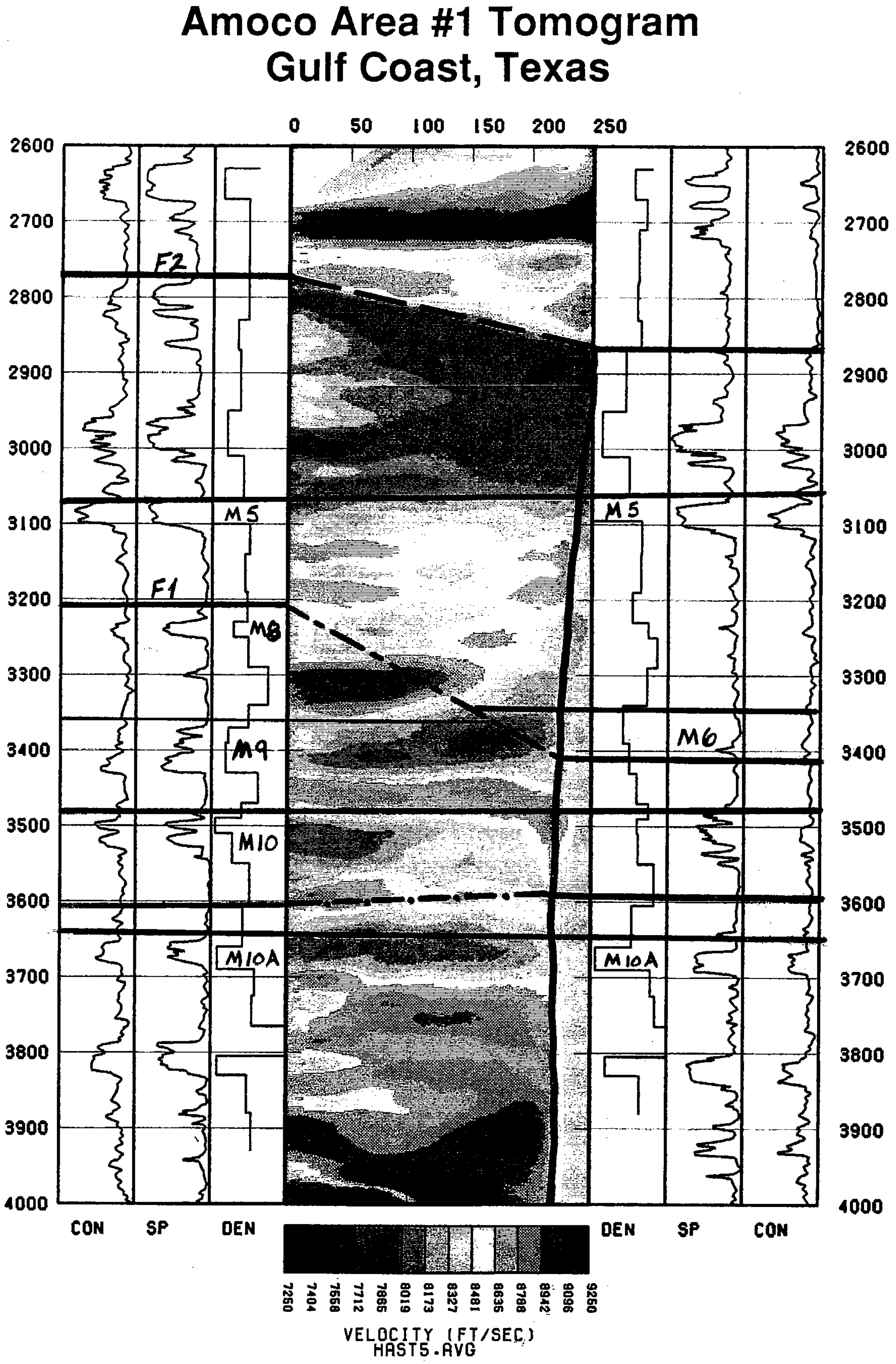


Fig. 7. Final tomogram with geological overlay.

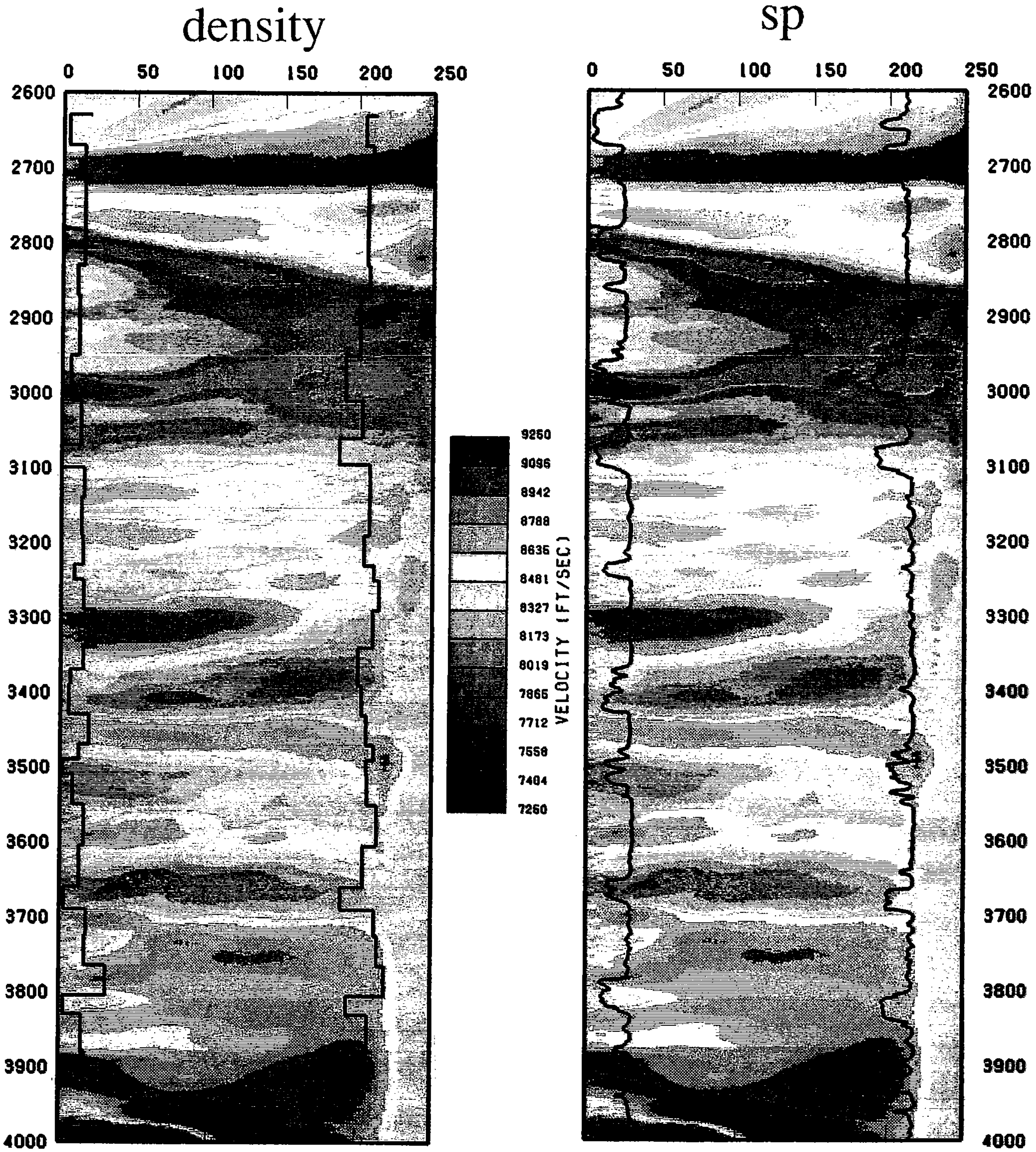


Fig. 8. Final tomogram with density and sp overlay.

In marked contrast to the expressions of the other sands, the position of the M-5 sand in the SP log is marked by a steep velocity gradient, rather than a low velocity zone. This sand is expressed in the electrical logs by a doublet (two sands separated by a less sandy unit?). Perhaps the M-5 is a permeability barrier, and the velocity gradient reflects differences in pore-fluid pressure or composition. The gravity data show that the M-5 sand has low density, similar to the densities of the other sands. Thus, the lack of a low velocity zone at the level of the M-5 may indicate that this sand is different microstructurally from the other sands in this section. This difference may also result in the M-5 acting as a permeability barrier

The M-10B is associated with a low velocity and density near the 06 well. Near the 35 well, the M-10B has low density and clear SP and resistivity expressions, but there is little velocity change. This lack of seismic expression may be due to lack of coverage near the bottom of the survey, or to stratigraphic microstructural changes between the two wells.

In summary, comparisons between petrophysical log data and the seismic velocity image show the synergy which can be achieved using multiple measurements. The combined data provide a consistent picture of geologic structure, and differences among these data give insight into the petrophysical properties of the individual units.

CONCLUSIONS

The results of cross-well experiments at the Southeast Texas site demonstrated the suitability of cross-well tomography for work at moderate depths and offsets in typical reservoir environments. A geologic target identified by interpretation of core and well log data and a priori structural information from log ties at other wells was successfully imaged. This is promising for two reasons. First, one never has complete confidence in well-to-well log correlations. Second, in many situations there is not enough data to produce even a testable interpretation of the structure between wells in a field.

The target was initially chosen based on the expectation that a large offset fault juxtaposed media with a pronounced velocity contrast. We imaged that target by resolving sand beds with thicknesses on the order of 15-20 feet that were truncated against a thick shale section. We also imaged two small offset faults. One of these had a throw of 35' and a strike sub-parallel to the section. The other fault had a throw of only 10' and was perpendicular to the section. Although this is a remarkable example of the ability of cross-well tomography to image features close to the wavelength of the seismic signal, it is unlikely that these features would have been interpreted based on the tomographic image alone.

The data were acquired with a high frequency piezoelectric downhole source. The frequency content of the detected seismic signal was bounded on the low end by source inefficiency and on the high end by attenuation in the Earth materials. Even so, frequencies to nearly 2000 Hz were recorded at slant path distances exceeding 400 feet. At velocities ranging down to 7000 ft/sec, this slant path distance corresponds to over 100 wavelengths.

Near wellbore low velocity intervals seen in the tomogram could be tied to individual sands identified by SP and conductivity logs. These sands had low densities relative to the surrounding shales, and therefore densities computed from borehole gravimetry correlated remarkably well with the tomographic image. Interestingly, there were similarities and differences between velocities computed from sonic waveforms recorded through casing, and cross-well seismic velocities. The differences, concentrated in a 300' depth interval, are intriguing but need verification.

We consider that the high frequency cross-well technique has been validated for these experimental conditions - that is imaging small-scale (on the order of 10-50 feet) structural features of a sand/shale sequence at moderate offsets (300 feet). We are planning further studies at the same site to extend the offset between wells, to investigate the out-of-plane 400' fault, and to allow evaluation of interpretation techniques for 3-D data.

REFERENCES

Bregman, N.D., Bailey, R.C., and Chapman, C.H., 1989, Cross-hole seismic tomography: Geophysics, 54, 200-215.

Dines, K. A., and Lytle, R. J., 1979, Computerized geophysical tomography, Proc. IEEE, 67, pp. 1065-1073.

Harris, J., Lazaratos, S., and Michelena, R., 1990, Tomographic string inversions, 1990 SEG meeting, San Francisco.

Mavko, G. and Lazaratos, S., 1990, Cross-well tomographic analysis in the pick domain, Paper C, this volume.

McMechan, G. A., Harris, J. M., and Anderson, L., 1987, Cross- well tomography for strongly variable media with application to scale model data, Bulletin of the Seismological Society of America, Vol. 77, No. 6, pp. 1945-1960.

Michelena, R, and Harris, J. M., 1990, Tomographic inversion using natural pixels, Paper F, this volume.

Justice, J. H., Vassiliou, A. A., Singh, S., Logel, J. D., Hansen, P.A., Hall, B.R., Hett, P.R., and Solanki, J.J., 1989, Acoustic tomography for monitoring enhanced oil recovery: The Leading Edge, 8, no. 2, 12-19.

ACKNOWLEDGEMENTS

The authors thank Amoco Production Company for its support in acquiring the data, and the Gas Research Institute for its financial support of the cross-well tomography project at Stanford University.

PAPER B

TOMOGRAPHIC STRING INVERSION

Jerry M. Harris, Spyros Lazaratos, and Reinaldo Michelena

Seismic Tomography Project

SUMMARY

When square pixels are used to parameterize the slowness field in traveltime tomography, the problems of discretization for inversion purposes and discretization for display purposes are inextricably mixed. A "high resolution" result demands many pixels or model parameters, thus burdening tomography with the inversion of a large and sparse projection matrix though simplifying the display problem by using regions of constant slowness. In this paper, we present a method of separating the tomography problem into two distinct steps - first inversion, then imaging. To reduce the complexity of the inversion step, we select a more natural set of pixels which are derived from the raypaths used to model the process of creating the traveltime data. The support of the string function is the raypath itself, thus the strings and raypaths are orthogonal (except when both equal the same line) and the projection matrix is diagonal and invertable in closed form. We are then left to image, i.e., synthesize the inversion result for display purposes. The method is tested on cross-well synthetic and field data requiring iterative curved raytracing. It is demonstrated to be a fast and robust technique of traveltime tomography.

INTRODUCTION

A solution to the well-known non-linear inversion problem of traveltime tomography in strongly refracting media involves three distinct steps: first, we must pick observed traveltimes; second, we calculate traveltimes for an assumed slowness model; third, we invert a matrix equation where the data is given by the traveltime residuals (calculated minus observed) to obtain corrections to the assumed slowness model. Steps two and three are performed iteratively and constitute linearization of the original non-linear problem of finding both the slowness field and the dependent raypaths. The iterations are usually stopped when an acceptable match between the calculated and the observed traveltimes is achieved.

The calculation of traveltimes, step 2, requires a mathematical model capturing albeit approximately the physical process of generating the data. For this purpose, we use ray tracing described by the equation

$$t_i = \int_{\Omega} S(x, z)\, \phi_i(x, z)\, dxdz. \tag{1}$$

In general, the function $\phi_i(x, z)$ represents the 2D beampath which describes the area of the slowness field influencing the traveltime (Michelena and Harris,1990). However, in the formulation presented here, we take the function $\phi_i(x, z)$ to be nonzero only along the geometrical acoustic raypath, i.e., a line, thus equation (1) describes conventional raytracing in two spatial dimensions. The third step is the inversion of the residual traveltimes for perturbations to the slowness field. For the nonlinear problem, we separate the slowness into a known background component $S_o(x, z)$ and a perturbation $\delta S(x, z)$. To set up the system of equations, we then parameterize the perturbations, that is, we devise a representation of the slowness field as a discrete superposition of known functions whose coefficients are to be determined by the inversion:

$$S(x, z) = S_o(x, z) + \delta S(x, z) \tag{2a}$$

$$\delta S(x, z) = \sum_{j=1}^{N} \alpha_j \psi_j(x, z) \tag{2b}$$

With equation (2b) representing the perturbations, the tomography problem is reduced to inverting the linear system of equations

$$\delta t = \mathbf{W}\alpha , \tag{3}$$

where $\mathbf{W}$ is the (MxN) projection matrix, α is the (Nx1) column vector of unknown coefficients parameterizing the slowness perturbations, and δt is the (Mx1) column vector of residual traveltime data. The elements of the projection matrix are given by the inner product of the basis function $\psi_j(x, z)$ with the raypath $\phi_i(x, z)$ computed for the background medium:

$$W_{ij} = \int_{\Omega} \psi_j(x, z)\, \phi_i(x, z)\, dxdz \tag{4}$$

When orthogonal pixels of constant slowness (Dines and Lytle, 1979; McMechan, 1988), are used for the basis functions $\psi_j(x, z)$, the projection matrix is typically very large and sparse, each element of it representing the segment of the ith ray projected on the jth pixel. These segment lengths are expensive to compute. Moreover, for the highly nonlinear problems encountered in seismic tomography the elements must be recomputed for each raytrace. Due to the large matrices, the solution of (3) by inversion of W is impractical, and (3) is often solved by iterative row-action methods (Censor, 1981). The choice of orthogonal pixels inextricably ties the inversion problem to the problem of sampling (and displaying) the estimated slowness field. A high resolution display requires smaller pixels, i.e., more model parameters, thus greatly increasing the dimensions of W and further increasing the cost of solving equation (3).

However, we are free to choose the basis functions and some choice other than orthogonal pixels of constant slowness may provide some advantages. This idea isn't new for splines and other parameterizations are often used to describe smoothly varying media (Pereyra, 1988). Michelena and Harris (1990) suggested using the beampaths as basis functions, thus forming a set of natural pixels wherein the matrix elements become the area of intersection of the ith beam with the jth beam. Conventional rays are used throughout our formulation; therefore, a judicious choice of basis functions might be to take $\psi_j(x, z) = \phi_i(x, z)$ or to use "strings" derived from the raypaths as the basis set.

STRING BASIS FUNCTIONS

When rays are used to model the traveltimes in equation (1) and strings are used for the basis set in equation (2), all off-diagonal elements of the projection matrix are identically zero and the diagonal elements simply reduce to the lengths of the rays. In this case, equation (3) simplifies greatly to become

$$\begin{pmatrix} l_1 & 0 & \cdots & 0 \\ 0 & l_2 & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & l_N \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} \delta t_1 \\ \delta t_2 \\ \vdots \\ \delta t_N \end{pmatrix} \qquad (5)$$

Each of the coefficients α_i represents the average slowness along the path, i.e., $\alpha_i = \delta t_i \,/\, l_i$.. The fact that the off-diagonal elements of the projection matrix are zero implies that the correlation between different traveltime perturbations δt_i generated by equation (5) is zero. This is expected because once the raypaths are found in the background medium $S_o(x, z)$, the traveltime perturbations are not affected by the slowness field in the neighborhood of the ray. If only one measurement is available, the result of the inversion is $t \,/\, l$ (a constant slowness along the one string). For many different measurements the slowness estimate is just the superposition of the strings according to equation (2):

$$\delta S(x, z) = \frac{\delta t_1}{l_1}\,\phi_1(x, z) + \frac{\delta t_2}{l_2}\,\phi_2(x, z) + \frac{\delta t_3}{l_3}\,\phi_3(x, z) + \cdots + \frac{\delta t_N}{l_N}\,\phi_N(x, z) \qquad (6)$$

A strength of the string method is that once a ray is traced, an update to the slowness model along that string can be calculated immediately. The cost of the inversion is only a small increment to the cost of raytracing. The structure of the string inversion method is illustrated in Figure 1.

Inversion along strings avoids two important and time consuming steps in tomography: 1) the computation of the matrix coefficients W_{ij} and 2) the inversion of the system of equations. However, a new problem arises: the imaging and displaying of the sum given by equation (2). Normally, with orthogonal pixels, we are defining the model description for display puposes as well as the correlation of points in the image with the data. Inversion along strings does not consider these aspects, a fact which keeps the inversion simple but shifts some burden to the display. Decisions about filtering and display of the reconstruction are made after inversion, thus giving more freedom to perform post-inversion processing of $\delta S(x, z)$ for interpretation purposes. The string formulation is in its infancy. We have yet to investigate all the benefits and limitations of the method.

The essential feature of the string inversion is that the basis function has nonzero support only along a line. Thus, the intersection of a basis function with a raypath is null except when they are identically the same line, thus resulting in a projection matrix which is diagonal. Moreover, the string need not have a uniform amplitude as formulated above. An interesting example is to define the string to have the geometry of the raypath and the amplitude distribution of the background slowness model $S_o(x, z)$ used to find the raypath. In this case, we have

$$\psi_j (x, z) = \phi_j (x, z) \bullet S_o(x, z) \tag{7}$$

The coefficients parameterizing the slowness then become $\alpha_i = \delta t_i / \tau_i$, where τ_i is the calculated traveltime in the background slowness model. Furthermore, the inversion simplifies even further because the calculation of ray length is no longer required:

$$\delta S(x, z) = \frac{\delta t_1}{\tau_1} \psi_1(x, z) + \frac{\delta t_2}{\tau_2} \psi_2(x, z) + \frac{\delta t_3}{\tau_3} \psi_3(x, z) + \cdots + \frac{\delta t_N}{\tau_N} \psi_N(x, z) \tag{8}$$

where $\psi_j (x, z)$ is given by equation (7). This dimensionless formulation for the coefficients α_i makes it easier to enforce bounds on the magnitude of the perturbations $\delta S(x, z)$ by simply bounding α_i.

SYNTHETIC DATA

As described above, iterative tomography, i.e., inversion and imaging, involves two well-defined solution steps - raytracing and matrix inversion. Our implementation of these steps has some special features: first, we use vectorized ray tracing to calculate traveltimes, raylengths, and ray trajectories for a fan of rays. This avoids the time-consuming operation of linking the ray but creates a new problem of comparing observed and calculated traveltimes. For comparison purposes, the fan of observed traveltimes are interpolated onto the calculated raypaths (McMechan, Harris, and Anderson, 1987). Presently, we use 1D interpolation for each common-receiver-gather. We plan to improve upon this by using 2D interpolation on the pick data organized in the shot-receiver domain. String inversion according to equation (6) is then performed on a super fine two-dimensional grid. After all rays are traced, the result given by equation (6) is imaged, that is to say, equation (6) is filtered for display and for use as the background model for the next raytrace iteration.

The string method was coded and tested on several synthetic examples simulating cross-well surveys. Results from two of these, a fault model and a layer model, are shown in Figures 2 and 3. The datasets consisted of traveltimes generated using equation (1):

Fault model

A synthetic dataset corresponding to a fault model was created using 100 source points spaced 10 feet apart. Each source radiated 50 rays in a 90 degree fan (+/- 45 degrees wide from horizontal). Five raytrace iterations were run from a constant velocity starting model. The results are shown in Figure 2. The mean absolute traveltime error after five iterations was 0.003 ms. The displayed images are sampled at 2.5-foot vertical and horizontal intervals. However, each velocity value is computed for a 10-foot by 10-foot cell about the sample point. Some artifacts are visible. However, these represent errors of less than 2% and are due mostly to the limited view and the asymmetry introduced by the raytracing.

Layer model

The geometry for the layer model is similar except that 60 rays were used in a 120 degree fan. The results after five iterations are shown in Figure 3. As with the fault model, some limited view artifacts are evident. In particular, as the edges of the layers are better defined with iterations, anomalies are erroneously placed in the reconstruction. We suspect that these types of anomalies can be reduced by controlled damping of the inversion or with the application of constraints. Overall, the errors in velocity are less than 2%. The mean absolute traveltime error after five iterations was 0.0256 ms.

FIELD DATA

The string tomography method was then applied to a real field dataset with over 5000 traveltimes. This dataset was acquired jointly by Amoco and Stanford University. The inversion results for two different starting models are shown in Figures 4 and 5. Again, the display is sampled at 2.5 feet with 10-foot square overlapping pixels. The mean absolute error for the two results after five iterations are 0.606 ms for the constant velocity starting model and 0.599 ms for the fault starting model. The two tomograms are very similar and more importantly lead to a common interpretation. These data were also inverted using a square pixel method that produced a similar tomogram (Harris and Tan, 1990a). An interpretation of this tomogram can be found elsewhere (Harris, Mavko, et. al, 1990b).

REFERENCES

Censor, Y. 1983, Finite series-expansion reconstruction methods, Proc. IEEE, 71, pp. 409-419.

Dines, K. A., and Lytle, R. J., 1979, Computerized geophysical tomography, Proc. IEEE, 67, pp. 1065-1073.

Harris, J.M., Tan, H., et. al., 1990a, Cross-well tomographic imaging of geological structures in Gulf coast sediments, 1990 SEG Annual Convention, San Francisco.

Harris, J.M., Mavko, G., Moos, D., Nolen-Hoeksema, R., 1990b, Cross-well tomographic imaging of geological structures in Gulf coast sediments, Paper A, this volume.

Mavko, G. and Lazaratos, S., 1990, Cross-well tomographic analysis in the pick domain, Paper C, this volume.

McMechan, G. A., Harris, J. M., and Anderson, L., 1987, Cross- well tomography for strongly variable media with application to scale model data, Bulletin of the Seismological Society of America, Vol. 77, No. 6, pp. 1945-1960.

Michelena, R, and Harris, J. M., 1990, Tomographic inversion using natural pixels, Paper F, this volume.

Acknowledgements

The authors thank Amoco Production Company for its support in acquiring the data, and the Gas Research Institute for its financial support of the cross-well tomography project at Stanford University.

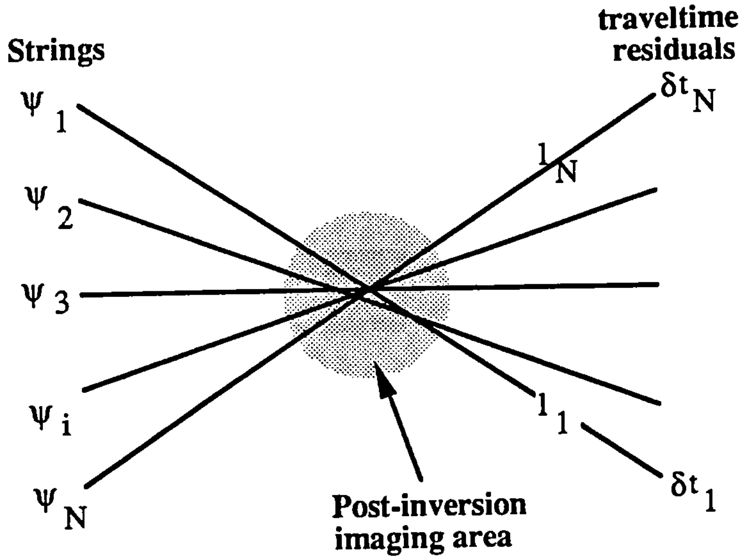


Figure 1. Schematic illustration of string reconstruction method: The traveltime residuals are normalized by the length of the string and backprojected uniformly. The result for many strings is summed, then imaged, i.e., averaged, by whatever scheme or size of display pixel desired.

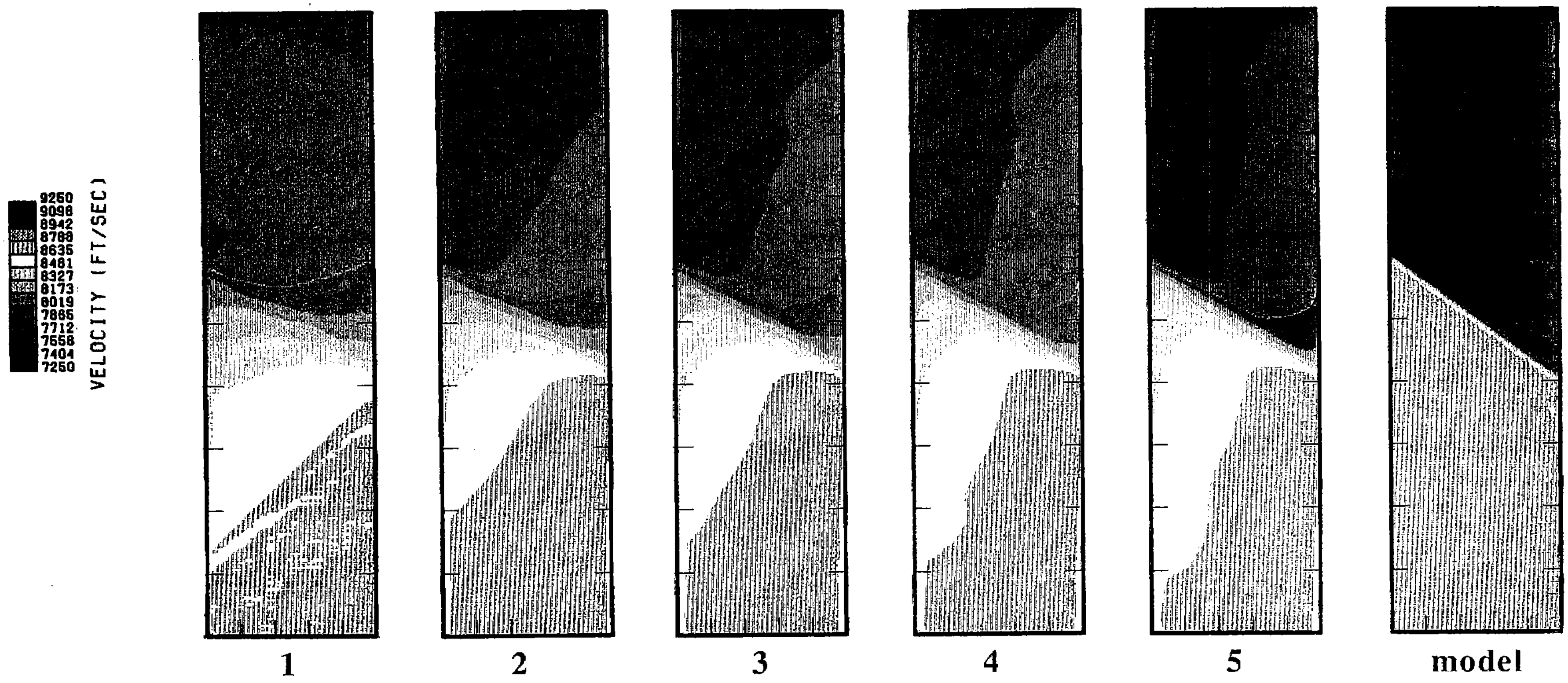


Figure 2. Five iterations of the string inversion applied to synthetic data. Distance between wells is 250 feet. A depth interval of 1000 feet is displayed.

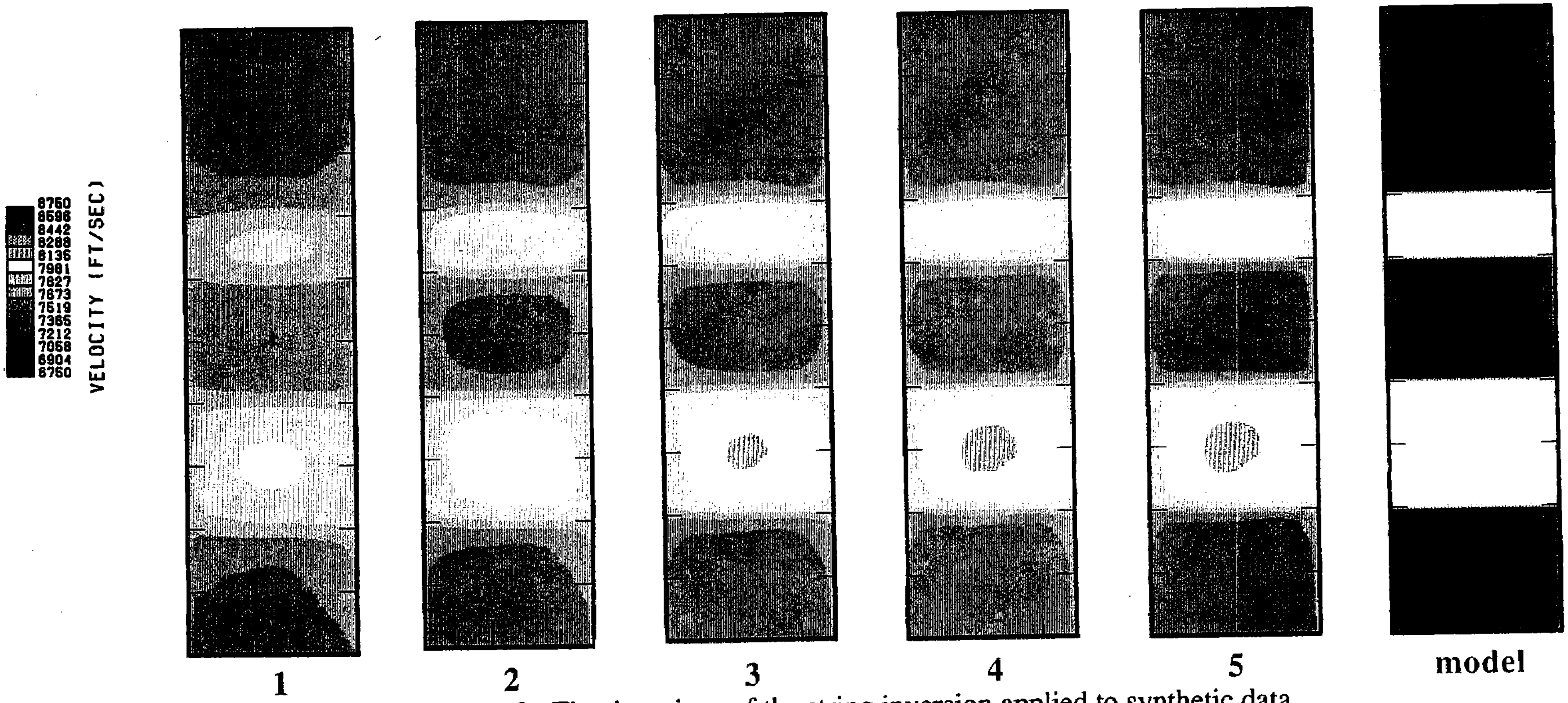


Figure 3. Five iterations of the string inversion applied to synthetic data. Distance between wells is 250 feet. A depth interval of 1000 feet is displayed.

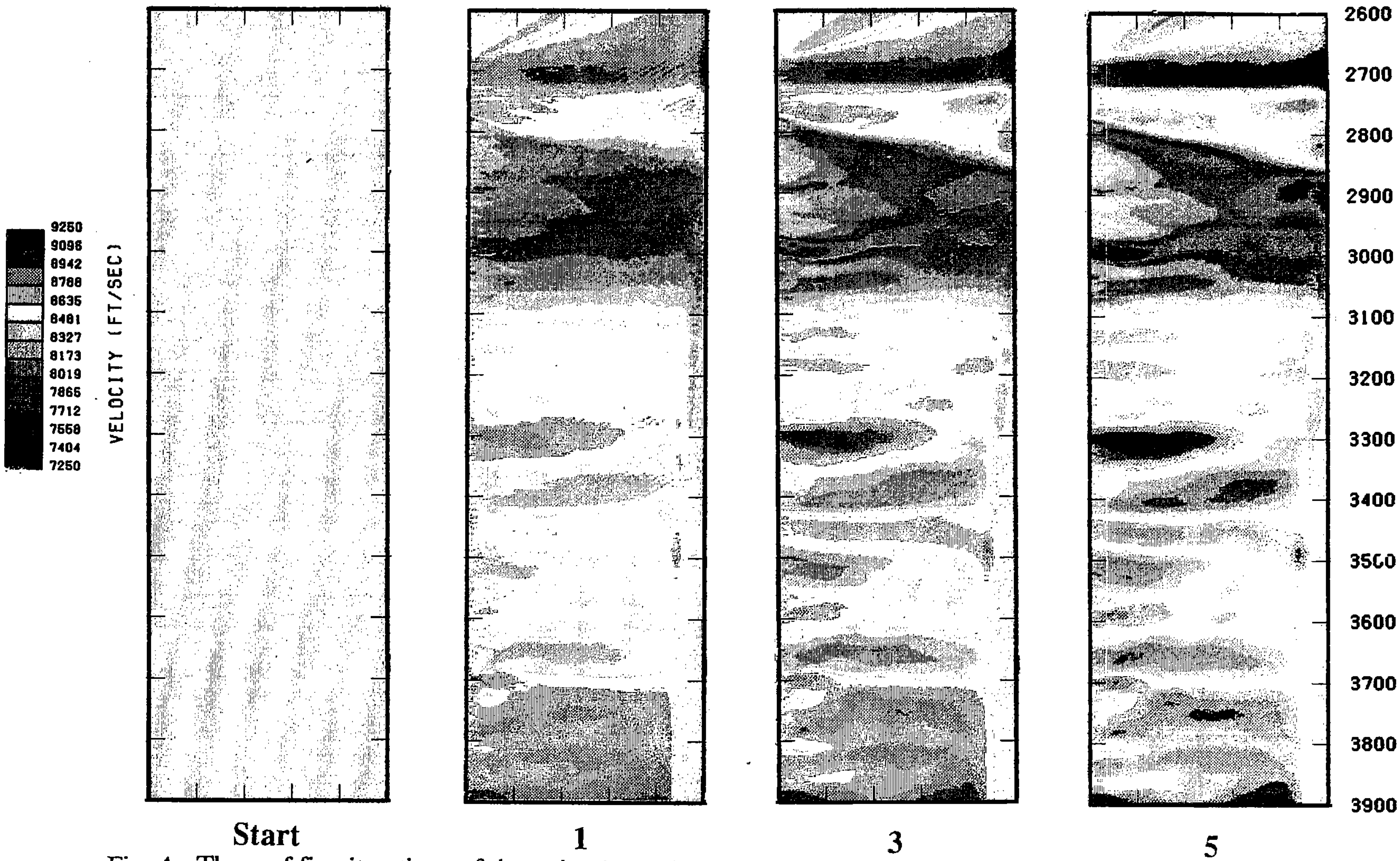


Fig. 4. Three of five iterations of the string inversion applied to field data. Distance between wells is approx. 250 feet. A depth interval of 1300 feet is displayed. The unimaged zone to the lower right of figures is outside deviated well.

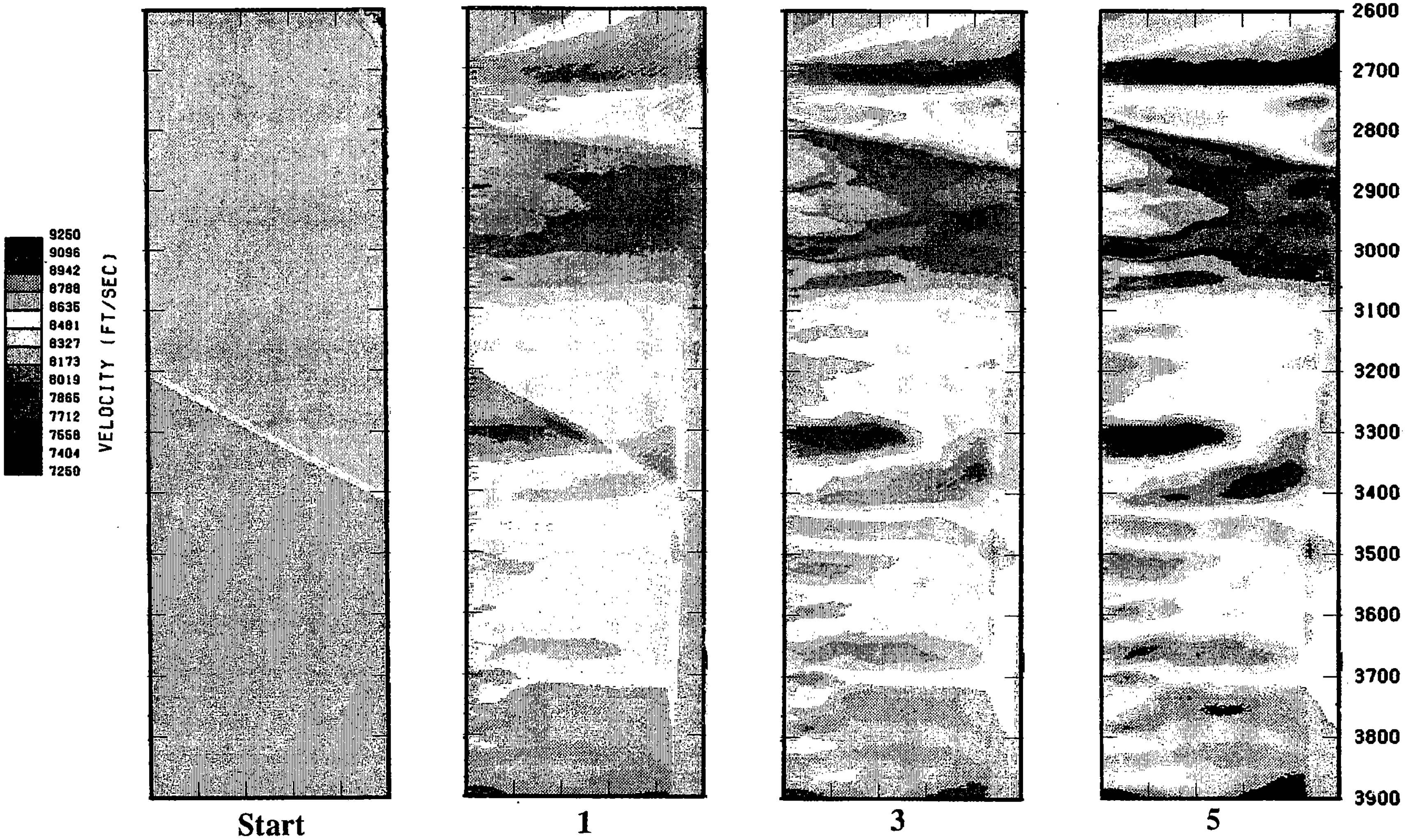


Fig. 5. Three of five iterations of the string inversion applied to field data. Distance between wells is approx. 250 feet A depth interval of 1300 feet is displayed. The unimaged zone to the lower right of figures is outside deviated well.

to the starting model. The single thing in common among all travel time inversion schemes is that they are limited by the quality of the picks, and they are all subject to errors in geometry.

In this paper we discuss a method of displaying travel time data that provides a quick and effective quality control over survey geometry and pick quality. We also present a fast method of extracting an approximate slowness image directly from the picks. The method is equivalent to straight ray tomography, and has the simple interpretation of a radon transform of the pick space. Although these techniques cannot replace a thorough analysis and inversion of the data, they will complement any further analysis, and can, in principle, be performed very quickly in the field in order to get feedback on experiment quality.

QUALITY CONTROL OF EXPERIMENTAL GEOMETRY AND PICKED TIMES

A typical cross-well survey geometry is shown in Figure 1. A series of source positions lies within one well and a series of receiver positions lies in the second. Usually there is only one source tool and a small number of receivers. The receivers typically are held stationary while the source occupies many different depths; the receivers are then moved to a slightly different depth and the source again is swept through a range of depths. When the survey is complete, the data can be sorted into high fold common source gathers or "fans" and common receiver fans, as is often done with surface seismic reflection data.

A useful way to display the geometry is to plot shot depth vs. receiver depth as shown in Figure 2. Each dot represents a recorded trace, exactly as in a "stacking chart" for surface reflection data. All of the traces along any vertical line correspond to a common source fan, and all of the traces along any horizontal line correspond to a common receiver fan. Traces along lines at +45° correspond to a common midpoint fan, which is less familiar in cross-well terminology. Similarly, traces along lines at −45° correspond to common offset sections. When displayed this way, one can quickly see the fold, aperture dimensions, and variations of coverage with depth.

Traces from a typical high quality common receiver fan from an actual field experiment are shown in Figure 3, from which first arrival times were picked. A convenient way to display the times is by plotting them in the form of a colored stacking chart, as in Figure 4. For more sensitivity, the times plotted here are actually the differences relative to travel times in a uniform velocity medium, $\Delta t = t_{pick} - r/V_0$, where t_{pick} is the picked arrival time, r is the distance between the source and receiver, and V_0 is the uniform reference velocity. The positions of the colored dots in Figure 4a again show the survey geometry, as well as highlighting traces for which there are no picks. Figure 4b shows the same picks interpolated to make more clear the travel time patterns. One can immediately see systematic variations of residual travel time with depth, as well as with offset. Figure 4c shows the first attempt at creating the display in Figure 4b. The striking horizontal stripes show systematic shifts within

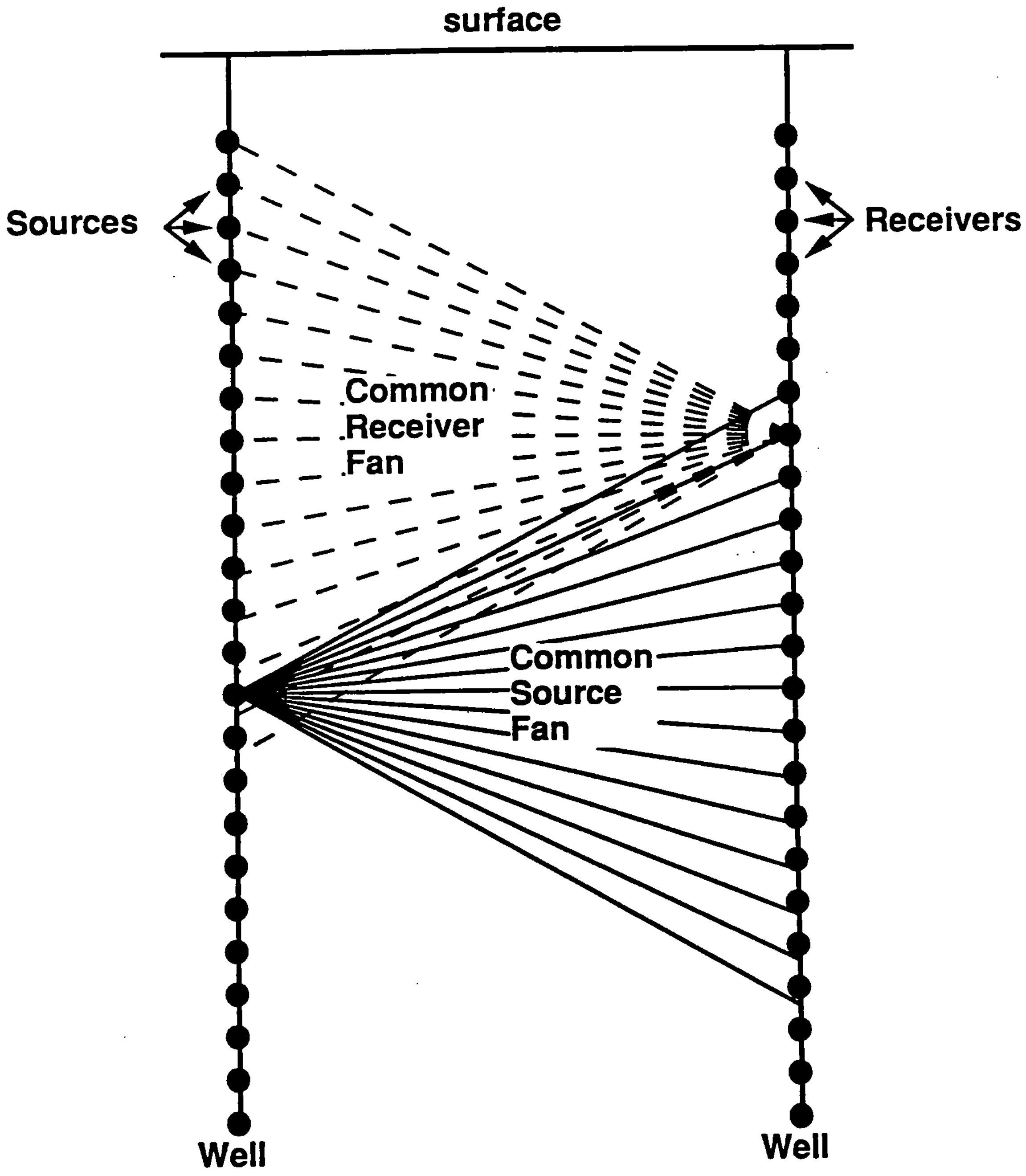


Figure 1: Typical crosswell survey. Sources lie in one well, receivers in the other. As with surface seismic data, traces can be sorted by common source, common receiver, common midpoint, or common offset.

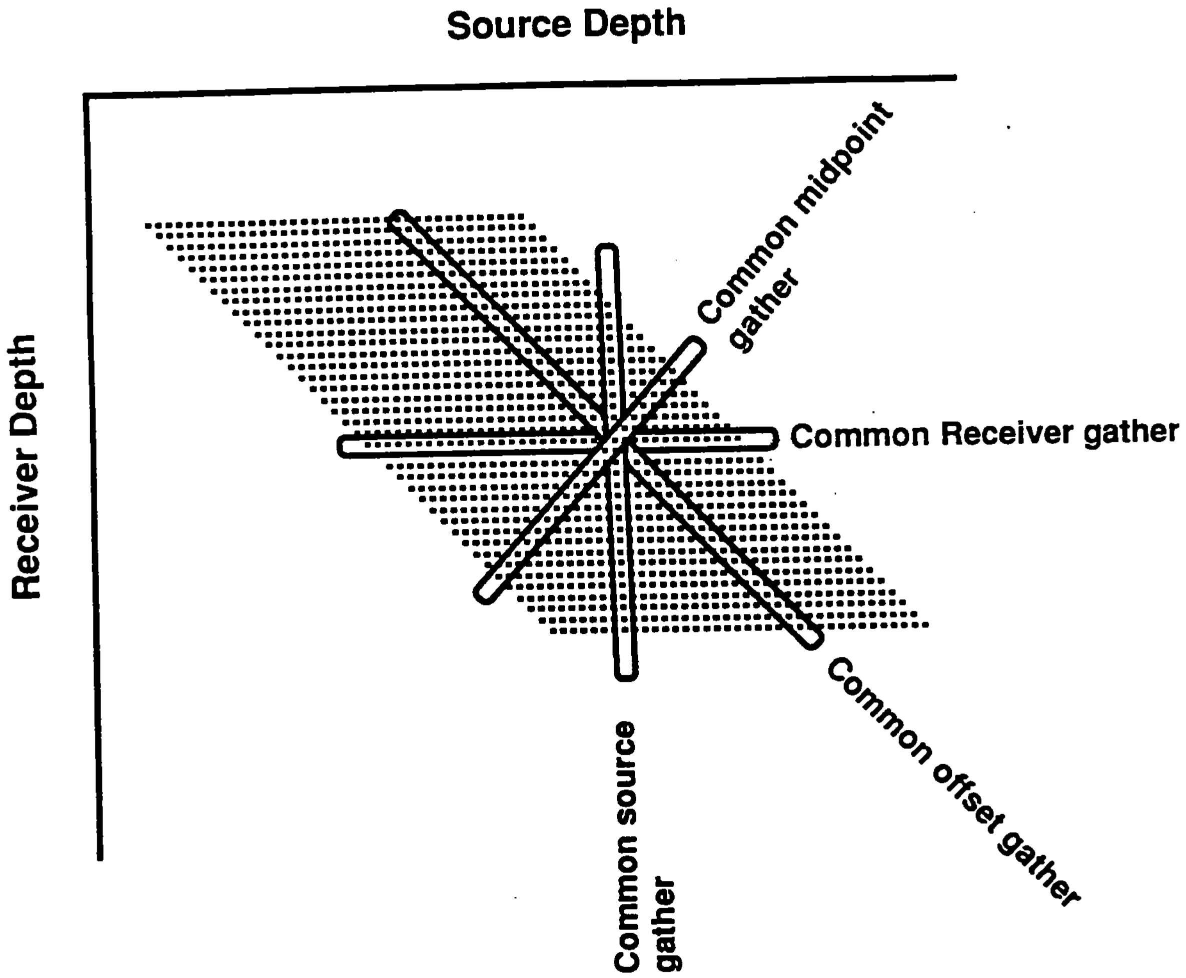


Figure 2: Stacking chart display of crosswell survey geometry.

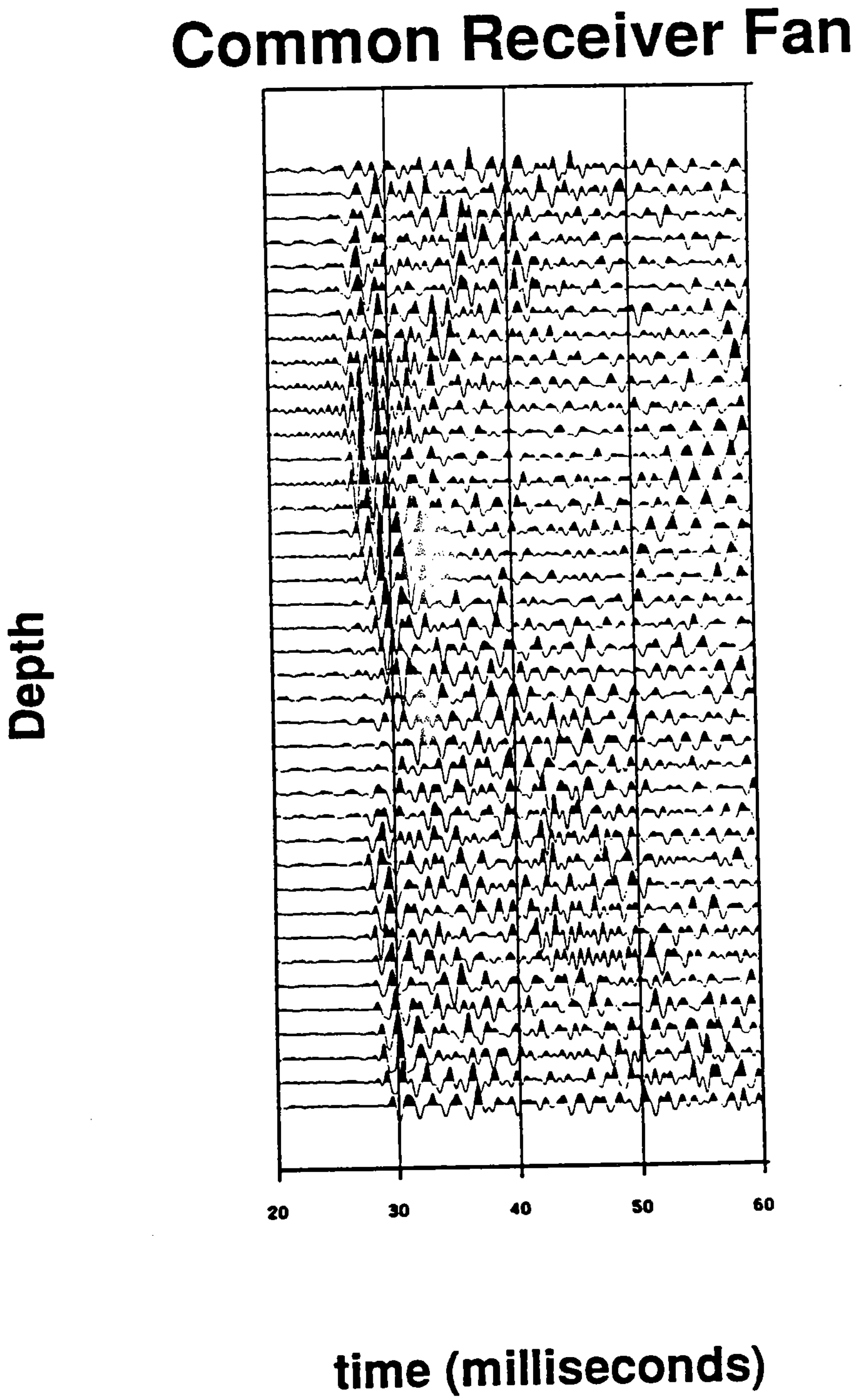


Figure 3: Near traces of a cross-well common receiver fan.

common receiver fans, that turned out to be a geometry error in extracting the receiver depths from the observer's logs. The unusual pattern was an immediate clue that something was wrong, and the horizontal nature indicated that it was systematically related to the receivers. The corrected display in Figure 4b shows very subtle horizontal streaks. The arrival times were picked by hand within common receiver fans; we suspect that occasional leg jumps were made when switching from one fan to the next, leading to slight receiver-consistent errors.

INTERPRETATION OF ARRIVAL TIMES

The travel time patterns in the corrected display, Figure 4b, directly indicate velocity variations in the formation between the wells. In general vertical streaks are related to a velocity anomaly near a source; horizontal streaks are related to an anomaly near a receiver, and streaks at other angles are related to anomalies between the source and receiver. One can also see the variation from slow to fast with depth. It would be useful if more detailed patterns can be recognized from these displays without the need for formal inversions. We explore this using synthetics.

Figure 5b shows a model which consists of 5 laterally uniform zero dip layers. Two wells are placed 100 meters apart. Sources are positioned every 4 m in the well on the left, between depths of 0 m and 400 m. Receivers are positioned every 4 m in the well on the right, between depths of 100 m and 300 m. Except at the top and bottom ends, each source is recorded by receivers ranging in depth from +/- 100 m of the source depth, and each receiver records sources ranging in depth from +/- 100 m of the receiver depth.

Rays were traced through the model to get travel times which we call the synthetic "picks". These are plotted in the stacking chart form in figure 5a, and have been divided by the straight line distance between source and receiver, to give them dimensions of slowness. The picks along the line $Z_{source} = Z_{receiver}$, which is the *zero offset* line, correspond to a cross-well log. Although rays may bend, this has the simple interpretation of horizontally propagating rays that sample the average slowness as a function of depth. The slownesses along this line agree very well with the model velocity vs. depth, as we expect. Other common offset gathers of picks give the average slowness between pairs of sources and receivers having a fixed vertical offset. The symmetric pattern of picks about the zero offset line indicates lateral symmetry in the model. The pattern of squares arranged along the zero offset line is characteristic of uniform flat layers; the square width is equal to the layer width. Although the synthetic rays were not straight, these patterns can be qualitatively compared with the Radon transform of simple slowness functions, as illustrated for example by Deans (1983).

A second model is shown in Figure 6b and consists of a single dipping fault separating two homogeneous layers. The source and receiver geometries are the same as before. The picks, converted to slowness, are plotted in Figure 6a. Again, the zero offset line shows the average slowness as a function of depth, which now has a

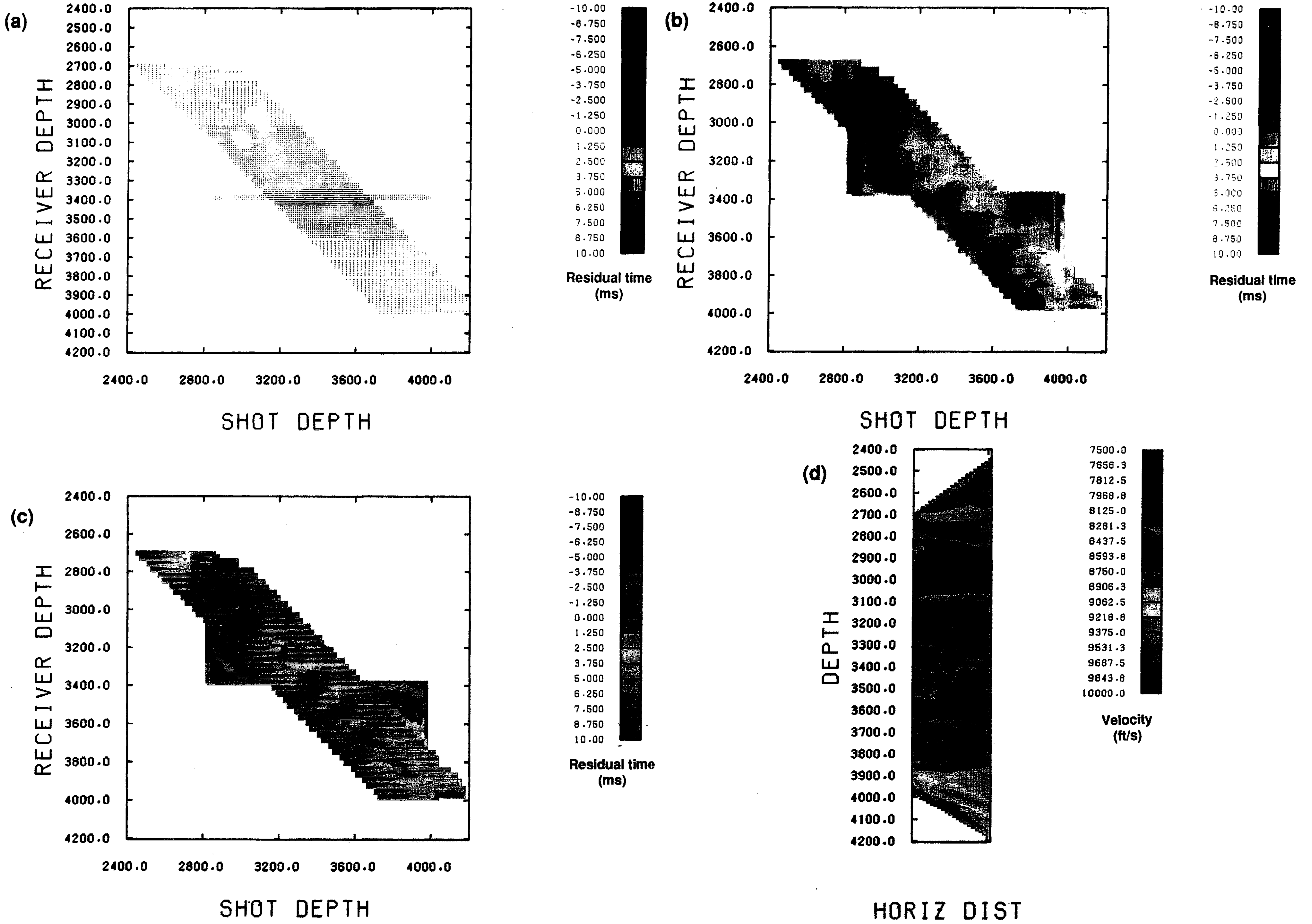


Figure 4: (a) Stacking chart display of crosswell survey trace attributes, in this case the picked first arrival times. Dot positions describe survey geometry; dot colors show residual travel times relative to an average velocity. (b) Residual travel times interpolated to better show the colored patterns of the travel time residuals. (c) Interpolated travel time residuals revealing a receiver-consistent geometry error. (d) Tomographic image obtained from the Radon transform of the picks in (b).

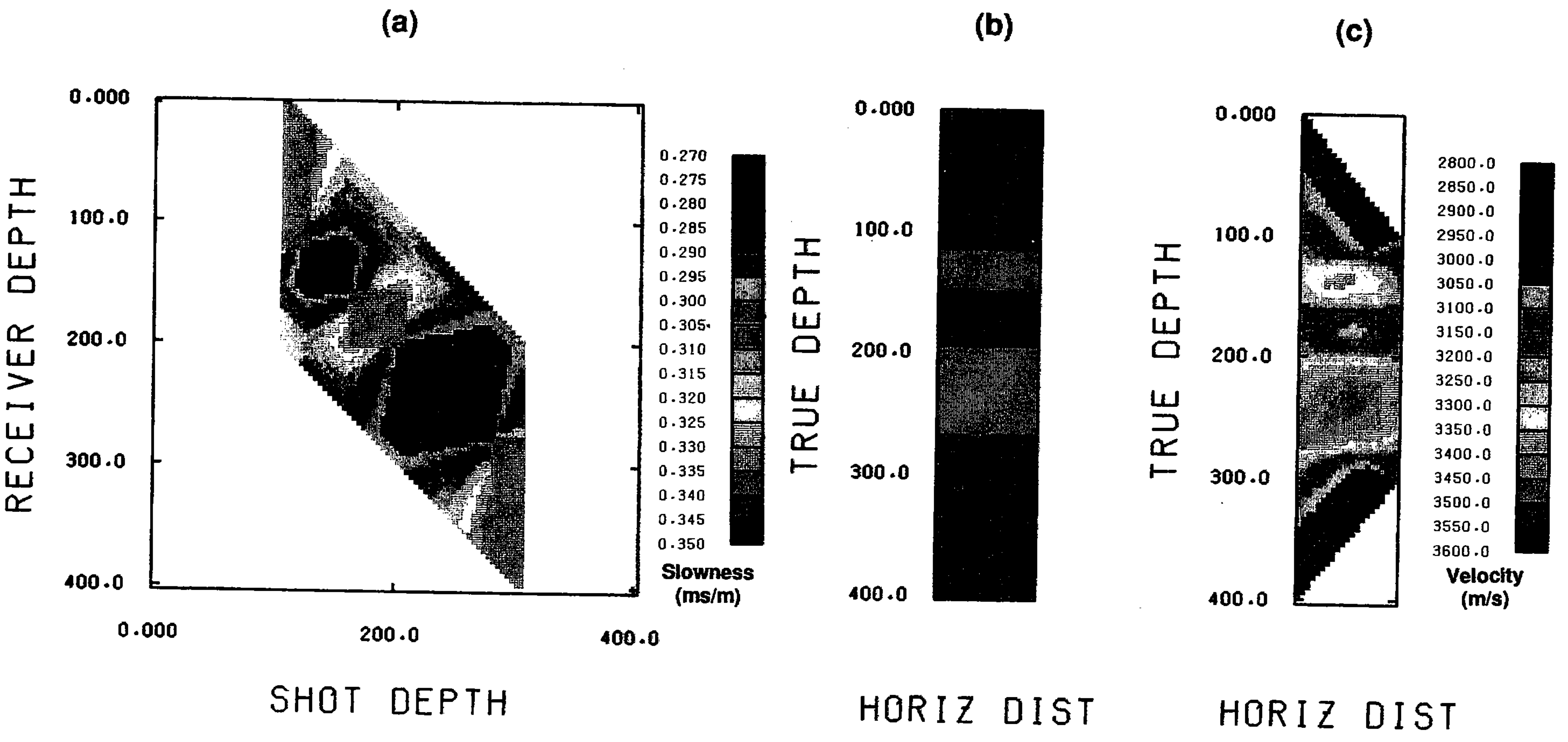


Figure 5: Flat layer synthetic study. (a) Picks converted to slowness obtained from curved rays traced through the model (b). (c) Image obtained from the Radon transform of the picks in (a)

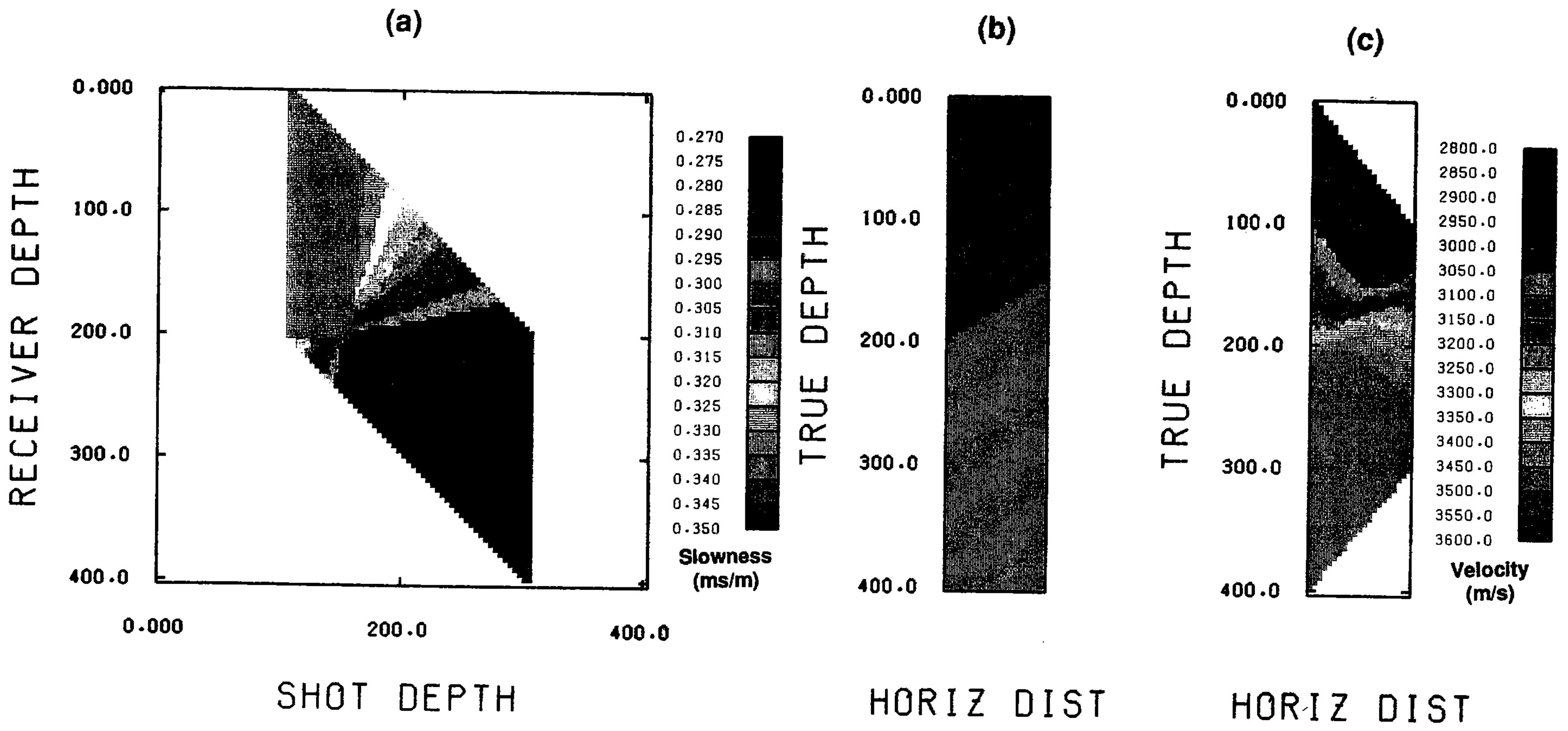


Figure 6: Faulted model synthetic study. (a) Picks. (b) Model. (c) Image obtained from the Radon transform.

gradient between depths of 150 m and 200 m which is the depth range spanned by the fault. Any lateral heterogeneity will tend to introduce gradients along the zero offset line. There is once again the pattern of squares, but it is offset because of the dipping fault. Since the pick domain plot is approximately the Radon transform of the model space, a rotation of the velocity boundary corresponds to a shift in the pick domain.

Our intuitive understanding of these patterns is based on averages of slowness along various ray paths. These interpretations have amounted to mentally inverting the Radon transform. We now discuss the process of more formally transforming the pick domain.

RAPID STRAIGHT RAY INVERSION IN THE PICK DOMAIN

Consider first a heurist arguement. As already discussed, all of the traces along any vertical line in the pick domain lie within a common source fan (Line A, in Figure 7a). These rays have in common the region immediately adjacent to the source at the well, and otherwise sample very different parts of the interwell region. One can imagine, then, that averaging all of these picks will tend to yield the near-source slowness and average out the variations away from the source. We therefore plot this average slowness at a point next to the source (point A in Figure 7b). Similarly, we imagine that averaging all of the picks along any horizontal line (line B in Figure 7a) gives the average slowness near the corresponding receiver, and we plot the average slowness at a point near the receiver (point B in Figure 7b). An average of the picks along a line at +45° (line C in Figure 7a) represents the average of all straight rays passing though a particular mid point (point C in Figure 7b). Finally, all of the straight rays passing through a point at location (x, z) between the wells (point D in Figure 7b), lie along a line with angle $0^o \leq \theta \leq 90^o$ in the picks domain (line D in Figure 7a), where

$$\theta = tan^{-1}(\frac{X_0 - x}{X_0 + x})$$

and X_0 is the well spacing. This process of averaging is equivalent to taking a Radon transform of the pick domain. Each vertical line $x = constant$ in the image corresponds to one view of the transform.

Figures 4d, 5c, and 6c show the images that result from this process applied to the picks in Figures 4b, 5b and 6b, with slowness converted to velocities. It is clear that the Radon transform does a good job of estimating the various images. In each case the "inversion" took only seconds of cpu time on a Sun4 workstation.

We now discuss why this method works so well.

JUSTIFICATION OF THE INVERSION PROCEDURE

Each travel time measurement represents a line integral of the slowness along the raypath connecting source and receiver. For straight rays this is simply the Radon

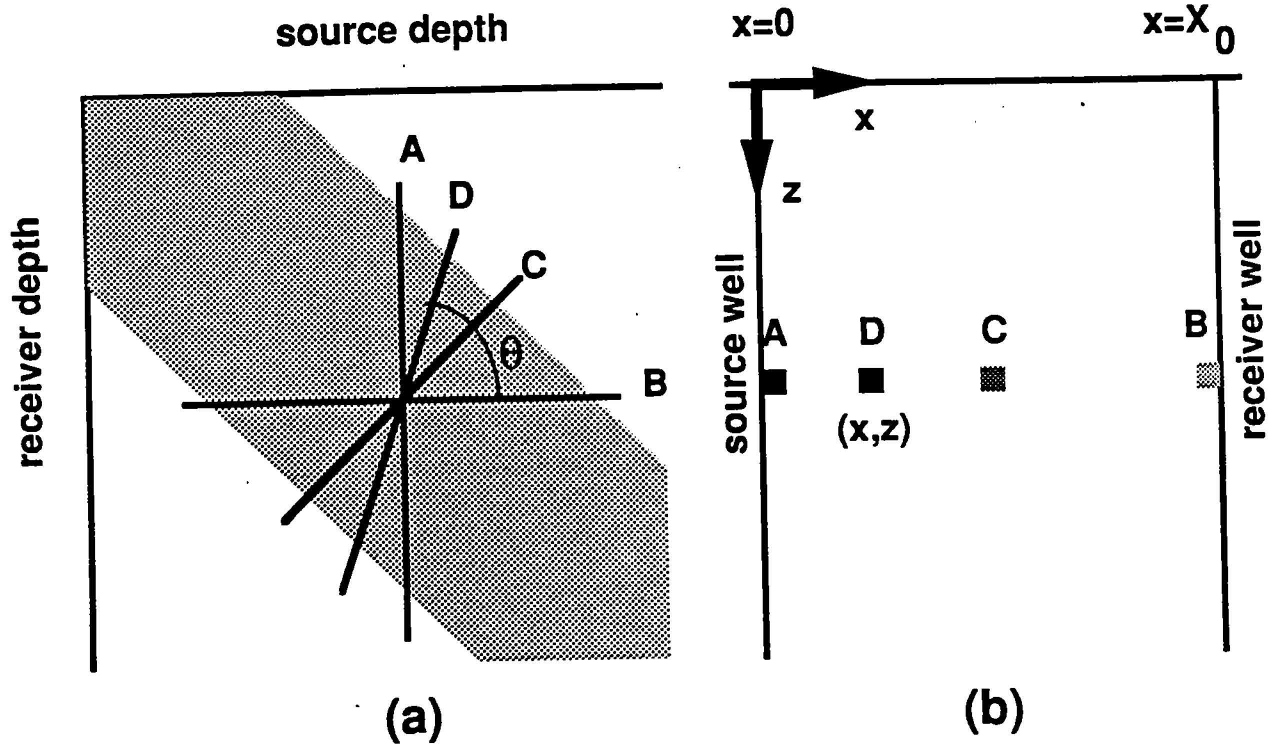


Figure 7: The imaging technique. Pick domain is shown on the left; space domain is shown on the right. Picks along line A (left) all pass through the region A (right), so the average of the slowness picks is plotted at pixel A. Similar for B, C, D, etc.

transform of the slowness and can be written formally as

$$t_\phi(l) = \int\int f(x,y)\delta(l - xsin\phi + zcos\phi)dxdz \tag{1}$$

where $\delta(l)$ is the Dirac delta function (Deans, 1983). The quantity t_ϕ is the travel time associated with a ray whose distance from the origin of the spatial coordinate system is $l = xsin\phi - zcos\phi$ and which forms an angle ϕ with the x axis. The two dimensional function $f(x,z)$ is the slowness that we wish to reconstruct.

The inverse of the Radon transform can be written formally as (Deans, 1983):

$$f(x,z) = -\frac{1}{4\pi}\int_0^{2\pi} \mathcal{H}\{\frac{dt_\phi}{dl}\}d\phi \tag{2}$$

$$= \frac{1}{\pi}\int_0^{\pi} t_\phi(l) * R(l)d\phi \tag{3}$$

where the operator $\mathcal{H}$ represents the Hilbert transform. The combination of the Hilbert transform and the derivative is known as the rho-filter, $R(l)$. We will demonstrate that our pick domain imaging scheme is a low frequency, discrete approximation of the exact inverse of the Radon transform. Unlike the familiar matrix inversion algorithms (ART,SIRT), it is not an iterative procedure. It is similar to the backprojection algorithms that are popular in medical imaging and radio astronomy. A good review of these techniques can be found in Lewitt (1983).

In the spatial frequency domain, the Radon transform inverse equation (3) can be written as

$$F(\rho,\theta) = \rho\hat{T}(\rho,\theta) \tag{4}$$

where ρ and θ are frequency domain polar coordinates, $F(\rho,\theta)$ is the two dimensional Fourier transform of the slowness image $f(x,z)$, and $\hat{T}(\rho,\theta)$ is the two dimensional Fourier transform of the unfiltered backprojection image $\hat{t}(x,z)$, defined as (Deans, 1983)

$$\hat{t}(x,z) = \frac{1}{\pi}\int_0^{\pi} t_\phi(l)d\phi \tag{5}$$

For a finite number of view angles, ϕ_i, $i = 1,2,...N$, equations (3) and (5) can be approximated as

$$f(x,z) = \frac{1}{\pi}\sum_{i=1}^{N} t_{\phi_i}(l) * R(l)\,\Delta\phi \approx \frac{1}{N}\sum_{i=1}^{N} t_{\phi_i}(l) * R(l) \tag{6}$$

$$\hat{t}(x,z) = \frac{1}{\pi}\sum_{i=1}^{N} t_{\phi_i}(l)\,\Delta\phi \approx \frac{1}{N}\sum_{i=1}^{N} t_{\phi_i}(l) \tag{7}$$

where $\Delta\phi = \pi/N$. Note that f has units of slowness and $\hat{t}$ has units of time. The difference between the images in equations (6) and (7) is the rho-filter, expressed as a one dimensional convolution of the individual views with $R(l)$ in the space domain, or multiplication with the spatial frequency ρ in the Fourier domain.

The rho-filter is the theoretically accurate way of transforming the time image $\hat{t}$ into the slowness image f. It accomplishes both a change of dimensions, by dividing by a length, and an enhancement of the high frequencies. However, the rho-filter is optimal for these purposes only when the rays are straight and when the projection angles ϕ_i range from 0 to π. With limited views and noisy data, a more conservative approach is to approximate the rho-filter with a filter $\tilde{R}(l) \approx R(l)$ that changes the dimensions of the projections to slowness with little or no boosting of the high frequencies. Our inversion method uses the simplest filter that accomplishes this, which is a constant $\tilde{R}(l) = \delta(l)/L_\phi$ where L_ϕ is the length of the ray path. The approximate image can then be written as:

$$f(x,z) \approx \frac{1}{N}\sum_{i=1}^{N} t_{\phi_i}(l) * \tilde{R}(l) \tag{8}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} t_{\phi_i}(l)/L_{\phi_i} \tag{9}$$

This filter converts each pick to the average slowness along the ray path, and its frequency response is flat. In practice, this is implemented by dividing each pick by the source-receiver distance, $L_\phi(l)$, which may be different for each ray, so it is not actually a convolution. For a very small finite number of views this gives a better estimate of the slowness image than using the rho-filter. For example, a single zero offset view will yield a laterally uniform image with the correct average slowness.

A different interpretation of our approximate method can be made in terms of direct Fourier reconstruction. The desired slowness image $f(x,z)$ can be written as

$$f(x,z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(k_x,k_z) e^{i2\pi(x\ k_x+(z\ k_z)} dk_x dk_z \tag{10}$$

$$= \int_0^{\pi}\int_0^{\infty} F(\rho,\theta) e^{i2\pi r\rho\ cos(\theta-\phi)} \rho d\rho d\theta \tag{11}$$

where F is the two dimensional Fourier transform of the slowness image expressed in either cartesian or polar coordinates. Polar coordinates are a natural choice for inverting tomographic data, because the travel times for each view determine spatial frequency components of the image that lie along a radial line (constant θ) in the frequency domain. A discrete approximation of the integral in equation (11) as a summation involves multiplying each measured Fourier component $F(\rho,\theta)$ by the polar coordinate element of 'area' $dA = \rho\ d\rho\ d\theta$, which has the effect of weighting high frequency (high ρ) measurements more than low frequency ones. This is equivalent to assuming that the unmeasured frequency components which lie between measured angles θ are equal to the nearest measured value at the same ρ. Our method instead weights all measured components equally by an element of 'area' $dA = d\rho\ dk_\perp$ where $dk_\perp$ is a tangential increment in the frequency domain. This resembles more a cartesian discrete fft of length M_{fft}, with local coordinates oriented parallel with a particular radial line θ, in which

$$dk_\perp = \frac{1}{M_{fft}\Delta s} \tag{12}$$

where Δs is the sample increment in the direction of the ray path. Then $M_{fft}\Delta s$ has the interpretation of the ray length L. In summary, the low frequency aspect of using the Radon transform as an inverse of the Radon transform can be viewed as a result of sampling the high frequencies less densely than the low frequencies, and reconstructing the image using only the measured Fourier components.

We have shown that except for the enhancement of high frequencies, our algorithm is an approximate one-step inversion of the Radon transform. If we ignore the rho-filter, the inversion extends naturally to the curved ray case. Hence, we can easily perform the same type of inversion in the real (x, z) space by back-projecting along the curved rays. This corresponds to a generalized Radon transform along curved trajectories in the pick domain. To implement the curved ray case requires knowing these curved trajectories.

CONCLUSIONS

We have presented a method of displaying cross-well seismic travel time data that provides a quick and effective quality control over survey geometry and pick quality. Times plotted in stacking chart form clearly show localized anomalies as linear streaks, which can be investigated as either geometry errors or true slowness anomalies. Other simple patterns corresponding to layers and faults can be qualitatively interpreted from published Radon transforms of simple functions. The method of display can obviously be extended to other trace attributes such as amplitude, pick quality, phase, or frequency.

We have also shown a very fast method of extracting an approximate slowness image directly from the pick domain. The operation is a Radon transform of the picks, and except for not enhancing the high frequencies, it is a one-step inversion of straight ray travel times. Although the method is not intended to replace a more thorough inversion of the data using curved rays, its speed and simplicity make it well suited for an initial assessment of the data. Such an imaging scheme might be incorporated, for example, into an interactive picking program, so that the image can evolve real time as picks are made. If performed in the field, the imaging can also give rapid feedback on the data acquisition.

The inversion method extends naturally to the curved ray case. Hence, we can easily perform the same type of inversion in the real (x, z) space by back-projecting the average slowness along the curved rays. This corresponds to a generalized Radon transform along curved trajectories in the pick domain. To implement the curved ray case requires knowing these curved trajectories, which is a topic of current study.

REFERENCES

Deans, S.R., 1983, The Radon Transform and Some of Its Applications, John Wiley & Sons, New York.

Devaney, A.J., 1984, Geophysical diffraction tomography, IEEE Trans. Geoscience and Remote Sensing, GE-22, 3-13.

Harris, J.M., 1987, Diffraction tomography with arrays of discrete sources and receivers, IEEE Trans. Geoscience and Remote Sensing, GE-25, 448-455.

Justice, J.H., Vassiliou, A.A., Yale, D.P., and Lin., C.J., 1989, Rock properties and tomographic imaging in reservoir surveillance, 59th Ann. Internat. Mtg., Soc. Explor. Geophys., Expanded Abstracts, 760-762.

Justice, J.H., Vassiliou, A.A., Singh, S., Logel, J.D., Hansen, P.A., Hall, B.R.,Hutt, P.R., and Solanki, J.J., 1989, Acoustic tomography for monitoring enhanced oil recovery, The Leading Edge, 8, 2, 12-19.

Lewitt, R.M, 1983, Reconstruction algorithms: transform methods, Proceedings of the IEEE, 71, 3, 390-408.

Lines, L.R., and LaFehr, E.D., 1989, Tomographic modelling of a cross-borehole data set, Geophysics, 54,10, 1249-1257.

Lo, T-W., 1987, Seismic borehole tomography, Ph.D. dissertation, Massachusetts Institute of Technology.

PAPER D

High Resolution Cross-well Seismic Data Acquisition System

Jerry M. Harris and Daniel Moos
Seismic Tomography Project

Summary

A new hardware system for acquisition of high frequency cross-well seismic data has been developed for the Seismic Tomography Project. The receiving system employs a 9-element array of hydrophone detectors and an optional 3-component accelerometer detector. Detected signals are digitized downhole at sample rates of 50 to 1000 microseconds. Downhole filters, gain, stacking depth, and sample rates are controlled from the surface. The data are recorded on disk and tape in SEG Y format. Each detector sonde operates as a node on a local-area-network; therefore, additional sondes can be easily added to the system. The receiver system is operationally specified to 10,000 psi and 125 C on standard oil field 7-conductor wirelines. The system is designed to operate with the high frequency piezoelectric source, e.g., 200 Hz to 4,000 Hz, but can be used with minor changes with other sources. The pieozelectric source is the cylindrical bender design and has two active elements. The two elements are assembled into a balanced mechanical structure which can be operated in phase for a dominant monopole or out of phase as a vertical dipole. Power is delivered to the source via standard industry wireline. The source is designed for deep hole operation to 5,000 psi and 125 C. This new hardware is available for experiments ranging from cross-well imaging surveys and vertical seismic profiles to in-situ measurements of rock properties.

Introduction

The hardware described herein addresses two related data acquisition problems which plague the volume and quality of cross-well seismic data - the speed of acquiring the data and the fidelity of the recordings. This aspect of cross-well seismology is important for it affects our assessment of the potential and limitations of the cross-well method for imaging and in-situ measurements of rock properties.

As cross-well seismology emerges as a tool for reservoir characterization and reservoir monitoring, the problem of high speed data acquisition continues to limit its use. Even today, after many successful applications of the method, routine surveys are rarely larger than 2500 shot-receiver points because of the long time its takes to acquire each record. Not only does this limit the size of the dataset, ultimately it limits access to wells, particularly in producing fields where shut down for seismic tomography is, more often than not, difficult to obtain. Although speed itself is not the primary need of the Tomography Project, there is the need to obtain larger more complete datasets for 2D and 3D imaging purposes. We believe the new system is capable of recording a 10,000 shot-receiver survey in 24-48 hours, perhaps two to ten times faster than other systems routinely available today.

Perhaps even more important is the quality or fidelity of the recordings. Fidelity refers to the accuracy at which the recording reproduces the input seismic signal. It is measured in terms of detector sensitivity, distortion, signal-to-noise ratio, bandwidth, dynamic range, and dynamic resolution. Our new system makes improvements in two areas. First, the signals are digitized downhole, thus eliminating the contamination and distortion experienced by analog transmission on long and varying quality wirelines. Second, the system is capable of recording to a Nyquist frequency of 10,000 Hz, though a 4000 Hz high cut filter is available for the control of aliasing. These and other features (e.g, floating point gain, low harmonic distortion, etc) combine to give the system a dynamic range of over 120 dB with 16 bits of instantaneous resolution. This high resolution recording system is designed and built to out specifications by Century Geophysical Corporation, a leading manufacturer of digital logging equipment for years.

All of this detection and recording bandwidth is useless unless a broadband downhole source is available. For this purpose, we have acquired a cylindrical piezoelectric bender source from Southwest Research Institute. The source radiates in the band of about 300 Hz to over 5000 Hz and operates on standard wireline. We have added a feature allowing it to be driven either a monopole for enhanced p-waves or dipole for enhanced s-waves.

General System Overview

A general system block diagram is shown in Figure 1. The downhole system consist of a 9-element hydrophone array and a 2-element piezoelectric source. Two surface computers (one in the source truck and one in the receiver truck) provide joint control and monitoring of all aspects of the data acquisition survey. These computers communicate via a high-speed optical link (19,200 baud) and share in common the survey parameters.

The entire downhole receiver system functions in a sense as a local area network (LAN), in that each element is a "node" which can be polled using its hard-wired ID, or can transmit data at its own behest. Software running in parallel on the two surface computers controls system configuration, including selecting the source signature, receiver sampling rates, analog and digital gains, stacking depth, and filter settings. These control computers also monitor cable depth and provide display of user selected recorded data traces. The data is recorded in SEG-Y format.

For purposes of description, the system is conveniently separable into two major subsystems, each having both surface and downhole components:

(1) Downhole receiver & surface control system
(2) Downhole source & surface control system

The surface hardware includes two Compu-log computers with the following general specifications:

- 25 MHz 80386
- 25 MHz DSP for digital communications and real-time processing
- MS/DOS with multi-tasking
- High-speed optical link between computers
- 120 MByte disk (max. 32 MByte file size)
- 60 MByte cartridge backup
- SEG-Y format recording

System Operation

The system can be operated in two basic modes. Set-up mode allows, under manual control, repeated testing and display of various survey parameters . In this mode the system does not use the information about survey geometry. Each shot is acquired, displayed, and written to disk under the operator's control. Spectral analysis, correlation, and display on the receiver system provide additional quality control on the survey set-up. Data can be offloaded to other systems for further enhancement and display.

For operations in automated mode, the survey software provides a acquisition design for a fixed detector array and a moving or scanning source. Each source point is stored as an individual record (common-receiver-gather) in a SEG-Y data file. The ranges of source and detector depths and coverage angles are specified prior to the start of the survey. Wireline moves are orchestrated by the software after completion of source scan. Source and detector positions are monitored by the computers during the move, and an alarm sounds when the next scheduled depths is reached. When the operator is ready the new recording sequence is then initiated to acquire the next shot gather. Thus data is recorded and written to disk with little or no operator intervention.

All survey parameters can be modified by the operator while data is being recorded. These changes are written into a log file and into the SEG-Y trace headers. Although system limitations do not allow SEG-Y files to exceed 32 mbytes, in practice this is not a severe limitation, as typical acquisition strategies would store 320 kbytes per source-gather (1 source record and 9 detector records, each 32 kbytes long). Thus each SEG-Y file could contain up to 100 source points, sufficient in most cases to record an entire source scan.

The new system substantially decreases the amount of time required to complete a cross-well survey. Data is double buffered in the downhole sondes, and thus transmission of data from a completed shot can continue during recording of the subsequent shot. At 38.4 kbaud, all nine 16 kword signals can be transmitted in 1 minute. In principle, this means that the complete shoot/move/ready-to-shoot sequence can be completed in the time it takes to shoot and move, without waiting for data transmission delays. For typical waveform (sweeps or pulse codes) lengths and recording times, a 64-fold vertical stack can be completed in approximately one minute. Moves take less than 30 seconds. Therefore, a fan of 100 source positions would require about two hours to complete. Data backup to streaming tape requires about 15 minutes, and can be accomplished during re-positioning of the receiver string and the source. Later versions of the software will do backup on removeable disks during data acquisition, and thus require no additional time for this process. The increase in number of receivers to nine from our earlier 3-receiver system translates in to a 3-fold decrease in acquisition time. Thus, we can complete surveys at least three times faster, and with considerably better quality control, than previously.

Description of the Subsystems

Receiver Sub-System

The receiver subsystem consists of a computer controller and, in its present configuration, nine separate and independent downhole receiver sondes, each of which acquires data digitally and communicates with the surface computer. Power for the downhole array is supplied from the surface. The system is interfaced to a digital encoder for monitoring wireline depth.

A master control and telemetry unit interface to the cable with a Schlumberger 7-conductor cablehead. High speed 16-bit digital signal processors (DSPs) are distributed at each hydrophone. These DSPs control the digitization rate and amplifier gain, and stacks the signal in one of two 16 kword buffers. The buffers allow data transmission during acquisition and during array movement for greatly increased throughput. Two low and two high cut filters (or out) can be set by from the surface. Additional filter settings can be obtained by switching the filter cards.

Surface control & recording system: Traces may be monitored on the receiver truck and plotted on a dot matrix printer. A custom software package displays and prompts the operator according to a user-specified shooting pattern. From the surface, the operator may configure the downhole digitizer as follows:

- sampling rate
- record length
- low cut filters
- high cut filters
- downhole digital gains
- downhole static gains
- recording delays
- vertical stack depth

Due to the high speed of the DSP at each detector sonde, the number of stacks per second is limited only be the record length and the sampling rate. Data are recorded in SEG-Y format on a 120 megabyte hard disk. A 60 megabyte tape drive is used for data archive. Future plans will replace the tape drive with a removable hard disk for faster transfer.

Downhole sondes. The detectors are OAS deep ocean hydrophones. The sondes are designed to allow easy upgrade to 3-component geophones or accelerometers. Each hydrophone is housed in a slotted stainless steel enclosure connected with 7-conductor logging cable. As described above, each sonde has a dedicated DSP for digitization. The design specifications call for total harmonic distortion of less than 0.02% with fixed gain, a noise floor at the input to the digitizer gain stage of less than 100 microvolts RMS, and input noise at the preamplifier of less than 10 nanovolts/root Hz. These specs are either met or exceeded in the delivered system. Downhole sonde specifications summary:

- OAS model E-4SD hydrophone; (can use one 3-comp VSP sonde)
- 16-bit DSP;
- 16 bit A/D;
- 50μs sample interval from 50μs to 1ms;
- 16,384 sample records;
- stacking to 32,768 signals;
- hi-cut filter at 2KHz or 4KHz;
- lo-cut filter: none; lo (250 Hz); hi (350 Hz);
- analog gains 60dB; 100dB;
- digital gain x1, x10, x100, or Tracking;
- programmable delay in 50μs intervals to 1s;
- communicate w/ surface at 38.4kbaud
- operation to 10,000 psi
- design to 125 C; operational to 110 C

Data are telemetried to the surface at a maximum rate of 38,400 baud over up to 20,000 feet of 7-conductor (15/32") wireline, starting at the end of the stacking period and continuing through array movement and the next acquisition period.

The housings are less than 3 inches in diameter, about 4.5 feet long, and capable of withstanding pressures up to 10,000 psi. Seven conductor cableheads for interconnecting sondes have a minimum breaking strength of 2,500 lbs. The cableheads are rated to 5,000 psi.

Source Sub-System

Surface control system: The source truck controller is equipped with a digital signal generator (DAC) and monitor (ADC) linked to the power amplifier driving the downhole source. The source Compu-log also monitors the wireline and maintains communications with the receiver system via the optical link. This unit provides the source wavelet using an input table:

- amplitudes specified in 20 µs intervals
- 8-bit resolution
- 32K samples (total time 1.6384 s)
- pulses or sweep signal
- variable repeat / delay (incl. quasi-random) within 1.6 sec. table
- variable repeat / delay for stacking of full source pulse
- +/- 10V D/A w/ smoothing filters

A 24KVA linear power amplifier is used to drive the piezoelectric source. The present amplifier has one channel but is being upgraded (summer 1990) to three independent channels for driving up to three source elements with multiphase wavelets. The linearity of the amplifier is an essential feature for it allows the reproduction of sweeps, complex waveforms, and pulse code sequences.

Downhole source. This source design has been used repeatedly over the last four years and has seen hundreds if not thousands of hours of operation. Our source is slightly modified from the original version built for Standard Oil nearly four years ago. In particular, the two active ceramic sections are symmetrically balanced between high power transformers which allow either in-phase or out-of-phase drives. Also, the structure is mass balanced to reduce spurious modes of vibration associated with vertical asymmetry. The far-field wavelet radiated by this source is proportional to the current drawn, thus it can be shaped (open loop) from within a linear system. The source is probably best suited for pulse operation though it has been successfully used in the sweep mode.

The source is approximately 4.5 inches in diameter and nearly twelve feet long. It is packaged for standard oil field operation and resembles standard logging tools.

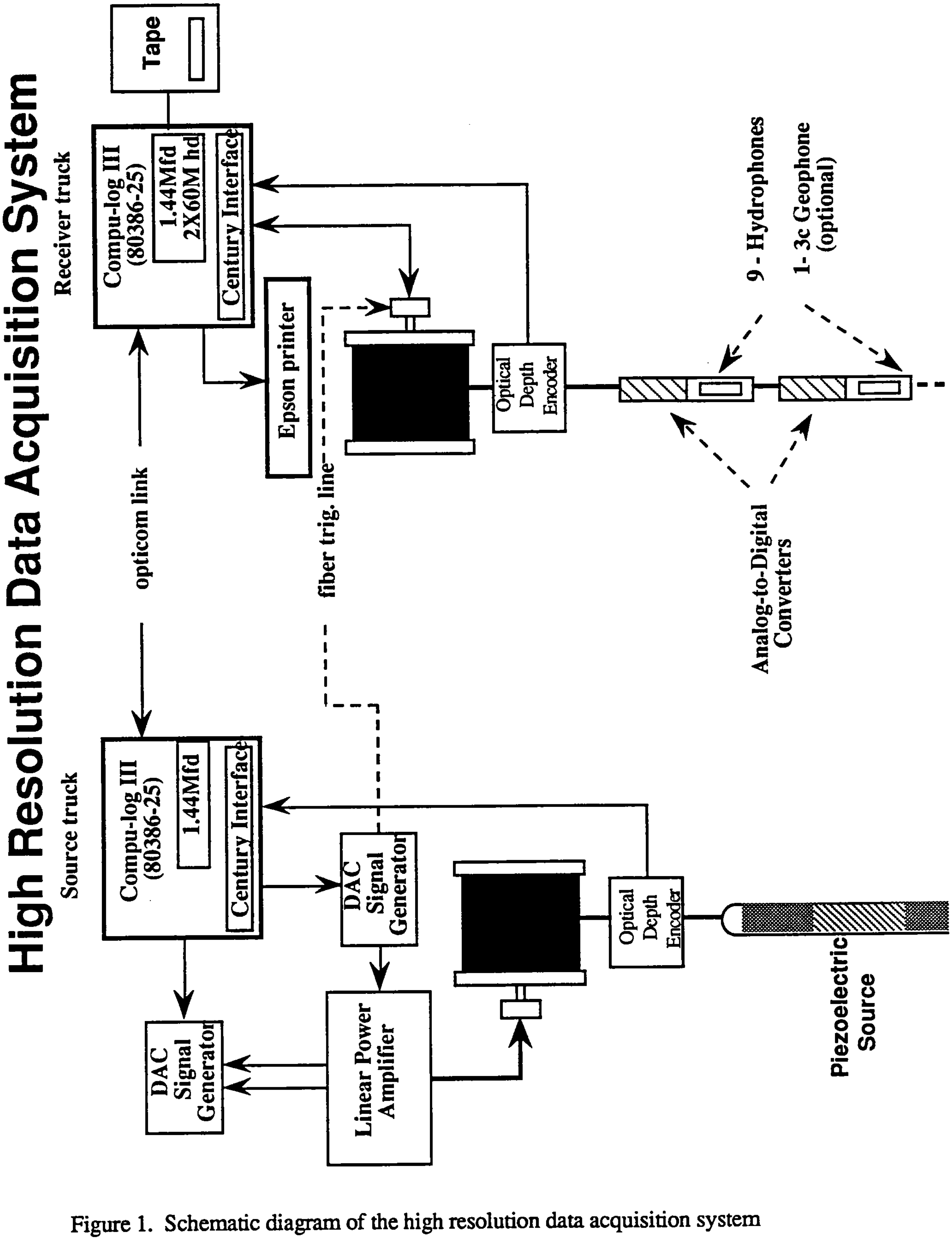


Figure 1. Schematic diagram of the high resolution data acquisition system

PAPER E

TOMOGRAPHIC IMAGING AND MODELING SOFTWARE

Gary Mavko, Mike Fitzpatrick, Caroline Lambert,
Pierre Samec, Jerry M. Harris

Seismic Tomography Project

SUMMARY

During the last few years we have been developing software for displaying, modeling, processing, inverting, and interpreting crosswell tomographic data. Part of this software, which we call the **TIMS** (Tomographic Imaging and Modeling Software) **System**, is directly aimed at providing the tools needed to process raw crosswell seismic data into interpreted tomograms.

An important part of our philosophy has been to keep the TIMS software simple and portable and to rely on features supplied by widely available standard software including FORTRAN, C, LINPACK, and the UNIX shell. Data is stored in a machine-independent and network compatible form called netCDF. We have attempted to keep numerical algorithms relatively separate from the I/O, recognizing that in the future we may wish to use these algorithms in other environments such as Landmark Openworks, which has been heavily developed at Stanford by Pierre Samec. A working prototype TIMS system exists and has been used successfully to process a real field data set. Additional work over the next few months will focus on standardizing a number of stand alone programs that have been developed for specialized plotting, modeling, and inversion experiments, as well as completing documentation.

The TIMS development was largely sponsored by the Gas Research Institute. Additional work, including much of the forward modeling and integration work done by Samec, and the inversion and image enhancement studies by Harris, Lazaratos, Mavko, and Michelena also received support from other projects, including the SRB (Stanford Rock and Borehole Project). Some of the research in these areas is reported elsewhere in this volume and in the SRB volume. In this note we describe primarily the TIMS system.

INTRODUCTION

TIMS was developed to address the growing need both in industry and academia to be able to handle, display, process, and interpret single well and crosswell seismic data. Many of the needs are the same as in surface seismic, and well known techniques for signal processing like spectral analysis, filtering, gaining, and agc are immediately applicable. However, there are enough differences -- including trivial ones like the source and receiver geometry and fundamental ones like the emphasis on direct arrivals instead of reflected arrivals -- that a new system is desirable.

When we started work on TIMS in the Spring of 1989 we already had crosswell data in house, so we had an immediate need for working software. At the same time we recognized the importance of planning the software for future growth, consistency, and flexibility. We have tried to design an approach that allows tools to be added quickly and that still will evolve gracefully as our needs and the available technology change. For example, the algorithms are modularized so that in the future they can be used with different I/O routines and/or with a different database strategy. A prototype system is now working that we have used successfully to process, display, and invert a real data set.

FEATURES OF THE TIMS SYTEM

The System has both **interactive** and **batch** capability (See Figure 1). **Interactive** capabilities are desirable for quick and convenient processing and display of small pieces of data. As with surface seismic processing, there is no single sequence that works well on all data sets. Testing for optimum filtering, gaining, noise rejection, and so on, is almost always necessary, and the testing is most effective when done interactively. For travel time analysis, there is no substitute for interactive picking and pick editing. **Batch** capabilities allow subsequent processing of very large pieces of data, and it is most convenient when the batch programs apply the identical algorithms that are used in the testing. At present batch processing streams are constructed using UNIX shell scripts to string together the various applications modules. The System allows the various processes to extract subsets of the data in a file or to default to reading and processing all of the data in a file.

The System is able to handle **heterogeneous data** sets. Although we expect source and receivers to be nominally equally spaced, they won't be so in practice. The research nature of crosswell seismic surveys naturally results in variable spatial coverage, variable trace lengths, and variable source types. Three component data are also likely to be common. For the seismic traces, sample rate is about the only thing we assume to be fixed for any given data set. Well logs, travel time picks, velocity models, and tomographic images will also be necessary parts of the data set.

A critical aspect of modeling, interpretation, data analysis, and processing is the data formatting and i/o. A typical situation in industry and academia is that networking hardware and software are in place to allow free access to data among various researchers, but differences in data format required by various analysis, display and processing tools end up hampering the exchange. In order to reduce this handicap early on, we (Pierre Samec) undertook a study of data handling strategies.

After considering several alternatives we adopted a public domain data access interface which was developed under the sponsorship of the National Science Foundation (see Figure 2). This purpose of this interface, called the Network Common Data Form (**netCDF**), is to allow one to create, access, and share scientific data in a form that is self-describing and network-transparent. "Self-describing" means that a file includes information defining the data it contains. "Network-transparent" means that a file is represented in a form that can be accessed by computers with different ways of storing integers, characters, and floating-point numbers. Using the netCDF interface in new software for scientific data access, management, analysis, and display can improve the reusability of the software for other data sets and by other users. The netCDF software provides common C and FORTRAN interfaces for applications and data, and it has been tested on both UNIX and VMS operating systems, so it appears to be highly portable. netCDF is an interface to a library of data access subroutines for storing and retrieving data. In simple terms, to a programmer, netCDF looks like a set of working subroutines somewhat analogous to the the FORTRAN statements OPEN, READ, and WRITE, but of course more powerful.

Another issue is portability. Since we have affiliations with a broad group of potential users and contributors to the System, and because we have a wide variety of hardware to work with, we have attempted to maintain wide portability to the extent that we do not seriously impact performance. To accomplish this, the system is based on:

- UNIX
- X-windows
- netCDF
- FORTRAN and C

We have attempted to keep the software modular, so that the basic numerical algorithms are independent of the input, output, and trace handling. Similarly many of the primitives called by the numerical algorithms are public domain standards such as LINPACK. These can always be optimized for speed on a particular machine by vectorizing, parallelizing, or coding in machine language; at the same time reliable FORTRAN equivalents will be immediately available for any new machine. The biggest portability flaw that we struggle with is the variability in the way that FORTRAN programs call C subroutines, and vice versa, and machine-dependent requirements when FORTRAN and C are linked together with I/O routines.

We have also tried to keep the software development simple. To the extent possible we have relied on existing features of UNIX, the netCDF common data form, and the C programming language, rather than to duplicate them with a complex central program. For example, C allows for dynamic memory allocation and flexible input options via command line parsing. UNIX shell scripts allow varied job streams to be constructed from the same stand-alone programs used for testing. The flexible machine-independent storage format of netCDF coupled with UNIX Network File System (NFS) allow networked computer resources to be efficiently shared.

DESCRIPTION OF AVAILABLE SOFTWARE

TIMS will include (1) **software** and (2) **documentation**. The software, which is described below, will include high level applications programs, low level utilities, graphics tools, and example shell programs for creating and running batch jobs. We envision that at some point in the future, we may evolve toward a main driver routine in place of or in addition to the shell programs if i/o performance requires it. The documentation will include (1) user manuals explaining how to install and use the delivered software and (2) programming manuals explaining how to write applications modules that will be compatible with the system.

SOFTWARE

There will be six major types of software associated with the System:

- **Applications modules**
- **Picker**
- **Plotting software**
- **Utility software**
- **netCDF i/o software**
- **shell utilities**

Applications modules. Applications modules perform the generically recognized, high level seismic processing tasks. These are what the user of the system will see and use. There are two broad categories: Trace processing modules that perform signal processing tasks very much like those in surface reflection seismology, and post-trace modules that create and process travel time picks, perform ray tracing, and do inversions. The trace processing applications include, for example, correlation, agc, deconvolution, dip-filter, Radon transform, up and down going wave separation, corridor stack, migration, and trace plotting. The post-trace applications include first break picking, ray tracing, 2-D inversion, and tomogram and log plotting. All of the trace processing applications will be oriented toward processing a "slice" of data: a source gather, receiver gather, a common offset panel or a time slice. Some will operate on only one trace at a time within the slice, so that the order of data going in does not matter. Examples of these modules are bandpass filtering, agc, deconvolution. Other modules will operate on entire slices at a time. Examples of these are fk-filtering, fk-spectral analysis, and multichannel deconvolution. Most commercial surface seismic processing packages are trace oriented and require special care to "collect" traces into slices for multichannel processing. Our approach has tremendously simplified the process of creating new applications.

Picker. The picker is an interactive program that allows the user to display crosswell seismic data on the video screen, identify wave arrival events, and 'pick' the arrival times by use of the mouse. The user can pick the time of each event on each trace, one at a time, or can instruct the program to interpolate or extrapolate the user's picks automatically, under the user's guidance.

Plotting Software includes tools for plotting on various devices including terminals, laser printers, and versatec color plotters. A variety of data types can be displayed including seismic traces, rays, travel time picks, well logs, and inverted velocity tomograms.

Utilities. Utility software includes a wide range of low level routines for handling and manipulating data within applications and plotting modules. We sometimes refer to these also as primitives. The user of the System will generally not be aware of utilities, but a programmer will use them routinely. The utility routines are intended to be useful for commonly repeated tasks and will generally not be unique to any one application module. Examples of utility software include:

- **i/o routines**. These are subroutines to read and write the commonly used subsets of data, like source gathers, receiver gathers, and time slices, as well as more general data like scalars and 1-D, 2-D, and 3-D arrays.
- **vector routines**. Seismic processing is highly vector-oriented. We are building a set of vector routines using for example LINPACK. These are written in FORTRAN, and when compiled can take advantage of vector and parallel operations, if available.
- **signal processing primitives**. These include commonly used functions like convolution, fft, filter design, correlation.
- **plotting primitives**. These are the commonly used inner routines of the plotting software listed above.

netCDF software is the basis of our i/o. netCDF is a public domain item. Although we will not 'deliver' it, we will supply information on obtaining it, on using it, and on interfacing it with our System.

shell utilities. These will include example scripts for creating batch jobs as well as tools for creating and compiling new applications modules.

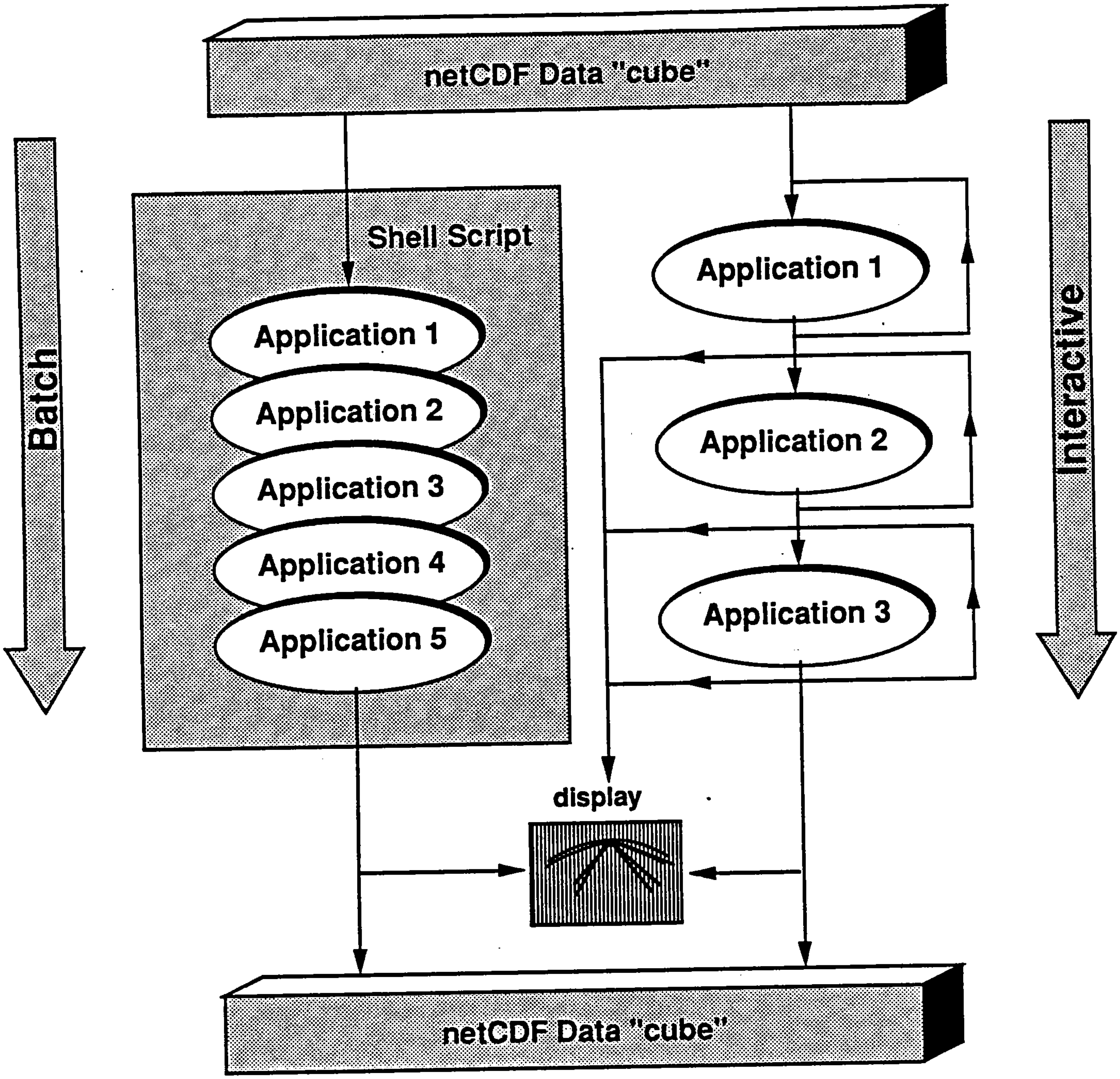


Figure 1. The TIMS system consists of a set of applications modules which read data from the netCDF file and deposit results back in the netCDF format. Applications can be called individually in an interactive mode, or combined using a UNIX shell script to create a batch processing stream.

netCDF data interface

- network transparent / machine independent
- self describing / self documenting
- heterogeneous data sets
- public domain
- implemented in C and Fortran, VMS and Unix

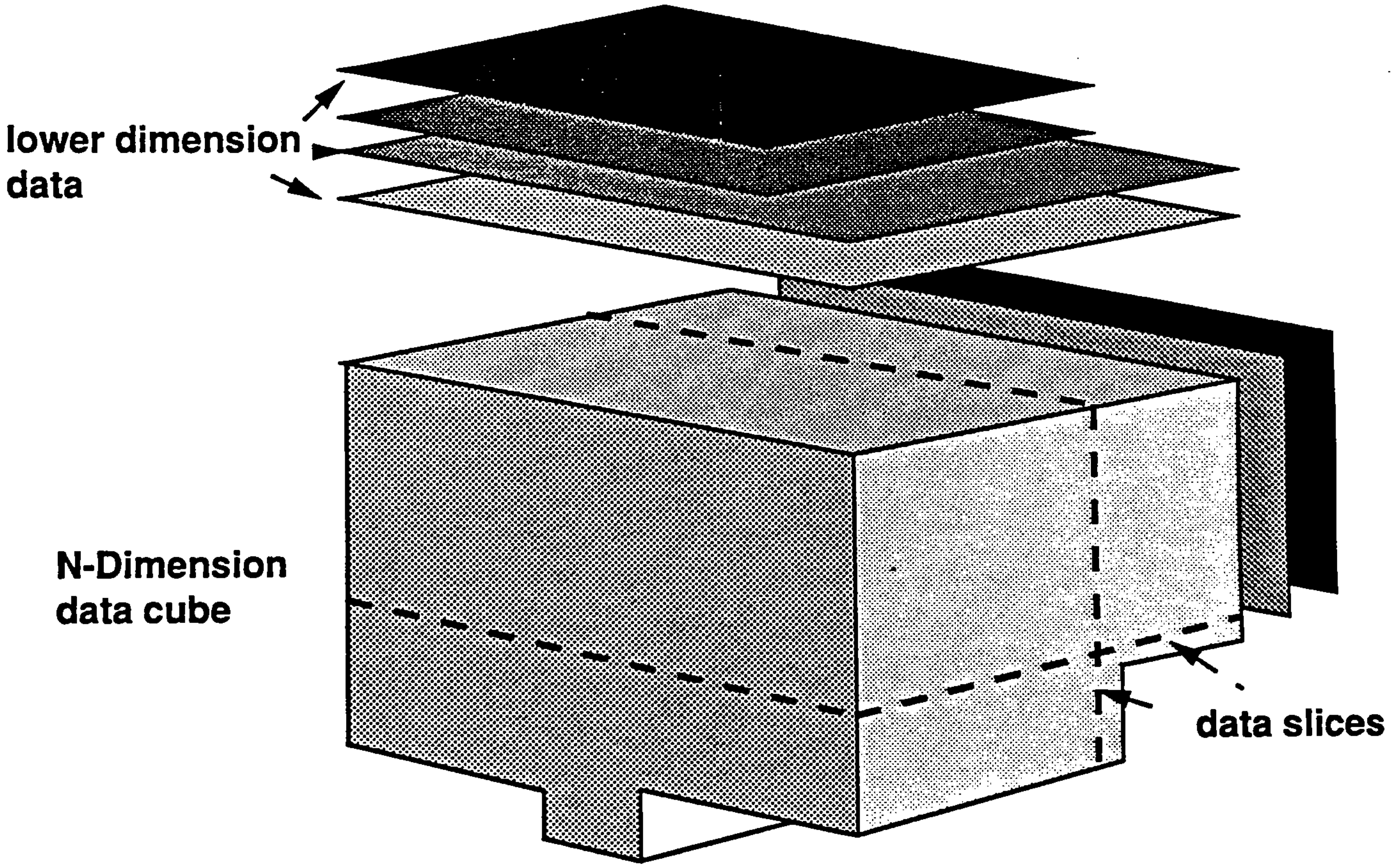


Figure 2. The netCDF common data form allows flexible, self-documenting, machine independent storage of heterogeneous data.

PAPER F

TOMOGRAPHIC TRAVELTIME INVERSION USING NATURAL PIXELS

Reinaldo J. Michelena and Jerry M. Harris
Seismic Tomography Project

ABSTRACT

Traditionally in the problem of tomographic traveltime inversion, the model is divided into a number of rectangular cells of constant slowness. Inversion consists of finding these constant values using the measured traveltimes. The inversion process can demand a large computational effort if a high resolution result is desired.

We show in this paper how to use a different kind of parametrization of the model based on beam propagation paths. This parametrization is obtained within the framework of reconstruction in Hilbert spaces by minimizing the error between the true model and the estimated model. The traveltimes are interpreted as the projections of the slowness along the beam paths. Although the actual beam paths are described by complicated spatial functions, we simplify the computations by approximating these functions with functions of constant width and height, i.e., "fat" rays, which collectively form a basis set of natural pixels.

With a simple numerical example we demonstrate that the main advantage of this parametrization, compared with the traditional decomposition of the model in rectangular pixels, is that 2D reconstructed images of similar quality can be obtained with considerably less computational effort. This result suggests that the natural pixels can provide considerable computational advantage for 3D problems.

INTRODUCTION

The process of reconstructing an image using line integrals through it is called tomography. In traveltime tomography the image to be reconstructed is the slowness model $S(\mathbf{r})$. The reconstructed model $\tilde{S}(\mathbf{r})$ is usually represented as a linear combination of functions $\beta_n(\mathbf{r})$ in the form

$$\tilde{S}(\mathbf{r}) = \sum_{n=1}^{M} a_n \beta_n(\mathbf{r}). \tag{1}$$

The problem consists of determining the unknown coefficients a_n from the measured traveltimes. Once these coefficients have been calculated, the computation of the sum (1) is straightforward.

The representation (1) has two important degrees of freedom that influence decisively the kind of results obtained. These are the number M and kind of functions

$\beta_n(\mathbf{r})$ to be used. The most common choice for the functions $\beta_n(\mathbf{r})$ is orthogonal cells (square or cubic pixels) and in that case the coefficients a_n represent the slowness within each cell (McMechan, 1983; Ivansson, 1985). Although this is the most popular basis function used for estimating the slowness model, others have been suggested recently. Harlan (1989) defines the velocity function as a sum of smooth basis functions (Gaussians), and Van Trier (1988) defines the functions $\beta_n(\mathbf{r})$ as cubic B-splines multiplied by functions that reproduce the expected structure of the model. The number of functions M is also arbitrary but is usually wanted to be "small" to avoid having to solve a huge system of equations.

The kind and number of functions used for expanding the slowness model determine many of the general features of the final image. With the same data set it is possible to obtain different results just because different parametrizations have been used. However, the goal is to obtain a reconstructed model free from these artifacts derived from the parametrization. This means that the selection of the basis function is a critical step in the inversion process and then should be considered more carefully, as described below.

There is no general criteria for deciding which representation is the best, although some may have clear advantages for solving specific problems. Our selection of the basis function will be based on the minimization of the expression that estimates the norm of the null space of the problem

$$||S(\mathbf{r}) - \sum_{n=1}^{M} a_n\beta_n(\mathbf{r})|| \tag{2}$$

where $S(\mathbf{r})$ is the true slowness model. Due to the nature of the measurements in traveltime tomography (integral along beam paths) we show in this paper that the minimum of (2) can be reached when the functions $\beta_n(\mathbf{r})$ describe the beam paths and when M equals the number of measurements available (because there is only one measurement per beam path). In the first part of the paper, this fact is demonstrated within the framework of reconstruction in Hilbert spaces. The remainder of the paper presents a comparison of the inversion of synthetic data using the traditional representation of the model in square pixels and the proposed representation in constant regions along the beam paths called natural pixels.

RECONSTRUCTION IN HILBERT SPACES

A Hilbert space is a linear space on which an inner product is defined. For example, the inner product for the Hilbert space L^2 of the Lebesque square-integral functions of support Ω is

$$< f(x), \beta(x) > = \int_\Omega f(x)\beta(x)dx. \tag{3}$$

We can assume that the particular function $f(x)$ that we want to estimate belongs to a Hilbert space H. Let's assume also that the information we have about $f(x)$, i.e., data, is a set of inner products of the function $f(x)$ with a finite set of known functions $\beta_m(x) \in H$

$$d_m = < f(x), \beta_m(x) > \qquad m = 1, ..., N. \tag{4}$$

In this context, the data can be interpreted as the projections of the unknown function $f(x)$ onto the "sampling" functions $\beta_m(x)$.

If F_1 is a closed linear subspace of the Hilbert space H, then $H = F_1 \oplus F_1^{\perp}$ (Berberian, 1976), where $F_1^{\perp}$ is called the orthogonal complement of F_1. From the projection theorem (Stakgold, 1979), we can always decompose $f(x)$ into $f_1(x)+f_2(x)$ (Fig. 1) where $f_1(x) \in F_1$ and $f_2(x) \in F_1^{\perp}$. $f_1(x)$ is called the orthogonal projection of $f(x)$ in F_1.

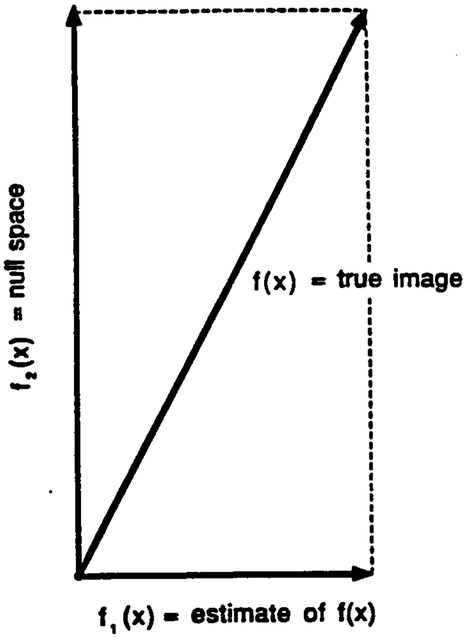


Figure 1: Orthogonal projections of the true image.

If we assume that the functions $\beta_n(x)$ form a basis of the space F_1, we can write

$$f(x) = \sum_{n=1}^{N} a_n \beta_n(x) + f_2(x). \tag{5}$$

We can understand the meaning of the function $f_2(x)$ by multiplying both sides of (5) by $\beta_m(x)$ and integrating in Ω

$$d_m = \sum_{n=1}^{N} a_n < \beta_n(x), \beta_m(x) > + < f_2(x), \beta_m(x) > . \tag{6}$$

Since $< f_2(x), \beta_m(x) >= 0$, we can say that $f_2(x)$ contains the information about $f(x)$ that does not affect the measurements made by the sampling functions $\beta_m(x)$.

Finally, the estimate $\tilde{f}(x)$ of $f(x)$ can then be written as

$$\tilde{f}(x) \; = \; f_1(x) \; = \; \sum_{n=1}^{N} a_n \beta_n(x), \tag{7}$$

where the coefficients a_n are calculated from the forward equation for the data d_m

$$d_m \; = \; \sum_{n=1}^{N} a_n < \beta_n(x), \beta_m(x) > \qquad m = 1, ..., N. \tag{8}$$

Note that the system of equations (8) is square. If all the sampling functions are independent, the problem is well conditioned and therefore, the system (8) has a unique solution. A unique solution is expected because the orthogonal projection of $f(x)$ in the space F_1 expanded by the functions $\beta_m(x)$ is unique (Fig. 1). If any of the sampling functions can be expressed approximately as a linear combination of the others, the problem becomes ill-conditioned since the rows (or columns) in the matrix of Eqn. (8) are not independent any more.

Minimum Norm Solution

This same result for $\tilde{f}(x)$ can be obtained through minimization of the norm $\|f_2(x)\|$ with respect to the unknown coefficients a_n (Darling et al., 1983),

$$min \; \|f_2(x)\|^2 \; = \; min \; \|f(x) - \sum_{n=1}^{N} a_n \beta_n(x)\|^2. \tag{9}$$

For this reason, the estimate $\tilde{f}(x)$ is called the minimum norm estimate of the unknown function $f(x)$. This estimate $\tilde{f}(x)$ is unique and consistent with the data (Eqn. (8)). It is also strongly related to the way the modeled data are generated, because the unknown function $f(x)$ is expressed as a linear combination of the sampling functions $\beta_m(x)$ used to compute the forward modeling (Eqn. 4). This means that each experiment will suggest "naturally" the reconstruction procedure which produces the minimum norm solution.

Examples of different sampling functions in different problems are the beam paths in the problem of traveltime tomography, complex exponentials when the measurements are the frequency components of $f(x)$, etc.. When the measurements are the frequency components of $f(x)$ the minimum norm estimate is simply the Fourier series expansion of $f(x)$ (Stakgold, 1979). In fact, Eqn. (7) can be interpreted as a generalized Fourier series for basis sets not necessarily orthogonal.

If we expand $\tilde{f}(x)$ in any other set of basis functions $\{\alpha_n(x,y), \; n = 1, ..., M\}$, the norm of $f_2(x)$ is

$$\|f_2(x)\| \; = \; \|f(x) - \sum_{n=1}^{M} c_n \alpha_n(x)\|. \tag{10}$$

Minimizing this expression with respect to the unknown coefficients c_n, we obtain

$$< f(x), \alpha_m(x) > = \sum_{n=1}^{M} c_n < \alpha_n(x), \alpha_m(x) > \qquad m = 1, ..., M. \tag{11}$$

Note that the independent term on the left-hand side in the system of equations (11) is formed by the inner products of the selected basis function with the unknown function $f(x)$. Note also that the matrix elements are the inner products among the different elements of the selected basis. The independent term is equal to the measurements only if the basis set used for expanding $\tilde{f}(x)$ ($\{\alpha_n(x,y),\ n = 1, ..., M\}$) is the same one used for generating the data, i.e., $\alpha_n = \beta_n$. If any other basis set is used, $\alpha_n \neq \beta_n$, then the independent term must be computed from the measurements. Therefore, the choice $\alpha_n = \beta_n$ is a convenient one among many others basis sets because the solution obtained is still minimum norm and the independent term in Eqn. (11) represents directly the measurements. We will see later that in the problem of traveltime inversion the functions $\beta_n(x,y)$ are the beam paths and then the minimum norm solution can be obtained easily from Eqn. (7).

Let's illustrate these ideas with a simple example. Although this example is not a geophysical one, it can help to understand why square orthogonal pixels are a convenient basis set when the measurements represent the average of the unknown over the same square regions. When the measurements have a different meaning, it might be computationally easier to get the minimum norm solution by using a different basis set. Consider a two dimensional object $O(x,y)$ (a photograph, for example). The process of spatial discretization of the object can be interpreted as the convolution of $O(x,y)$ and the sampling function $R(x,y)$ which describes the shape of the pointer of the digitizer (for simplicity the function $R(x,y)$ is assumed to have unit area)

$$d(x,y) = O(x,y) * R(x,y), \tag{12}$$

where $d(x,y)$ is the digitized image. This is equivalent to superimposing a square grid, for example, over the object and calculating the function $d(x,y)$ from the volume of the object in the support of each pixel surrounding the grid point. This procedure can be expressed as

$$d_i = \int_{\Omega_i} O(x,y) R_i(x,y)\, dx\, dy \tag{13}$$

where

$$R_i(x,y) = \begin{cases} 1 & \text{if } (x,y) \text{ is in the pixel } i \qquad (i = 1, ..., N) \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

d_i, $i = 1, ..., N$ are the data points and Ω_i is the support of the i-pixel.

Given the inner product (13), the minimum norm estimate of $O(x,y)$ according to (7), is

$$\tilde{O}(x,y) = \sum_{n=1}^{N} a_n R_n(x,y). \tag{15}$$

The coefficients a_n can be found from the equation

$$d_m = \sum_{n=1}^{N} a_n < R_n(x,y), R_m(x,y) > \qquad m = 1, ..., N. \tag{16}$$

According to the definition of R_i

$$< R_n(x,y), R_m(x,y) > = \delta_{nm}. \tag{17}$$

Then, $a_n = d_n$ and the estimate becomes

$$\tilde{O}(x,y) = \sum_{n=1}^{N} d_n R_n(x,y). \tag{18}$$

As expected, the reconstructed object is formed with a superposition of N cells, each with constant height d_n and located where the measurements were taken. If we use any other basis set instead of $R_i(x,y)$, it is possible to get an estimate that does not reproduce the data, or requires of more model parameters to get a better representation.

As we said before, Fourier analysis is another example where the sampling functions $\exp(ik_n x)$ are the same ones used to expand the estimate of the unknown. The result is also a minimum norm estimate.

The theory of reconstruction in Hilbert spaces generates consistent estimates of the unknown in the sense that the same basis set used to sample the function is used to expand it. The two previous examples (orthogonal sampling and Fourier reconstruction) confirm that in some situations this might be a convenient choice among many other possibilities. In the following sections we will exploit this idea of consistency in the problem of traveltime tomography where the data are generated in a very specific way (integrals along beam paths).

TOMOGRAPHIC TRAVELTIME INVERSION

The traveltime along a ray l_m in a medium where the slowness is $S(x,y)$, is traditionally given as

$$t_m = \int_{l_m} S(x,y) dl_m \qquad m = 1, ..., N, \tag{19}$$

where dl_m is the incremental distance along the ray path l_m. In general, the ray path depends on the slowness distribution. For sake of simplicity, let's assume that the variations in the slowness are just a few precent. Then we can safely consider that the ray paths are straight lines and independent of the slowness. The general case will be discussed later.

Although the expression (19) simplifies the mathematics considerably, it fails to convey the fact that the traveltimes between two points are affected by velocities in the region called the Fresnel zone, which is infinitely narrow only when the wavelength

λ is infinitely small, $\lambda \rightarrow 0$ (Nolet, 1987). To account for the finiteness of this effect, we can say that the traveltime between two points can be better described by the equation

$$t_m = \int_\Omega S(x,y)\phi_m(x,y)dxdy, \tag{20}$$

where $\phi_m(x,y)$ is a two dimensional function or "beam" of finite support centered along the ray path and Ω is the support of $S(x,y)$. The functions $\phi_m(x,y)$ can be interpreted as the wavepaths introduced by Woodward (1989).

With the forward modeling equation written in this way, the estimation of the slowness from the traveltimes can be seen as a reconstruction problem in a Hilbert space where the inner product is defined by (20). According to (7), the minimum norm estimate of the slowness $S(x,y)$ is

$$\tilde{S}(x,y) = \sum_{n=1}^{N} a_n\phi_n(x,y), \tag{21}$$

where N is the number of traveltimes.

We can transform (19) into a two dimensional integral of the form of (20), if we describe the ray path with a two dimensional delta function $\delta_m(x,y)$. However, the problem of reconstruction of the slowness from such an expression cannot be seen as a reconstruction problem in a Hilbert space because the inner product $< \delta_i(x,y), \delta_i(x,y) >$ is not defined.

From Eqn. (8), we find that the coefficients a_n can be calculated through the system of equations

$$t_m = \sum_{n=1}^{N} a_n < \phi_n(x,y), \phi_m(x,y) > \qquad m = 1,...,N \tag{22}$$

where

$$< \phi_n(x,y), \phi_m(x,y) > = \int_\Omega \phi_n(x,y)\phi_m(x,y)\, dxdy. \tag{23}$$

In contrast with the traditional reconstruction using square pixels as basis function (Eqn. (18)), the reconstruction described above is based on a discretization of the model along the beam paths. In the example discussed in the previous section we showed that the square orthogonal pixels are a convenient discretization that lead to minimum norm estimators when the data are "point" orthogonal samples of the two dimensional function that we want to reconstruct. The discretization along the beam paths comes from the fact that they are the regions sampled with each measurement in traveltime tomography.

When the beam paths are used, the discretization of the model will depend in general on the particular data set that is going to be inverted, because it will reflect the propagation of the energy in the medium. It does not have the advantage of other parametrizations that can reflect some prior knowledge we may have about the model. In that sense, some flexibility is lost. The discretization along the beam paths is similar to the discretization of the model in square pixels in the way the prior information is handled, since both assume no prior information about the model. The difference is that when the problem is nonlinear and it is solved as a sequence of linearized steps, the discretization along the beam paths adapts progressively to the real model.

It is important to note that although the solution of the system of equations (22) is unique, the null space of the problem has not been suppressed. It has been just separated at the beginning of the formulation in the form of an space orthogonal to the beam paths (function $f_2(x)$, Eqn. (5)). The same situation occurs in discrete Fourier reconstruction, where for a given data set the estimate of the unknown is unique even though the null space (Fourier components above the Nyquist frequency) is not zero.

Natural Pixels

As a first approximation, we can describe the basis function $\phi_i(x, y)$ as functions of width λ' and heigth $1/\lambda'$

$$\phi_i(x,y) = \begin{cases} 1/\lambda' & \text{if } (x,y) \text{ is in the region of width} \\ & \lambda' \text{ centered along the ray path } i \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

Therefore, the matrix coefficients $< \phi_n(x,y), \phi_m(x,y) >$ are

$$< \phi_n(x,y), \phi_m(x,y) > \lambda'^2 = \begin{cases} \textit{area of the beam path} & \text{if } m = n \\ \textit{area of the intersection} & \text{if } m \neq n. \end{cases} \tag{25}$$

A natural pixel for a single ray is shown in Fig. 2. Even when the rays curve or when reflections are included, the natural pixels are "tubes" centered on the ray path. The shape of the tubes may vary. For example, they can have variable cross section that gives more weight to the center than to the sides.

Buonocore, et al., (1981) and Buonocore (1981), without working within the framework of reconstruction in Hilbert spaces, define an estimator identical to (21) and call it "natural pixel" decomposition of the two dimensional image, where the natural pixels are the functions $\phi_m(x, y)$. They study extensively the properties of such a reconstruction and the theoretical advantages of it compared with the traditional reconstruction using square pixels. They show that the matrix of coefficients $< \phi_i(x,y), \phi_j(x,y) >$ represent the measurement covariance matrix if there is no measurement noise. If the measurement noise is nonzero but uncorrelated, only the diagonal elements of the matrix are different from those of the measurement covariance matrix. According to Buonocore et al., (1981), square pixels errors are caused by the inaccurate estimation of the measurement covariance. These errors can be eliminated only if the size of the pixels could be made infinitely small.

An example of a set of natural pixels is shown in Fig. 3, for the case of a cross borehole geometry in a medium of constant slowness.

The number of cells in the square-pixel-based inversion is commonly determined by a trade off between the required resolution and cost of the inversion. With square

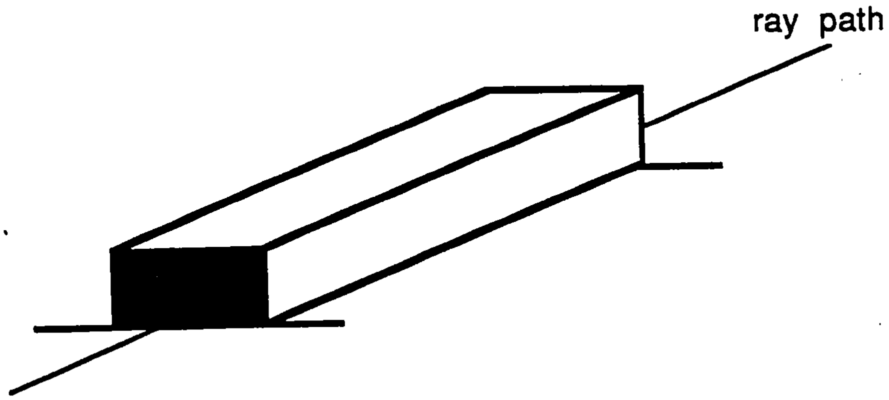


Figure 2: Natural pixel for a single ray.

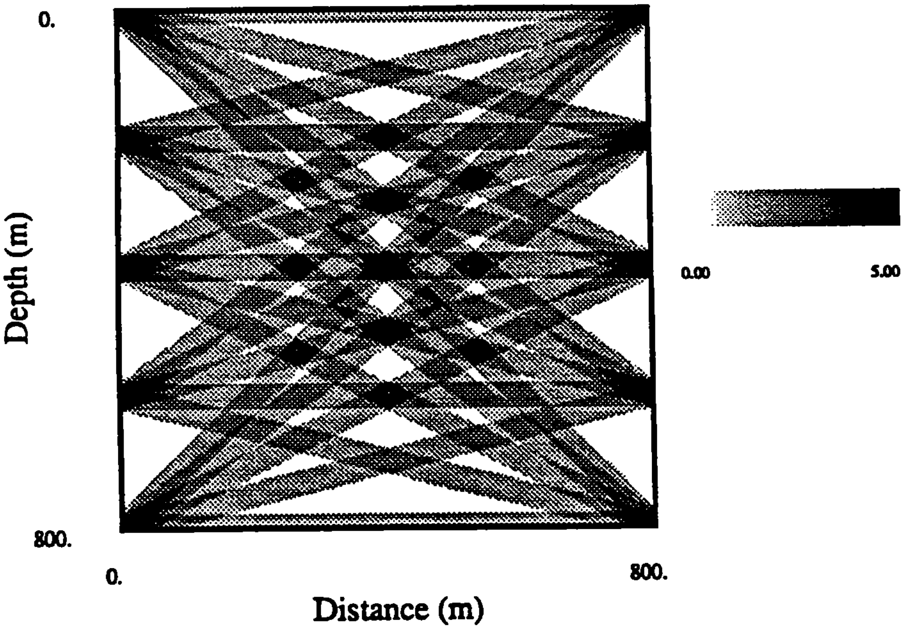


Figure 3: Natural pixels in a constant slowness medium for a cross-well configuration of five sources and five receivers.

pixels we are forced to establish that compromise because they do not differentiate between model parameters and display parameters, although both have opposite purposes: we want many display parameters for an accurate representation but, at the same time, we want few model parameters for an inexpensive inversion. This trade off is particular of the square pixels. An example of this will be explained below. If we decide to use any other basis function instead of square pixels, we always have to discretize fine enough to ensure the *numerical* accuracy in the summation (21), but this fact does not change the number of model parameters.

The coefficients a_n in Eqn. (21) represent all the information gained from the measurements, which can be seen easily when only one measurement is available. In this case, the coefficient a_1 is proportional to the average slowness in the region of the beam path. If we want to invert that measurement using square pixels, the problem in general will be underdetermined or in the best case (using only one pixel in the inversion) we can obtain information not present in the data, e.g., information outside the spatial support of the observation.

Each traveltime measurement *does not* contain information about variations in the slowness along the beam path. The variations are averaged into a single number. It is the combination of them (if they overlap) that gives the information about these variations along the path. This basic fact is contradicted when the model is discretized into square pixels that potentially introduce variations along each beam path and demand from the data more information than it contains. Introducing in the problem more degrees of freedom might be convenient if there is enough information to resolve all of them. The extra information needed to solve the problem is introduced in the form of constraints, some of them result of independent data and others are simply "reasonable" ones. Smoothness is an example of a reasonable constraint that can help to solve the problem of the extra information needed. However, it is not clear how the solution may depend on the various ways of introducing the smoothness (Claerbout, 1976) or other reasonable constraints that don't come from independent observations.

To this point, the inversion is strictly linear, this is, the sampling functions do no depend on the slowness. This is analogous to Fourier reconstruction where the sampling functions (complex exponentials) do not depend on the properties of the unknown. No iterations are needed after the estimate is found. This is not the situation in traveltime tomography where the sampling functions may strongly depend on the unknown slowness. The next section addresses this topic.

Iterative inversion

As we said before, the traveltime along a ray in a medium of slowness $S(x,y)$ is

$$t = \int_l S(x,y)dl, \tag{26}$$

where the traveltimes as well as the ray paths depend on the slowness. When the ray paths are straight lines, like in X-ray tomography or when the variations in slowness are small, we can derive $S(x,y)$ from the traveltimes using this expression.

In geophysical applications, however, straight rays are rarely found and as a result, the problem (26) becomes highly nonlinear since the unknown $S(x, y)$ is also implicitly present in the ray path (Nolet, 1987).

If the medium is perturbed to $S'(x, y) = S(x, y) + \Delta S(x, y)$, the new traveltime calculated along the new ray path l' is

$$t' = \int_{l'} S'(x, y) dl', \qquad (27)$$

Using the Fermat principle, it can be shown (Aki and Richards, 1980) that the difference in traveltime between the two media is

$$\Delta t = \int_{l} \Delta S(x, y) dl \qquad (28)$$

where $\Delta t = t' - t$.

The nonlinear problem is then solved as a sequence of linearized steps that seeks to minimize the difference between real and calculated traveltimes.

If the perturbations in traveltimes are calculated as integrals along the beam paths, (28) becomes

$$\Delta t_m = \int_{\Omega} \Delta S(x, y) \phi_m(x, y) dx dy \qquad (29)$$

The beam paths are centered in the rays traced in the unperturbed model. Substituting $\Delta S(x, y)$ by $\tilde{S}(x, y)$ in Eqn. (21) and Δt_m by t_m in Eqn. (22), we can get the estimate of the slowness perturbation $\Delta S(x, y)$ after solving (22). This estimate reproduces the perturbations in traveltimes and because of this, when the problem is linear, it converges in one iteration. This is the situation that is going to be studied in the examples that follow.

NUMERICAL EXAMPLES

We will now show synthetic inversion examples comparing natural pixels and square pixels as basis functions. Our aim is to compare the results of the inversion when both are used with the same data set. This goal can be achieved with synthetic data for a cross borehole geometry generated from the model shown in Fig. 4. The example is simplified considerably by assuming that the contrast between the circular disc ($S = 2.02$) and the background ($S = 2.00$) is 1%. Therefore, straight rays adequately describe the propagation of the energy in the medium.

The data are generated from strip integrals across the model of Fig. 4. The integrals are calculated from the analytical expressions of the intersection of the strips with the circle. In this way the numerical errors in the forward modeled data have been minimized. In Appendix A it is explained how to calculate numerically strips integrals across any two dimensional slowness model. For these examples 289 traveltimes were computed, which corresponds to the 17 sources and 17 receivers used. Another simplification is made assuming that the width of the strips $\lambda' = 40\ m$ is the same during both the forward modeling and the inversion.

When the model is discretized into square pixels the estimate of $\Delta S(x,y)$ (Eqn. 29) is obtained after solving a system of linear equations where the matrix coefficients represent the area of the intersection of the strip with each pixel. We are going to solve this system and the one obtained with the natural pixels (Eqn. 22) using the LSQR variant of the conjugate gradient method (Nolet, 1987) that has been proved to be faster than SIRT methods (Nolet, 1985; Van der Sluis and Van der Vorst, 1987).

Fig. 5 and 6 show the results of the inversion when the model is discretized with two different pixel sizes. The starting model has a constant slowness $S_0(x,y) = 2$. The inversion produces directly the slowness value in each pixel, and therefore, reducing the size of the pixels (for better resolution) increases the number of model parameters and consequently the size of the system of equations to solve. In the examples shown, the size of the system of equations solved is 289 X 1681 (grid size = 41 X 41; Fig. 5) and 289 X 25921 (grid size = 161 X 161; Fig. 6) respectively. Evidently, the quality of the reconstruction and the amount of information about the model contained in the image increases with the number of model parameters. Note that no interpolating or smoothing process has been applied to the images. The coarse discretization and the limited view of the data are the causes of the artifacts in Fig. 5.

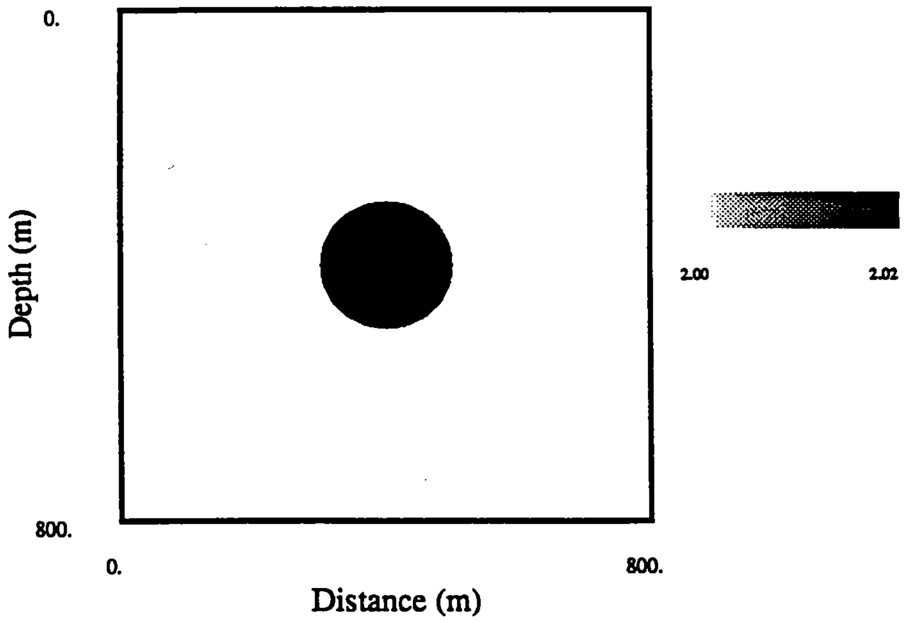


Figure 4: Slowness perturbation. 17 sources are located on the right hand side of the model and 17 receivers are located on the opposite side. The radius of the disc is $r = 100\ m$. The width of the natural pixels is $\lambda' = 40\ m$. The vertical separation between adjacent sources and/or receivers is 50 m.

The result of the inversion using natural pixels is shown in Fig. 7. This image is represented with a grid identical to the one used in Fig. 6 (161 X 161) and then, both results can be compared directly. The system of equations solved with the natural

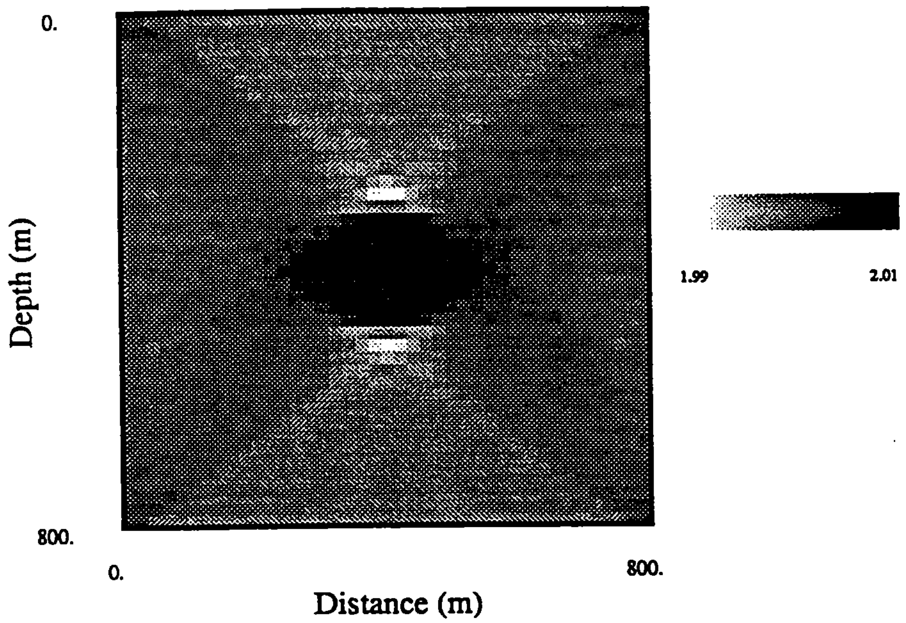


Figure 5: Inversion when a grid of 41 X 41 square pixels is used.

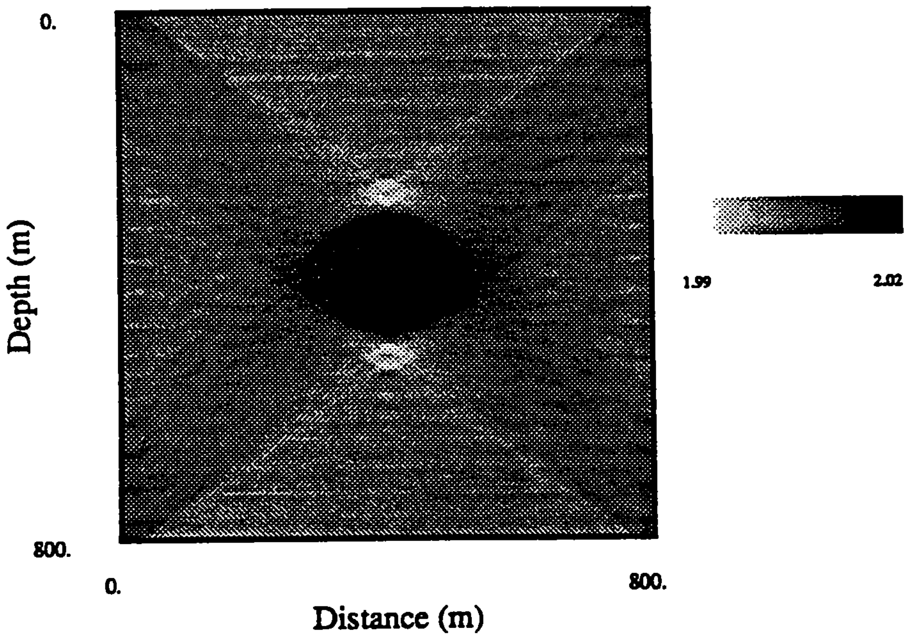


Figure 6: Inversion when a grid of 161 X 161 square pixels is used.

pixels is 289 X 289 and these dimensions are independent of the level of resolution of the image.

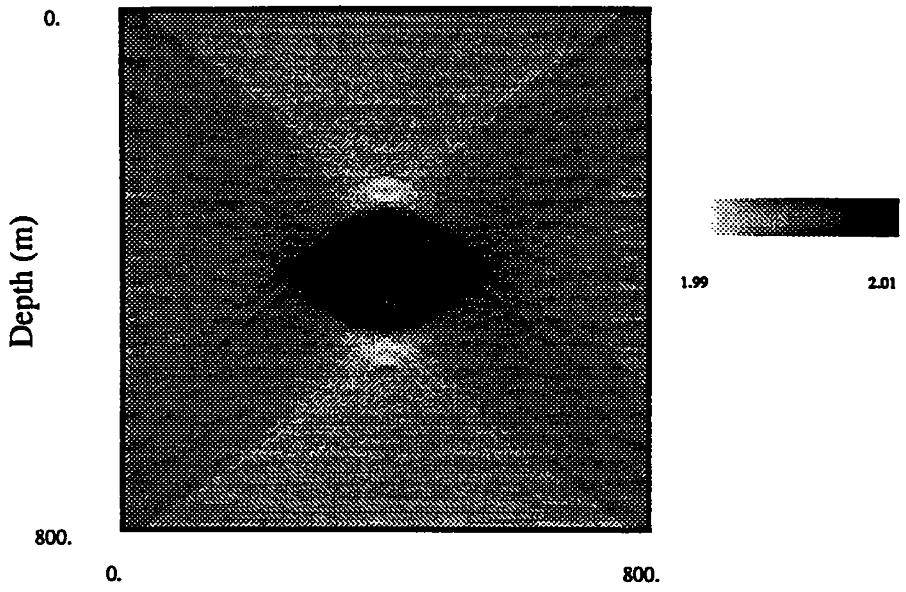


Figure 7: Inversion when the model is discretized in natural pixels. The image is displayed in a grid of 161 X 161 cells.

Both images look almost identical in terms of resolution. The artifacts produced by a coarse sampling of the model (Fig. 5) has been reduced. We observe that the discretization along the natural pixels does not contribute to eliminate limited view problems in the inversion, since we can see them in both cases. The main difference between the two solutions is related with the smoothness of the image. The reconstruction with the square pixels produces an slightly smoother image than the reconstruction with the natural pixels.

For comparing the results of the inversion with square and natural pixels, we plotted the absolute value of the difference between the original image (Fig. 4) and the inverted ones (Fig. 6 and 7). The results are shown in Fig. 8 and 9. The maximum error obtained with natural pixels is the same as obtained with square pixels. We expect both images to be roughly the same since both discretizations, minimize the expression (2). This expression represents the norm of the null space of the problem $f_2(x,y)$. If we discretize the original image very densely (grid size $=$ 889 X 889), we can calculate this norm in both reconstructions. For the square pixels the result is $||f_2|| = 2.383$ and for the natural pixels $||f_2|| = 2.397$. The norm of the null space for the inversion with less square pixels (Fig. 5) is $||f_2|| = 2.471$. This means that although the dimensions of the null space increases by sampling the image more densely in the square pixels based inversion, its norm decreses.

The noisy appearance of the inversion with natural pixel reconstruction can be

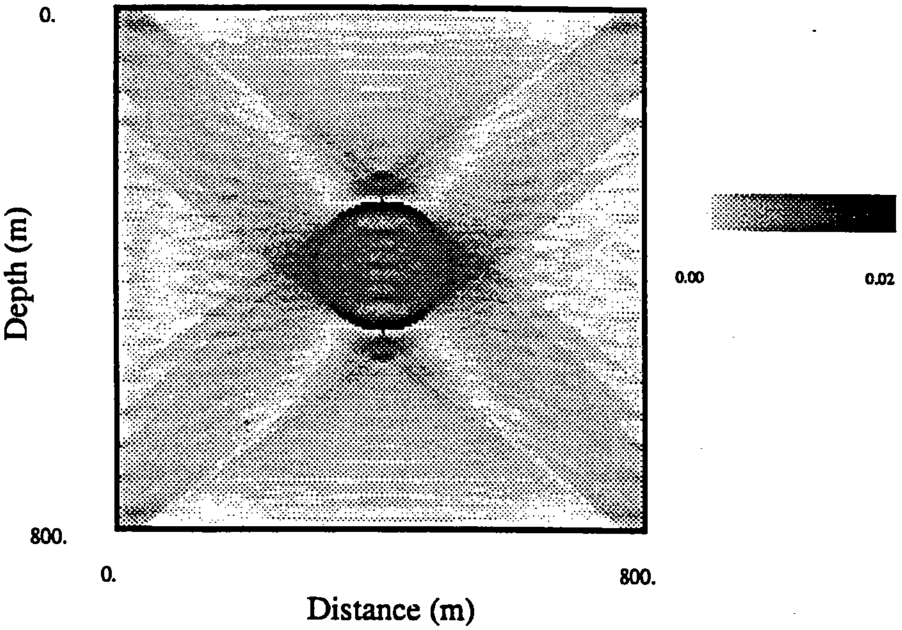


Figure 8: Error in the inversion with square pixels. This image represents the absolute value of the difference between the original model (Fig. 4) and the inverted one (Fig. 6).

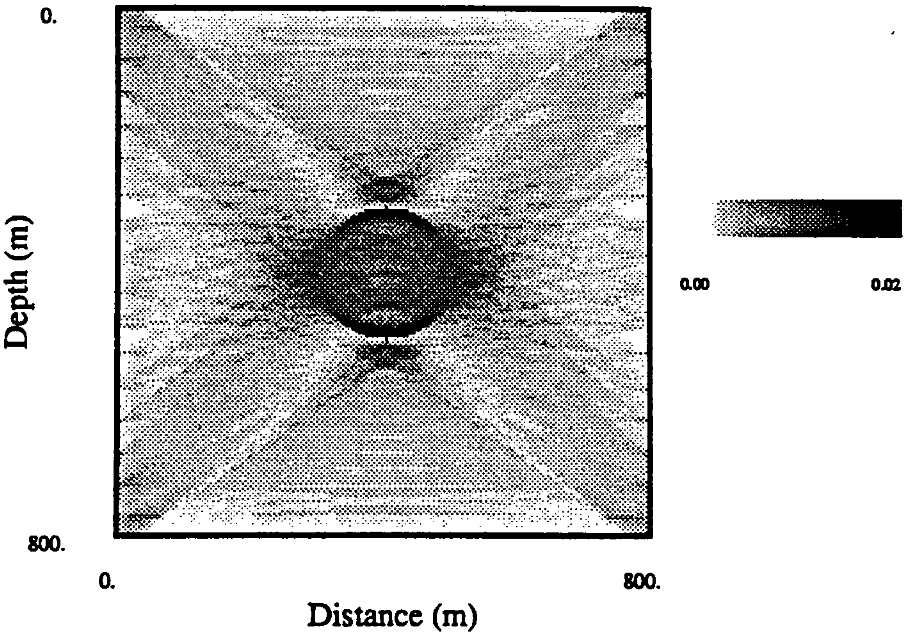


Figure 9: Error in the inversion using natural pixels. This image represents the absolute value of the difference between the original model (Fig. 4) and the inverted one (Fig. 7).

reduced by sampling the beam paths more densely. The result is shown in Fig. 10, where a grid of 401 X 401 points has been used. The norm of the null space of this image is $\|f_2\| = 2.394$. We have to say here that the extra computations necessary for producing the image from the digitized beam paths (Eqn. (21)), is negligible compared with the computation of the matrix elements and the solution of the system of equations for the coefficients of the square pixels. Once the coefficients a_n for the natural pixels are obtained, the image can be displayed using different grid sizes without having to build a new matrix and solve the system of equations again.

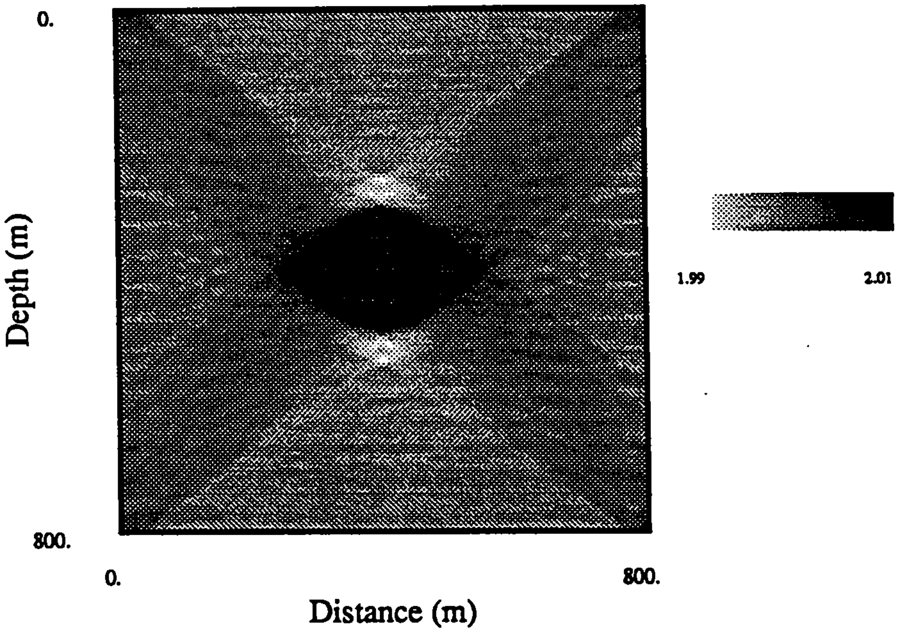


Figure 10: Inversion when the model is discretized in natural pixels. The image is displayed in a grid of 401 X 401 cells.

We can compare also the mean absolute error in both images from the following expression

$$error = \frac{1}{N}\sum_{i=1}^{N} |\ (original\ pixel)_i - (reconstructed\ pixel)_i\ |, \qquad (30)$$

where N in this case represents the total number of cells.

For the natural pixels as well as for the square pixels, the mean error is 2.0 10^{-3}. However, remember that although the quality of the inversion is basically the same for both basis functions, the computational effort necessary in the whole process is roughly two orders of magnitude smaller when natural pixels are used and both images are densely sampled with the same number of points.

As we said before, the sizes of the matrices involved in the previous inversions are 289 X 1681 and 289 X 25921 for the square pixels and 289 X 289 for the natural pixels. The first and the last matrices allow the computation of the singular values in a reasonable time. The results are shown in Fig. 11, were the curve labeled with "3" (upper curve) refers to the natural pixels discretization and the curve "1" to the 289 X 1681 matrix obtained when square pixels are used. Curve "2" represents the singular values of the matrix obtained when the model is discretized in a grid of 17 X 17 square pixels (size of the matrix 289 X 289, the same as the matrix in curve "3"). We notice that the matrix computed as intersections of natural pixels (curve "3") is significantly better conditioned that the other two matrices. Sampling the image more densely also makes the square pixels matrices (curves "1" and "2") better conditioned.

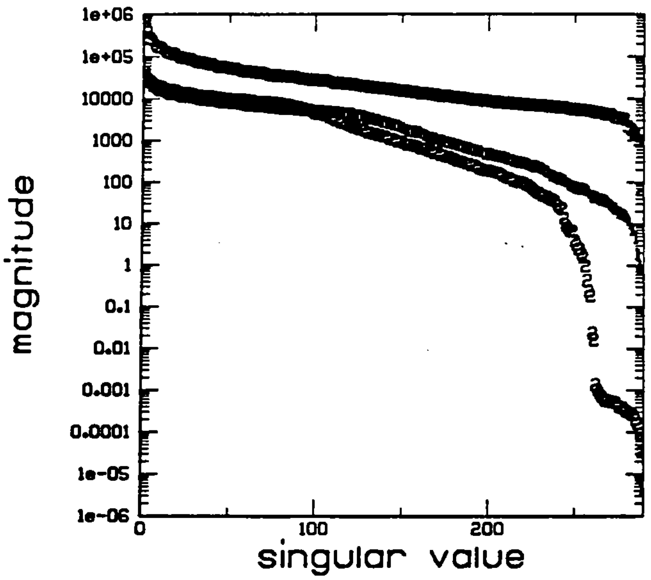


Figure 11: Singular value decomposition for the matrices obtained with different the discretizations: 1) 41 X 41 square pixels, 2) 17 X 17 square pixels and 3) natural pixels.

CONCLUSIONS

We have shown that the natural pixels provide an efficient way of discretizing the slowness model in the problem of traveltime tomographic inversion. In the examples studied, images of similar quality were obtained using natural pixels compared with the traditional reconstruction of square pixels. The main advantage of the natural pixels is that the number of model parameters needed is two orders of magnitude smaller, which means a proportional reduction on the computational effort. Besides that, the inversion with the natural pixels is better conditioned than the inversion with square pixels when comparable amount of parameters are used.

For obtaining the estimate $\tilde{S}(x,y)$ with natural pixels, two different minimizations problems are involved. The first one is implied by Eqn. (9), where the function to be minimized is the error between the true *model* and the estimated one. The second minimization problem is related to the solution of the system of equations (22), where the function to be minimized is the error between observed and calculated *data.* We have shown in this paper the importance of the first minimization in terms of the computational effort required to obtain the the final estimate.

The number of natural pixels equals the number of data points. It means that the number of model parameters in the inversion remains *constant* for a fixed amount of data, regardless the spatial dimensions of the problem or the resolution of the display. Consequently, the natural pixels provide a direct procedure for inversion in three dimensions, problems that can be computationally impossible to attack if the model is described with orthogonal three dimensional pixels (boxes).

The discretization of the model along the beam paths will change in general from one experiment to another and from one iteration to another one within the same inversion, since the beam paths may change in each of these situations. In this sense, we can say that the discretization is flexible because it depends on the given data. However, some flexibility is lost if we want to introduce in the model information not described by beam paths such as known boundaries or slowness in some areas.

More research has to be done for determining the most appropriate function that approximates the beam paths (instead of the natural pixels described herein), depending on the characteristics of the data set. Woodward (1989) gives important indications about this problem describing the beam paths (wavepaths) as elliptical, multiple-Fresnel-zone patterns, analogous to the migration ellipses. The width of her wavepaths is inversely proportional to the bandwidth and it is independent on the central frequency. Woodward's wavepaths are calculated using finite differences which might not be convenient when large data sets are inverted with our method of discretization.

ACKNOWLEDGMENTS

We would like to thanks Spyros Lazaratos for many interesting discussions and important suggestions. The first author thanks INTEVEP, S.A. for the financial support.

REFERENCES

Aki, K., and Richards, P. G., 1980, Quantitative seismology: W. H. Freeman and Co.

Berberian, S. K., 1976, Introduction to Hilbert space: Chelsea Publ. Co.

Buonocore, M. H., 1981, Fast minimum variance estimator for limited angle computed tomography image reconstruction: Ph.D thesis, Stanford University.

Buonocore, M. H., Brody, W. R., and Macosvski, A., 1981, A natural pixel decomposition for two dimensional image reconstruction: IEEE Trans. Biomedical Engineering, **BME-28**, 69-78.

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

Darling, A. M., Hall, T. J., and Fiddy, M. A., 1983, Stable noniterative object reconstruction from incomplete data using a priori knowledge: J. Opt. Soc. Am., **73**, 1466-1469.

Harlan, W. S., 1989, Tomographic estimation of seismic velocities from reflected raypaths: presented at the 59th Ann. Internat. Mtg., Soc. Expl. Geophys.

Ivansson, S., 1985, A study of methods for tomographic velocity estimation in the presence of low velocity zones: Geophysics, **50**, 969-988.

McMechan, G. A., 1983, Seismic tomography in boreholes: Geophys. J. Roy. Astr. Soc., **74**, 601-612.

Nolet, G., 1985, Solving or resolving inadequate and noisy tomographic systems: J. Comp. Phys., **61**, 463-482.

Nolet, G., 1987, Seismic wave propagation and seismic tomography, *in* Nolet, G., Ed., Seismic Tomography: D. Reidel Publ. Co., 1-23.

Stakgold, I., 1979, Green's functions and boundary value problems: John Wiley & Sons, Inc.

Van der Sluis, A., and Van der Vorst, H. A., 1987, Numerical solution of large, sparse linear algebraic systems arising from tomography problems, *in* Nolet, G., Ed., Seismic Tomography: D. Reidel Publ. Co., 49-84.

Van Trier, J., 1988, Migration velocity analysis using geological constraints: presented at the 58th Ann. Internat. Mtg., Soc. Expl. Geophys.

Woodward, M. J., 1989, Wave equation tomography: Ph.D thesis, Stanford University.

APPENDIX A: TRAVELTIME CALCULATION ALONG NATURAL PIXELS

We explain in this appendix how to compute accurately the integrals along the natural pixels in any slowness model after the rays has been traced.

Equation (20) is a surface integral (in 2D). Surface integrals are in general difficult to evaluate, specially if there is no analytical expression for the integrand. A conceptually simple way of evaluating this integral is based on a fine discretization of the model, evaluating and summing the contribution of points that belong to the support of each particular natural pixel. The final sum is divided by the width of the natural pixel. The main disadvantage of this method is that a very fine discretization of the model is needed to ensure accuracy in the traveltimes, even when the medium is homogeneous. Fig. A1 (thick curve) shows an example of the traveltimes calculated with this method in an homogeneous medium.

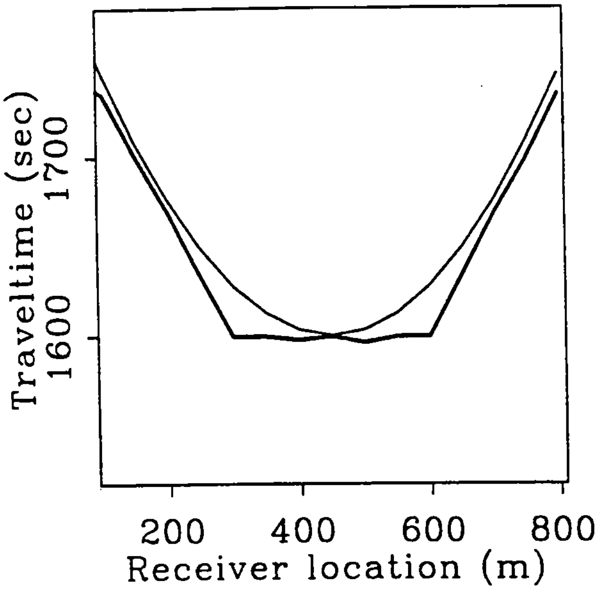


Figure A1: Traveltimes in a homogeneous medium calculated using natural pixels. The thin curve is calculated multiplying the step of the ray by the average velocity in a perpendicular segment. The thick curve is calculated by discretizing and summing for all the points that belong to the support of each natural pixel.

A more precise way of evaluating numerically Eqn. (20) is simply by summing the products of the length of the segment between two consecutive points of the ray by the average slowness in a line perpendicular to that segment. The line has *length = width of the tube.* This is simply a generalization of the traveltimes calculated from expression (19) as the sum of products *slowness x length.* An example of this technique is shown in Fig. A1. Note that the shape of the curve is the expected hyperbola (thin curve), which is not the case (thick curve) in the previously explained method, where a very fine discretization is needed for the hyperbolic shape to be reproduced. In the example of Fig. A1 the discretization used for calculating the thick curve is approximately 60 times denser than the one used for the other curve.

Fig. A2 compares the traveltimes anomalies (i.e. *travetime in heterogeneous medium - traveltime in homogeneous medium*) calculated using Eqn. (19) and Eqn. (20). The thick curve represents anomalies calculated as integrals along lines (Eqn. (19)) and the thin curve represents anomalies calculated as integrals along natural pixels. The source is located at 400 m in the model of Fig. 4. Note that the anomaly calculated with the natural pixels is smoother than the one calculated integrating along the ray path. Usually, the model is smoothed before tracing the rays, and this difference might be eliminated. Another difference between the two cases (which cannot be solved by smoothing the model) is that the traveltimes calculated with natural pixels affect a larger range of receivers than the ones calculated with rays. This behavior better resembles the real measurements. The thin curve is not completely symmetric because inaccuracies in the interpolation for calculating the borders of the natural pixels and also due to roundoff errors.

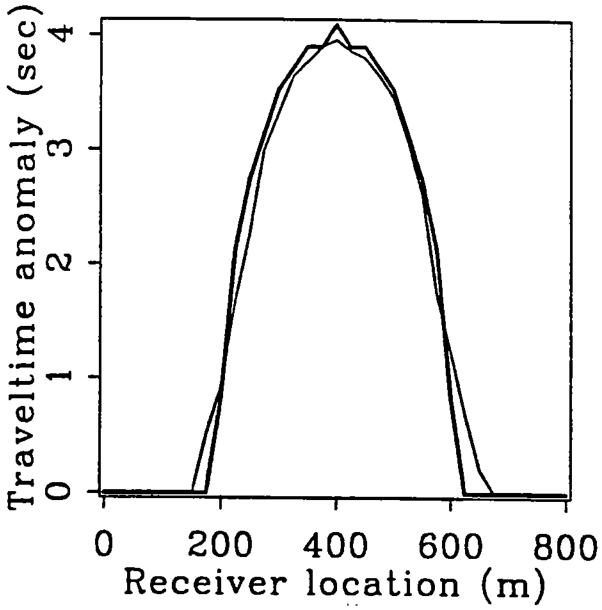


Figure A2: Traveltime anomalies calculated for the model of Fig. 4. The source is in the left of the model at 400 m depth. Receivers are in the opposite well. The thick curve is calculated with conventional ray tracing. The thin curve is calculated multiplying the step of the ray by the average velocity in a perpendicular segment.

PAPER G

RADON TRANSFORM / GAUSSIAN BEAM MIGRATION

Spyros K. Lazaratos and Jerry M. Harris
Seismic Tomography Project

ABSTRACT

Migration algorithms usually treat each point in the subsurface as an isolated diffractor. Yet, dicrete point diffractors are rare; reflections from continuous beds usually dominate the data. Regions of the subsurface could be adequately imaged through the use of a limited range of dips around a dominant local dip. It is often the case that several migrations will be performed, in order to select the optimal velocity model that will produce a well focused image. In this case an estimate of the local dip will be available after the first migration. Subsequent migrations could be very efficient, if this estimate could be incorporated into the imaging algorithm.

Motivated by this observation we derived a migration algorithm based on the reconstruction of the image from its dip spectrum. In more formal terms, we form an image by inverting its Radon transform. The Radon transform is formed by applying appropriate stacking operators to the data.

The Radon transform is a global concept. In order to be able to treat different areas of the image independently and migrate through general velocity models, we need to localize our imaging principle. We achieve this through the use of Gaussian beam wavefields as a basis for imaging. This represents an intermediate approach between point source wavefields, forming the basis for Kirchoff migration, and plane waves associated with slant-stack migrations.

The implementation of the algorithm is efficient. The subsurface can be imaged through the use of sparse fans of rays, each of them effectively imaging the neighbourhood of its origin. So, although the algorithm represents a bottom-up approach, excessive ray tracing is avoided.

The ability of the algorithm to image portions of the subsurface with relatively few rays and the possibility to concentrate on specific dips might have interesting applications in the optimization of the velocity model used for migration.

INTRODUCTION

In recent years there has been renewed interest in the use of ray methods for the solution of the seismic inverse problem (Clayton and Stolt (1981), Carter and Frazer (1984), Stolt and Weglein (1985), Cohen, Hagin and Bleistein (1985), Bleistein, Cohen and Hagin (1987), Bleistein (1987), Beydoun and Keho (1987), Keho and Beydoun (1988), Beydoun and Mendes (1989), Ikelle (1989)). The work of Beylkin has been

In the above definition A is the spreading-free amplitude defined as

$$A = (\frac{v_0 \rho_0}{v \rho})^{1/2} v_0^{1/2} \tag{2}$$

where v,ρ are the velocity and density and v_0,ρ_0 their values at the point where the initial conditions for the beam (width, curvature) are specified. The quantity Q is a complex solution of the dynamic ray tracing equations. It is a complex spreading, describing both amplitude change and phase change. So, as the beam propagates and spreads, both the amplitude and the wavelet change. The traveltime τ is complex

$$\tau = \tau_r + i\tau_i \tag{3}$$

The real part of the traveltime is given by the paraxial approximation

$$\tau_r(\vec{r}) = \tau_0(\vec{r}_0) + \frac{1}{2} M_R x^2 \tag{4}$$

where $\tau_0(\vec{r}_0)$ is the traveltime at the projection $\vec{r}_0$ on the ray of the point $\vec{r}$ (see Figure 1), x is the distance of $\vec{r}$ from the ray and M_R is the curvature of the wavefront defined from the initial curvature and width of the beam through the solution of the dynamic ray tracing equations. The form of the last equation shows that the wavefronts of Gaussian beams are parabolic. The imaginary part of the traveltime is given by

$$\tau_i(\vec{r}) = \frac{1}{2} M_I x^2 \tag{5}$$

where the imaginary curvature M_I is also defined through the dynamic ray tracing equations. So the imaginary traveltime is parabolic and its exponential is a Gaussian, hence the name Gaussian beams. An excellent reference on Gaussian beams and dynamic ray tracing is the paper by Červený (1985).

Gaussian beams are generally presented in the frequency domain. In fact the amplitude decay of a Gaussian beam away from the central ray is Gaussian only for a fixed frequency. Since Gaussian beam wavefields are the basis for our algorithm, and time domain images of Gaussian beam wavefields are not common in the literature, we show a few examples in Figure 2.

RADON TRANSFORM

We give here the expressions for the 2-D Radon transform and its inverse, mainly to introduce the notation that we are going to use. The Radon transform $\hat{f}(\vec{p},l)$ of a 2-D function $f(\vec{r})$ is formed by the line integrals of $f(\vec{r})$ along straight lines, identified by their direction (defined by the normal unit vector $\vec{p}$) and their distance from the origin l. So:

$$\hat{f}(\vec{p},l) = \int d\vec{r} f(\vec{r}) \delta(l - \vec{p}\vec{r}) \tag{6}$$

The inverse of the Radon transform is

$$f(\vec{r}) = \frac{1}{4\pi} R(\vec{r}) * \int_0^{2\pi} dp \hat{f}(\vec{p}, l = \vec{p}\vec{r}) \tag{7}$$

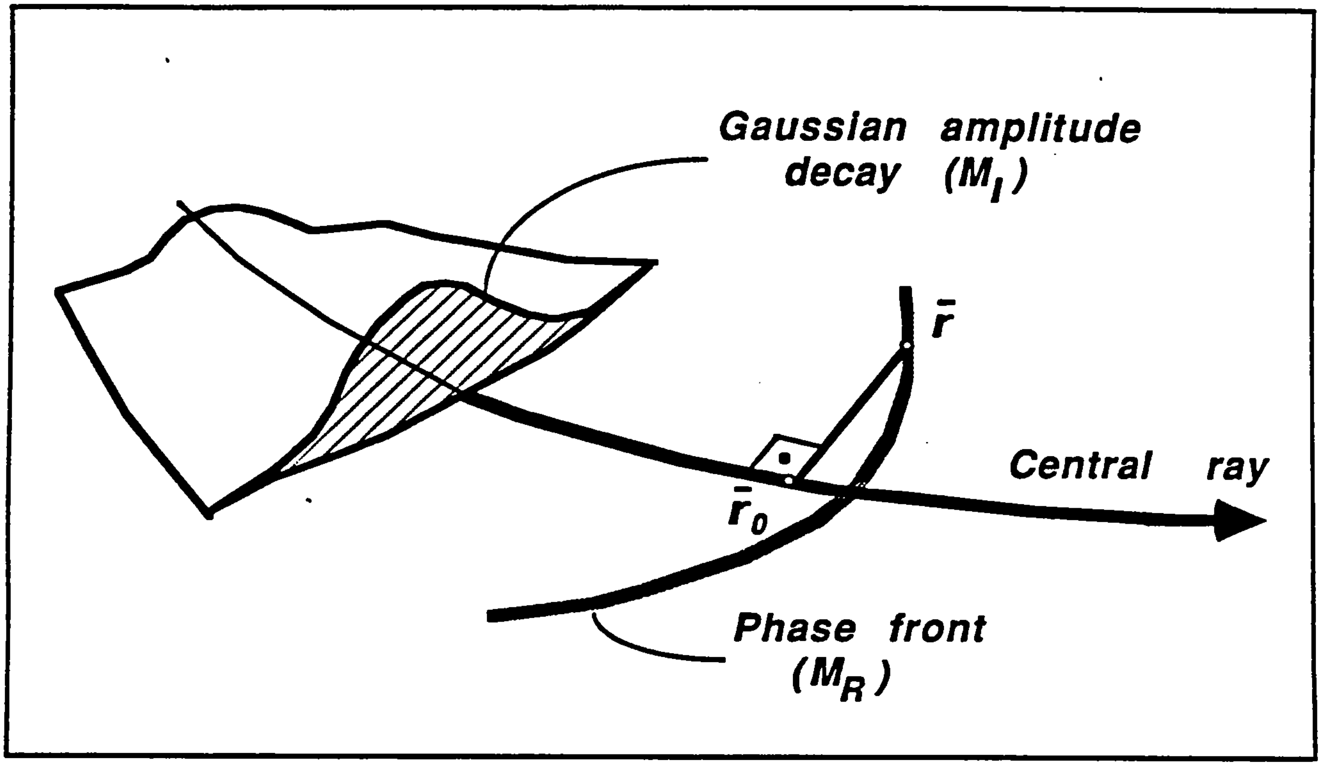


Figure 1. Gaussian beams

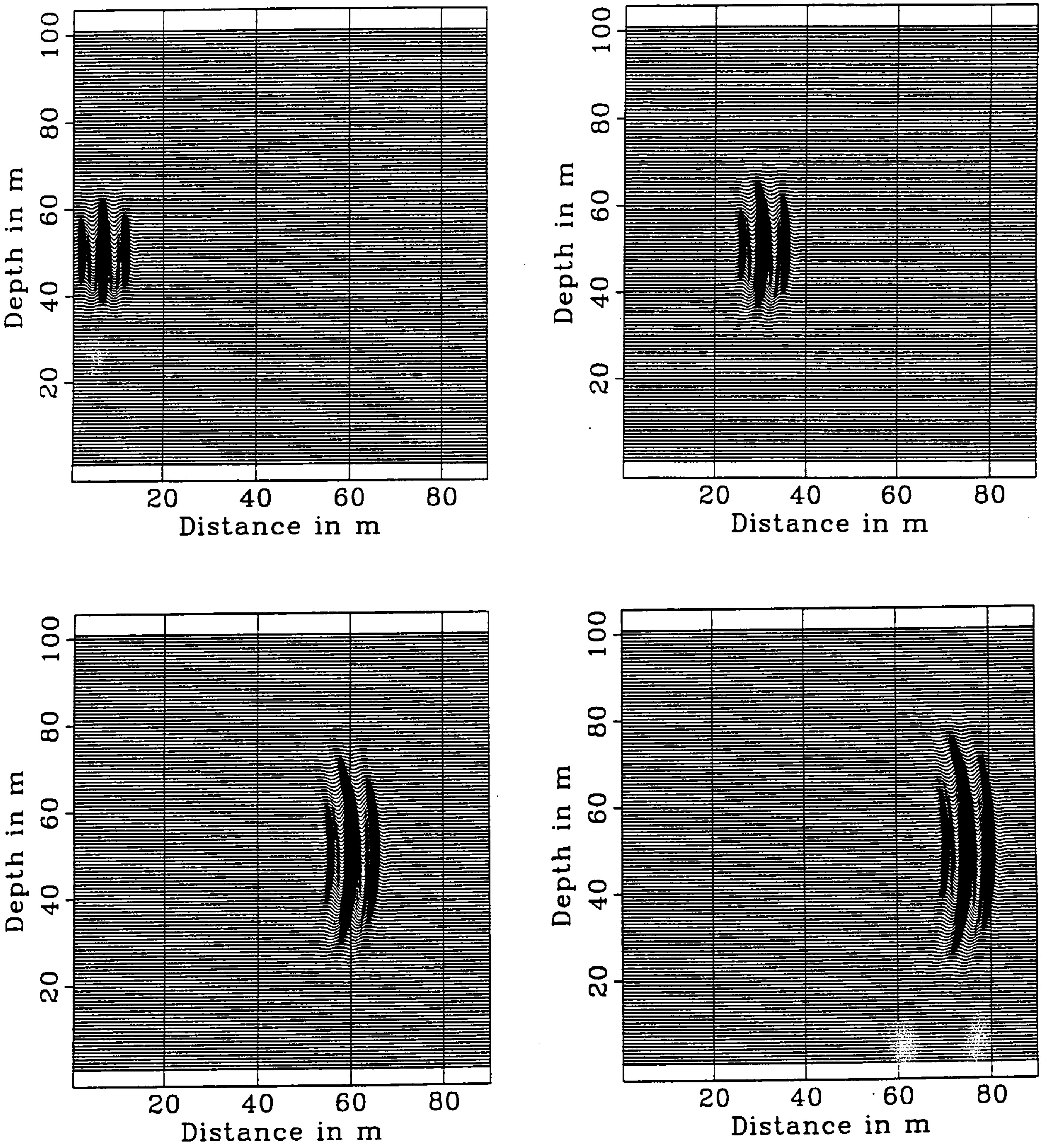


Figure 2. Snapshots of a propagating Gaussian beam. The wavefield has been calculated from a single horizontal ray.

where $*$ denotes convolution and $R(\vec{r})$ is the two-dimensional rho-filter (a spatial filter whose frequency response is $\sqrt{k_x^2 + k_z^2}$).

HOW TO SAMPLE THE RADON TRANSFORM

Since the Radon transform can be inverted, it follows that, if we can form the Radon transform of the image from the data, we have essentially solved the imaging problem. In this section we describe the appropriate way to operate on the data in order to sample the Radon transform.

To simplify the presentation we are going to describe the zero offset case, corresponding to CDP stacked data. We are going to outline the generalization to pre-stack migration in a later section.

We base our derivation on the Kirchoff integral (Schneider (1978), Berkhout (1981)), which allows us to calculate the scattered wavefield from its values at the surface. For a constant density acoustic medium the appropriate expression is

$$U_s(\vec{r};\omega) = \frac{1}{4\pi}\int [U_s(\vec{r}_R;\omega)\frac{\partial G_0(\vec{r};\vec{r}_R;\omega)}{\partial z_R} - \frac{\partial U_s(\vec{r}_R;\omega)}{\partial z_R}G_0(\vec{r};\vec{r}_R;\omega)]d\vec{r}_R \tag{8}$$

In the above equation $U_s(\vec{r};\omega)$ is the scattered wavefield at point $\vec{r}$ of the subsurface; $U_s(\vec{r}_R;\omega)$ is the scattered wavefield recorded at the location $\vec{r}_R$ on the surface. The Green's function $G_0(\vec{r};\vec{r}_R;\omega)$ represents the pressure field recorded at surface location $\vec{r}_R$ due to a monopole at $\vec{r}$. The partial derivative with respect to z_R is the partial derivative in the direction normal to the receiver line, which, for simplicity, is assumed to be the line $z = 0$.

According to the exploding reflector imaging principle the image of the subsurface is given by the value of the scattered wavefield at time zero. If we call $f(\vec{r})$ this image, representing the result of migration, we have

$$f(\vec{r}) = \frac{1}{4\pi}\int d\omega \int d\vec{r}_R[U_s(\vec{r}_R;\omega)\frac{\partial G_0(\vec{r};\vec{r}_R;\omega)}{\partial z_R} - \frac{\partial U_s(\vec{r}_R;\omega)}{\partial z_R}G_0(\vec{r};\vec{r}_R;\omega)] \tag{9}$$

The Radon transform of the image $f(\vec{r})$ is then

$$\hat{f}(\vec{p},l) = \frac{1}{4\pi}\int d\omega \int d\vec{r}_R[U_s(\vec{r}_R;\omega)\frac{\partial W_{\vec{p},l}(\vec{r}_R;\omega)}{\partial z_R} - \frac{\partial U_s(\vec{r}_R;\omega)}{\partial z_R}W_{\vec{p},l}(\vec{r}_R;\omega)] \tag{10}$$

where the quantity

$$W_{\vec{p},l}(\vec{r}_R;\omega) = \int d\vec{r}G_0(\vec{r};\vec{r}_R;\omega)\delta(l - \vec{p}\vec{r}) \tag{11}$$

is the wavefield at the surface due to a plane wave source located into the medium along a line whose orientation is given by the unit vector $\vec{p}$ and whose distance from the origin is l (see Figure 3). In Appendix A we show that a high-frequency, far-field approximation to equation (10) is

$$\hat{f}(\vec{p},l) = \frac{1}{2\pi}\int d\omega \int d\vec{r}_R U_s(\vec{r}_R;\omega)\frac{\cos\theta(\vec{r}_R)}{v(\vec{r}_R)}[-i\omega W_{\vec{p};l}(\vec{r}_R;\omega)] \tag{12}$$

where $\theta(\vec{r}_R)$ is the angle formed by the normal to the wavefront of $W_{\vec{p},l}$ and the vertical. The quantity $\cos\theta(\vec{r}_R)$ is an obliquity factor similar to the one encountered in Kirchoff migration.

So, in order to find the value of the line integral along the image, we need to

1. Calculate the wavefield $W_{\vec{p},l}(\vec{r}_R;\omega)$.
2. Apply to the data the stacking operator

$$\int d\omega \int d\vec{r}_R S_{\vec{p},l}(\vec{r}_R;\omega) \tag{13}$$

where

$$S_{\vec{p},l}(\vec{r}_R;\omega) = \frac{1}{2\pi}\frac{\cos\theta(\vec{r}_R)}{v(\vec{r}_R)}[-i\omega W_{\vec{p},l}(\vec{r}_R;\omega)] \tag{14}$$

In a homogeneous or layered velocity background it is easy to propagate a plane wavefield. This is the basic idea behind $\tau - p$ domain inversions and Fourier domain algorithms like the phase shift and $f - k$ migrations. Diffraction tomography algorithms (Devaney (1984), Harris (1987)) are also generally formulated in the Fourier domain. Yet, when lateral variation is introduced in the velocity model, originally plane wavefields do not propagate in a simple way. We propose to handle the propagation issue by decomposing the plane wavefields into a superposition of Gaussian beams.

The fact that Gaussian beams can be used as a basis for the synthesis of wavefields has been established through the work of Červený and his co-workers (Červený, Popov and Pšenčik (1982), Červený and Pšenčik (1983), Červený (1983), Červený (1985), Červený (1985a)). Klimeš (1984) showed how a high frequency time-harmonic wavefield specified on an initial surface can be expanded into a sum of Gaussian beams. Raz (1987) proposed the Gabor transformation as a way of performing this decomposition.

Our basic idea of decomposing into Gaussian beams the originally plane wavefields which sample the Radon transform of the medium is shown schematically in Figure 3. In a laterally varying medium the originally plane wavefield arrives at the surface having a different shape. We essentially propose to use spatially limited parabolic wavefields (the Gaussian beams) to approximate this wavefield. The exact way this decomposition is carried out and the resulting imaging algorithm are the subject of the next two sections.

GAUSSIAN BEAM WAVEFIELDS AS A BASIS FOR IMAGING

The inverse of the Radon transform given in equation (7) can be written as

$$f(\vec{r}) = \frac{1}{4\pi}R(\vec{r}) * \int d\vec{p} \int dl\,\delta(l - \vec{p}\vec{r})\hat{f}(\vec{p},l) \tag{15}$$

The δ function appearing in the last expression can be viewed as the wavefield of a plane wave at a particular instant. We claim that this plane wave wavefield can be

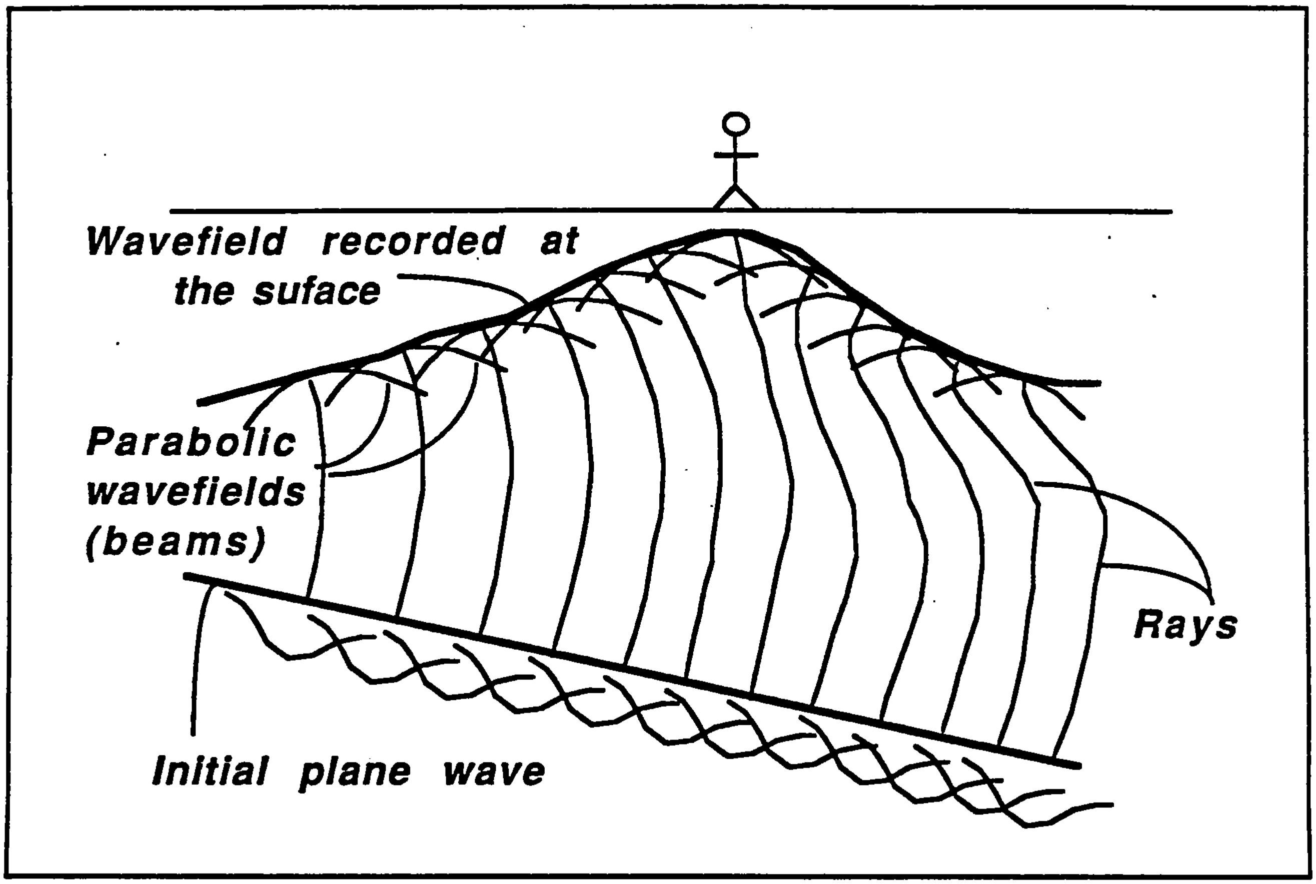


Figure 3. Basic idea for migration algorithm. To find the line integral of the image along a line, we need to calculate the wavefield recorded at the surface when a plane wave source is placed on the line. We calculate this wavefield by decomposing the plane wave into Gaussian beams. This is equivalent to decomposing the wavefield into parabolic wavefields, resulting from the propagation of the Gaussian beams to the surface.

approximated by a sum of Gaussian beam wavefields. We use plane Gaussian beams. An example of the wavefields we use is given in Figure 4. We show a bandlimited wavefield to make it easier to visualize. We are going to use such wavefields as a basis for the construction of the image.

Plane Gaussian beam wavefields are not exactly truncated plane waves. The support of a plane Gaussian beam has some thickness, as shown in Figure 5. Yet, we observed that, when we use beams whose widths are a few times larger than the wavelength for the frequency band of interest, this effect is insignificant (see for example Figure 4). Another approximation we make to simplify the decomposition is to ignore the change in the wavelet as we move away from the central ray. This change occurs because different frequencies attenuate with different rates. The synthetic examples shown later support the fact that these approximations are not severe. We expect that with real data their effect will be much less significant than the effect of noise.

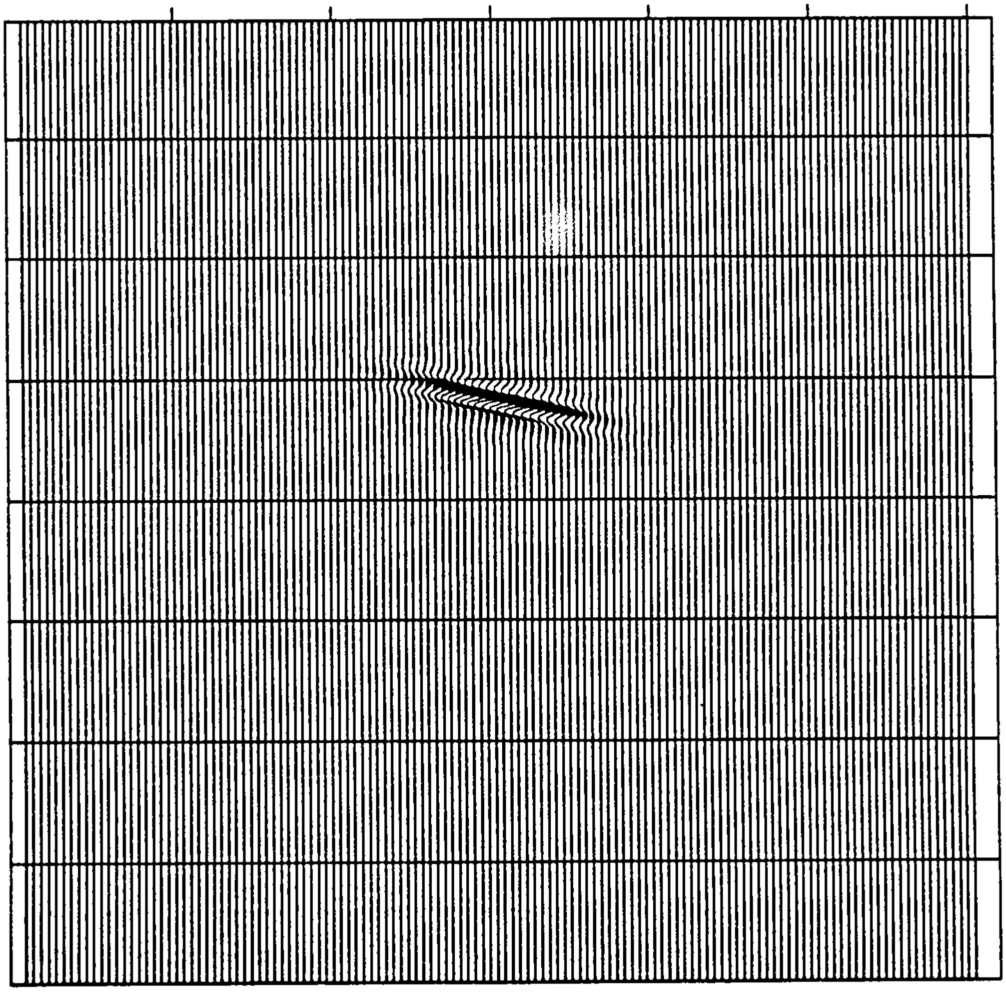


Figure 4. Plane Gaussian beam wavefield. Such wavefields are our basis for imaging.

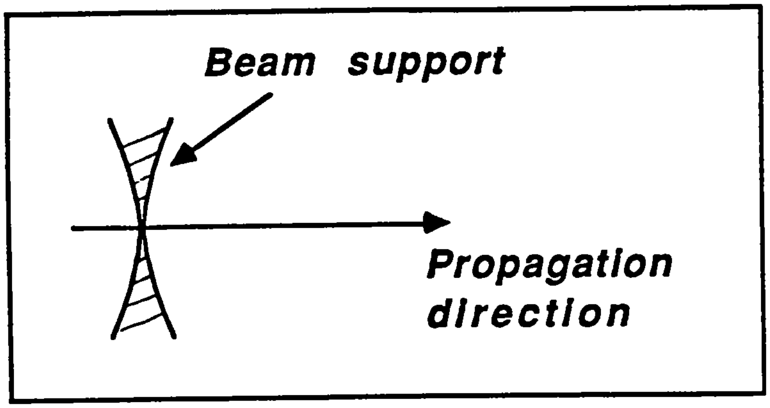


Figure 5. Shape of the support of a plane Gaussian beam wavefield.

So, we essentially approximate the Gaussian beam basis wavefields by short plane waves. The decay away from the central ray is not Gaussian for a bandlimited wavelet. It is Gaussian only for a monochromatic beam. If the spectrum of the wavelet is $S(\omega)$, the decay away from the central ray is

$$\int S(\omega) \exp(-\frac{1}{2}\omega M_I x^2) d\omega \tag{16}$$

where M_I is the imaginary curvature and x the distance from the central ray.

Under the above approximations the decomposition of the plane wavefield into a sum of Gaussian beam wavefields becomes a one dimensional problem. It essentially means the decomposition of a constant function into a sum of functions representing the amplitude tapering of the beams. This can be performed with controlled accuracy, as shown in Figure 6.

Indeed, summing identical translated tapering functions is equivalent to convolving the tapering function with a series of delta functions. In the frequency domain, this means approximating a delta function (the Fourier transform of a constant) by the product of the Fourier transform of the tapering function and a series of delta functions (the Fourier transform of the series of delta functions). We can easily see that we can do that with controlled accuracy provided the tapering has a Fourier transform with effectively finite extent.

We call the Gaussian beam wavefields that form our imaging basis $B_{i,\vec{p},l}(\vec{r})$. The meaning of the indices is shown in Figure 7. The unit vector $\vec{p}$ and the distance from the origin l describe the particular line where the Gaussian beam is originally defined. The index i runs along this line. Beams with the same $\vec{p}$ are originally parallel. Beams with the same l start from the same line.

The decomposition of a plane wave $\delta(l-\vec{p}\vec{r})$ into Gaussian beams, can be described as

$$\delta(l-\vec{p}\vec{r}) = \sum_i B_{i,\vec{p},l}(\vec{r}) \tag{17}$$

Substituting this into equation (15) and approximating the integrations with summations we get

$$f(\vec{r}) = \frac{1}{4\pi} R(\vec{r}) * \sum_{\vec{p}} \sum_{l} \sum_{i} \hat{f}(\vec{p},l) B_{i,\vec{p},l}(\vec{r}) \tag{18}$$

This expression suggests that we can write $f(\vec{r})$ in the form

$$f(\vec{r}) = R(\vec{r}) * \sum_{\vec{p}} \sum_{l} \sum_{i} a_{i,\vec{p},l} B_{i,\vec{p},l}(\vec{r}) \tag{19}$$

The calculation of the coefficients in this expansion is done through the minimization of a least-squares norm. The derivation is done in Appendix B. The final result of this decomposition is

$$f(\vec{r}) = R(\vec{r}) * \sum_{\vec{p}} \sum_{l} \sum_{i} B_{i,\vec{p},l}(\vec{r}) \int d\vec{r} f(\vec{r}) B'_{i,\vec{p},l}(\vec{r}) \tag{20}$$

where

$$B_{i,\vec{p},l}(\vec{r}) = R(\vec{r}) * B'_{i,\vec{p},l}(\vec{r}) \tag{21}$$

Equation (20) is the inversion of the Radon transform in the basis that we introduced. What remains to be done in order to obtain an imaging formula is to show how the integrals in the right hand side of equation (20) can be calculated from the data. This is done in the next section.

THE IMAGING EQUATION

We now show how the coefficients in the expansion of $f(\vec{r})$ given in equation (20) can be calculated from the data. Using again the Kirchoff integral as we did in equations (8) to (12) we can show that

$$\int d\vec{r} f(\vec{r}) B'_{i,\vec{p},l}(\vec{r}) = \int d\vec{r}_R \int d\omega C_{i,\vec{p},l}(\vec{r}_R;\omega) U_s(\vec{r}_R;\omega) \tag{22}$$

where

$$C_{i,\vec{p},l}(\vec{r}_R;\omega) = \frac{1}{2\pi} \frac{\cos\theta(\vec{r}_R)}{v(\vec{r}_R)} [-i\omega W_{i,\vec{p},l}(\vec{r}_R;\omega)] \tag{23}$$

and $W_{i,\vec{p},l}(\vec{r}_R;\omega)$ is the Gaussian beam wavefield at the surface when the wavefield at time zero is $B'_{i,\vec{p},l}(\vec{r})$. It can be calculated through dynamic ray tracing from just one ray. The wavefields $W_{i,\vec{p},l}$ are the spatially limited parabolic wavefields used to approximate the wavefield in Figure 3. So our final imaging equation is:

$$f(\vec{r}) = R(\vec{r}) * \sum_{p} \sum_{l} \sum_{i} B_{i,\vec{p},l}(\vec{r}) \int d\vec{r}_R \int d\omega C_{i,\vec{p},l}(\vec{r}_R;\omega) U_s(\vec{r}_R;\omega) \tag{24}$$

The imaging algorithm described by the last equation works as follows:

- After selecting $B_{i,\vec{p},l}$ (widths and positions of beams to be used) we define $C_{i,\vec{p},l}$ through dynamic ray tracing.
- We stack the data as defined by $C_{i,\vec{p},l}$ and form the quantity

$$\int d\vec{r}_R \int d\omega C_{i,\vec{p},l}(\vec{r}_R;\omega)U_s(\vec{r}_R;\omega) \tag{25}$$

- We spread the above quantity on the image weighting it by $B_{i,\vec{p},l}(\vec{r};\omega)$.
- Repeat for all beams adding all the contributions to form the final image.

In the next section we are going to show how this procedure is actually carried out.

IMPLEMENTATION

We showed before that each component of the Radon transform is defined by the propagation of an initially plane wave to the surface (see Figure 3). This propagation is performed through a group of Gaussian beams, into which the initial plane wave is decomposed. So, we need a very large number of beams to form an image, since we need to sample all the components of the Radon transform (all the combinations of the $\vec{p}$, l parameters). Since every beam requires a central ray, it would seem that our algorithm would require a prohibitively large amount of ray tracing. This is not true. The crucial point about the implementation of the algorithm is that a whole group of beams can be propagated to the surface by just one ray.

We know that, through dynamic ray tracing, a ray defines how a plane Gaussian beam propagates to the surface. In Figure 8a we show the situation. The plane Gaussian beam wavefield is schematically shown as a short linear segment normal to the ray. Yet, this ray contains a lot more information. It also defines how a similar plane Gaussian beam starting at a different position along the ray propagates to the surface. This is shown in Figure 8b. In fact the whole family of plane Gaussian beams starting at different points along the ray can be propagated to the surface through this single ray. This is shown in Figure 8c. Each of those Gaussian beam wavefields is defined by a different linear combination of the two fundamental solutions (point source and plane wave initial conditions) of the dynamic ray tracing equations.

If we shoot a fan of rays, as shown in Figure 8d, we can image an area in the neighbourhood of the center of the fan. The size of the neighbourhood depends on the width of the beams. The rays also give information outside this neighbourhood. Yet, in order to form an image, we need beams intersecting in all possible directions. This is only true around the center of the fan.

The conclusion is that, through a single fan of rays we can image a whole area, the neighbourhood of the fan center. For Kirchoff migration we would need a fan of rays for each image point.

As described in the previous section our migration essentially consists of two basic operations:

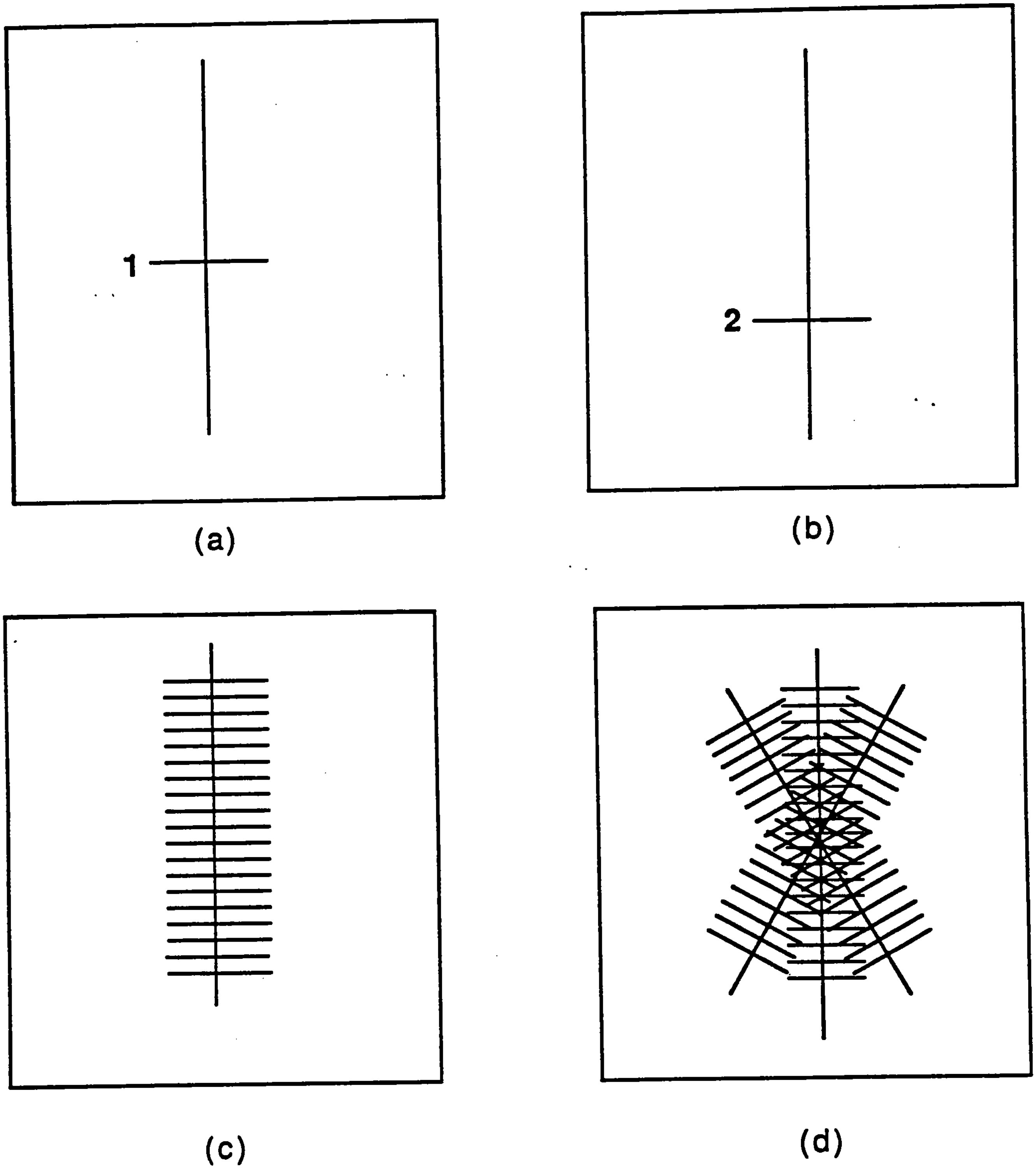


Figure 8. (a) The Gaussian beam starting at position 1 can be propagated to the surface through one central ray. (b) The Gaussian beam starting at position 2 can also be propagated to the surface through the same central ray. (c) In fact a whole group of beams can be propagated to the surface through the same central ray. (d) Imaging through a fan of rays. The neighbourhood of the center can be correctly imaged through this single fan.

- A stacking operation, governed by $C_{i,\vec{p},l}$.
- A spreading operation, governed by $B_{i,\vec{p},l}$.

The stacking operator operates on the data. The result of stacking is spread on the image by the spreading operator. The stacking operator is the wavefield recorded at the receivers when the wavefield at time zero is the wavefield of a single locally plane Gaussian beam given by the spreading operator.

An example of stacking and spreading operators is given in Figures 9 and 10. There is a one-to-one correspondence between the stacking and spreading operators shown in these Figures. Each short linear segment in Figure 9 (a spreading operator) corresponds to a short parabolic segment in Figure 10 (a stacking operator). The spreading operator at the shallowest depth corresponds to the stacking operator at the earliest time. Spreading operators at larger depths correspond to stacking operators at later times. All the stacking/spreading pairs in Figures 9a and 10a have been calculated by a single vertical ray. All the stacking/spreading pairs in Figures 9b and 10b have been calculated by a single ray forming an 8^0 angle with the vertical.

The widths of the stacking and spreading operators are of course related through the wave equation. The spreading operator essentially defines a wavefield on a linear aperture. The rate at which this wavefield spreads depends on the ratio of the wavelength to the size of the aperture. Our objective is to keep the beams as spatially limited as possible. Given the velocity model, the frequency band and a range of depths, we can choose the initial beam width (the width of the spreading operator) in the optimal way, so that the spreading of the beams is minimized.

Let us describe a little bit better the nature and calculation of the stacking and spreading trajectories. Since they represent Gaussian beam wavefields, their frequency content changes as we move away from the central ray. The stacking operation has the form

$$\int d\vec{r}_R \int d\omega \exp(-\omega\tau_i)\exp(i\omega\tau_r)U_s(\vec{r}_R;\omega) \tag{26}$$

where τ_r the real traveltime, calculated by the paraxial approximation and τ_i the imaginary traveltime, proportional to the square of the distance of the receiver from the central ray. In the time domain, the stacking operation becomes

$$\int d\vec{r}_R[U_s(\vec{r}_R;t) * \frac{\tau_i}{t^2+\tau_i^2}]/t = \tau_r \tag{27}$$

When we are close to the central ray $\tau_i \rightarrow 0$ and

$$\frac{\tau_i}{t^2+\tau_i^2} \rightarrow \delta(t) \tag{28}$$

So, the stacking trajectory starts narrow close to the central ray and widens as we move further away. The way we implemented the stacking is the time domain version given in equation (28). We found that, for reasonably broadband wavelets, this was much faster than the usual frequency domain implementation of Gaussian beams, indicated by equation (27).

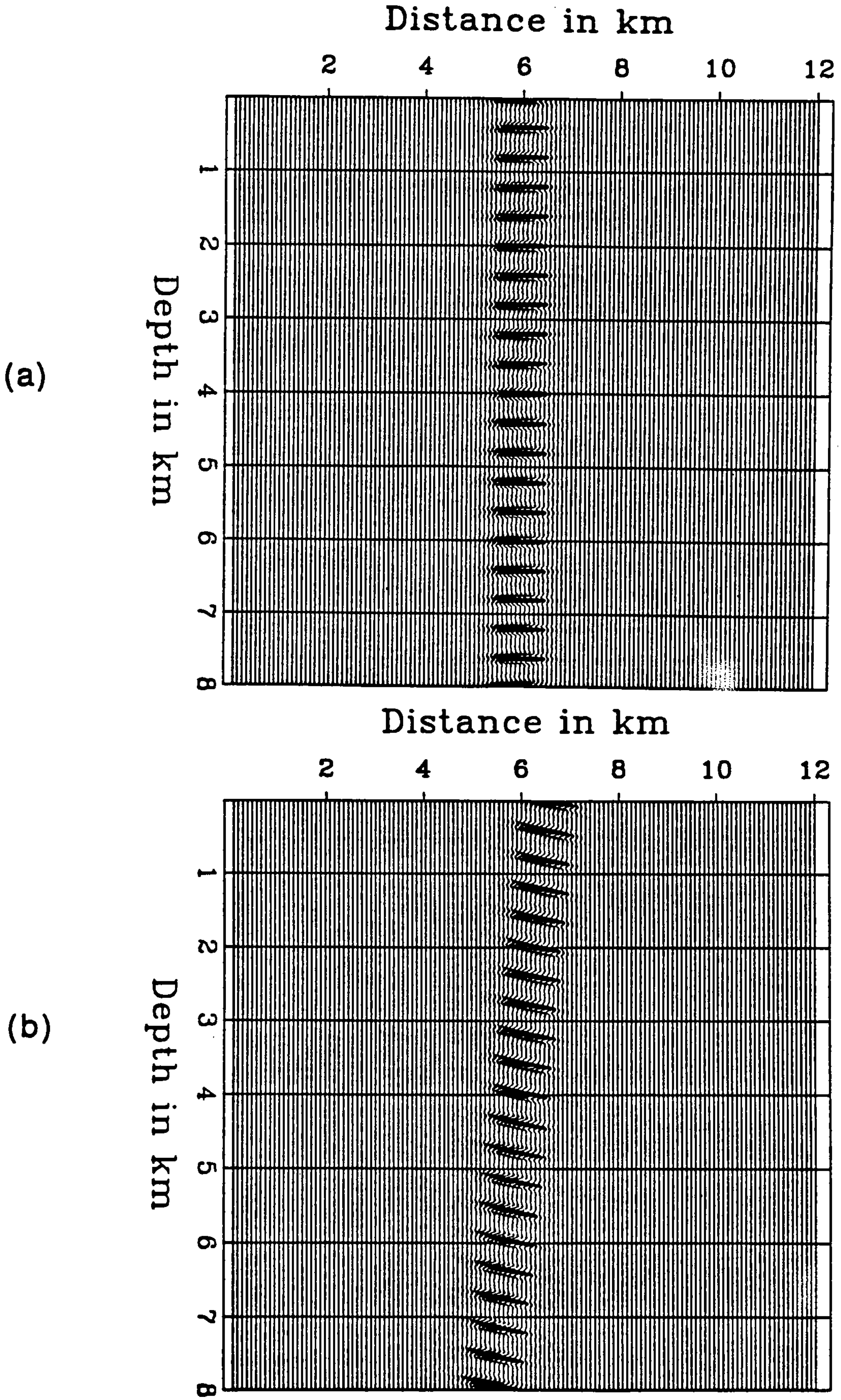


Figure 9. (a). Spreading operators calculated by a vertical ray. (b). Spreading operators calculated by a ray forming an 8^0 angle with the vertical.

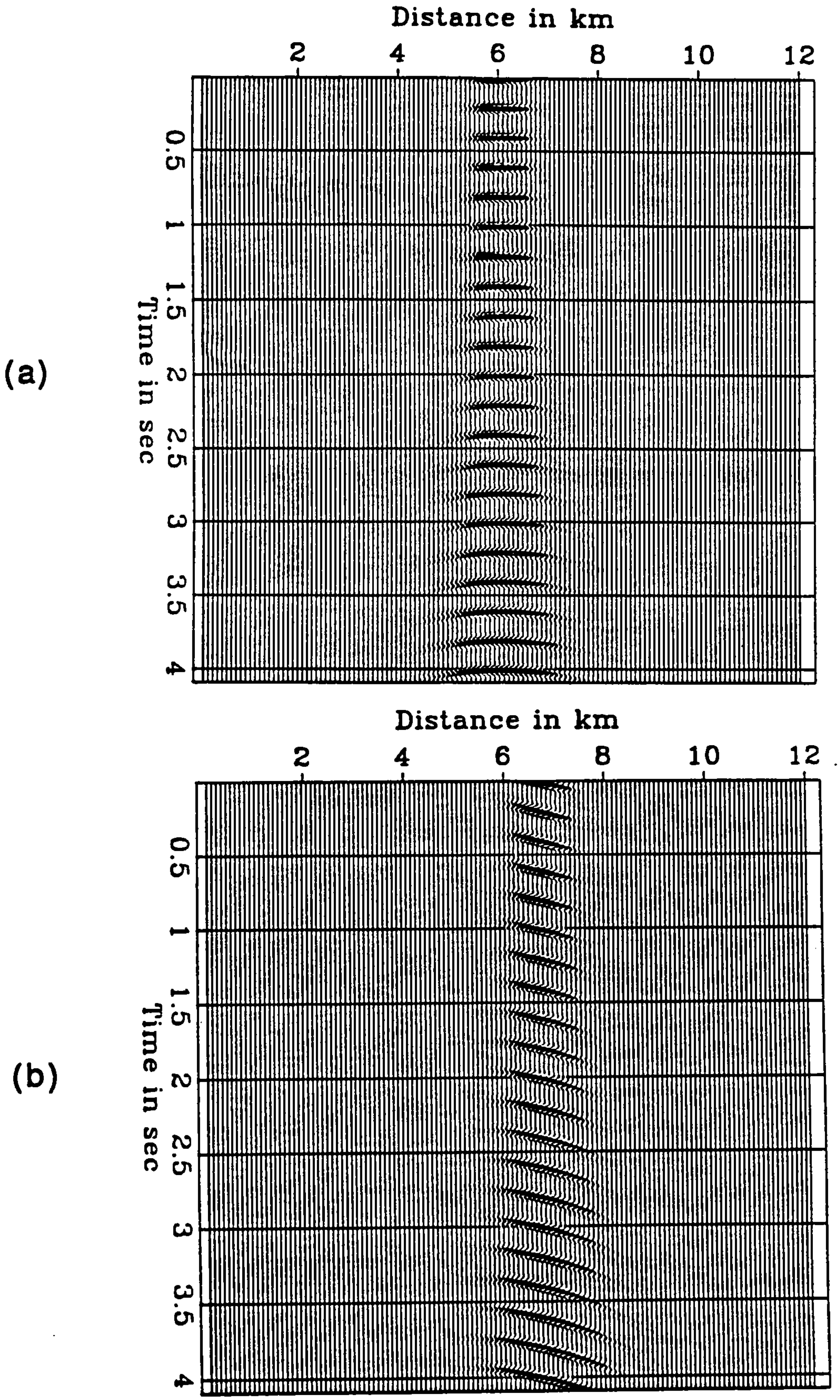


Figure 10. (a). Stacking operators corresponding to the spreading operators in Figure 9a. (b) Stacking operators corresponding to the spreading operators in Figure 9b.

To summarize, our migration works by summing along short parabolic trajectories (which are very close to short hyperbolic trajectories) and spreading the result of the summation along short linear trajectories. The actual way this is implemented is the following:

- From few points in the area of interest we shoot fans of rays.
- Each ray defines a family of stacking trajectories. Stacking the data along these trajectories we create a trace.
- Each ray also defines a family of spreading trajectories. There is a one-to-one correspondence between stacking and spreading trajectories. We spread around the ray along the spreading trajectories the trace generated by stacking.
- We repeat this procedure for all the rays in each fan.
- We combine the images generated by all fans in the way defined by the decomposition of plane waves into Gaussian beams.

PRESTACK MIGRATION

The imaging operator described before extends naturally to the prestack case. In this case, every element of the Radon transform is sampled through two beams: a 'source beam' corresponding to a stacking operation on the sources and a 'receiver beam' corresponding to a stacking operation on the receivers (see Figure 11). The stacking of the sources generates a synthetic source and similarly the stacking of the receivers a synthetic receiver. This synthetic source - synthetic receiver pair samples an element of the Radon transform. This element can be viewed as a reflector. The source beam reflects on this reflector and produces the beam that is received by the synthetic receiver.

Of course there is usually more information about each reflecting element. The same element can be sampled by several synthetic source - synthetic receiver pairs, provided that the geometric requirement of the law of reflection (incidence and reflection angles are equal) is satisfied. In other words there is usually amplitude vs. offset information in the data. The way to use this information to recover the elastic properties of the subsurface under the framework of Born inversion has been shown in Lazaratos (1989).

For the zero-offset case we have seen that the stacking operator implies summing along lines (short pieces of parabolas) in the data space. In the prestack case, the double stack described before (to create the source and receiver beams) implies summing along surfaces in the three-dimensional (source-receiver-time) data cube.

We presented the outline of prestack migration in the source-receiver coordinate system. Through a simple linear transformation we can get to the midpoint-offset coordinate system. Then we would have a 'midpoint beam' and an 'offset beam'.

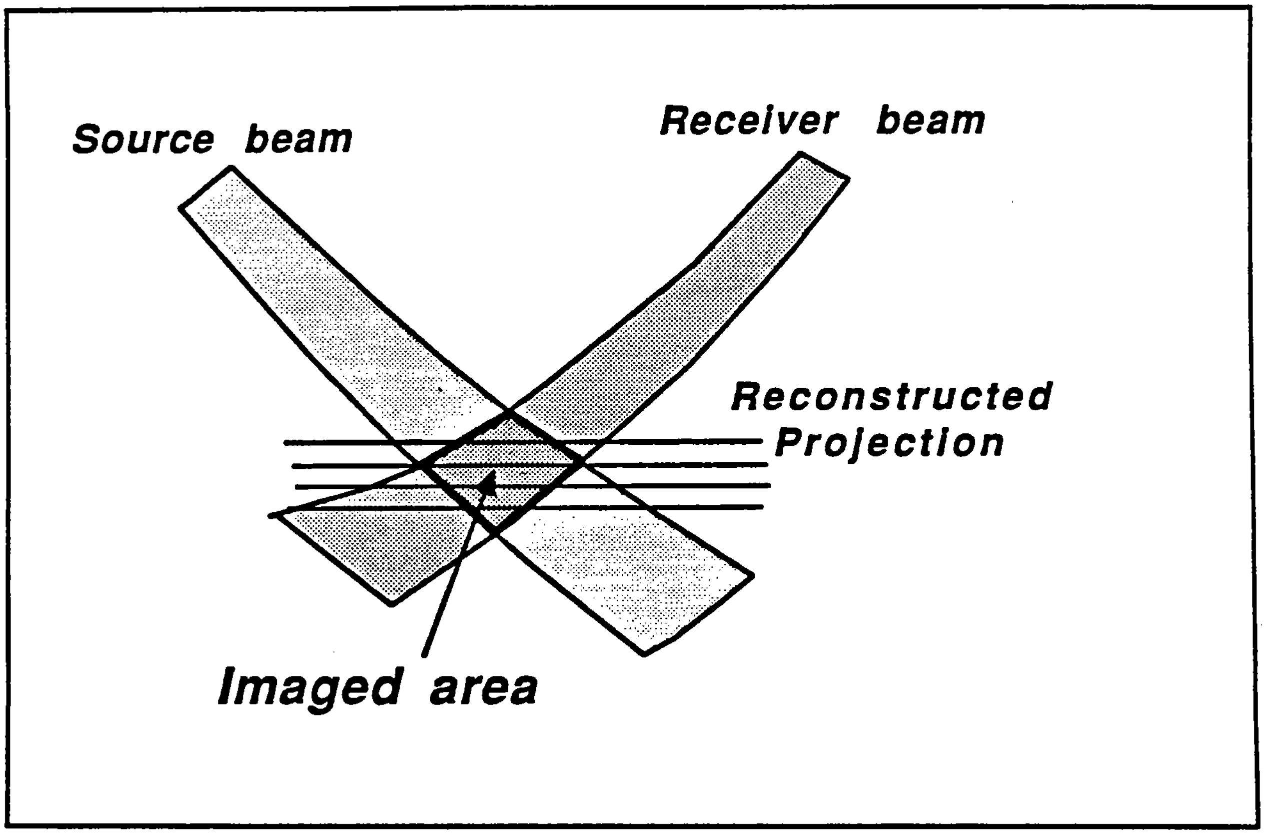


Figure 11. Principle of prestack migration.

The offset beam would be created through a short CDP stack and the midpoint beam through summing along a short parabola, approximating a short hyperbola. So the operation implied by a pair of beams is a partial CDP stack and a partial migration.

The details of the implementation of the prestack version of our migration are a subject of current research.

A SYNTHETIC EXAMPLE

A very simple synthetic example of zero-offset migration is presented in this section. The model is a syncline shown in Figure 12. Synthetic zero-offset data have been generated through the Kirchoff integral. They are shown in Figure 13. A velocity of 4.0 km/s was used. In Figure 15 we show the part of the image constructed through the single ray fan shown in Figure 14. In Figure 17 we show the part of the image constructed through the single ray fan shown in Figure 16. We can see that even a single ray fan images correctly the structure in its neighbourhood. The size of this neighbourhood depends of course on the width of the beams. For these examples the half-width of the beams, for the central frequency of the wavelet, was 1 km in the first case and 2.4 km in the second case.

Of course the amplitudes are not correct. Since the beam fades away from the central ray, the image fades as we move away from the center of the fan. Yet the correct amplitudes can be recovered by combining nearby fans. This part is currently being implemented.

Although the migration has to be tested with more complicated models and velocity backgrounds, we believe that the results from this very simple example are very encouraging. The really important point is that parts of the total image can be reconstructed with relatively few rays. The rays contain the information about the velocity model. They define how the image is affected by changes in the velocity model. Doing the imaging with as few rays as possible means that changes to the image due to changes in the velocity model can be efficiently calculated. We can calculate an exact gradient (of the image with respect to the velocity model) since the effect of each ray on the image is well defined. In the next section we are going to better explain this fact and its possible use for the optimization of the velocity model through which we migrate.

CHARACTERISTICS - POSSIBLE APPLICATIONS

In this section we outline some applications for which our migration algorithm seems to be suitable by its nature. Most of the topics presented below will be the subject of future research.

1. Efficient imaging of restricted areas. One of the things our migration has in common with Kirchoff migration is that the computational effort is inversely proportional to the size of the area that is imaged. This is not true for other types

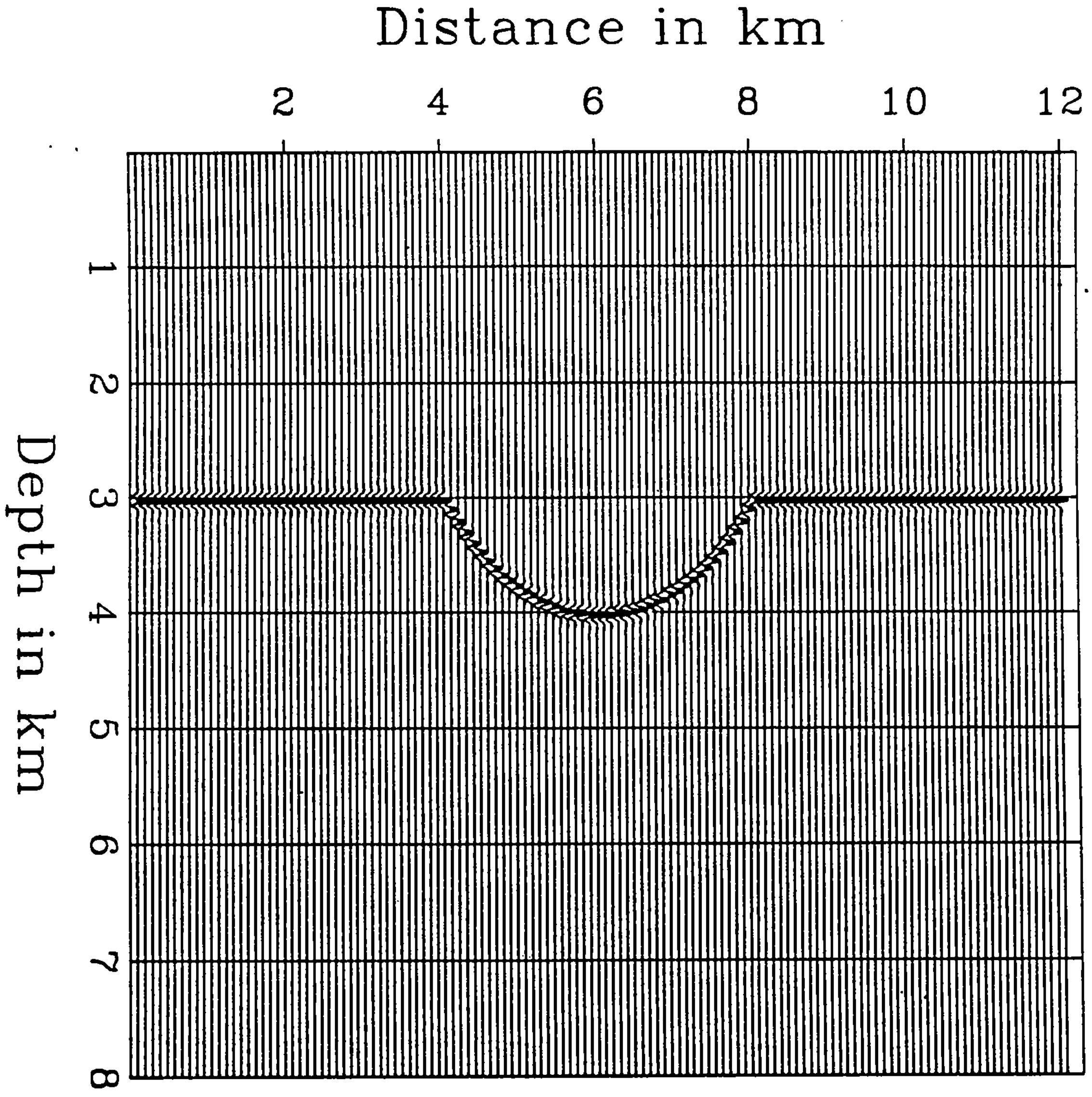


Figure 12. Synthetic model.

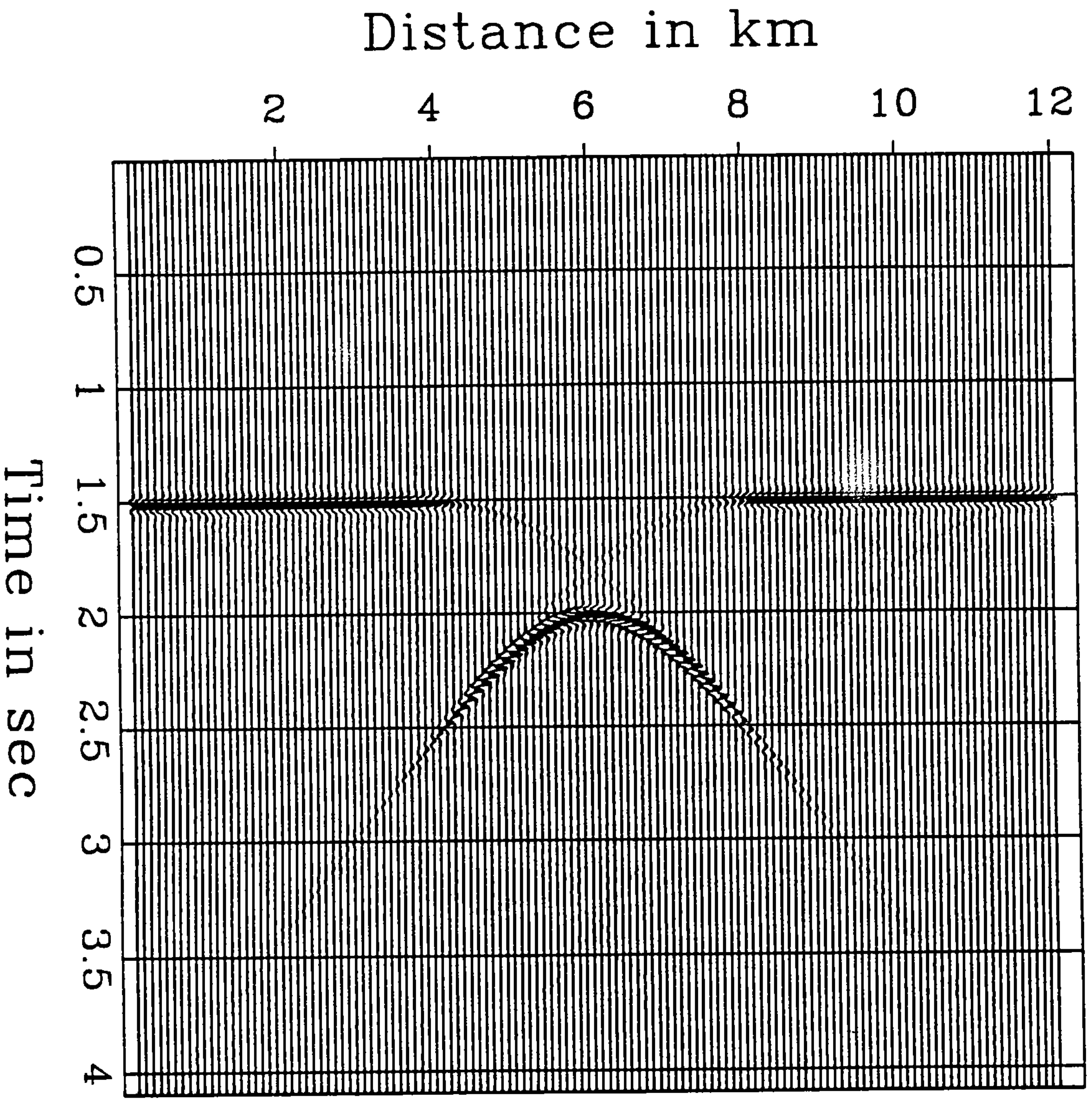


Figure 13. Synthetic zero-offset data.

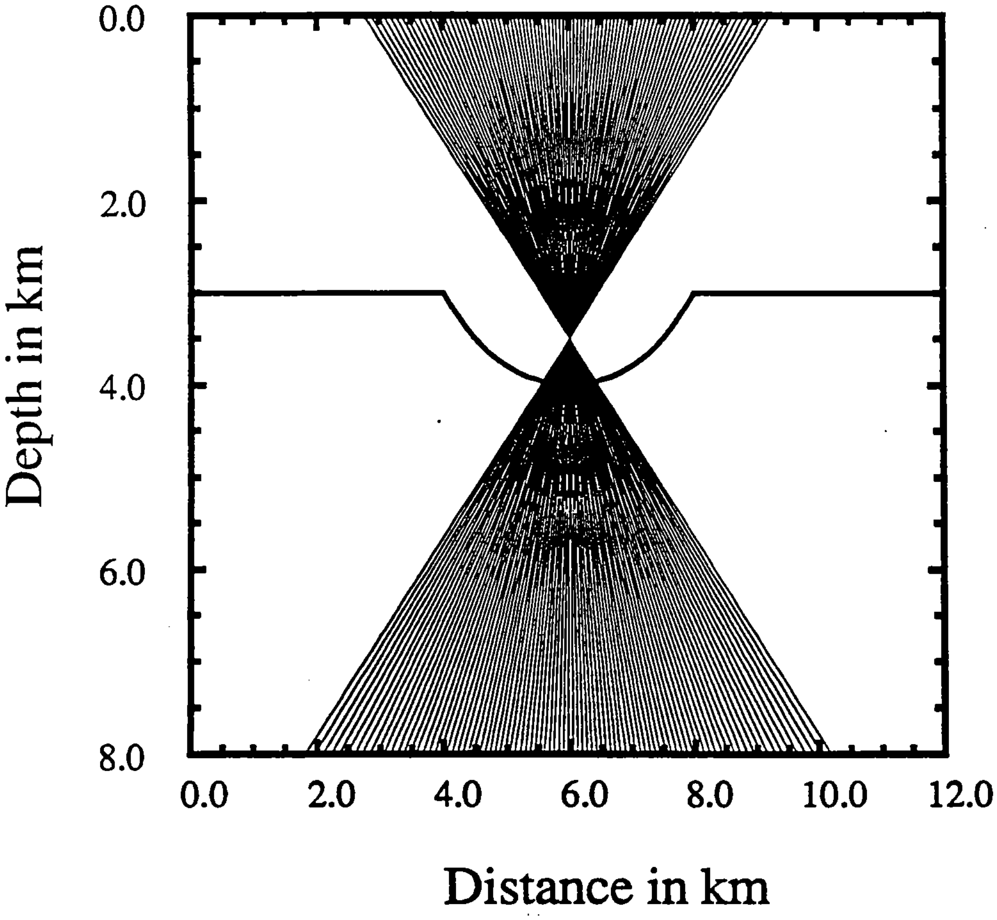


Figure 14. Rays used for imaging.

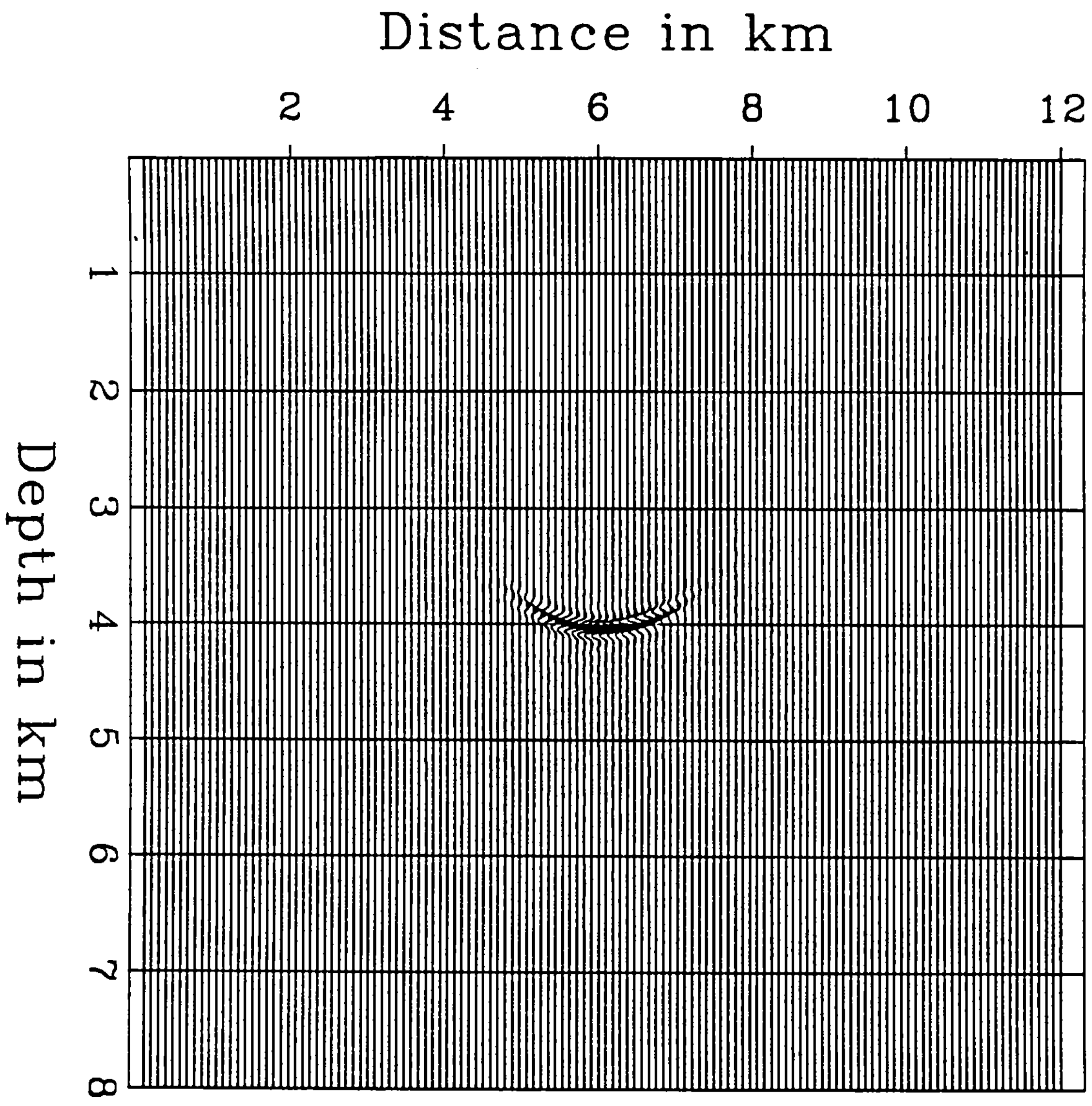


Figure 15. Image constructed from rays shown in Figure 14.

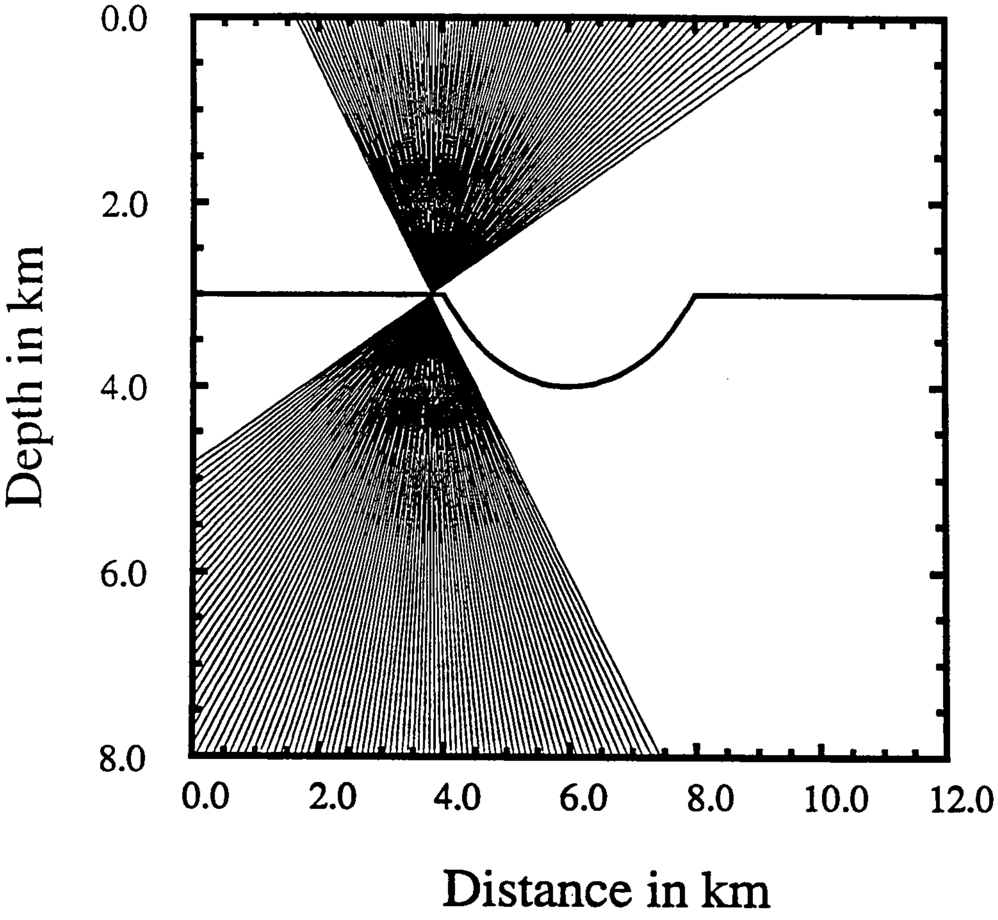


Figure 16. Rays used for imaging.

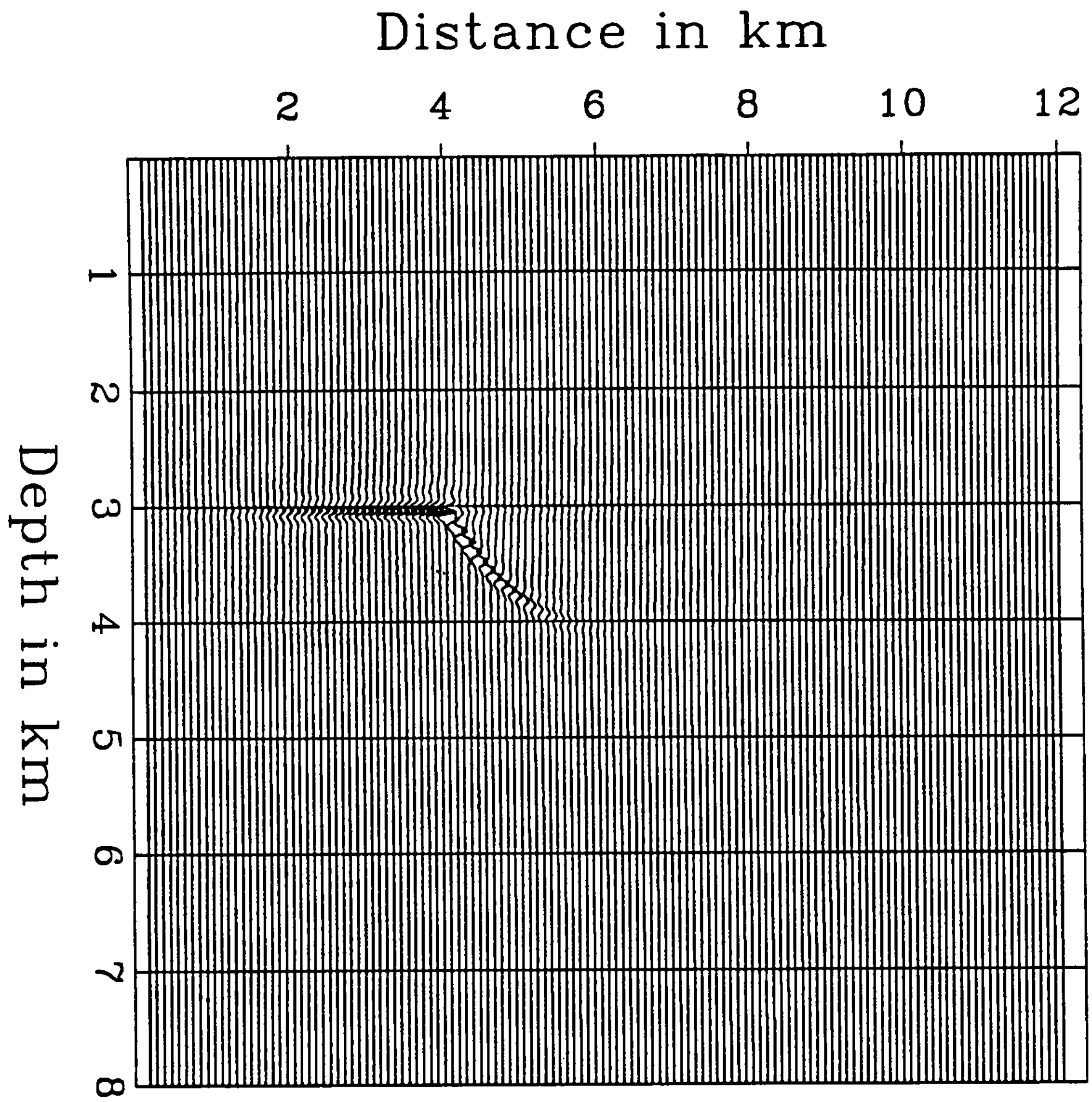


Figure 17. Image constructed from rays shown in Figure 16.

of migration like finite-difference migrations.

2. Flexibility in the choice of imaging parameters. Some of the parameters that can be selected are:

- Beam widths and positions of ray fans.
- Range of dips to be imaged. Some dips might be associated with noise. We might want to image a few dips at a time (like for example in the case of VSP migrations and VSP-CDP transformations). It is often the case that several migrations will be performed, in order to select the optimal velocity model that will produce a well focused image. In this case an estimate of the local dip will be available after the first migration. In subsequent migrations we could just image a small range of dips around the main local dip. Since the cost of our migration is proportional to the number of dips we are migrating for, these subsequent migrations could be performed very efficiently.
- Sampling rate. Steep dips might need higher sampling rates to avoid aliasing artifacts, while coarser sampling might perform satisfactorily for small dips.
- Range of offsets used in prestack migration. Information about a reflector can be extracted from all available angles of incidence or from a subset of them. We might want to use the AVO information in different ways for different ranges of incidence angle. For example, we may want to separate pre- and post-critical reflections, small from large offsets etc.

The choice of these parameters could be different for different areas.

3. Optimization of velocity model. Probably the most difficult problem associated with imaging is not how to migrate correctly given a velocity model. It is how to define this velocity model. Or, given an imperfectly focused migrated image, how should we update the velocity model to achieve better focusing?

Prestack migration could be used as a tool for velocity analysis. The underlying idea is that, if the velocity model is correct, the images obtained from different offsets should be the same independent of structure. This is a very active field of current research activity.

Consider an area that we want to image. Suppose the condition for optimal focusing stated above can be expressed as the maximization of an objective function (Denelle, Trézéguet and Tarantola (1987)). To maximize the objective function through some of the classic optimization methods we will need to calculate its gradient with respect to the velocity model. Our imaging approach should provide us with an efficient way to calculate this gradient.

Look at the situation shown schematically in Figure 18. Ray 1 has a well defined contribution to the image of the target area: this contribution is calculated by summing the data along the stacking trajectories A and spreading the result along the spreading trajectories P. If the velocity model changes, Ray 1 is going to change into Ray 2 (we keep the initial angle of the ray the same to keep the same spreading

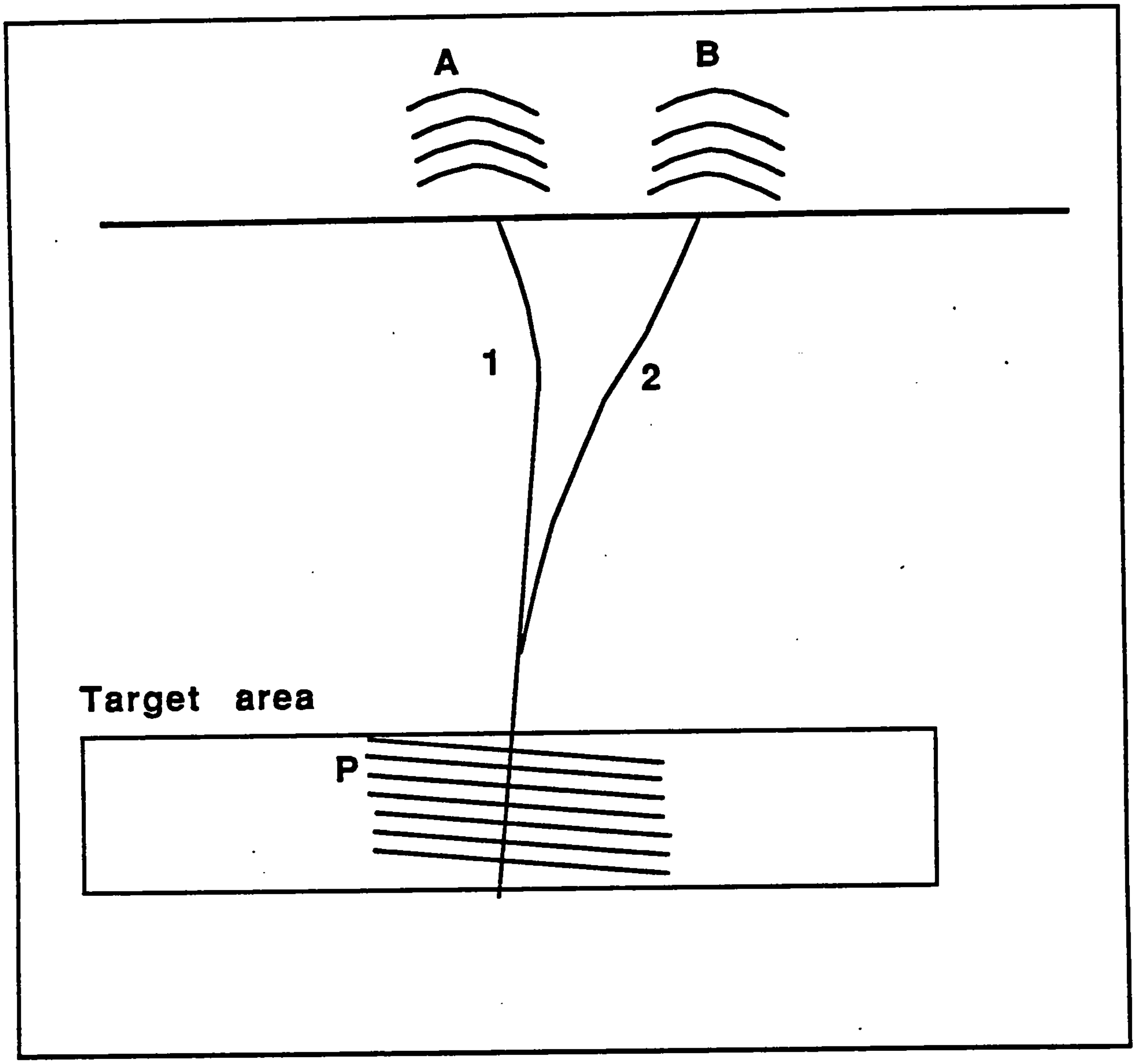


Figure 18. Change of the image of a target area due to a change in a ray.

trajectories). The stacking trajectories A are replaced by the stacking trajectories B. So, to calculate the change in the image due to the change in the model which produced this particular ray perturbation, we just have to do the following:

1. Stack along the old stacking trajectories and spread the result along P with a negative sign (to eliminate the effect of the old stacking trajectories on the image).
2. Stack along the new stacking trajectories and spread the result along P with a positive sign.

For small perturbations this procedure will essentially be the calculation of the gradient. This gradient can be described as a chain of two effects: the change of the raypath due to a perturbation in the velocity model and the change of the result of the stacking due to a perturbation in the stacking trajectories. The first part can be calculated by a slightly more expensive ray tracing. The second part is completely data dependent. It can be calculated by finite difference operations on the data.

Obviously the ability of our algorithm to image with relatively few rays is very important, since the connection between image and velocity model is established in an efficient way. It is also important that the effect of each ray on the image is well defined through the use of the Radon transform, since this allows for a well defined gradient. The ability of the algorithm to concentrate on specific dips is significant as well, since the iterative migrations could be performed efficiently, imaging a small range of dips around the dominant local dip.

4. Born inversion Born inversion uses the amplitude vs. offset information to produce estimates of the high wavenumbers of the elastic properties of the subsurface (P-wave velocity, S-wave velocity, density). In Lazaratos (1989) we showed how to perform a Kirchoff type elastic Born inversion and Gaussian beam elastic Born inversion by combining the outputs of three slightly different migrations.

5. 3-D. This is too far down the road yet. Still, it is safe to predict that algorithms that can concentrate on areas of interest and optimize their imaging would be of even greater significance in 3-D.

CONCLUSION

We presented a new migration algorithm. It is based on the synthesis of an image from its dip spectrum. The calculation of this spectrum is performed by stacking operators, defined through Gaussian beams. Some initial experiments with simple zero-offset synthetic examples have been encouraging. Still further testing with more complicated models and real data is needed, particularly for the prestack version of the algorithm. The new migration provides the ability to efficiently image regions of interest with significant flexibility in the choice of the imaging parameters. It might have important applications in migration velocity optimization and Born inversion.

REFERENCES

Berkhout, A. J., 1981, Wave field exptrapolation techniques in seismic migration, a tutorial: Geophysics, **46**, 1638-1656.

Beydoun, W. B., and Keho, T. H., 1987, The paraxial ray method: Geophysics, **52**, 1639-1653.

Beydoun, W. B., and Mendes, M., 1989, Elastic ray-Born l_2-migration/inversion: Geophysical Journal, **97**, 151-160.

Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: J. Math. Phys., **26**, 99-108.

Beylkin, G., and Burridge, R., 1987, Multiparameter inversion for acoustic and elastic media, in Expanded Abstracts of 57th Annual International SEG Meeting, Society of Exploration Geophysicists, New Orleans, 747-749.

Bleistein, N., 1987, On the imaging of reflectors in the earth: Geophysics, **52** 931-942.

Bleistein, N., Cohen, J. K., and Hagin, F. G., 1985, Computational and asymptotic aspects of velocity inversion: Geophysics, **52**, 26-36.

Carter, J. A., and Frazer, L. N., 1984, Accommodating lateral velocity changes in Kirchoff migration by means of Fermat's principle: Geophysics, **49**, 46-63.

Červený, V., 1983, Synthetic body wave seismograms for laterally varying layered structures by the Gaussian beam method: Geoph. J. R. Astr. Soc., **73**, 389-426.

Červený, V., 1985, The application of ray tracing to the propagation of shear waves in complex media: Handbook of Geophysical Exploration, Vol. 15A, Seismic Shear Waves, Helbig K. and Treitel S., eds, 1-124.

Červený, V., 1985a, Gaussian beam synthetic seismograms: Journal of Geophysics, **58**, 44-72.

Červený, V., and Pšenčik, I., 1983, Gaussian beam and paraxial ray approximation in three-dimensional elastic inhomogeneous media: Journal of Geophysics, **53**, 1-15.

Červený, V., Popov, M. M., and Pšenčik, I., 1982, Computation of wave fields in inhomogeneous media - Gaussian beam approach: Geoph. J. R. Astr. Soc., **70**, 109-128.

Claerbout, J. F., 1984, Imaging the Earth's interior: Blackwell Scientific Publications Ltd.

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

Cohen, J. K., Hagin, F. G., and Bleistein, N., 1986, Three-dimensional Born inversion with an arbitrary reference: Geophysics, **51**, 1552-1558.

Costa, C. A., Raz, S., Kosloff, D., 1989, Gaussian beam migration, in Expanded Abstracts of 59th Annual International SEG Meeting, Society of Exploration Geophysicists, Dallas, 1169-1171.

Deans, S. R, 1983, The Radon transform and some of its applications, J. Wiley and Sons, Inc.

Denelle, E., Trézéguet, D., and Tarantola, A., 1987, Nonlinear inversion by depth extrapolation in the shots-geophones domain, in Expanded Abstracts of 57th Annual International SEG Meeting, Society of Exploration Geophysicists, New Orleans, 753-754.

Devaney, A. J., 1984, Geophysical diffraction tomography: IEEE Trans. on Geoscience and Remote Sensing, **22**, 3-13.

Gabor, D., 1946, Theory of communication: J. Inst. Electr. Electron. Eng. (London), **93 III**, 429-491.

Harris, J. M., 1987, Diffraction tomography with arrays of discrete sources and receivers: IEEE Trans. on Geoscience and Remote Sensing, **25**, 448-455.

Hildebrand, S., and Carroll, R., Slant-stack depth migration: dynamic ray theory approach, in Expanded Abstracts of 59th Annual International SEG Meeting, Society of Exploration Geophysicists, Dallas, 1190-1191.

Ikelle, L. T., 1989, GRT Inversion: Practical Aspects, in Expanded Abstracts of 59th Annual International SEG Meeting, Society of Exploration Geophysicists, Dallas, 1009-1012.

Keho, T. H., and Beydoun, W. B., 1988, Paraxial ray Kirchoff migration: Geophysics, **53**, 1540-1546.

Klimeš, L., 1984, Expansion of a high frequency time-harmonic wavefield given on an initial surface into Gaussian beams: Geoph. J. R. Astr. Soc., **79**, 105-118.

Lazaratos, S., 1989, Gaussian beam inversion for cross-borehole seismic data, in Stanford Rock and Borehole Project, Rep. No. 37, I1-66.

Miller, D., Oristaglio, M., and Beylkin, G., 1987, A new slant on seismic imaging: Migration and integral geometry: Geophysics, **52**, 943-964.

Radon, J., 1917, Über die Besttimmung von Funktionen durch ihre Integralwerte längs gesisser Mannigfaltigkeiten: Ber., Verh. Sächs. Akad., **69**, 262-277.

Raz, S., 1987, Beam stacking: a generalized preprocessing technique: Geophysics, **52**, 1199-1210.

Schneider, W. A., 1978, Integral formulation for migration in two and three dimensions: Geophysics, **43**, 49-76.

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

APPENDIX A - FAR FIELD, HIGH FREQUENCY APPROXIMATION OF KIRCHOFF INTEGRAL

In this appendix we show that equation (12) is an approximate form of equation (10). In the far field and for high frequencies the wavefield $W_{\vec{p},l}(\vec{r}_R;\omega)$ can be approximated locally by a plane wave propagating at the direction of the normal to the wavefront. So, under these approximations

$$W_{\vec{p},l}(\vec{r}_R;\omega) \simeq A(\vec{r}_R)\exp[i\frac{\omega}{v(\vec{r}_R)}(x\sin\theta(\vec{r}_R) - z\cos\theta(\vec{r}_R)] \tag{29}$$

In this equation $A(\vec{r}_R)$ is a geometric spreading factor, $\theta(\vec{r}_R)$ is the angle between the vertical and the normal to the wavefront and $v(\vec{r}_R)$ the velocity at the receiver location $\vec{r}_R$. All three are assumed to be very slowly varying functions of $\vec{r}_R$.

The first term of the integral in equation (10) is

$$\int d\vec{r}_R U_s(\vec{r}_R;\omega)\frac{\partial W_{\vec{p},l}(\vec{r}_R;\omega)}{\partial z_R} \equiv \int dx U_s(x;\omega)\frac{\partial W_{\vec{p},l}(x,z=0;\omega)}{\partial z} \tag{30}$$

since the receivers are on the line $z = 0$. Using the inverse Fourier transform expression for $U_s(x;\omega)$ and interchanging the order of the integrations we can write this term as:

$$\frac{1}{2\pi}\int dk_x \int dx U_s(k_x;\omega)A(\vec{r}_R)[-i\frac{\omega}{v(\vec{r}_R)}\cos\theta(\vec{r}_R)]\exp\{i[k_x + \frac{\omega}{v(\vec{r}_R)}\sin\theta(\vec{r}_R)]x\} \tag{31}$$

Using the fact that

$$\frac{\partial U_s(x;\omega)}{\partial z} = \frac{1}{2\pi}\int dk_x U_s(k_x;\omega)[i\frac{\omega}{v(\vec{r}_R)}\sqrt{1 - k_x^2 v^2(\vec{r}_R)/\omega^2}]\exp(ik_x x) \tag{32}$$

the second term in equation (10) can be written as

$$\frac{1}{2\pi}\int dk_x \int dx U_s(k_x;\omega)A(\vec{r}_R)[i\frac{\omega}{v(\vec{r}_R)}\sqrt{1 - k_x^2 v^2(\vec{r}_R)/\omega^2}]\exp\{i[k_x + \frac{\omega}{v(\vec{r}_R)}\sin\theta(\vec{r}_R)]x\} \tag{33}$$

It can be easily shown that the stationary phase approximations of the expressions in equations (31) and (33) are the same with opposite signs. So, the Kirchoff integral of equation (10) can be approximated with twice the quantity

$$\int d\vec{r}_R U_s(\vec{r}_R;\omega)\frac{\partial W_{\vec{p},l}(\vec{r}_R;\omega)}{\partial z_R} \tag{34}$$

where, according to equation (29)

$$\frac{\partial W_{\vec{p},l}(\vec{r}_R;\omega)}{\partial z_R} = -i\omega\frac{\cos\theta(\vec{r}_R)}{v(\vec{r}_R)}W_{\vec{p},l}(\vec{r}_R;\omega) \tag{35}$$

Substituting this last expression into equation (34) we end up with the right-hand side integral in equation (12).

APPENDIX B - DETERMINATION OF THE COEFFICIENTS FOR THE GAUSSIAN BEAM DECOMPOSITION

We demonstrate here how to calculate the coefficients for the expansion of $f(\vec{r})$ into the basis defined by the Gaussian beam wavefields $B_{i,\vec{p},l}(\vec{r})$. The expansion is of the form

$$f'(\vec{r}) = \sum_{i}\sum_{\vec{p}}\sum_{l} a_{i,\vec{p},l}B_{i,\vec{p},l}(\vec{r}) \tag{36}$$

where

$$f(\vec{r}) = R(\vec{r}) * f'(\vec{r}) \tag{37}$$

and $R(\vec{r})$ is the two-dimensional rho-filter. We determine the coefficients $a_{i,\vec{p},l}$ by minimizing the error

$$\int d\vec{r}\,|f'(\vec{r}) - \sum_{i}\sum_{\vec{p}}\sum_{l} a_{i,\vec{p},l}B_{i,\vec{p},l}(\vec{r})|^2 \tag{38}$$

Minimizing this objective function leads to a system of equations

$$Ax = y \tag{39}$$

where x is the vector of unknown coefficients $a_{i,\vec{p},l}$ and y is the vector consisting of the elements

$$\int d\vec{r} f'(\vec{r})B_{i,\vec{p},l}(\vec{r}) \tag{40}$$

This quantity is the value of the convolution of the spatial functions $f'(\vec{r})$ and $B_{i,\vec{p},l}(-\vec{r})$ at $\vec{r} = 0$. The function $f'(\vec{r})$ is the convolution of $f(\vec{r})$ and the inverse of the rho-filter. Instead of applying the inverse rho-filter on $f(\vec{r})$, we can apply it on $B_{i,\vec{p},l}(\vec{r})$. So the elements of y can be written as

$$\int d\vec{r} f(\vec{r})B'_{i,\vec{p},l}(\vec{r}) \tag{41}$$

where

$$B_{i,\vec{p},l}(\vec{r}) = R(\vec{r}) * B'_{i,\vec{p},l}(\vec{r}) \tag{42}$$

The elements of the matrix A are

$$\int d\vec{r} B_{i,\vec{p},l}(\vec{r})B_{i',\vec{p}',l'}(\vec{r}) \tag{43}$$

The matrix A is close to being diagonal. Keeping only the diagonal terms we simply have

$$a_{i,\vec{p},l} = c^{-1}\int d\vec{r} f(\vec{r})B'_{i,\vec{p},l}(\vec{r}) \tag{44}$$

where the constant c is

$$c = \int d\vec{r} B^2_{i,\vec{p},l}(\vec{r}) \tag{45}$$

So the final result is

$$f(\vec{r}) = R(\vec{r}) * \sum_{i}\sum_{\vec{p}}\sum_{l} B_{i,\vec{p},l}(\vec{r})\int d\vec{r} f(\vec{r})B'_{i,\vec{p},l}(\vec{r}) \tag{46}$$

The constant c is omitted, since it produces no effect on the migrated image.

PAPER H

TRAVELTIME INVERSION WITH CONSTRAINTS

Reinaldo J. Michelena
Seismic Tomography Project

ABSTRACT

In this paper, I present a new method for using any given slowness image (prior information) in tomographic traveltime inversion. In contrast with the common practice of using the given image just as starting model, the method I propose uses that image as data points (constraints) simultaneously with the traveltimes. The result is that in the iterative inversion both the misfit in the traveltimes and the misfit between the given image and the updated one are minimized. Minimizing the misfit with the prior model is important when that model contains information from independent data. I show in this paper that this minimization helps to reduce the null space of the problem.

INTRODUCTION

An old problem in geophysics has been the integration of different data sets into the same inverse process with the aim of producing an image consistent with all the information available. It is well known that different methodologies can contribute with complementary insights in the solution of one particular problem, by eliminating the particular ambiguities of each method. For example, Devaney(1984) and Wu and Toksöz (1987) show that in the problem of diffraction tomography the spatial Fourier components of the object to be reconstructed are sampled differently depending upon the geometry used (cross-well, surface seismic or VSP). In exploration with electrical methods is also well known that DC data refers to shallow depths whereas MT data refers to depths of maybe several kilometers (Jupp and Vozoff, 1975). Sonic logs provide accurate depths of the layers boundaries, information that might not be present in surface seismic data.

The inverse problem usually consists on finding some model parameters that minimize the misfit between real and synthetic data. Any other source of information different to the given data set and used in the solution of the problem is called prior information.

A common practice when dealing with data sets that belong to different experiments is to invert them independently. This might be convenient for several reasons. One of them is that we don't have to worry about how to model the coupling between experiments. Another is that we don't have to worry about weights applied to each data set. However, for understanding the different results simultaneously, the interpreter has to introduce his/her own judgement (another form of prior information) to keep from each inversion the most relevant features and discard from them the "absurd" features caused essentially by the limited nature of each data set.

The most common prior information used by the interpreter when analyzing some inversion results is the smoothness of the solution. The interpreter tries to eliminate "noise" in the image and focus his/her attention on the bigger features. This assumption of good behavior of the solution can also be used for inverting the data, and this is the most common way of dealing with the nonuniqueness of the problem. Smoothness is introduced in the problem when we want some filtered version of the unknown to vanish and at the same time it reproduces the data. This is the equivalent of choosing a model norm (Claerbout, 1976). Damped least squares is an example of this technique where the filter used is the identity matrix. In general, the filter is a roughening operator and because of that the output is smooth. A disadvantage of the procedure is that, besides the subjectivity involved in the selection of the filter, the solution can be very sensitive to the type of filter applied.

Prior information is also used when we build an initial model for starting an iterative process of minimization. The model can be obtained from independent information (Lines et al., 1988) or it can be simply a guess. The objective function in the minimization is usually the misfit in the data and the speed of convergency depends not only on how close to the minimum the starting model is, but also on weather the problem is well posed or not. If the initial model contains useful information (result of an independent inversion), it won't be necessarily reproduced in the final result that minimizes the misfit in the data. This is because the inversion does not control the misfit between the calculated and the given model. Tarantola and Vallete (1982) and Tarantola (1984) propose to minimize both the misfit in the data and the misfit between the starting model and the updated one. Both misfits are weighted by the data and model covariance matrices. Their formulation reduces to those given by Franklin (1970) and Jackson (1979) for the linear case. Tarantola and Vallete (1982) show that minimizing the misfit in the model is the right condition to be applied when a nonlinear problem is going to be solved by iteration of linearized problems, in contrast with the common practice of minimizing the norm between successive models. I will show later in the paper that minimizing the misfit between the initial model and the updated one also helps to reduce the the null space of the problem.

The error in the model is calculated through the expression $(\mathbf{m}-\mathbf{m}_0)^T C_{\mathbf{m}}^{-1}(\mathbf{m}-\mathbf{m}_0)$ where $\mathbf{m}$ is the calculated model, $\mathbf{m}_0$ is the starting model, and $C_{\mathbf{m}}$ is the model covariance matrix. Examples of covariances functions used for calculating the model covariance matrix are

$$C_{\mathbf{m}}(\mathbf{m},\mathbf{m}') = \sigma^2 \exp\{-\frac{\|\mathbf{m}-\mathbf{m}'\|^2}{2L^2}\}$$

or

$$C_{\mathbf{m}}(\mathbf{m},\mathbf{m}') = \sigma^2 \exp\{-\frac{\|\mathbf{m}-\mathbf{m}'\|}{2L}\}.$$

Nolet (1987) points out that $C_{\mathbf{m}}$ is an smoothing operator and then $C_{\mathbf{m}}^{-1}$ is then a roughening operator. This means that in the practice we just have to select the appropriate filter (choose a model norm) and minimize the power of the output (Claerbout, 1976).

The previous techniques introduce the prior information as probability distributions in data and model spaces. Information introduced in this way is called soft bounds. Although these techniques are completely general, they have the major problem that they cannot be applied to large scale problems such as tomography since they involve the inversion of large matrices. The size of the matrices is large because the covariance functions implicitly assume that the model is described as a superposition of small cells (square pixels), and in tomography usually many cells are needed for describing the model correctly.

Several attempts have been made to reduce the number of model parameters in the tomographic inversion. Most of them are based on prior assumptions about the structure of the model. For example, Chiu and Stewart (1987) modeled the earth by continuous curved boundaries (of approximately known position) separating regions of constant velocity. Van Trier (1988) uses two dimensional functions for describing the boundaries of the expected regions in the model and also allows smooth velocity variations within each region. Harlan (1989) describes the model as a superposition of smooth functions (Gaussians) since it is well known that fine details cannot be recovered from traveltimes measurements.

The discretization of the model based on prior knowledge about it certainly helps to reduce the number of model parameters in contrast with the discretization of the model into small cells that assumes no prior knowledge. The problem in the former case is that we can resolve only the features that the parametrization allows to be resolved. Any other information contained in the data is lost. In the second case, we resolve the information contained in the data but at the expense of many model parameters, because no prior information has been used.

From the examples just mentioned we conclude that a problem of major importance in tomographic inversion is the parametrization of the model. Techniques that introduce prior information as probability distributions in data and model space are not practical because a large number of model parameters is needed. Usually the number of model parameters is much larger than the amount of data available. When the number of model parameters is reduced using prior information about the model, some information present in the data might not be used. This problem raises the question about a parametrization of the model that allows the use of prior information in such a way that the information contained in the data is still completely recovered without having to use a large amount of model parameters, such as square pixels. Carrion (1989) and Carrion et al. (1988) developed a technique in which the number of cells can be made arbitrarily large but the size of the problem remains constant (and equal to the number of available rays). Prior constraints in terms of soft and hard bounds can be incorporated.

In this paper, I propose a different approach to solve the problem just mentioned. It is based on image reconstruction in Hilbert spaces. I'll show that a parametrization of the medium with the desired characteristics can be obtained directly from the theoretical expression that describes the generation of the measurements. Prior images have to be converted into data points before using them for constraining the inversion. I'll also show that the concept of consistency between forward modeling and parametrization keeps the number of model parameters equal to the number of data points and the result is an estimate of slowness that reproduces both data and prior images.

With the methodology proposed in this paper it will be possible to combine consistently different prior images or data that belong to different experiments related with slowness measurements. The result should be an improved version of all the previous results obtained by analyzing independently each data set.

CONVENTIONAL CROSS-WELL TOMOGRAPHIC RECONSTRUCTION

The traveltime along a ray path in a medium whose slowness is $S(x,y)$ is calculated with the expression

$$t_m = \int_{l_m} S(x,y) dl_m \qquad m = 1, ..., N, \tag{1}$$

where dl_m is the incremental distance along the ray path l_m. If the variations in the slowness are small, we can consider that the problem is linear (like X-rays tomography). For large variations in slowness, the problem becomes highly nonlinear because the unknown ray path depends on the slowness. This problem is usually solved by a sequence of linearized steps (using Fermat's principle) about some starting model.

The estimation of the slowness $\tilde{S}(x,y)$ usually starts with the discretization of the model in square orthogonal pixels (McMechan, 1983):

$$\tilde{S}(x,y) = \sum_{n=1}^{M} b_n R_n(x,y). \tag{2}$$

where

$$R_i(x,y) = \begin{cases} 1 & \text{if } (x,y) \text{ is in the pixel } i \qquad (i = 1, ..., M) \\ 0 & \text{otherwise} \end{cases}$$

Let's assume for simplicity that the problem is linear. Integrating both sides of Eqn. (2) along the ray l_m we get

$$t_m = \sum_{n=1}^{M} M_{mn} b_n \tag{3}$$

where

$$M_{mn} = \int R_n dl_m. \tag{4}$$

The system of equations (3) is then solved by use of the SVD technique or iterative procedures such as ART, SIRT or conjugate gradients. The coefficients b_n are substituted into (2) to obtain the slowness estimate.

The Problem with the Parametrization

Many different basis functions besides square pixels $R_n(x, y)$ (Eqn. (2)) can be used for discretizing the slowness model. The discretization in square pixels is just one way that is convenient for displaying purposes. Once the system (3) is solved the displaying of the image is immediate after a simple rearrange of the indexes. The selection of the size of the pixels (and consequently the number of them) is also arbitrary, but it is usually based on "resolution" considerations. From Eqn. (3), it is easy to conclude that for reproducing each measurement t_m accurately, the size of the pixels has to be made infinitely small, which means that in the practice we should use a large amount of small pixels. Large amounts of pixels means large systems of equations that are difficult to handle. The common solution to this problem consists simply in discretize the image more coarsely. Thus, a tradeoff between resolution and the computational effort is required.

Discretizing the image coarsely may produce some problems. As I said before, when the problem is nonlinear, it is usually solved as a sequence of linearized steps. The common goal is to minimize the misfit between real and synthetic data. The discretization of the model into square pixels can introduce by itself a misfit different to the one expected when the problem is linearized. This is because the matrix M in Eqn. (4) does not describe the measurements correctly unless many pixels are used. The problems arise because in one hand we try to minimize the misfit between real and calculated data and, in the other hand, the modeling operator M is incorrect. Because of this problem the convergency in the linear case won't be reached in one iteration as expected.

Discretizing the model into a large amount of square pixels means essentially that we don't know anything about it. Because the number of model parameters is much larger than the number of data points, we need to introduce in the problem extra information to resolve all of them. Some information may be introduced with the model covariance matrix but it translates principally into smoothing. Other discretizations might be more convenient in some situations since they can reflect with fewer parameters some prior knowledge about the model (smoothness, structure, etc.). However, most of them fail when used to build a modeling operator that reproduces the measurements correctly.

Therefore, using many or few model parameters can create problems. I will show in the next section that the formalism of reconstruction in Hilbert spaces suggest a discretization of the model where the number of degrees of freedom is same as the number of data points and all the parameters can be resolved without using smoothness conditions. The discretization is suggested by the nature of the measurements.

RECONSTRUCTION FROM PROJECTIONS IN HILBERT SPACES

In this section, I will review the fundamentals of reconstruction in Hilbert spaces. Let the original image $S(x,y)$ that we want to estimate be an element of the Hilbert space H and suppose that the data available about $S(x,y)$ is in the form of inner products with a finite set of functions $\phi_m(x,y) \in H$

$$d_m = < S(x,y), \beta_m(x,y) >= \int_\Omega S(x,y)\beta_m(x,y)dxdy \qquad m = 1,...,N. \tag{5}$$

This information can be interpreted also as projections of the unknown function $S(x,y)$ in the set of "sampling" functions $\beta_m(x,y)$. It can be shown (Darling et al., 1983, Michelena and Harris, 1990) that the minimum norm estimate $\tilde{S}_1(x,y)$ of $S(x,y)$ is

$$\tilde{S}_1(x,y) = \sum_{n=1}^{N} a_n\beta_n(x,y), \tag{6}$$

where the coefficients a_n are calculated from the system of equations

$$d_m = \sum_{n=1}^{N} a_n < \beta_n(x,y), \beta_m(x,y) > \qquad m = 1,...,N. \tag{7}$$

The name minimum norm estimate comes from the fact that the function $\tilde{S}_1(x,y)$ minimizes the expression

$$\|S(x,y) - \sum_{n=1}^{M} a_n\alpha_n(x,y)\|, \tag{8}$$

where $\alpha_n(x,y)$ and M are arbitrary. Although other choices of $\alpha_n(x,y)$ also minimizes (8), the particular choice $\alpha_n(x,y) = \beta_n(x,y)$ is a convenient one since in that case the independent term d_m in Eqn. (7) is form directly by the data points.

The estimate (6) is unique and obviously depends on the definition of the inner product in the Hilbert space.

The inner products $< \beta_n(x,y), \beta_m(x,y) >$ estimate the correlation between two different sampling functions. If they do not overlap, the inner product is zero (orthogonal sampling functions) and we can say that the measurements are uncorrelated. Strictly speaking, the number $< \beta_n(x,y), \beta_m(x,y) >$ is the cross correlation at the origin between the two functions $\beta_n(x,y)$ and $\beta_m(x,y)$. The larger the inner products, the more the overlap between the sampling functions.

The formulation of reconstruction in Hilbert spaces is completely general and can be applied directly any time the measurements are expressed in a way like (5) (the sampling functions $\beta_n(x,y)$ have to be square integrable in the support of the unknown $S(x,y)$). Fourier analysis is an example of reconstruction in Hilbert spaces (Stakgold, 1979). Another example is given by Michelena and Harris (1990). They show that in the problem of traveltime tomography, if the functions $\beta_m(x)$ are interpreted as the beam paths, the traveltimes can be interpreted as projections of the slowness along the beam paths. Another example of the application of this technique is in the problem of diffraction tomography reconstruction with constraints. This is shown in another paper in this volume.

Weighted Inner Products

We can define another Hilbert space H_2 with inner product

$$< S(x,y), \psi_n(x,y) >_{S_0} = \int_\Omega S(x,y)\psi_n(x,y)\frac{1}{S_0(x,y)}, dxdy. \tag{9}$$

where $S_0(x,y)$ is a weighting function that is nonzero in the support of $S(x,y)$, and $\psi_n(x,y)$ is a set of functions to be defined according to the generations of the measurements (5). If we multiply and divide the integrand in (5) by $S_0(x,y)$ we can say that the data can also be described as

$$d_m = < S(x,y), \psi_m(x,y) >_{S_0}, \tag{10}$$

where $\psi_m(x,y) = \beta_m(x,y)S_0(x,y)$.

According to (6) the minimum norm estimate of $S(x,y)$ is

$$\tilde{S}_2(x,y) = \sum_{n=1}^{N} a_n\psi_n(x,y) = S_0(x,y)\sum_{n=1}^{N} a_n\beta_n(x,y). \tag{11}$$

The weighting function $S_0(x,y)$ can be chosen to represent broad expected features of $S(x,y)$. In this way, $\tilde{S}_2(x,y)$ not only reproduces the data but also important features present in $S_0(x,y)$. The coefficients a_n in Eqn. (11) satisfy the system of equations

$$d_m = < \psi_m(x,y), \psi_n(x,y) >_{S_0} a_n \qquad n,m = 1,...,N \tag{12}$$

where

$$< \psi_m(x,y), \psi_n(x,y) >_{S_0} = \int_\Omega \beta_n(x,y)\beta_m(x,y)S_0(x,y)dxdy. \tag{13}$$

This technique has been applied successfully in the past by Hall et al (1982) to introduce prior information in medical imaging problems and by Michette et al. (1984) in the problem of enhancing the resolution of synthetic seismograms. In the appendix A I show the basic equations that result when this technique is applied to diffraction tomography inversion. The main conclusion of this appendix is that the weighting function contributes to reduce the null space of the problem by increasing the area of coverage of the unknown in the frequency domain.

I am going to illustrate now how this technique can be used in the problem of traveltime tomographic inversion. As I said before, in this problem the traveltimes (d_m in the previous equations) can be interpreted as the projections of the slowness along the beam paths $\phi_n(x, y)$. In the previous equations we have to identify $S(x, y)$ as the real slowness model and $S_0(x, y)$ as a given prior image. In the Fig. 3 shows an example of the application of this technique in the reconstruction of the anomaly of Fig. 1. The prior image used in this case is shown in Fig. 2. The results of the inversion when no prior information is used is shown in Fig. 4, which is taken from Michelena and Harris (1990). They show that the norm of the null space for this reconstruction is $||f_2|| = 2.397$ and the mean absolute error is 2.0 10^{-3}. For the prior image $||f_2|| = 2.188$ and the mean error is 1.6 10^{-3}. When prior information is used (Fig. 3), these two quantities are reduced to 1.582 and 0.8 10^{-3} respectively. As expected, combining both types of information helps to reduce the null space of the problem. The starting model used in both cases is the same. Figures 5 and 6 represent the absolute value of the difference between the original model (Fig. 1) and the inversion with and without prior information (figures 3 and 4 respectively). Note that there are more zeros in Fig. 5 which means that the inversion with constraints reproduces better the values of the original image.

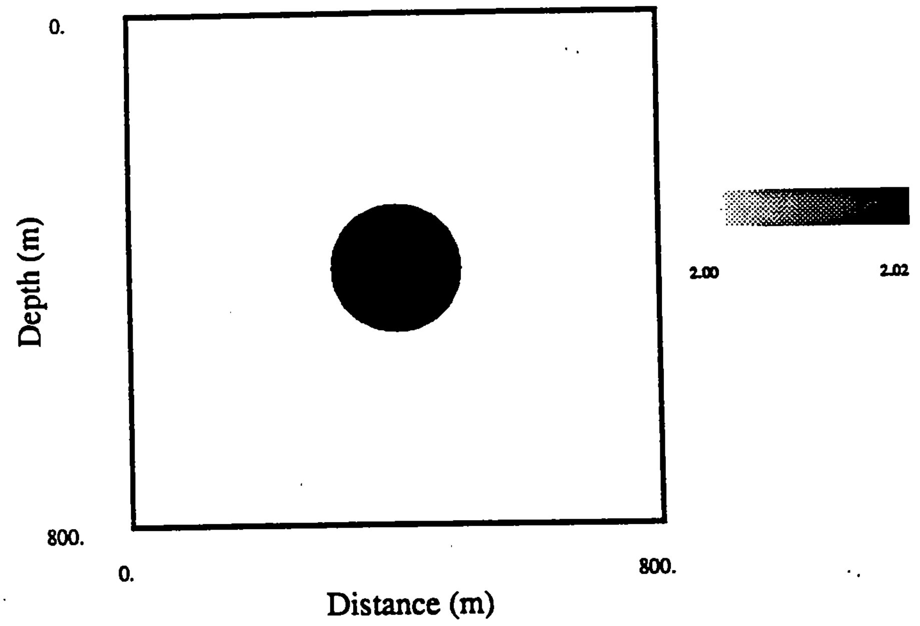


Figure 1: Slowness model. The slowness of the disk of 2.02 and the slowness of the background is 2.00.

The slowness in the prior image might be better constrained in some areas than in others. In this case we want the reconstructed model to reproduce that information. Fig. 3 shows that this technique fails in this case, because it only reproduces the "important" features in the present in the prior image but not the details (that might not be necessarily noise). The reason of this failure is that the term $\sum_{n=1}^{N} a_n \beta_n(x, y)$ in Eqn. (11) cannot be equal to one in limited view problems.

Both sources of information (prior slowness and traveltimes) are used in a different

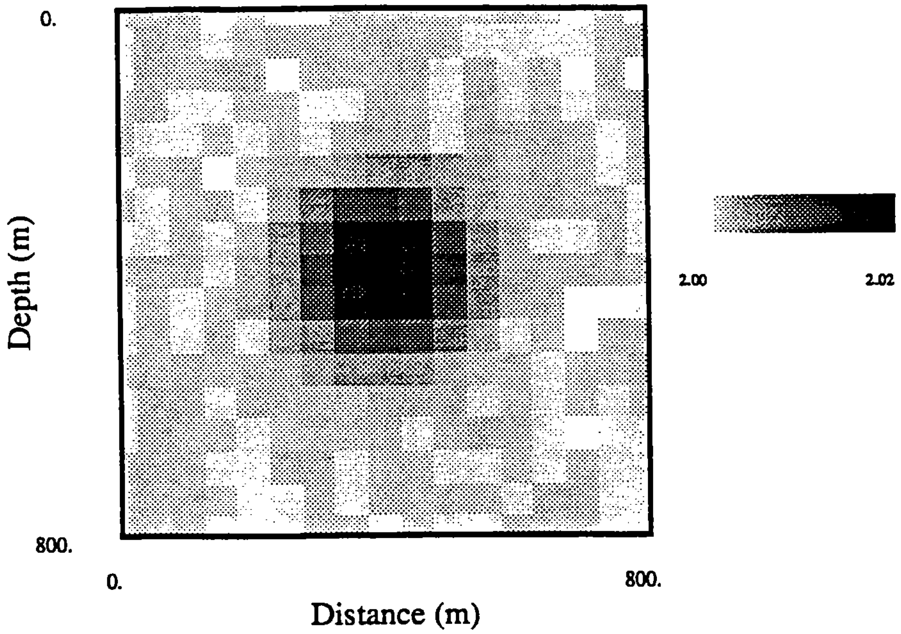


Figure 2: Slowness model used as prior information for the inversion of Fig. 3.

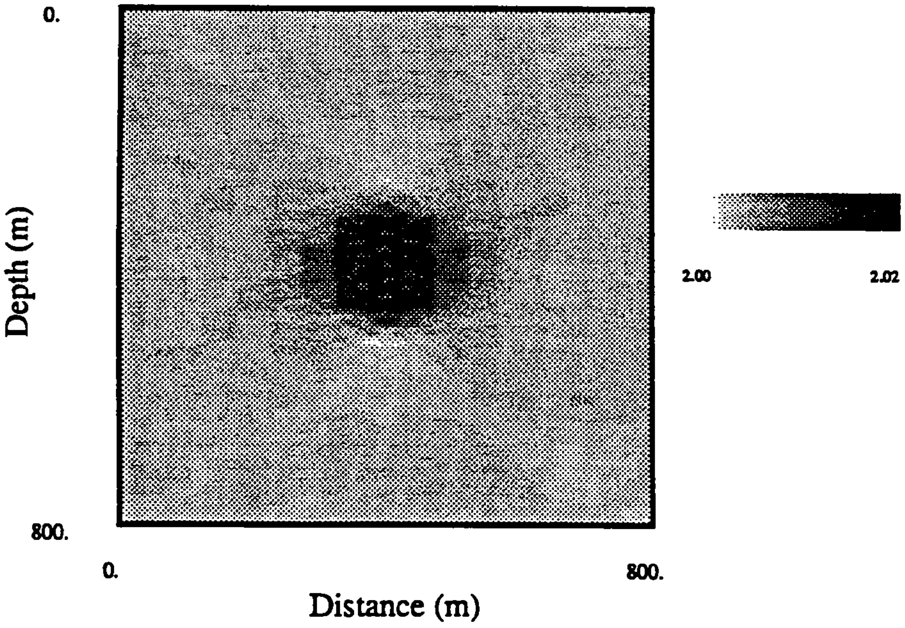


Figure 3: Natural pixels based inversion, using Fig. 2 as prior image.

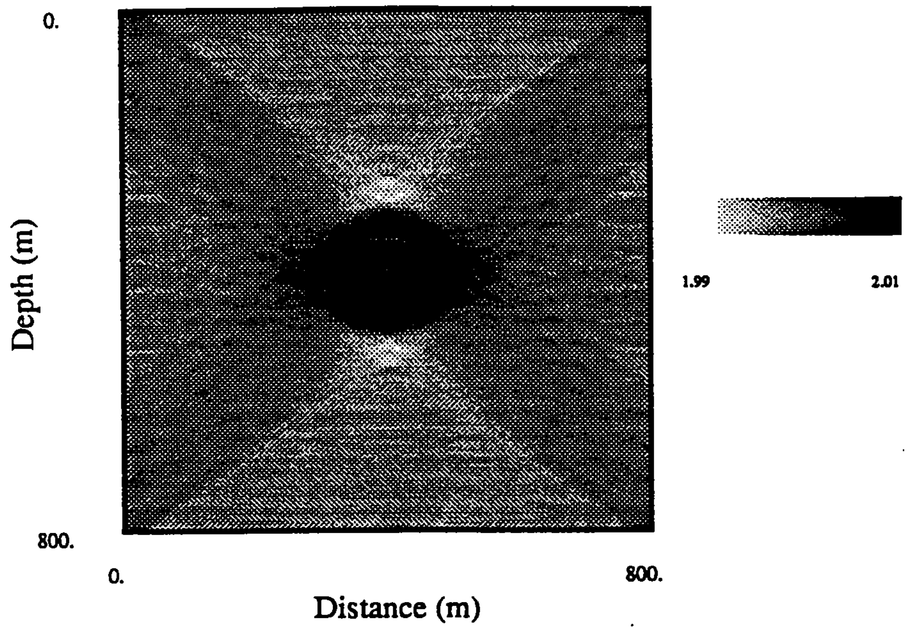


Figure 4: Natural pixels based inversion without using prior information.

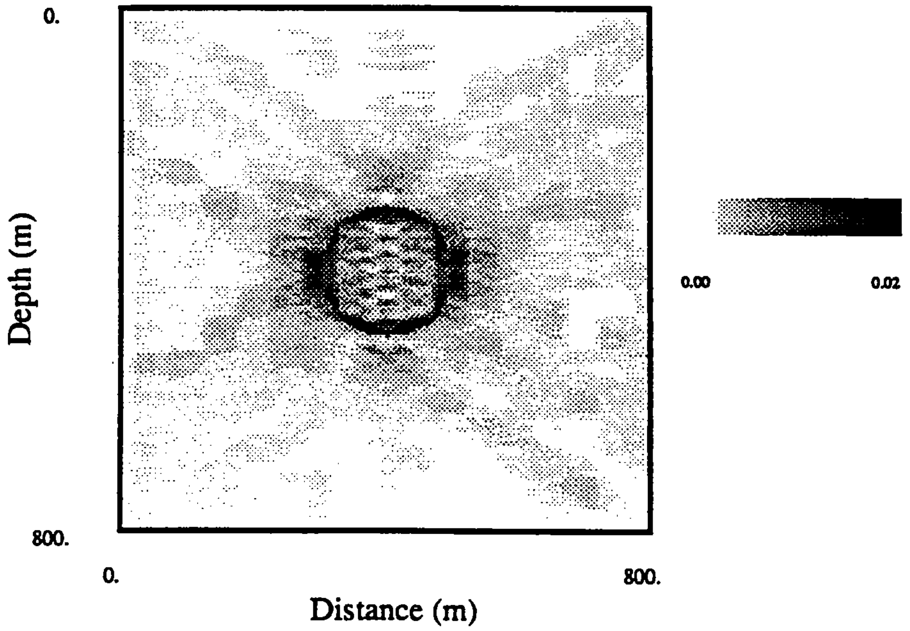


Figure 5: Error in the inversion when prior information is used

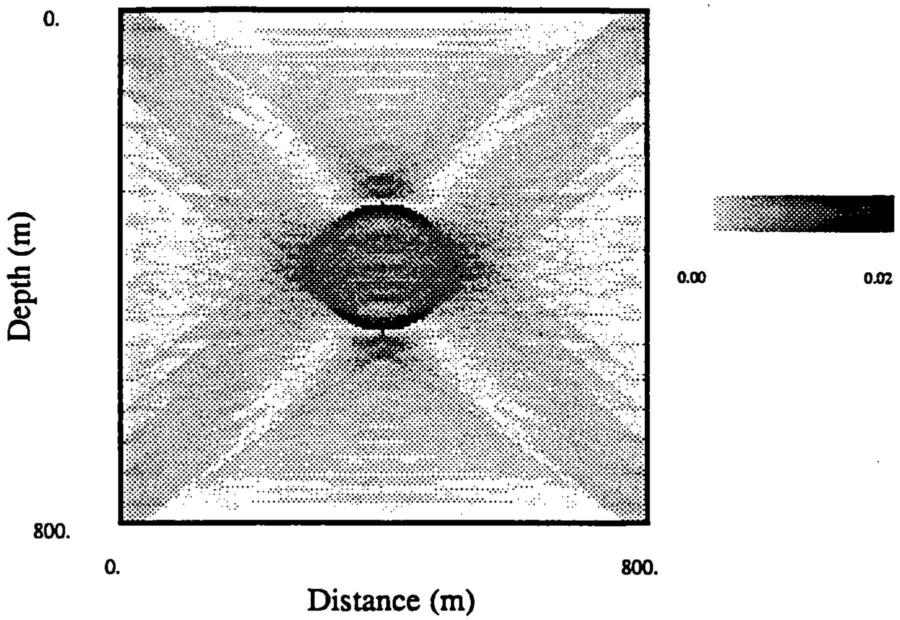


Figure 6: Error in the inversion when no prior information is used

way in this technique. The traveltimes are considered as data points in Eqn. (12), whereas the prior image is a weight the affects the calculation of the matrix coefficients and the final display of the image. If we want the prior model to be reproduced in the reconstruction, it is necessary to consider it as *data* simultaneously with the traveltimes, not as a weighting function. In the next section, I'll study the way of doing this.

Hard Constraints in the Form of Sharp Boundaries

For simplicity, let's assume that the prior image can be decomposed into M *non overlapping* regions $R_n(x,y)$ (layers) of average slowness s_n

$$S_0(x,y) = \sum_{n=N+1}^{N+M} s_n R_n(x,y) \tag{14}$$

where

$$R_i(x,y) = \begin{cases} 1/A_i & \text{if } (x,y) \text{ is in the pixel } i \qquad (i = 1,...,N) \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

A_i is the area is the i^{th} region. The reason why the index n in the summation is shifted N positions will be evident later.

Michelena and Harris (1990) show that this is the minimum norm estimate when the measurements are given by

$$s_n = \int_\Omega S(x,y) R_n(x,y) dx dy.$$

Remember that the traveltimes are projections of the slowness along the beam paths. In the same way, the average slowness are the projections of the slowness in the basis set defined by the functions $R_n(x,y)$. Because both kinds of information are just projections, we can treat them simultaneously if we define a new basis set $T_n(x,y)$ in the following way

$$T_n(x,y) \;=\; \begin{cases} \phi_n(x,y) & \text{if n=1,...,N} \\ R_n(x,y) & \text{if n=N+1,...,N+M} \end{cases} \tag{16}$$

The minimum norm estimate of $S(x,y)$ from the projections over the basis set $T_n(x,y)$ is

$$\tilde{S}(x,y) \;=\; \sum_{n=1}^{N+M} c_n T_n(x,y) = \sum_{n=1}^{N} c_n \phi_n(x,y) \;+\; \sum_{n=N+1}^{N+M} c_n R_n(x,y) \tag{17}$$

which becomes after reordering the indexes

$$\tilde{S}(x,y) \;= \sum_{n=1}^{N} a_n \phi_n(x,y) \;+\; \sum_{n=1}^{M} b_n R_n(x,y). \tag{18}$$

This is simply the superposition of the slowness estimate from traveltime measurements and the slowness estimate from average slowness indirect measurements.

The coefficients a_n and b_n are determined simultaneously with the following system of equations

$$\begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} <\Phi,\Phi> & <R,\Phi> \\ <R,\Phi>^T & <R,R> \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \tag{19}$$

where

$$<\Phi,\Phi>_{ij} = \int \phi_i(x,y)\phi_j(x,y)dxdy$$

$$<R,\Phi>_{ij} = \int R_i(x,y)\phi_j(x,y)dxdy$$

$$<R,R>_{ij} = 1/A_i \;\delta_{ij}$$

and

$$\mathbf{t} = \begin{pmatrix} t_1 \\ \vdots \\ t_N \end{pmatrix} \quad \mathbf{s} = \begin{pmatrix} s_1 \\ \vdots \\ s_M \end{pmatrix} \quad \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_M \end{pmatrix}.$$

The matrix in the system of equations (19) estimates the correlations between the different regions sampled by the basis set $T_n(x,y)$, which is a way of taking into account the redundancy between the different kinds of information. If the prior image is completely contained in the null space of the beam paths, all the terms $<R,\Phi>_{ij}$ vanish and the two sources of information can be inverted independently. This is equivalent to the minimization of the expression

$$\| <\Phi,\Phi> \mathbf{a} - \mathbf{t}\|^2 + \| <R,R> \mathbf{b} - \mathbf{s}\|^2 \tag{20}$$

assuming that **a** and **b** are uncorrelated. In general this is not the case. In the next section I'll study in more detail the implications of treating one given image as a starting independent model in an iterative procedure (sequential inversion) or as constraint in the same procedure (joint inversion).

Sequential Inversion vs. Joint Inversion

Traveltime tomography is essentially a nonlinear problem because the beams paths depend on the slowness model to be determined. The problem is usually solved in a sequence of linearized steps where, starting from an initial model, the forward modeled data is compared with the real measurements. The starting model is modified and the process is repeated until the misfit between the real and calculated data is below a certain level (or until we like the image obtained). This is an example of sequential inversion, since the starting model has not been used as data. Considering the starting model as data means that the projections of the given model in the basis set $\{R_n(x,y)\}$ (average slowness) are treated in the same way as the projections of the slowness in the beam paths (traveltimes).

Let's say that the starting model is $S_0(x,y)$. The correction that has to be made to that model in order to reproduce the real slowness model is

$$\Delta S(x,y) = S(x,y) - S_0(x,y) \tag{21}$$

In the sequential inversion, $\Delta S(x,y)$ is expressed as a linear combination of natural pixels

$$\Delta S(x,y) \;=\; \sum_{n=1}^{N} a_n \phi_n(x,y) + f_{null}(x,y), \tag{22}$$

where the function $f_{null}(x,y)$ describes the null space of the traveltime anomalies

$$\int_\Omega f_{null}(x,y)\phi_n(x,y)dxdy = 0. \tag{23}$$

In other words, $f_{null}(x,y)$ represents regions in the model that produce no traveltime anomalies.

If the starting model is used as data also (joint inversion), ΔS is expanded as follows

$$\Delta S(x,y) \;=\; \sum_{n=1}^{N} a_n \phi_n(x,y) + \sum_{n=1}^{M} b_n R_n(x,y) + f'_{null}(x,y). \tag{24}$$

Note that the null space is different in both formulations ($f_{null}(x,y) \neq f'_{null}(x,y)$). The reason is that if any of the functions $\{R_n(x,y)\}$ cannot be expanded in terms of $\phi_n(x,y)$, we can say that the expression (24) contains terms that belong to the null space $f_{null}(x,y)$ in Eqn. (22). Therefore, the main difference between the two ways of using $S_0(x,y)$ is that using it as data effectively contributes to reduce the null space of the problem. The coefficients a_n and b_n are the solution of the following system of equations

$$\begin{pmatrix} \Delta \mathrm{t} \\ 0 \end{pmatrix} = \begin{pmatrix} <\Phi,\Phi> & <R,\Phi> \\ <R,\Phi>^T & <R,R> \end{pmatrix} \begin{pmatrix} \mathrm{a} \\ \mathrm{b} \end{pmatrix} \tag{25}$$

The zero in the independent term of the previous equation is because the projections of the perturbation in the regions $R_n(x,y)$ have to be zero. In the linearized inversion based on Eqn. (24) the traveltimes has to be matched while the mismatch with the original one remains always zero. Remember that what remains zero is the *mean* slowness mismatch in each region. The final output is consistent with both sets of projections, average slowness and traveltimes.

If we have no information about the velocities in the regions (layers) $R_n(x,y)$ but we know the boundaries, the Eqn. (24) remains valid, but we have to eliminate the second row in Eqn. (25). The problem is not square any more because of the incomplete information. A typical situation where only the boundaries are known is when the given image is a migrated section.

The reader may wonder about what happen if the prior image is not a simple superposition of independent regions, but a superposition of non orthogonal functions, such as Gaussians. The output of the inversion of independent data may be represented in this form. In other cases we may have images formed as superposition of *many* independent regions (square pixels, for example) and we might want to reduce the number of parameters involved in the description of that image before using it as prior information for inverting another data set. This two problems will be the topic of the next section.

Soft Constraints in the Form of Fuzzy Boundaries

In a previous section, I explained a way of using any image as prior information considering it as a weighting function in the definition of the inner product that describes the generation of the measurements. The main disadvantage of this approach is that, although the null space is reduced, the reconstructed image reproduce only the general features contained in the prior one. Another way of using that first image is as starting model of an iterative procedure. In this case the prior image does not contribute to reduce the null space of the problem. The conclusion of the previous section is that the way to avoid this problem is to consider the prior image also as data. Let's see how we can convert any given image into data, regardless it is expressed as a combination of independent regions or not.

Let's say that we have an image represented as a linear combination of functions $\{C_n(x,y)\}$

$$S_0(x,y) = \sum_{n=1}^{L} c_n C_n(x,y). \tag{26}$$

The representation of $S_0(x,y)$ in another basis set $D_n(x,y)$ is

$$\tilde{S}_0(x,y) = \sum_{n=1}^{K} d_n D_n(x,y) \approx S_0(x,y) \tag{27}$$

Multiplying both sides of Eqn. (27) by $D_m(x,y)$ and integrating, we obtain the system of equations for the coefficients d_n

$$\int S_0(x,y)D_m(x,y)dxdy \;=\; s'_m \;=\; \sum_{n=1}^{K} d_n \int D_n(x,y)D_m(x,y)dxdy \qquad m = 1,...,K \tag{28}$$

Therefore, the procedure for representing one given image in a different basis set is the following :

- Compute the projections of the given image in the desired basis set (data generation, left hand side of Eqn. (28)).
- Compute the inner products between all the possible combinations of functions in the new basis set (generation of the matrix).
- Solve the system of equations (28).

It is interesting that is we assume that $S_0(x,y)$ is the real slowness and if $D_m(x,y) = \phi_m(x,y) = (beam\ path)_m$, our problem of changing basis turns into the problem of traveltime tomography. The change of basis made in traveltime tomography is from the continuous representation of the real slowness to the discrete basis defined by the beam paths. Of course, a considerable amount of information is lost in that process. This tell us that in any process of changing basis information is lost if the new basis set does not expand the same space expanded by the original one.

If we want to use any given image as constraint in traveltime inversion we have to decide first the basis where we are going to represent it. This selection can be as arbitrary as the selection of the basis function explained at the beginning of the paper. The difference is that now the selected basis set is also used for calculating the data. Let's call this basis set $D_n(x,y)$. Then we have to add one more term to the summation (18) and the estimate of the slowness is now represented as

$$\tilde{S}(x,y) = \sum_{n=1}^{N} a_n\phi_n(x,y) + \sum_{n=1}^{M} a_n R_n(x,y) + \sum_{n=1}^{K} d_n D_n(x,y). \tag{29}$$

The unknown coefficients a_n, b_n and d_n satisfy

$$\begin{pmatrix} \mathbf{t} \\ \mathbf{s} \\ \mathbf{s}' \end{pmatrix} = \begin{pmatrix} <\Phi,\Phi> & <R,\Phi> & <D,\Phi> \\ <R,\Phi>^T & <R,R> & <D,R> \\ <D,\Phi>^T & <D,R>^T & <D,D> \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{d} \end{pmatrix}. \tag{30}$$

s' represents the data generated from the inner products of the given image with the selected basis set $\{D_n(x,y)\}$.

The slowness estimate in Eqn. (29) is a linear combination of different kinds of information. The first term in the right hand side of (29) is a superposition the beam paths. The term in the center (superposition of orthogonal functions) describes the regions in the model where the average slowness is known ("sharp boundaries") form information independent to the traveltimes. The third term models prior images where no clear boundaries are defined ("fuzzy boundaries"). The last two terms can of course be included into one, but I preferred to keep it that way to make clear the differences between the different kinds of information.

It is important to note that the system of equations (30) has unique solution and therefore, the slowness estimate satisfy all the given projections. That does not mean that the estimate is identical to the prior image. If the mean velocity in one area is given, the model estimated will reproduce that value, but it won't necessarily be homogeneous in that region.

For several reasons we might want the inversion to give more weight to one source of information than another. The next section addresses this topic.

Gradual Mixing of Images

In this section, I propose a way of weighting the traveltimes and the prior images to produce inversions that are influenced only by the traveltimes, only by the prior image or any by combination of both.

Equation (18) tells us that by multiplying the beam paths by a constant α and the regions R_n by a different constant β, we may achieve our goal. The slowness estimate is then

$$\tilde{S}(x,y) \;=\; \alpha \sum_{n=1}^{N} a_n \phi_n(x,y) \;+\; \beta \sum_{n=1}^{M} b_n R_n(x,y). \tag{31}$$

The system of equations for a_n and b_n is

$$\begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} \alpha < \Phi,\Phi > & \beta < R,\Phi > \\ \alpha < R,\Phi >^T & \beta < R,R > \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}. \tag{32}$$

We see that for $\beta = 0$, the inversion for only traveltimes is not reproduced. For $\alpha = 0$ the inversion for only slowness values is not reproduced either. The solution to this problem is to weight independently the system of equations (32) in the following way

$$\begin{pmatrix} \alpha I_{NN} & 0 \\ 0 & \beta I_{MM} \end{pmatrix} \begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} \alpha I_{NN} & 0 \\ 0 & \beta I_{MM} \end{pmatrix} \begin{pmatrix} \alpha < \Phi,\Phi > & \beta < R,\Phi > \\ \alpha < R,\Phi >^T & \beta < R,R > \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \tag{33}$$

where I_{NN} and I_{MM} are the identity matrices of dimensions NxN and MxM respectively.

After multiplying results

$$\begin{pmatrix} \alpha\mathbf{t} \\ \beta\mathbf{s} \end{pmatrix} = \begin{pmatrix} \alpha^2 < \Phi,\Phi > & \alpha\beta < R,\Phi > \\ \alpha\beta < R,\Phi >^T & \beta^2 < R,R > \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}. \tag{34}$$

If we constraint $\alpha+\beta = 1$, we get the desired result, and we can change gradually from inversion influenced only by traveltimes to inversion influenced only by the prior image. The final estimate of the slowness is given by Eqn. (31).

CONCLUSIONS

The theory of reconstruction in Hilbert spaces provides a unified framework for inverting simultaneously traveltimes and prior slowness information about the model. When the prior slowness image is used as a weighting function in the inner product that describes the measurements, the null space of the reconstruction is reduced when compared with the null space of each source of information (traveltimes and prior image). The disadvantage of this procedure is that only the main features in the prior image are reproduced by the the final result. I have shown, theoretically, that if the prior image is converted into data points (like the traveltimes) this problem can be solved and the result should be an image consistent with all the information available.

I plan to test the proposed theory in realistic synthetic examples as well as real data.

ACKNOWLEDGMENTS

I am grateful to INTEVEP, S.A. for supporting my study at Stanford. I am also grateful to Jerry Harris for many useful discussions and important suggestions.

REFERENCES

Carrion, P. M., 1989, Generalized non-linear elastic inversion with constraints in model and data spaces: Geophys. J., **96**, 151-162.

Carrion, P. M., Auyeung, C., and Mersereau, R., 1988, Constrained traveltime tomography: presented at the 58th Ann. Internat. Mtg., Soc. Expl. Geophys.

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

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

Darling, A. M., Hall, T. J., and Fiddy, M. A., 1983, Stable noniterative object reconstruction from incomplete data using a priori knowledge: J. Opt. Soc. Am., **73**, 1466-1469.

Devaney, A. J., 1984, Geophysical diffraction tomography: IEEE Trans. Geosci. Remote Sensing,**GE-22**, 3-13.

Hall. T, J., Darling, A. M., and Fiddy, M. A., 1982, Image compression and restoration incorporating prior knowledge: Optics Letters, **7**, 467-468.

Harlan, W. S., 1989, Tomographic estimation of seismic velocities from reflected raypaths: presented at the 59th Ann. Internat. Mtg., Soc. Expl. Geophys.

Jackson, D. D., 1979, The use of a priori data to resolve non-uniqueness in linear inversion: Geophys. J. Roy Astr. Soc., **57**, 137-157.

Jupp, D. L. B., and Vozoff, K., 1975, Stable iterative methods for inversion of geophysical data: Geophys. J. Roy. Astr. Soc., **42**, 957-976.

Lines, L. R., Schultz, A. K., and Treitel, S., 1988, Cooperative inversion of geophysical data: Geophysics, **53**, 8-20.

McMechan, G. A., 1983, Seismic tomography in boreholes: Geophys. J. Roy Astr. Soc., **74**, 601-612.

Michelena, R. J., and Harris, J. M., 1990, Tomographic traveltime inversion using natural pixels: paper F in this volume.

Michette, A. G., Darling, A. M., Fiddy, M. A., Ward, W. A., and Franklin, E., 1984, Resolution enhancement of well log and seismic data: Geophys. Prosp., **32**, 564-580.

Nolet, G., 1987, Seismic wave propagation and seismic tomography, *in* Nolet, G., Ed., Seismic Tomography: D. Reidel Publ. Co., 1-23.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259-1266.

Tarantola, A., and Valette, B., 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Rev. of Gephys. and Space Physics, **20**, 219-232.

Van Trier, J., 1988, Migration velocity analysis using geological constraints: presented at the 58th Ann. Internat. Mtg., Soc. Expl. Geophys.

Wu, R. S., and Toksöz, N., 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, **52**, 11-25.

PAPER I

DIFFRACTION TOMOGRAPHY RECONSTRUCTION USING CONSTRAINTS

Reinaldo J. Michelena and Jerry M. Harris
Seismic Tomography Project

ABSTRACT

Adding independent information to the inversion of a given data set helps to reduce the nonuniqueness of the solution by reducing the null space of the measurements. We show in this paper, within the framework of reconstruction in Hilbert spaces, that prior information can be used in diffraction tomography reconstruction and the result should be an image whose support in the frequency domain has been expanded when compared with the support of the image obtained without using any prior information. This means that the null space of the reconstruction is reduced by adding high frequencies not present in the original data.

INTRODUCTION

Image reconstruction in Hilbert spaces is a general formulation that can be adapted to the solution of many inverse problems where the data are obtained by integrating the product of the unknown by a certain set of functions (sampling functions). The nature of the unknown and the set of functions define, of course, the kind of problem that we have to solve. In others papers in this volume (Michelena and Harris, 1990, paper F; Michelena, 1990, paper H) it is shown that when the unknown is the slowness model and the set of functions are the beam paths, the problem we are dealing with is tomographic traveltime inversion. It can be shown also that if the unknown is a signal and the sampling functions are shifted versions of another given 1D function, the problem we are facing is deconvolution. In Fourier reconstruction the sampling functions are complex exponentials and the unknown can be the Fourier transform of a particular function. In any of these cases, the formulation allows the use of prior information about the unknown in the form of images by simply using the given image as a weighting function in the definition of the inner product that describes the measurements. Examples of this procedure are given by Darling et al., (1983) and Michette et al., (1984) and Hall et al., (1982). More details can also be found in paper H in this volume.

Diffraction tomography reconstruction can be seen essentially as a problem of Fourier reconstruction where only the propagating waves are considered. The sampling functions in this case are complex exponentials with real wave numbers and the unknown is related with the velocity perturbation in an homogeneous background.

Therefore, this problem can also be considered as a reconstruction problem in Hilbert spaces. We show in this paper that when the problem is considered this way, prior information about the object profile can be easily incorporated, improving the frequency domain coverage of the reconstructed image. If successfull on this simple example, we expect the method can be applied to the more general problems of diffraction tomography for inhomogeneous background models on pre-stack migration.

PRIOR INFORMATION IN DIFFRACTION TOMOGRAPHY

The basic problem in diffraction tomography is the reconstruction of the object profile $O(x,z)$ from the scattered data. Fig. 1 shows the geometry for the cross well configuration. The velocity of propagation $C(x,z)$ in the object profile is related with the velocity of propagation in the homogeneous background through the equation

$$O(x,z) = 1 - \frac{C_0^2}{C^2(x,z)},$$

where C_0 is the velocity of the background medium, $C_0 = \omega/k$ (ω = angular frequency, k = wavenumber of the incident wave).

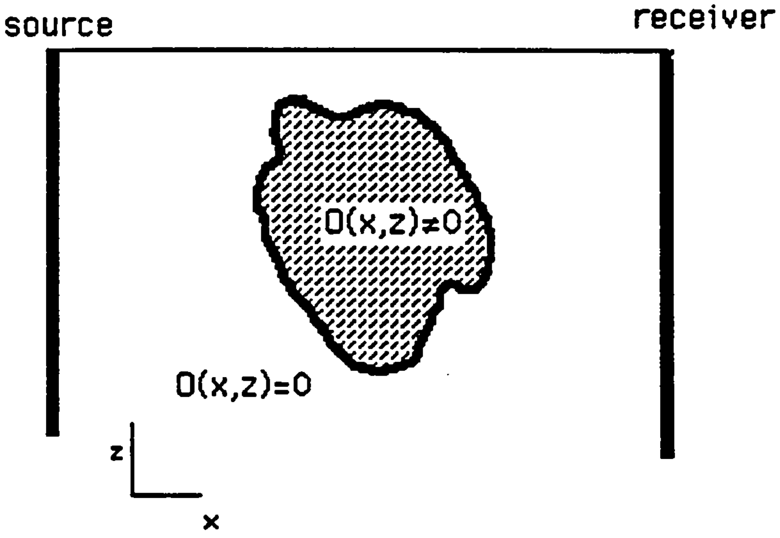


Figure 1: Tomographic configuration

Assuming that the scattered field is much smaller than the background field and applying the Born approximation (or Rytov in the same way) we obtain the following relationship between the spectrum of the scattered field and the 2D Fourier transform of the object profile $O(x,z)$ (Harris, 1987):

$$\hat{U}(x_g, k_g; x_s, k_s) = \pi k^2 (\gamma_s \gamma_g)^{-1} \hat{O}(\gamma_g - \gamma_s, k_g + k_s) U_0(\omega) \exp(-i\gamma_s x_s + i\gamma_g x_g) \quad (1)$$

where k_g and k_s are the wavenumbers along the geophone line and source line respectively and γ_g and γ_s the corresponding perpendicular wavenumbers ($\gamma_g^2 + k_g^2 = k^2$). The variable x_g identifies the position of the receiver line and x_s the position of the source line. $\hat{O}(\gamma_g - \gamma_s, k_g + k_s)$ is the 2D Fourier transform of $O(x, z)$

$$\hat{O}(\gamma_g - \gamma_s, k_g + k_s) = \int_{-\infty}^{\infty} dz dx O(x, z) \exp(-i(\gamma_g - \gamma_s)x - i(k_g + k_s)z). \quad (2)$$

$\hat{U}(x_g, k_g, x_s, k_s)$ is the spectrum of the scattered field along the source line and geophone line, or simply the "data". These equations were derived in 2D for line sources.

Substituting Eqn. 2 into Eqn. 1 we obtain,

$$\hat{U}(x_g, k_g; x_s, k_s) = \pi k^2 U_0(\omega) (\gamma_s \gamma_g)^{-1} \exp(-i\gamma_s x_s + i\gamma_g x_g) \int_{-\infty}^{\infty} dz dx O(x, z) \exp(-i(\gamma_g - \gamma_s)x - i(k_g + k_s)z). \quad (3)$$

Defining $\hat{\mathbf{U}}(k_x, k_z)$ as

$$\hat{\mathbf{U}}(k_x, k_z) = \hat{U}(x_g, \gamma_g; x_s, \gamma_s)(\gamma_s \gamma_g)/(\pi k^2 U_0(\omega)) \exp(ik_s x_s - ik_g x_g) \quad (4)$$

and $k_x = \gamma_g - \gamma_s$, $k_z = k_g + k_s$, we obtain the following compact relation between the object $O(x, z)$ and the data

$$\hat{\mathbf{U}}(k_x, k_z) = \int_{-\infty}^{\infty} dz dx O(x, z) \exp(-ik_x x - ik_z z). \quad (5)$$

Two important restrictions should be satisfied before applying the previous equations. The first one relates, as we said before, with the magnitude of the scattered fields relative to the direct field. The other restriction is that for preserving the Fourier transform relationship between the object profile and the data (Eqn. 5), it is necessary to consider only the case in which k_x is real, which means that we have to discard the effect of nonpropagating fields. This translates into limitations on the coverage of the spectrum of the object profile, as shown in Fig. 2 (Wu and Toksöz, 1987).

The conventional way of obtaining $O(x, z)$ as a function of $\hat{\mathbf{U}}$ is performing an inverse Fourier transform of the expression 5 considering only propagating waves (Devaney, 1984; Harris, 1987). A change of variables from the (k_x, k_z) domain to the (k_s, k_g) domain is needed. The inverse problem can be considered also as a problem of reconstruction in Hilbert spaces if we consider that the measurements are the projections of the object profile in the functions $\exp(-ik_x x - ik_z z)$.

$$\hat{\mathbf{U}} = < O(x, z), \exp(-ik_x x - ik_z z) > . \quad (6)$$

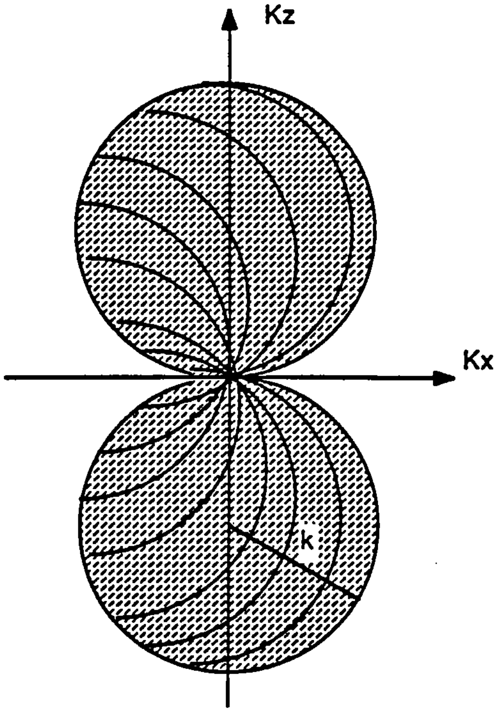


Figure 2: Region in the Fourier space in which the object profile is determined.

Let's denote $p(x,y)$ the prior information we have about the object $O(x,z)$. That information can be included in the reconstruction process as a weighting function in the inner product 6 that describes the generation of the measurements $\hat{\mathrm{U}}$. These measurements for each particular k_{x_i} and k_{z_j} can be described also as

$$< O(x,z),\beta_{ij}(x,z) >_p \; = \int_{-\infty}^{\infty} dxdzO(x,z)\beta^*_{ij}(x,z)/p(x,z), \tag{7}$$

where

$$\beta_{ij}(x,z) \; = \; p(x,z)\exp(ik_{x_i}x + ik_{z_j}z).$$

The minimum norm estimate of the object profile $O(x,z)$ is

$$\tilde{O}(x,z) \; = \; p(x,z)\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\exp(ik_{x_i}x + ik_{z_j}z). \tag{8}$$

If we multiply both sides of Eqn. 8 by $\exp(ik_{x_m}x + ik_{z_n})$ and integrate, we obtain that the coefficients b_{ij} must satisfy the system of linear equations

$$\hat{\mathrm{U}}(k_{x_m},k_{z_n}) \; = \; \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\hat{p}(k_{x_m} - k_{x_i}, k_{z_n} - k_{z_j}) \tag{9}$$

where $\hat{p}(k_x,k_z)$ is the 2D Fourier transform of $p(x,z)$. The independent term $\hat{\mathrm{U}}(k_{x_m},k_{z_n})$ is the scattered field compensated by a factor proportional to the deconvolution of the radiated source frequency spectrum $(4\pi k^2 U_0(\omega))^{-1}$ (Harris, 1987) and another factor which compensates for propagation effects $\exp(i\gamma_s x_s - i\gamma_g x_g)$. $\hat{p}(k_{x_m} - k_{x_i}, k_{z_n} - k_{z_j})$ is a block Toeplitz matrix.

If $p(x, z) \;=\; 1$, it results that the object is a simple Fourier inversion of the data. That is

$$\tilde{O}(x, z) \;=\; \frac{1}{(2\pi)^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\mathrm{U}}(k_{x_i}, k_{z_j}) \exp(ik_{x_i} x + ik_{z_j} z). \tag{10}$$

where only the case k_{x_i} real is considered (propagating waves).

According to 8, the estimated of the object profile is the product of the function $p(x, z)$ (that reflects our prior knowledge about $O(x, z)$) by an unknown function $f(x, z)$, where $f(x, z)$ is

$$f(x, z) \;=\; \sum_{i=1}^{N} \sum_{j=1}^{N} b_{ij} \exp(ik_{x_i} x + ik_{z_j} z). \tag{11}$$

The Fourier transform of $f(x, z)$ has a support that is the same as the one of the reconstructed object (Fig. 2) when $p(x, z) = 1$. When $p(x, z) \neq 1$ the Fourier transform of the reconstructed object is simply

$$\hat{\tilde{O}}(k_x, k_z) = \hat{p}(k_x, k_z) * * \hat{f}(k_x, k_z), \tag{12}$$

where $**$ means 2D convolution. In this equation, the support of the function $\hat{f}(k_x, k_z)$ equals the support of the object reconstructed without constraints. The effect of the prior information can be easily understood when the function $p(x, z)$ is band limited in both directions x and z. In this case the reconstructed object has the spectrum

$$\hat{\tilde{O}}(k_x, k_z) = \Pi(k_x/W, k_z/W) * * \hat{f}(k_x, k_z), \tag{13}$$

where $\Pi(.,.)$ is the rectangle function and W its bandwidth. This particular case is illustrated in Fig. 3. In general, Eqn. 12 tells us that when we use prior information, the support area of the reconstructed object will contain always higher frequencies than the support of the object imaged without constraints.

The dependency of $\hat{\mathrm{U}}(k_x, k_z)$ on k_x and k_z in Eqn. 4 is not explicit. For making that dependency explicit we have to solve the equations (Harris, 1987)

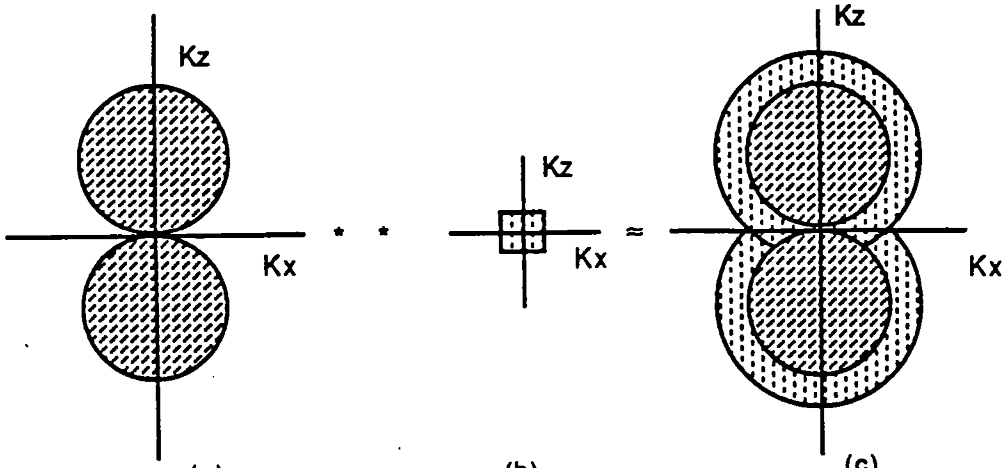
$$k_x = \gamma_g - \gamma_s \qquad\qquad k_z = k_g + k_s$$

for k_g and k_s. The result is

$$k_s = \frac{k_z}{2} \pm \frac{k_x}{2K_T}(4k^2 - K_T^2)^{1/2}$$

and

$$k_g = \frac{k_z}{2} \mp \frac{k_x}{2K_T}(4k^2 - K_T^2)^{1/2}$$



(a)
Support of the object imaged without constraints

(b)
Support of the contraints

(c)
Support of the object imaged using contraints

Figure 3: Effect of the prior information in the frequency domain coverage of the reconstructed object.

where $K_T^2 = k_x^2 + k_z^2$.

The effect of substituting these equations into 4 is to interpolate the data from the circles (Fig. 2) where it is originally acquired in the frequency domain to the regular grid (k_x, k_z). After making these substitutions, we obtain the independent term needed for solving the system of equations 9. This procedure contrasts with those proposed by Devaney (1984), Harris (1987) and Wu and Toksöz (1987), where the inversion is performed using directly the data from the circles of Fig. 2 and therefore, a tomographic filter (Jacobian) is needed before the backpropagation.

CONCLUSION

We have presented the way of introducing prior information in diffraction tomography when the problem is considered within the framework of reconstruction in Hilbert spaces. The effect of the prior information is to increase the frequency domain coverage of the reconstructed object. The data needed for the inversion is obtained from the original data by analytical interpolation in the frequency domain and then no tomographic filter is required.

ACKNOWLEDGMENTS

The first author is grateful to INTEVEP, S.A. for supporting his study at Stanford.

REFERENCES

Darling, A. M., Hall, T. J., and Fiddy, M. A., 1983, Stable noniterative object reconstruction from incomplete data using a priori knowledge: J. Opt. Soc. Am., **73**, 1466-1469.

Devaney, A. J., 1984, Geophysical diffraction tomography: IEEE Trans. Geosci. Remote Sensing,GE-**22**, 3-13.

Hall. T, J., Darling, A. M., and Fiddy, M. A., 1982, Image compression and restoration incorporating prior knowledge: Optics Letters, **7**, 467-468.

Harris, J.M., 1987, Diffraction tomography with a arrays of discrete sources and receivers: IEEE Trans. Geosci. Remote Sensing, GE-**25**, 448-455.

Michelena, R. J., 1990, Traveltime inversion with constraints: paper H in this volume.

Michelena, R. J., and Harris, J. M., 1990, Tomographic traveltime inversion using natural pixels: paper F in this volume.

Michette, A. G., Darling, A. M., Fiddy, M. A., Ward, W. A., and Franklin, E., 1984, Resolution enhancement of well log and seismic data: Geophys. Prosp., **32**, 564-580.

Wu, R. S., and Toksöz, N., 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, **52**, 11-25.

PAPER J

RELATIVE ENTROPY ENHANCEMENT OF GEOPHYSICAL TOMOGRAPHY IMAGES

Gary Mavko

Seismic Tomography Project

ABSTRACT

Maximum Entropy (ME) and minimum Relative Entropy (RE) methods provide a powerful variational approach to extracting objective, high resolution images from incomplete and noisy data. They have proven useful for a wide variety of image reconstruction and enhancement problems in radio astronomy, spectral analysis, medical imaging, and photography. ME and RE have been widely applied to geophysical spectral analysis problems, but their application to geophysical seismic imaging is relatively new. A typical seismic image is underdetermined, so that an infinite number of images are consistent with the available data. Of that infinite number, the Maximum Entropy image is most objective and maximally non-committal with regard to the unknown information. The minimum Relative Entropy image is closest to a prior estimate of the image while still exactly consistent with the available partial information. In this paper ME and RE methods are applied to several synthetic images representative of those obtainable from cross-well seismic tomography. The two common forms of the entropy expression are compared, and while both lead to improved resolution, the "$-S\ln S$" form was found to be more stable numerically. For the examples studied, ME enhancement slightly improved resolution compared to conventional Fourier reconstructions, but not nearly as much as the RE enhancement. An important innovation is the concept of weights applied to the prior estimate of the image. The weights express one's confidence in the prior estimate, and also allow the algorithm to "focus" on subsets of the image. In all examples, use of the weighting function gave improved resolution in the areas of interest.

INTRODUCTION

Geophysical tomography is an attempt to reconstruct an image of the subsurface from incomplete and noisy data. In the case of seismic diffraction tomography, seismic sources generate an incident wavefield which is scattered by velocity heterogeneities and sampled by receivers. Finite bandwidth, source and receiver geometry, and noise put fundamental limitations on the resolution of the reconstructed image.

A typical geometry for cross-well tomography is illustrated in Figure 1. A vertical array of sources lies within one well, and a vertical array of receivers lies within a second well. A region of anomalous seismic velocity, $V(x,z)$, relative to a fairly uniform background, V_0, lies in between. The desired product of the tomographic inversion is an image of the anomalous velocity or the slowness $S(x,z) = 1./V(x,z)$.

It is equivalent to determine the Fourier transform, $\tilde{R}(k_x, k_z)$, of the slowness squared $S^2(x,y)$ for *all* spatial frequencies, k_x and k_z, since $S^2(x,z)$ can be recovered from $\tilde{R}$ by Fourier synthesis. Most of the popular geophysical diffraction tomography reconstruction algorithms (Devaney, 1984; Harris; 1987; Lo, 1987) are derived in the spatial frequency domain.

The problem is that only some of the Fourier components can be measured with realistic bandwidths and viewing angles. For a fixed $k = f/V_0$ (where f is the temporal frequency) and very long vertical source and receiver arrays, the most that can be sampled is a region of the (k_x, k_z) plane lying within the two semicircular disks shown in Figure 2. The radius of each disk is given by $k_{max} = f_{max}/V_0$, where f_{max} is the maximum temporal frequency recorded. It is clear from the plot that the vertical resolution will be much better than the horizontal resolution, because the coverage along the k_z axis is $\pm 2k_{max}$ and the coverage along the k_x axis is zero.

If the non-zero Fourier components of the desired object lie entirely within the sampled disks, then with good signal to noise and careful processing one might approach a near-perfect reconstruction of the image. However, if components lie outside of the sampled disks, then no amount of clever processing can recover them. The information is simply not measured.

In general, an infinite number of different images can be found whose Fourier transforms exactly match the measured transform within the sampled disks. Each choice corresponds to a different extension of the Fourier transform into the unsampled region. A common approach is to guess some particular extension, and the most common guess is that the unmeasured components are zero. This is built into virtually all conventional reconstruction techniques, including the direct Fourier reconstruction, the backprojection, and the filtered backpropagation techniques (Devaney, 1984; Harris, 1987, Lo, 1987).

The effect of zeroing regions of the (k_x, k_z) domain is illustrated in Figure 3. A synthetic slowness function representing three localized scatterers (anomalously slow material) superimposed on a uniform background is shown in Figure 3a. The separation between centers of the scatterers is 6 meters both vertically and horizontally. If we know all of the Fourier components of this object we can perfectly reconstruct it. If, however, we zero out all of the (k_x, k_z) domain that lies outside of the sampled disks shown in Figure 2b, we construct only a low-passed or narrow band version of the image, as shown in Figure 3b. For this example k_{max} corresponds to a minimum wavelength $\lambda_{min} = 1/k_{max} = 10.6$ meters. As expected, the low resolution image is blurred predominantly in the horizontal direction with the shallower two scatterers merged into a single broad ridge.

A second example is shown in Figure 4a. In this case, an anomalously slow horizontal layer is simulated, corresponding to high porosity sediments. In the middle of the layer a permeability barrier, corresponding to a loss of porosity, is simulated by normal high velocity. The low resolution image, resulting from zeroing out all frequencies outside of the sampled disks, is shown in Figure 4b. In this example $\lambda_{min} = 1/k_{max} = 16$ meters. The sharp layer has become an easily recognizable rounded ridge, but the central anomaly is essentially gone with the loss of horizontal bandwidth.

A third example, shown in Figure 5a, is similar to the one in Figure 4a. In this case a very slow region is superimposed at the center of the slow layer. The corresponding low resolution image, resulting from zeroing out all frequencies outside of the sampled disks, is shown in Figure 5b. Again the central anomaly is essentially undetectable.

Part of the loss of resolution in these examples is a fundamental limitation of the experimental geometry, but part results from the rather specific and harsh assumption about the unknown Fourier components. We don't know the Fourier components outside of the disks, but for an arbitrary scatterer we can be fairly confident that those components are not equal to zero.

In contrast, Maximum Entropy (ME) and Relative Entropy (RE) methods (Jaynes, 1968; Burg, 1975; Shore and Johnson, 1980; Gull and Skilling, 1980; Gull and Skilling, 1984) use a variational approach to extend the partial information about the Fourier components in a physically more realistic way than assuming them to be zero. ME chooses the extension into the unsampled region that exactly incorporates all of the available information but is free of all other assumptions; RE chooses the extension that is closest to a prior estimate of the image, deviating only as necessary to satisfy the available measurements. In general, ME and RE enhanced images have higher resolution than conventional reconstructions while still agreeing exactly with the measured information. ME methods have been widely used for nongeophysical image processing (Wernecke and D'Addario, 1977; Skilling et al., 1979; Danielle and Gull, 1980; Gull and Skilling, 1984) and tomography (Lent, 1977; Mohammad-Djafari and Demoment, 1986). Both ME and RE enhancements have been applied to geophysical diffraction tomography by Lo (1987), but generally their use in geophysical tomography has been limited.

This paper illustrates and compares some of the features of ME and RE enhancement of images representative of cross-well diffraction tomography. We compare for several synthetic examples a conventional reconstruction with ME and RE enhanced reconstructions. A new feature is the incorporation of a weighting or confidence function applied to the prior estimate of the image. This is particularly useful for geophysical problems where the prior knowledge is good near boreholes and poor in the interwell region. ME has been shown to be useful for image enhancement when measurements are contaminated by noise and when the measurements are insufficient to uniquely determine the image. The discussion will be limited to the latter problem, specifically the bandwidth extrapolation and Fourier synthesis in cross-well situations. The starting point is to assume that some form of conventional reconstruction has been performed, so that the Fourier transform of the object is known within the sampled disks in the (k_x, k_z) plane. This bypasses the difficulties of the mapping or backpropagation from the receiver domain to the (k_x, k_z) domain, which are discussed by Lo (1987).

MAXIMUM ENTROPY RECONSTRUCTION

Of the infinite number of images that agree exactly with the partial Fourier information, the one choice that is maximally non-committal with regard to the unknown information is the one that maximizes the entropy (Jaynes, 1968). Two common forms of the entropy expression have been applied to problems of spectral analysis and image enhancement, and a fairly extensive literature is devoted to debating the preferred one (Kikuchi and Soffer, 1977; Johnson and Shore, 1984; and Mohammad-Djafari and Demoment, 1986; Gull and Skilling, 1985).

The first approach treats the acoustic wavefield as a stationary random process similar to the application by Wernecke and D'Addario (1977) to electromagnet fields in radio astronomy. The desired slowness squared function, S^2, is then regarded as the a two dimensional (nonnegative) spectral density representative of the random field (Johnson and Shore, 1984; Mohammad-Djafari and Demoment, 1986). S^2 must be estimated from the 'autocorrelation' $\tilde{R}(k_x, k_z)$ of the wavefield, known only within the sampled disks in the (k_x, k_z) domain. This is precisely the spectral estimation problem treated by Burg (1975) and Mavko and Burg (1987). With this interpretation the entropy of the image has the form:

$$H_1 = \int_{-X}^{X} \int_{-Z}^{Z} lnS^2(x,z)dxdz \qquad (1)$$

where X and Z are the dimensions of the image. The image that maximizes the entropy H_1 always has the functional form (Mavko and Burg, 1987):

$$S^2(x,z) = \frac{1}{\sum_{n,m} \lambda_{n,m} \exp\left(-i2\pi(xn\Delta k_x + zm\Delta k_z)\right)} \qquad (2)$$

The unknown coefficients $\lambda_{n,m}$ correspond to the same spatial frequencies $(n\Delta k_x, m\Delta k_z)$ where the measurements of $\tilde{R}(n\Delta k_x, m\Delta k_z)$ are made, and the $\lambda_{n,m}$ are chosen to ensure that the image agrees with each of the Fourier measurements:

$$\int_{-X}^{X} \int_{-Z}^{Z} S^2(x,z) \exp\left(i2\pi(xr\Delta k_x + zs\Delta k_z)\right)dxdz = \tilde{R}(r\Delta k_x, s\Delta k_z) \qquad (3)$$

It is assumed that the Fourier information $\tilde{R}$ can be found at or interpolated to an equally spaced rectangular grid within the sampled disks.

The second approach models the image itself $S^2(x,z)$, suitably normalized, as a probability density (Lo, 1987; Gull and Skilling, 1985). The entropy of the image in this case is given by:

$$H_2 = -\int_{-X}^{X} \int_{-Z}^{Z} S^2(x,z)lnS^2(x,z)dxdz \qquad (4)$$

The image that maximizes the entropy H_2 always takes the form (see Appendix):

$$S^2(x,z) = \exp\left(-\alpha + \sum_{n,m} \lambda_{n,m} \exp\left(-i2\pi(xn\Delta k_x + zm\Delta k_z)\right)\right) \qquad (5)$$

As before, the $\lambda_{n,m}$ are chosen to ensure that the image agrees with the Fourier measurements when S^2 is substituted into equation (3), and α is chosen to ensure the proper normalization for a probability density.

Maximum entropy enhancements of the blurred images in Figures 3b, 4b, and 5b are shown in Figures 3c, 4c, and 5c, using the H_2 entropy expression, equations (4) and (5). These were made by using a discrete approximation of (digitizing) the blurred images on a $1m \times 1m$ grid (32×64 samples) and fast Fourier transforming to obtain the known nonzero Fourier components $\tilde{R}(n\Delta k_x, m\Delta k_z)$ within the sampling disks. There were 135 "measured" components within the disks for the example in Figure 3, and 60 in Figures 4 and 5. The unknown coefficients $\lambda_{n,m}$ in the general form of the image $S^2(x,z)$ in equation (5) were found with a Newton-Raphson algorithm, similar to one suggested by Johnson(1983) for one dimensional problems and described in the Appendix. In all cases the ME images show slightly improved resolution, in both the vertical and horizontal directions, relative to the conventional reconstruction. In particular the two shallow scatterers in Figure 3c can begin to be resolved. The central anomalies in Figures 4c and 5c appear as a slight dip and peak, respectively, along the horizontal slowness ridge, but their "ringy" character makes them unconvincing and difficult to recognize as anomalies.

For comparison of the entropy forms H_1 and H_2, a ME enhancement of Figure 3b was also computed using the functional form for the image given by equation (2) and is plotted in Figure 3d. For this example there is little difference in the resolution performance of one form over the other, although the H_1 form is somewhat asymmetric in its estimate of the shallow peak amplitudes.

In general, the H_2 form appeared to be better behaved numerically than the H_1 form, at least using the Newton-Raphson algorithm. For both entropy forms, the resolution of the slowness image is improved by subtracting away most of the background average value $\Delta S^2 = S^2 - \alpha S_0^2$, where $0 < \alpha < 1$, treating the perturbation in slowness rather than the slowness itself. In terms of the scattering problem, only the perturbations in slowness are sources of the scattered field. When $\alpha = 1$ the perturbation is equivalent to the objective function $O(x,z) = 1 - S^2(x,z)/S_0^2$ that appears in the Born scattering theory. However, $O(x,z)$ can be negative, which violates both the spectral density and probability density interpretations of the image. Lo (1987) avoided that problem by treating the absolute value of $\mid O(x,z) \mid$, but this leaves an ambiguity in determining $\pm O(x,z)$. The effect of the background slowness on resolution is exactly the same as the effect of prewhitening in spectral estimation. The background numerically stabilizes or damps the problem, but at the expense of resolution. We found that the entropy form H_2 was stable with much less background damping, presumably because of the better behavior of H_2 as $S^2 \rightarrow 0$, compared with the behavior of H_1.

RELATIVE ENTROPY RECONSTRUCTION

Often in geophysical problems we have additional information in the form of a prior velocity model that we wish to incorporate into the tomographic inversion. This prior information might come from well logs, surface seismic models, or geologic information. The relative entropy of an image $S^2(x,z)$ relative to a prior estimate of the image $P(x,z)$, again has two common interpretations. The first expresses the relative entropy between the unknown distribution of Fourier components of the wavefield and the prior distribution and leads to the following functional form (Johnson et al., 1984), analogous to equation (2):

$$S^2(x,z) = \frac{1}{\frac{1}{P(x,z)} + \sum_{n,m} \lambda_{n,m} \exp\left(-i2\pi(xn\Delta k_x + zm\Delta k_z)\right)} \tag{6}$$

The second approach is analogous to equation (4) and gives a relative entropy of the form:

$$H_2 = \int_{-X}^{X} \int_{-Z}^{Z} S^2(x,z) ln \frac{S^2(x,z)}{P(x,z)} dx dz \tag{7}$$

The image that minimizes H_2 always takes the form (see Appendix):

$$S^2(x,z) = P(x,z) \exp\left(-\alpha + \sum_{n,m} \lambda_{n,m} \exp\left(-i2\pi(xn\Delta k_x + zm\Delta k_z)\right)\right) \tag{8}$$

In both cases the unknown coefficients $\lambda_{n,m}$ must be found to ensure that the images agree exactly with the measurements $\tilde{R}(n\Delta k_x, m\Delta k_z)$, equation (3).

Relative entropy enhancements of the blurred images in Figures 4b and 5b are shown in Figures 4d and 5d. Only the H_2 forms are shown, again because of the superior numerical behavior compared with the H_1 form. In both cases the prior image $P(x,z)$ was chosen as a linear interpolation of the slowness at the wells - a uniform background slowness and the higher slowness layer without a center anomaly. In these images the prior has guided the image to give a very smooth background and a sharp edged layer. The additional information in the measurements $\tilde{R}(n\Delta k_x, m\Delta k_z)$ about the center anomalies causes a broad shallow dip in Figure 4d and a broad low peak in Figure 5d. Both central anomalies are more convincing and less ringy than in Figures 4c and 5c, although the width and magnitude of the anomalies are impossible to resolve.

RELATIVE ENTROPY WITH WEIGHTING

Situations will arise when the prior information is more reliable in some portions of the image, for example near the wells, than in other portions. In these cases one might wish to give a spatially variable weight to the prior, similar to the frequency dependent weight applied to one dimensional speech spectra by Johnson et al. (1984). Extending their result to two dimensions gives the following generalization of the H_1 form of entropy, equation (8):

$$S^2(x,z) = \frac{1}{\frac{1}{P(x,z)} + \frac{1}{W(x,z)} \sum_{n,m} \lambda_{n,m} \exp\left(-i2\pi(xn\Delta k_x + zm\Delta k_z)\right)} \tag{9}$$

where $W(x,z)$ is the weight function. Large values of W express a large confidence in the prior estimate. Arbitrary multiplicative factors on W do not affect the result, because they can be absorbed by the $\lambda_{n,m}$.

We write the alternative H_2 form of the weighted entropy as an extension of equation (9):

$$H_2 = \int_{-X}^{X}\int_{-Z}^{Z} W(x,z)S^2(x,z)ln\frac{S^2(x,z)}{P(x,z)}dxdz \tag{10}$$

The image that minimizes H_2 can be written as (see Appendix):

$$S^2(x,z) = P(x,z)\exp\left(-\alpha + \frac{1}{W(x,z)}\sum_{n,m}\lambda_{n,m}\exp\left(-i2\pi(xn\Delta k_x + zm\Delta k_z)\right)\right) \tag{11}$$

Weighted relative entropy enhancements of the images in Figures 4 and 5 are shown in Figures 4e and 5e. In both cases, the same prior was used as for Figures 4d and 5d. A broad, circularly symmetric Gaussian weighting function, $W(r) = 1 - .9\exp(-(r/12)^2)$, where $r = \sqrt{(x^2+z^2)}$, was centered on the layer, giving the center 10 times less confidence than the edges near the wells. While the prior generally guided the image, the low weight allowed the central anomalies to deviate from the prior. In both cases the result is a narrower, better resolved central anomaly.

A similar weighting strategy was applied to the blurred image in Figures 3b. In this case we observe the narrow deep peak and the broad shallower one. We choose a prior that is uniform everywhere except for a single deep localized scatterer. Since we are uncertain about the shallower ridge we choose a broad weighting function $W(x,z) = 1. - .9\exp(-(x/12)^2 - (z/3)^2)$, centered on the shallow anomaly that weights it 10 times less than the rest of the image. The result is shown in Figure 3e, and has marked improvement in resolution of the two peaks.

DISCUSSION

Variational techniques provide a useful way of improving and stabilizing image reconstruction in problems with incomplete and noisy data. Common approaches are to choose the image that has, for example, the minimum energy (Devaney, 1984) or that is in some way the smoothest. The maximum entropy image is the one that is most objective, containing no more information than is actually in the limited measurements of the image. Relative entropy methods provide an objective way of incorporating prior information into the reconstruction. The RE image deviates from the prior no more than is necessary to satisfy the measured constaints.

ME and RE methods have been widely applied to problems in radio astronomy (Wernecke and D'Addario, 1977; Gull and Skilling, 1984) and spectral analysis (Burg, 1975; Mavko and Burg, 1987). Lo (1987) applied relative entropy to geophysical tomographic images reconstructed from laboratory measurements on physical models.

In this paper we compare the two common entropy forms, the "lnX" and the "$-XlnX$". Although the "lnX" is often superior for spectral analysis of both one-dimensional and two-dimensional problems, the $-XlnX$ is usually favored in image inhancement. Experiments here favor the latter form, because of slightly better resolution, but mostly because of better numerical stability.

The most important new point here is the introduction of a spatially dependent weighting function applied to the prior estimate. This is particularly valuable in cross-well tomography problems where information about the prior is heterogeneous – high confidence near the wells and in clean data areas, but low elsewhere. Low weights also seem to focus the algorithm for improved resolution in areas of low weight. In examples 3e, 4e, and 5e, the low weight features were resolved better than before.

APPENDIX

We wish to minimize the entropy, H_2, which is written in the form:

$$H_2 = \int_{-X}^{X} \int_{-Z}^{Z} W(x,z)Q(x,z) ln \frac{Q(x,z)}{P(x,z)} dxdz \tag{12}$$

where $Q(x,z)$ is the image, $P(x,z)$ is the prior estimate of the image, and $W(x,z)$ is the positive weighting function on the prior. We know that $Q(x,z)$ can be written in terms of its Fourier transform, $\tilde{R}(n\Delta k_x, m\Delta k_z)$, which for shorthand we write as $R_{n,m}$.

$$Q(x,z) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} R_{n,m} \exp\left(-i2\pi(xn\Delta k_x + zm\Delta k_z)\right) \tag{13}$$

However, only a small finite number of the $R_{n,m}$ are known. From equation (13) we see that

$$\frac{\partial Q}{\partial R_{n,m}} = \exp\left(-i2\pi(xn\Delta k_x + zm\Delta k_z)\right) \tag{14}$$

Using equation (14) we minimize H_2 with respect to the unknown $R_{n,m}$ by setting the derivatives equal to zero:

$$\frac{\partial H_2}{\partial R_{n,m}} = \int_{-X}^{X} \int_{-Z}^{Z} W(x,z)[1 + ln \frac{Q(x,z)}{P(x,z)}] \exp\left(-i2\pi(xn\Delta k_x + zm\Delta k_z)\right) dxdz = 0 \tag{15}$$

Now, assuming the integrand is a smooth, well behaved function, it can be written in terms of its Fourier coefficients $\lambda_{r,s}$:

$$W(x,z)[1 + ln \frac{Q(x,z)}{P(x,z)}] = \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \lambda_{r,s} \exp\left(-i2\pi(xr\Delta k_x + zs\Delta k_z)\right) \tag{16}$$

Now substituting equation (16) into equation (15) we get

$$\sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \lambda_{r,s} \int_{-X}^{X} \int_{-Z}^{Z} \exp\left(i2\pi(x(n-r)\Delta k_x + z(m-s)\Delta k_z)\right) dxdz = 0 \tag{17}$$

which leads us to the conclusion that $\lambda_{r,s} = 0$ when $r = -n$ and $s = -m$. The $\lambda_{r,s}$ are nonzero only for the (n,m) where measurements are available. Now solving equation

(16) for Q, and using only the nonzero $\lambda_{r,s}$:

$$Q(x,z) = P(x,z)\exp\left(-\alpha + \frac{1}{W(x,z)}\sum_{r,s}\lambda_{r,s}\exp\left(-i2\pi(xr\Delta k_x + zs\Delta k_z)\right)\right) \tag{18}$$

which is the same as equation (11). The parameter α represents the scaling parameter that must be used for the additional constraint that $\int\int Q dxdz = 1$. Setting the weighting function $W(x,z) = 1$ gives equation (8) and setting both $W(x,z) = 1$ and $P(x,z) = 1$ gives equation (5).

For the Newton-Raphson algorithm we need to evaluate the derivatives $\partial R_{n,m}/\partial\lambda_{r,s}$. Begin with

$$R_{n,m} = \int_{-X}^{X}\int_{-Z}^{Z} Q(x,z)\exp\left(i2\pi(xn\Delta k_x + zm\Delta k_z)\right)dxdz \tag{19}$$

$$\frac{\partial R_{n,m}}{\partial\lambda_{r,s}} = \int_{-X}^{X}\int_{-Z}^{Z}\frac{\partial Q(x,z)}{\partial\lambda_{r,s}}\exp\left(i2\pi(xn\Delta k_x + zm\Delta k_z)\right)dxdz \tag{20}$$

Substituting $\partial Q/\partial\lambda$ from equation (18) allows us to write

$$\frac{\partial R_{n,m}}{\partial\lambda_{r,s}} = \int_{-X}^{X}\int_{-Z}^{Z}\frac{Q(x,z)}{W(x,z)}\exp\left(i2\pi(x(n-r)\Delta k_x + z(m-s)\Delta k_z)\right)dxdz \tag{21}$$

The Newton-Raphson algorithm begins with an initial guess of the unknown $\lambda_{r,s}$ which are substituted into equations (18) to evaluate the corresponding estimate of Q. The Fourier sum in equation (18) is conveniently evaluated with a two-dimensional FFT. The corresponding autocorrelation is evaluated from equation (19) approximated with a two-dimensional FFT. The misfit $\Delta R_{n,m}$ with the measured autocorrelation is used to compute corrections $\Delta\lambda_{r,s}$ using the matrix of derivatives in equation (21). Note that the derivatives are also easily computed by taking an FFT of $Q(x,z)/W(x,z)$ and selecting the terms $(n-r, m-s)$.

REFERENCES

Burg, J.P., 1975, Maximum entropy spectral analysis, Ph.D. dissertation, Stanford University, Stanford, California.

Danielle, G.J. and Gull, S.F., 1980, Maximum entropy algorithm applied to image enhancement, IEE Proceedings, Pt. E., 127, 170-172.

Devaney, A.J., 1984, Geophysical diffraction tomography, IEEE Trans. Geoscience and Remote Sensing, GE-22, 3-13.

Gull, S.F. and Skilling, J., 1984, Maximum entropy method in image processing, IEE Proceedings, Pt. F, 131, 646-659.

Gull, S.F. and Skilling, J., 1985, The entropy of an image, in Maximum Entropy and Bayesian Methods in Inverse Problems, Reidel Pub. Co., Dordrecht.

Harris, J.M., 1987, Diffraction tomography with arrays of discrete sources and receivers, IEEE Trans. Geoscience and Remote Sensing, GE-25, 448-455.

Jaynes, E.T., 1968, Prior probabilities, IEEE Trans. Sys. Sci. and Cyber., SSC-4, 227-241.

Johnson, R.W., 1983, Algorithms for single-signal and multisignal minimum-cross-entropy spectrum analysis, NRL Report 8667, Naval Research Lab, Washington, D.C.

Johnson, R.W. and Shore, J.E., 1984, Which is the better entropy expression for speech processing: -S log S or log S?, IEEE Trans. Acoust., Speech, and Sig. Proc., ASSP-32, 129-137.

Johnson, R.W., Shore, J.E., and , J.P., 1984, Multisignal minimum-cross-entropy spectrum analysis with weighted initial estimates, IEEE Trans. Acoust. Speech, and Sig. Proc., ASSP-32, 531-539

Kikuchi, R. and Soffer, B.H., 1977, Maximum entropy image restoration. I. The entropy expression, J. Opt. Soc. Am., 67, 1656-1665.

Lent, A., 1977, A convergent algorithm for maximum entropy image restoration, with a medical X-ray application: Image analysis and evaluation, SPSE Conv. Proc., R. W. Shaw, Ed., Washington, D.C., Soc. Photogr. Sce. Eng., 249-257.

Lo, T-W., 1987, Seismic borehole tomography, Ph.D. dissertation, Massachusetts Institute of Technology.

Mavko, G.M. and Burg, J.P., 1987, Two-dimensional maximum entropy spectral analysis, Presented at SEG 57th Annual Meeting, New Orleans.

Mohammad-Djafari, A. and Demoment, G., 1986, Maximum entropy diffraction tomography, 1968, IEEE ICASSP 86 Tokyo, 1749-1752.

Shore, J.E. and Johnson, R.W., 1980, Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, 1980, IEEE Trans. Inf. Theory, IT-26, 26-943.

Skilling, J., Strong, A.W., and Bennet, K, 1979, Maximum-entropy image processing gamma-ray astonomy, Mon. Not. R. ast. Soc. 145-152.

Wernecke, S.J. and D'Addario, L.R., 1977, Maximum entropy image reconstruction, IEEE Trans. Computers, C-26, 351-364.

Borehole tomography

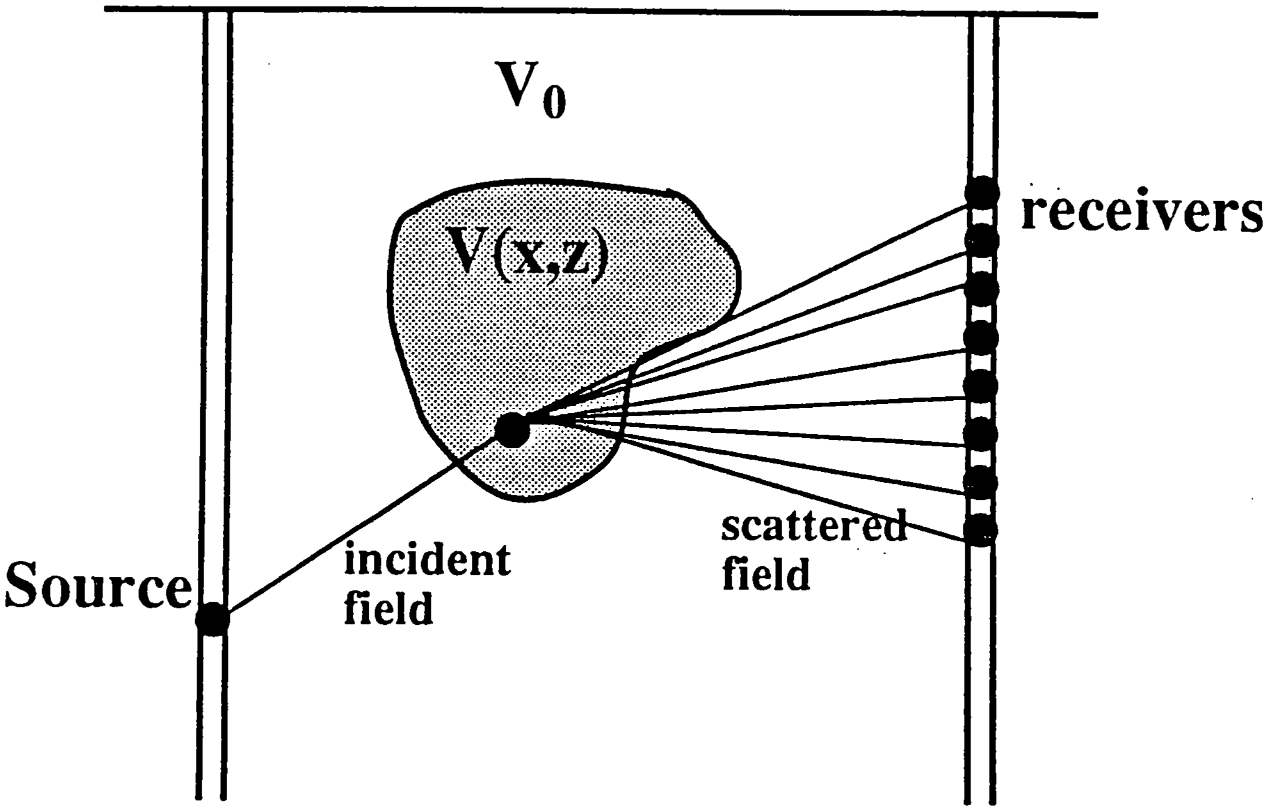
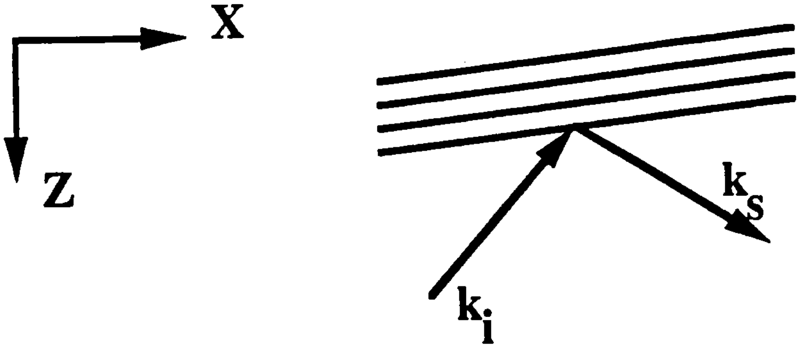


Figure 1: Typical crosswell survey geometry, in which we wish to image the velocity field V(x,z).

Limited Sampling in k-k domain



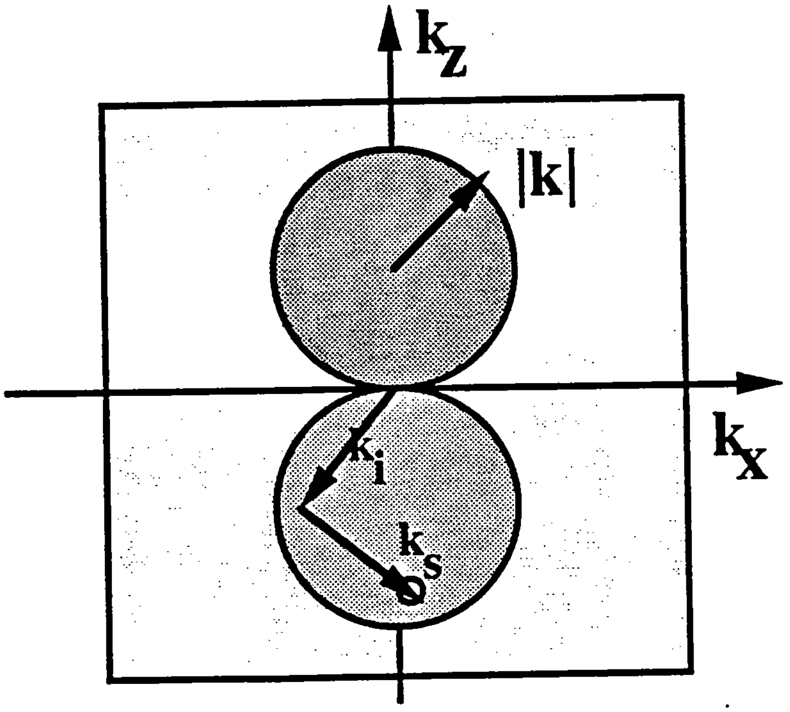


Figure 2: Only certain Fourier components of the desired image can be measured in a diffraction tomography experiment. For the crosswell geometry these components are limited to two circular disks as shown.

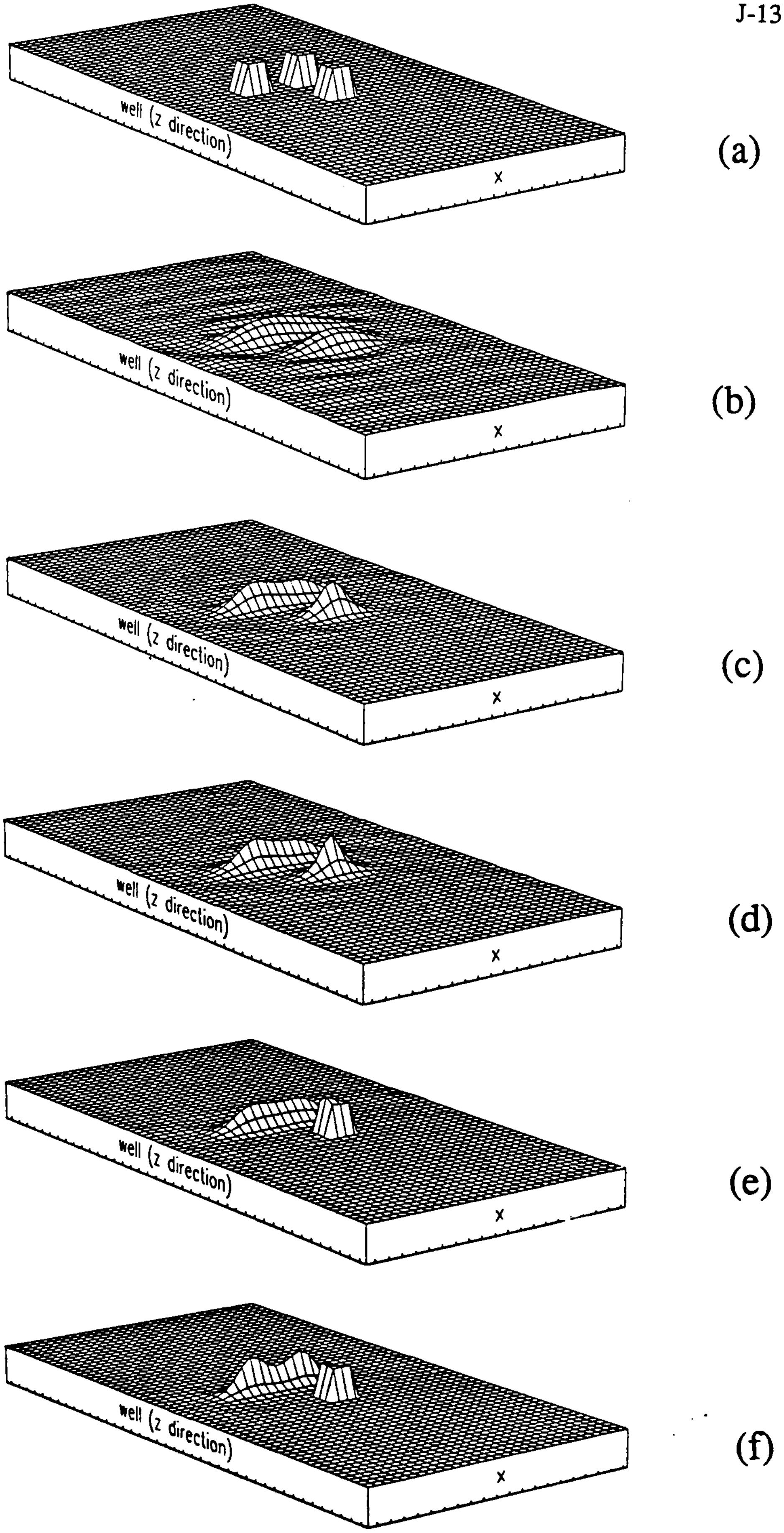


Figure 3: (a) Three scatterers. (b) Result of zeroing the unknown components. (c)-(d) ME enhancements. (e) RE enhancement with the deeper scatterer in the prior. (f) RE enhancement when the shallow anomaly is given a very low weight.

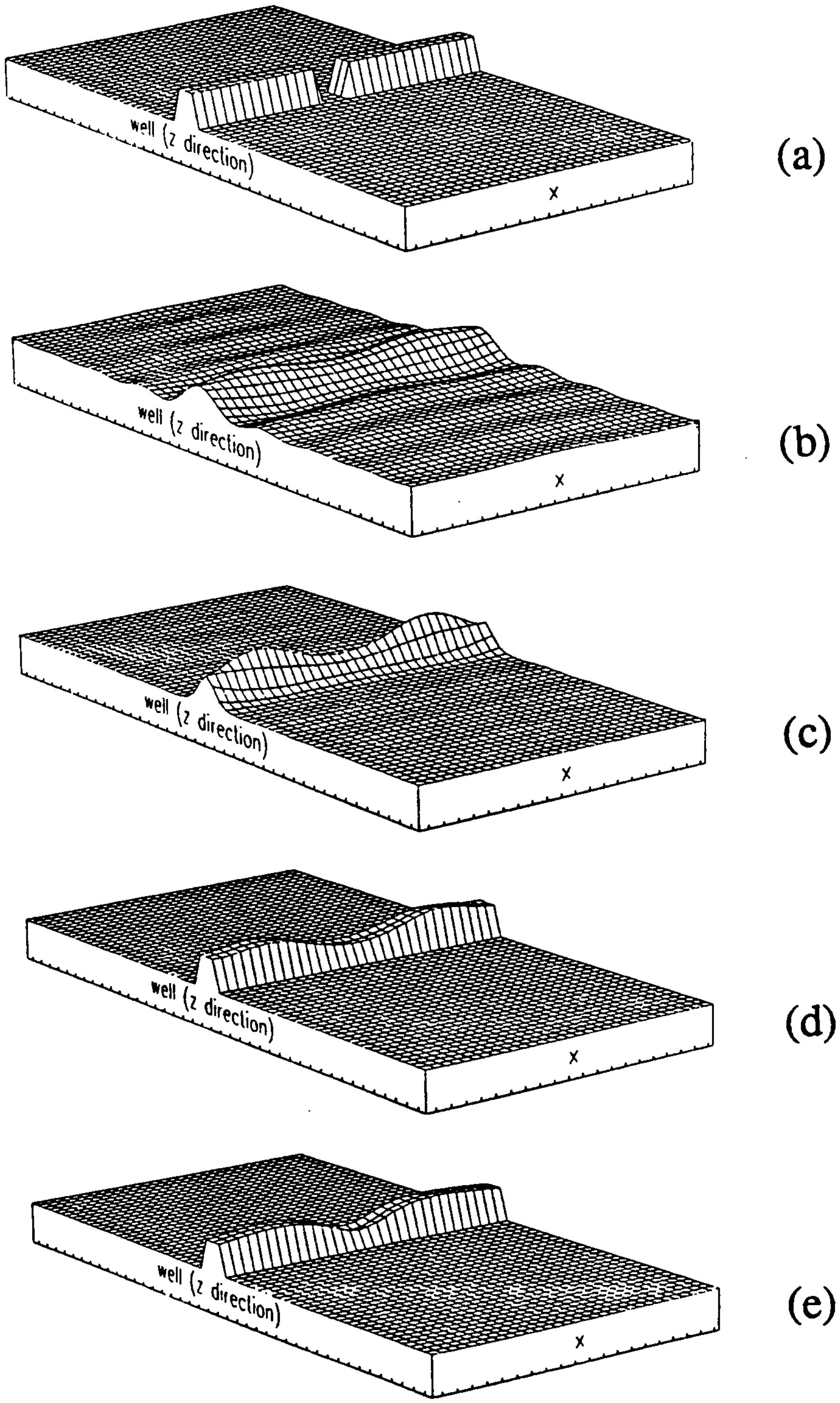


Figure 4: (a) A synthetic slow horizontal layer with a permeability barrier. (b) Result of zeroing out the unsampled components. (c) ME enhancement. (d) RE enhancement with a uniform layer as the prior. (e) Very low weight given to the region between the wells.

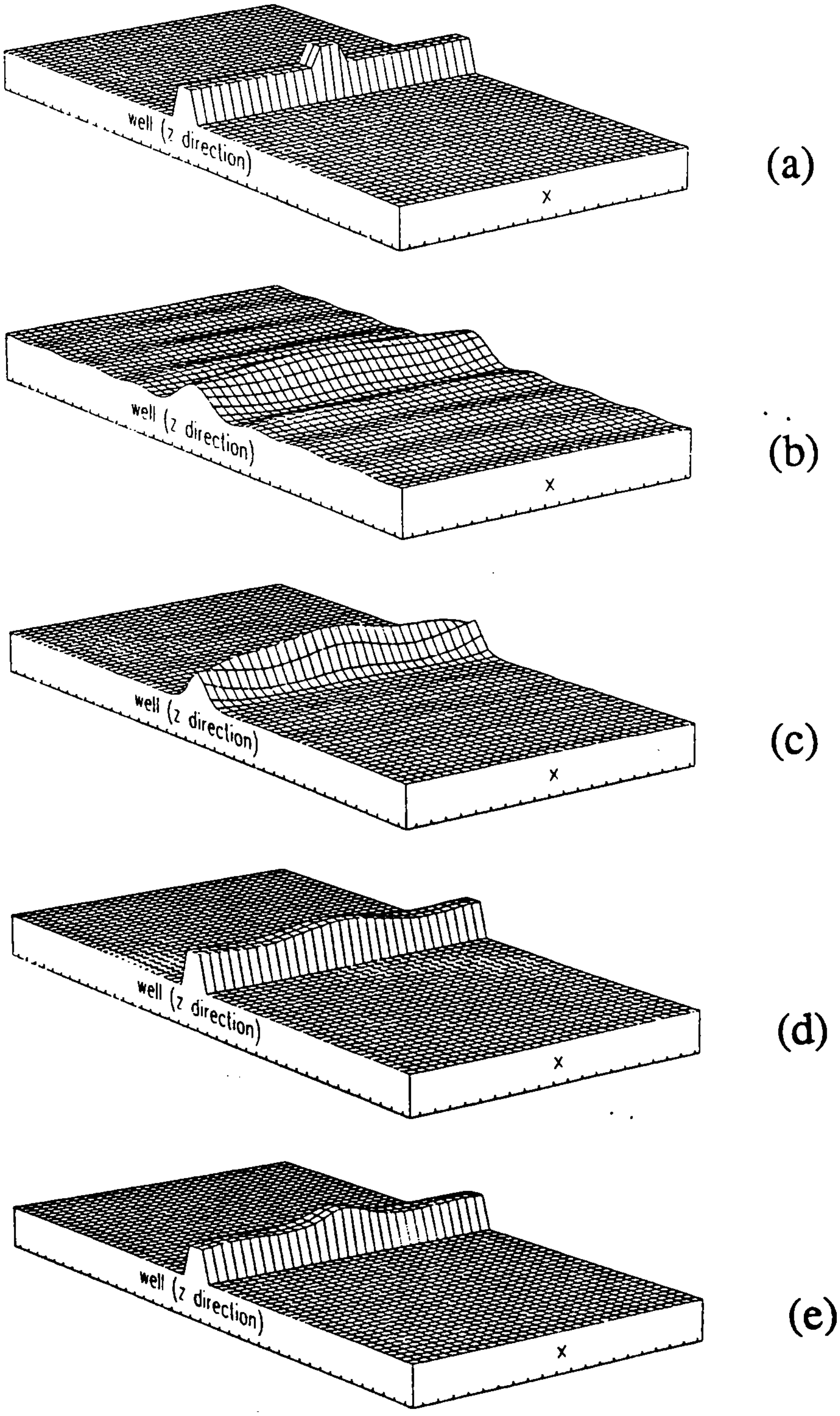


Figure 5: Similar to Figure 4 with the central anomaly slower instead of faster.

PAPER K

BAND-PASS DECONVOLUTION FOR SHORT WAVELETS

Luis L. Canales
Seismic Tomography Project

ABSTRACT

In this paper we present a method for band-pass deconvolution. Band pass deconvolution is very important for reliable estimation of the effective source wavelet, since it minimizes the effect of the out-of-band noise on the phase inside the band. The deconvolution operator is defined as the solution of a constrained least squares problem and an exact solution is given. This is a time domain method and will be particularly useful for short wavelets. The problem is present in the cross-well tomography data acquisition with vibroseis-like source, when the sweep is very short so that crosscorrelation does not effectively produce a compact wavelet. The method will also be useful for deconvolution after correlation, since the effective wavelet is normally very short.

INTRODUCTION

The source for the cross-well tomography project is often a high frequency vibroseis-like band-limited source. Deconvolution must take into account the lack of energy at low and high frequencies. Band-pass deconvolution will be desirable in order not to whiten the data in the noisy part of the spectrum and to minimize the effect of such noise in the phases in the high S/N band-pass.

Band-pass deconvolution has been studied by Deeming(1987). Deeming's method modifies the estimated autocorrelation so that the prediction filter is band-limited. Essentially the modification puts high energy outside of the frequency band of interest, so that the one-step-ahead prediction filter does not contain energy at those frequencies.

Here we will treat the band limited prediction problem as a constrained least squares problem and we give the exact solution to the problem.

MATRIX FORM OF THE PREDICTION PROBLEM

The prediction problem consists of finding a convolutional filter that predicts the data from its past. The one-step-ahead prediction filter turns out to be the inverse of the source wavelet and it forms the basis for spiking deconvolution.

Convolution can be represented with the following matrix form:

$$\begin{bmatrix} x_1 & 0 & \cdots & 0 \\ x_2 & x_1 & \ddots & 0 \\ x_3 & x_2 & \ddots & 0 \\ x_4 & x_3 & \ddots & x_1 \\ x_5 & x_4 & \ddots & x_2 \\ \vdots & x_5 & \ddots & x_3 \\ x_N & \ddots & \ddots & x_4 \\ 0 & x_N & \ddots & x_5 \\ 0 & 0 & \ddots & x_5 \\ 0 & 0 & \cdots & x_N \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_M \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ \vdots \\ y_N \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Where x_k are the elements of one sequence and f_k the elements of the other, normally x_k is a time series and f_k is a filter. The output y_k is normally represented in terms of z-transforms as:

$$X(Z)\,F(Z) = Y(Z)$$

And the z-transform of the error $E(Z)$ is then:

$$E(Z) = X(Z)\,F(Z) - Y(Z)$$

In terms of matrices the prediction problem can be stated as follows:

$$\mathbf{X\,f - y = e}$$

For spiking deconvolution (one-step-ahead prediction) the elements of the desired output are:

$$y_k = x_{k+1}$$

For gap deconvolution with prediction distance g the desired output is then:

$$y_k = x_{k+g}$$

The elements of the filter **f** are obtained by minimizing the power of the error.

THE BAND-PASS CONSTRAINT

In matrix form the Discrete Fourier Transform is:

$$\mathbf{W}\,\mathbf{f} = \mathbf{F}$$

Where $\mathbf{W}$ is a matrix whose coefficients are $e^{2\pi i(k-1)(j-1)/N}$, $\mathbf{f}$ is the vector representation of the filter in the time domain and $\mathbf{F}$ is its Discrete Fourier Transform. Of course $\mathbf{W}$ and $\mathbf{F}$ are complex matrix and vector respectively, but can be reorganized to involve only real arithmetic. It is also possible to use the cosine or sine transform, in which case $\mathbf{W}$ is made up with either the real part or the imaginary part of the initial transformation. We will use the cosine transform, leaving for later the decision as to which is the best strategy to use.

We now define a constraint matrix $\mathbf{Q}$, formed with the rows of $\mathbf{W}$ that represent the frequencies outside the desired band. Then the fact that $\mathbf{f}$ is band-limited is represented with the following constraint equation:

$$\mathbf{Q}\,\mathbf{f} = 0$$

In normal band-limited data, the rows of $\mathbf{Q}$ will be the first and last few rows of $\mathbf{W}$. This will effectively account for the undesired low and high frequencies.

BAND-PASS DECONVOLUTION

The band-limited deconvolution operator $\mathbf{f}$ is computed by solving the following Constrained Least Squares Problem:

$$\mathbf{X}\,\mathbf{f} \approx \mathbf{y}$$
$$\mathbf{Q}\,\mathbf{f} = 0$$

This is the same least squares problem defined by Deeming(1976), but with our constraints. The choice of $\mathbf{y}$ will either give spike deconvolution or gap deconvolution.

The solution is readily available (Claerbout, 1976). After some algebraic manipulation the solution is:

$$\mathbf{f} = \mathbf{f}_0 - \mathbf{R}^{-1}\mathbf{Q}^T(\mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^T)^{-1}\mathbf{Q}\;\mathbf{f}_0$$

where $\mathbf{f}_0$ is the solution to the unconstrained problem and $\mathbf{R} = \mathbf{X}^T\mathbf{X}$ is the autocorrelation matrix.

We note that in this method we are modifying the inverse of the autocorrelation matrix, whereas Deeming's method modifies the autocorrelation itself.

CONCLUSIONS

We have presented an exact deconvolution method that acts only in the pass-band of interest. The technique whitens the data only on the given pass-band, is implemented in the time domain and works with short filters.

We show that this method is equivalent to modifying the inverse of the autocorrelation function while Deeming's method modifies the autocorrelation itself.

In standard deconvolution the noise outside of the frequency band affects the phase in the frequency band. This is not the case with band-pass deconvolution, as the deconvolution operator will not introduce time shifts that are dependent on noise outside of the working frequency band. This is of course very important in seismic tomography in which timings are essential.

REFERENCES

Claerbout J. F. 1976, Fundamentals of Geophysical Data Processing, McGraw-Hill, New York.

Deeming T. J. 1987, Band-Limited Minimum Phase, Deconvolution and Inversion, M.H. Worthington Editor, Blackwell Scientific Publications.

STP DIRECTORY

Name:	Phone:	Login name:
Faculty:		
Jerry Harris	(415) 723-0496	harris
Gary Mavko	(415) 723-9580	gary
Research Associates:		
Dan Moos	(415) 723-3463	moos
Richard Nolen-Hoeksema	(415) 723-0092	richard
Graduate Students:		
Mike Fitzpatrick	(415) 723-9410	fitz
Caroline Lambert	(415) 723-9410	caroline
Spyros Lazaratos	(415) 723-9410	spyros
Reinaldo Michelena	(415) 723-9410	reinaldo
Tomography lab	(415) 723-9410	
Tomography staff:		
Robin Lewis	(415) 723-1099	lewis

Geophysics Department FAX:	(415) 725-7344
Stanford University FAX:	(415) 723-0010
Stanford University Telex:	348402 STANFORD STNU

RESEARCH PERSONNEL

Luis Canales - received his BS in civil engineering at the University of Mexico in 1970. He received his MS (1973) and Phd (1975) in geophysics at Stanford. He worked at CICESE (Center for Graduate Studies and Scientific Research of Ensenada Mexico) as a professor of geophysics and head of the computer center. He worked at Mobil Oil Co. as a research geophysicist in Dallas. Additionally, he spent eight years in England with Digicon Geophysical Inc. where he was the director of special projects. Luis invented the method for random noise reduction know as LUNA, FXPrediction or FXDeconvolution. He is a member of SEG, EAEG, SSA, AGU. His interests are exploration geophysics, time series analysis, and digital signal processing.

Mike Fitzpatrick - received his BS in electrical & computer engineering from Clarkson University in 1986. He worked at Texas Instruments from 1986-1988 in Houston designing and testing microcontrollers. He will receive his MS in Computer Science from Stanford in June, 1990. Mike's work with SRB began in April of 1989 and included system administration and developing the TIMS system software package.

Jerry M. Harris - received his BS (1973) from the University of Mississippi. His MS (1974) from the California Institute of Technology, and his PhD in (1980) electrical engineering from the California Institute of Technology. From 1974 to 1977, Jerry was with the Communications Satellite Corp., working mainly as an experimentalist on millimetre microwave scattering and attenuation in the lower atmosphere. After receiving his PhD, Jerry joined Exxon Production Research Company where he worked on adaptive beam steering and scattering inhomogenous media and became group leader of the experimental seismic methods group. In 1984, Jerry joined the Standard Oil Company to lead development of the seismic cross-well tomography project. Jerry joined Stanford in 1988 as an associate professor of geophysics. His current research interests include experimental methods in seismology and electromagnetics for imaging and measurement of in-situ rock properties, and more general topics involving wave physics and signal processing.

Caroline Lambert - received her BS from the University of Alberta, Canada, in 1980, in Geology and Geophysics. She received a MS in Geophysics from California Institute of Technology in 1982. After receiving her master's degree, she worked for a three years for Geometrics, located in Sunnyvale, as a geophysicist working on aeromagnetics. She came to Stanford in 1987 and will complete a MS in Computer Science in June this year, with a specialization in graphics and image processing. She has been writing software for the SRB group since May last year.

Spyros Lazaratos -received his diploma from the National Technical University of Athens, Greece, in 1985 and a MS from Princeton University in 1987, both in electrical engineering. He has been a graduate student at Stanford since 1987. He is working toward a PhD in geophysics. His current research interests include seismic inversion, migration and tomography. He is a member of SEG and IEEE.

Gary Mavko - received his PhD in geophysics from Stanford in 1977. He then joined the Tectonophysics branch of USGS in Menlo Park where he worked in areas of rock physics and earthquake fault mechanics. In 1984 Gary joined Entropic Geophysical Inc., in its first moths as a start-up reflection seismic processing company. Gary developed may of Enthropic's algorithms and software for reflection and refraction analysis, and eventually became their Vice President of research and development. He returned to Stanford in February, 1989 as an acting associate professor in geophysics, and has been working on modeling and analysis of the acoustic properties of rocks, cross-well seismic imaging and interpretation, and techniques of seismic reflection processing.

Reinaldo Michelena - received his B.S. in Physics in 1984 from Universidad Simon Bolivar, Venezuela. From 1985 to 1988, he worked as a research geophysicist for Intevep, S.A. He received his M.S. in Geophysics from Stanford University in 1990 and is currently working toward a PhD in geophysics under a Intevep, S.A. scholarship. He is a member of SEG.

Daniel Moos - received his bachelor's degree in geology from Cornell University in 1977, and his MS and PhD in geophysics from Stanford University. Dan spent three years with the Borehole Research Group at Lamont-Doherty Geological Observatory, working on the acquisition and analysis of geophysical logs for the Ocean Drilling Program, where he helped design and implement the research logging program and participated in three research cruises. Dan returned to Stanford in the fall of 1986 as a post doc, and has been working on the measurement of in situ stress sand physical properties in boreholes, both on the continents and in oceanic crust. Most recently, Dan has been working on the joint interpretation of cross-well seismic and geophysical well log data.

Richard C. Nolen-Hoeksema - received his B.A. with majors in physics and geology from Hope College and his PhD in geophysics from Yale University. After post-doctoral research at Yale, he worked in the Exploration Research Department at Cities Service Oil & Gas Corporation and in the Reservoir Engineering Division at Chevron Oil Field Research Company. His research interests are in rock physics, rock mechanics, and their application to subsurface characterization and engineering. He has been a Research Associate in Geophysics at Stanford University since 1988.

Pierre Samec - graduated from the Ecole des Mines de Paris in 1986. He then joined the SRP group and was sponsored by Alf Aquitaine for the first eighteen months. Pierre was awarded a Landmark interpretation station in 1989 for his proposal on "Interactive modeling and interpretation of cross well seismic data". Pierre has also been involved in the upgrading and management of the SRP computer facility for the past two years. His research activities are focused on numerical modelling of the wave equation, borehole geophysics, and cross well data interpretation. His activities also include software development on the Landmark Open Works platform. Pierre currently plans on finishing his dissertation by the end of this year.

STP SPONSORS

AMOCO Research Center
Henry Tan
4502 E. 41st St.
Tulsa, OK 74135

British Petroleum
Steve Hildebrand & Ed Witterholt
P O Box 4587 (5151 San Felipe)
Houston, TX 77210

Gas Research Institute
Bruce Smith
8600 Bryn Mawr Avenue
Chicago, IL 60631

INTEWEP
Dept. Tecnologia de Yacimientos
Section Petrofisica
Apartado 76343 - Intewep
Caracas 1070 A
Venezuela