Stanford Exploration Project, Report 82, May 11, 2001, pages 1–164

Datuming velocity from traveltime tomography

Dimitri Bevc[1]

ABSTRACT

Near-surface velocity anomalies cause kinematic and focusing effects which can be ameliorated by wave-equation datuming and layer replacement. The trick is to find the correct datuming velocity. When the anomalies are shorter in lateral extent than the cable length, residual times picked from deep non-hyperbolic events can be used for traveltime tomography. The optimal reconstruction is obtained by constraining the model to the near surface and by estimating the inverse covariance matrix with a prediction error filter. Downward continuation with the traveltimes from this model, followed by upward continuation with a replacement velocity, removes most of the unwanted distortions.

INTRODUCTION

Seismic data are often plagued by near-surface velocity variations which adversely affect processing, imaging and interpretation. Focusing and defocusing of the wavefield as it propagates through these anomalies causes kinematic and amplitude distortions which cannot be adequately compensated by conventional static shift methods. Kjartansson (1979) demonstrated that these types of anomalies can be detected by performing an inversion with either trace energy calculated in a time gate or with residual traveltimes calculated by crosscorrelating traces in a time gate. He chose his time gate to be substantially deeper than the location of the velocity anomalies and performed his inversion by assuming linear raypaths. This method can be implemented by performing iterative slant stacks of the trace energy or residual traveltime data in midpoint-offset coordinates (Claerbout, 1993). The resulting depth images locate the anomalies in midpoint and depth, but do not provide a method of estimating the velocity. I use traveltime tomography to determine the near-surface velocity for datuming and layer replacement. By restricting the tomographic solution to the near surface, I can use the assumption of straight raypaths in the near surface while implicitly allowing the raypaths in the deeper part of the model to be curved. This relieves some of the constraints imposed by Kjartansson's assumption that the entire raypath is straight. Another benefit of restricting the solution to the near surface is that the problem becomes more overdetermined. I implement the tomography so that I simultaneously solve for the prediction error filter (PEF) which estimates the inverse covariance matrix (Claerbout, 1994). This last feature of the inversion brings the inverted amplitudes up to the correct values and reduces tomographic artifacts. The method is designed to estimate near-surface velocity anomalies which are shorter in lateral extent than the cable

[1]**email:** dimitri@sep.stanford.edu

length. A basic assumption is that the wavefield can be propagated through the near-surface to some datum below the anomalies. If the appropriate velocity is used for datuming, the effects of the anomalies are removed from the data. I start out by reviewing the nature of problem and by describing how I generate synthetic data. I then describe the tomographic inversion and use the velocities obtained from it to generate traveltimes for Kirchhoff datuming and layer replacement.

NEAR-SURFACE VELOCITY ANOMALIES

The effect of near-surface velocity anomalies is seen in a CMP gather from Kjartansson's Grand Isle data set as non-hyperbolic moveout and amplitude variation with offset (Figure 1a). These effects are caused by anomalies which are shorter in lateral extent than the recording cable. Figure 1b is a synthetic shot gather generated using a Kirchhoff modeling program. I have not attempted to model the Grand Isle data exactly, but it is evident that the synthetic embodies the same non-hyperbolic moveout and amplitude variation with offset as the real data. The effect of the amplitude variation is more evident in the moveout corrected CMP's of Figure 2. These are the same CMP locations as in Figure 1. The top synthetic was generated with slow near-surface velocity anomalies, the middle synthetic was generated with fast near-surface velocity anomalies. The bottom frame is the CMP gather of Figure 1a after NMO. The slow synthetic matches the character of the data better than the fast synthetic. In Figure 3, I

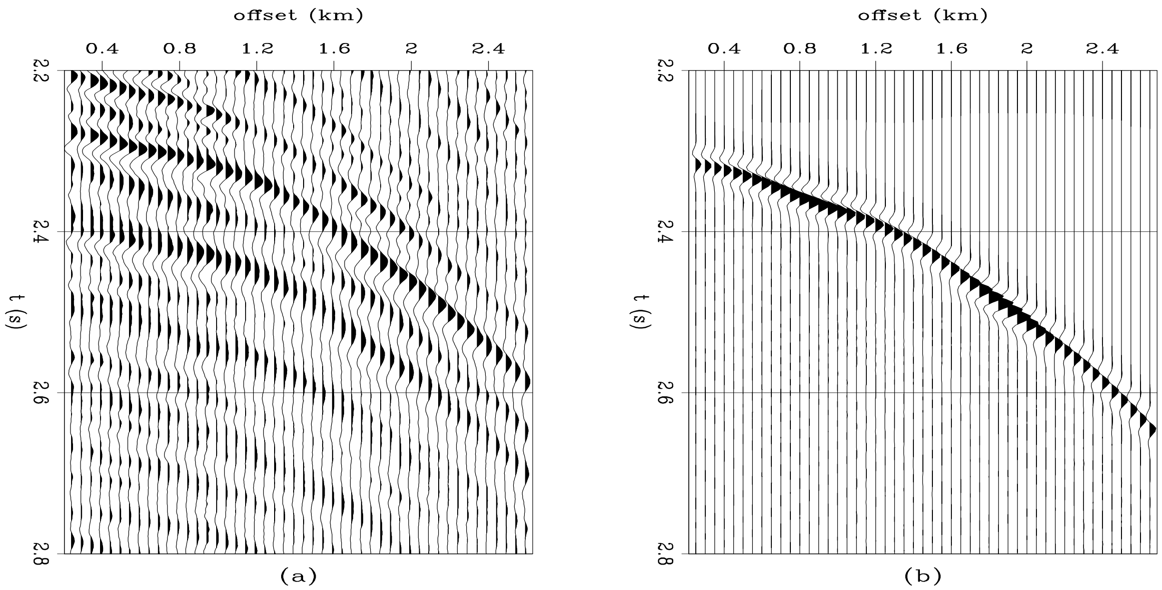


Figure 1: (a) A windowed portion of a CMP gather from the Grand Isle data set. The events have kinematic and amplitude distortions due to near-surface velocity anomalies. (b) A synthetic CMP gather with the same types of distortions. dimitri-realsyn [ER]

have applied Claerbout's (1993) implementation of Kjartansson style tomography. Performing iterative slant stacks of trace power in midpoint-offset coordinates has located the velocity anomalies laterally and in depth. Two prominent low velocity anomalies are located near 7

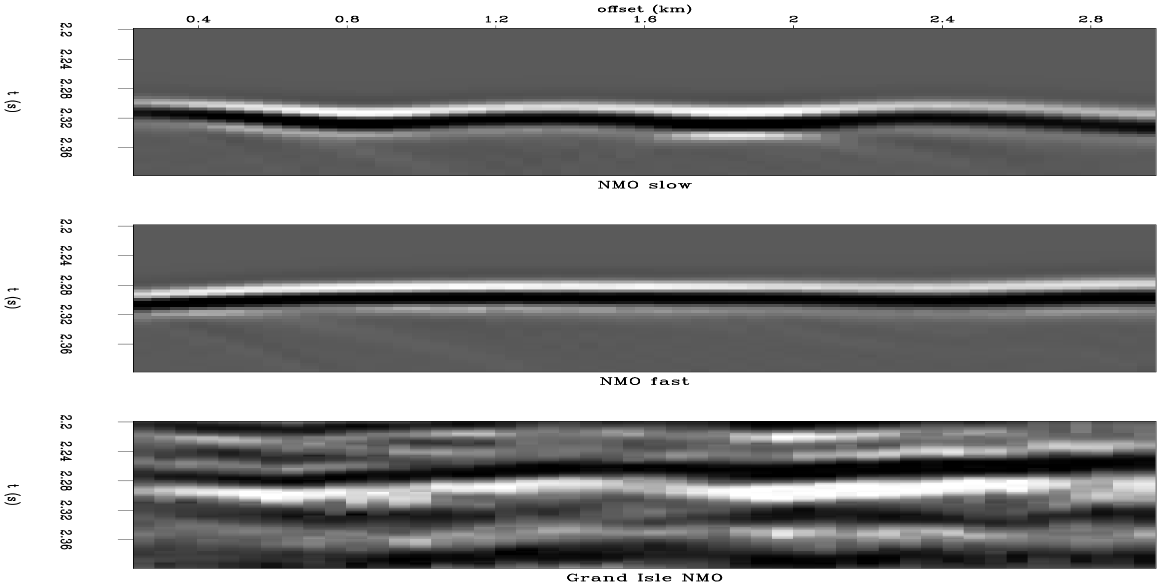


Figure 2: NMO corrected Synthetic data with slow (top) and fast (middle) near-surface anomalies, compared to a NMO corrected Grand Isle CMP gather (bottom). dimitri-realsynnmo [ER]

km at a depth of about 500 m. The result of performing the same type of tomography on the residual traveltime picks is presented in Figure 4. The traveltimes were generated by crosscorrelating the normal moveout corrected traces and picking the peak correlation amplitude. As in Kjartansson's case, the traveltime image is of poorer quality than the amplitude image.

Modeling near-surface anomalies

The synthetic gathers shown in Figures 1 and 2 are generated using Kirchhoff modeling and datuming codes similar to those described in previous reports (Bevc, 1993; Bevc, 1994). Finite difference traveltimes were used to calculate the necessary Green's functions for the Kirchhoff datuming algorithm (van Trier and Symes, 1991; Popovici, 1991). The model used to generate the synthetics consists of three near-surface velocity anomalies in a constant velocity background of 2.1 km/s (Figure 5). There is a single reflector at a depth of 2.4 km and the anomalies are centered at depths of 300 m, 500 m, and 700 m. Representative synthetic shot gathers are shown for slow and fast anomalies in Figures 6 and 7. The peak values of the velocity anomalies deviate about 15 % from the background velocity. Figure 8 displays the slow velocity model, the corresponding trace amplitude plot, and the result of performing the Kjartansson style tomography. The width and amplitude of the trace power streak increases with increasing depth because the Fresnel zone increases. This has the effect of making the image of the shallowest anomaly much weaker than the other two. The image displays typical tomographic artifacts such as elongation along the depth axis and "X" patterns due to back-projection trajectories. The imaged anomalies are laterally smaller than the actual anomalies because a straight ray assumption is used in the inversion. Slow anomalies appear smaller

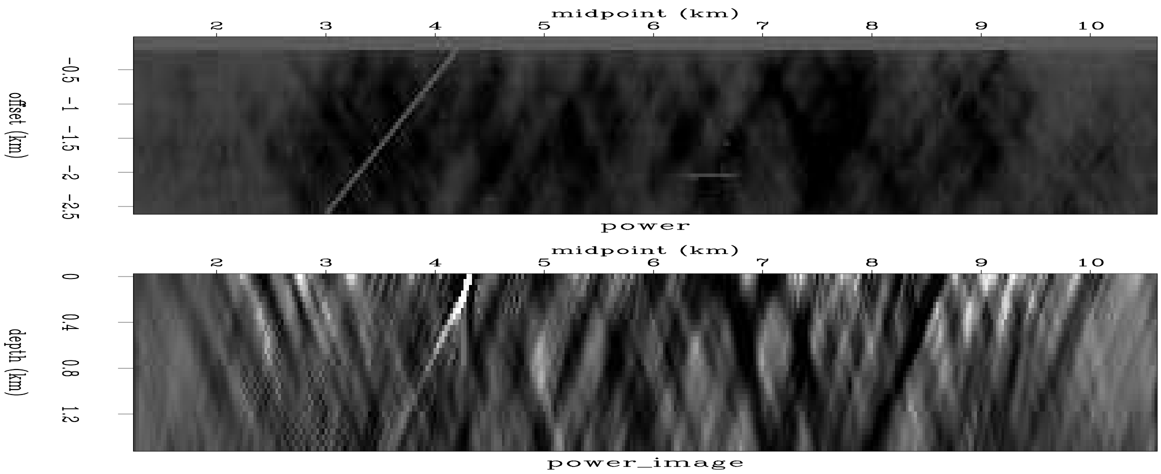


Figure 3: Kjartansson style tomography on the Grand Isle data trace power plots. dimitri-gitomo.power [ER]

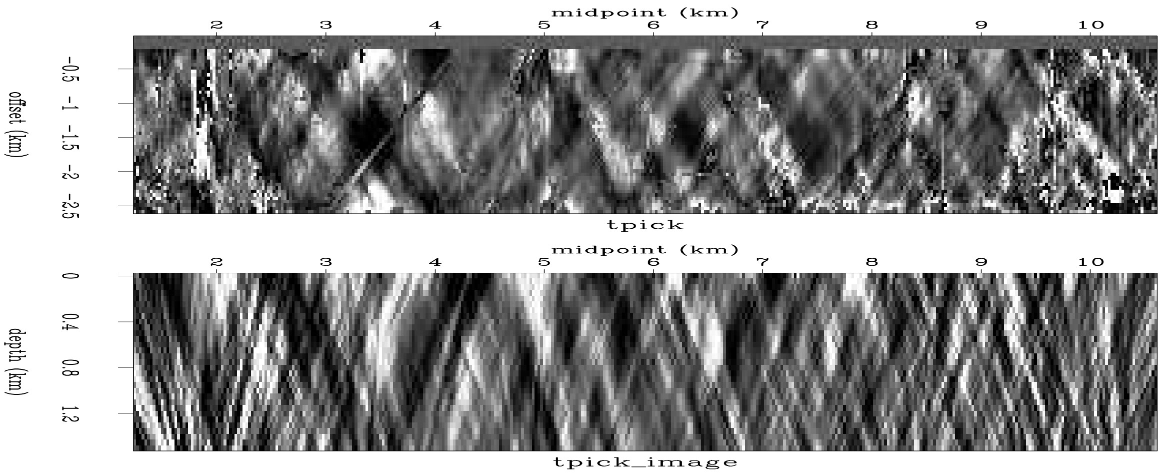
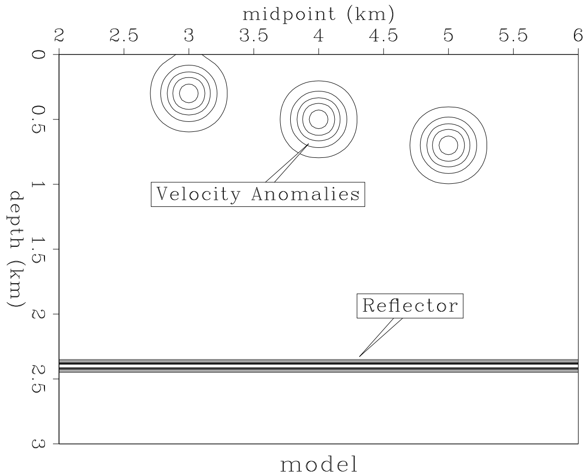


Figure 4: Kjartansson style tomography on the Grand Isle data residual traveltime picks. dimitri-gitomo.tpick [ER]

Figure 5: The model used to generate the synthetics has a reflector at a depth of 2.4 km and three anomalies near the surface. Synthetic gathers were generated for both fast and slow near-surface anomalies in an otherwise constant velocity medium. dimitri-model [ER]



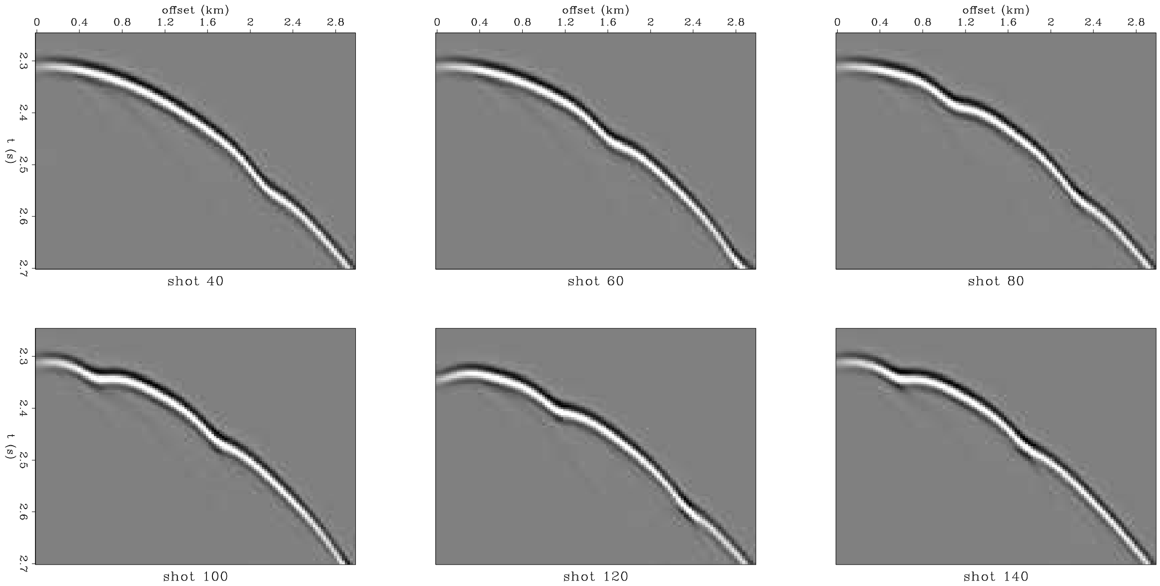


Figure 6: Synthetic data with slow near-surface anomalies. dimitri-syn.slow [ER]

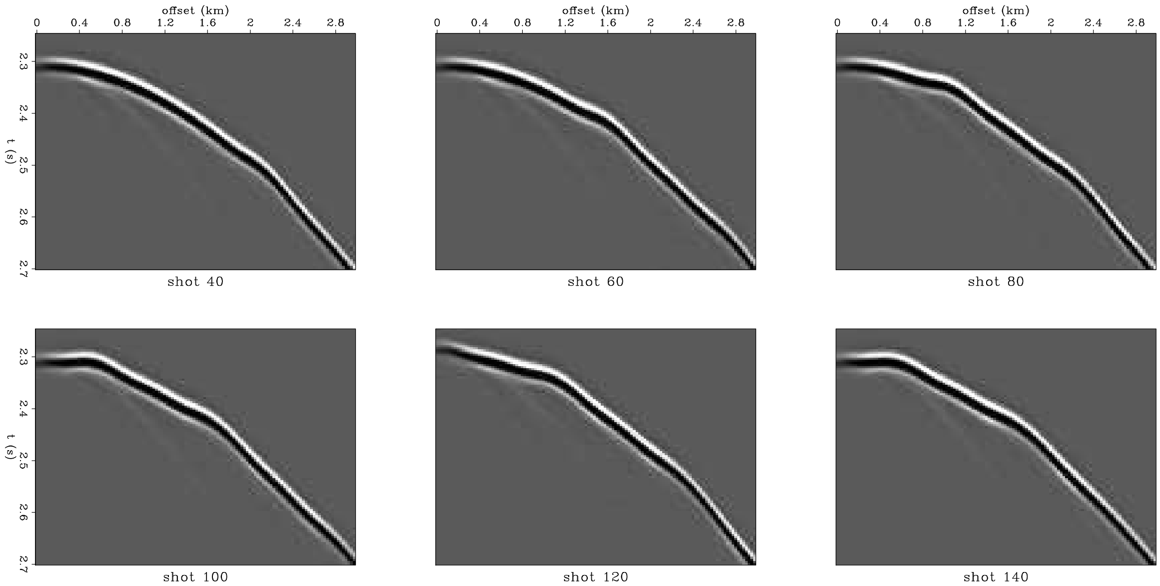


Figure 7: Synthetic data with fast near-surface anomalies. dimitri-syn.fast [ER]

because the true Fermat raypath will tend to go around the slow region; However, under the assumption of straight ray reconstruction, the residual is backprojected along a straight line so that the effects of a slow region are focused into a smaller region (e.g. Berryman, 1991). Figure 9 displays the same results for the fast near-surface velocity anomalies. The images

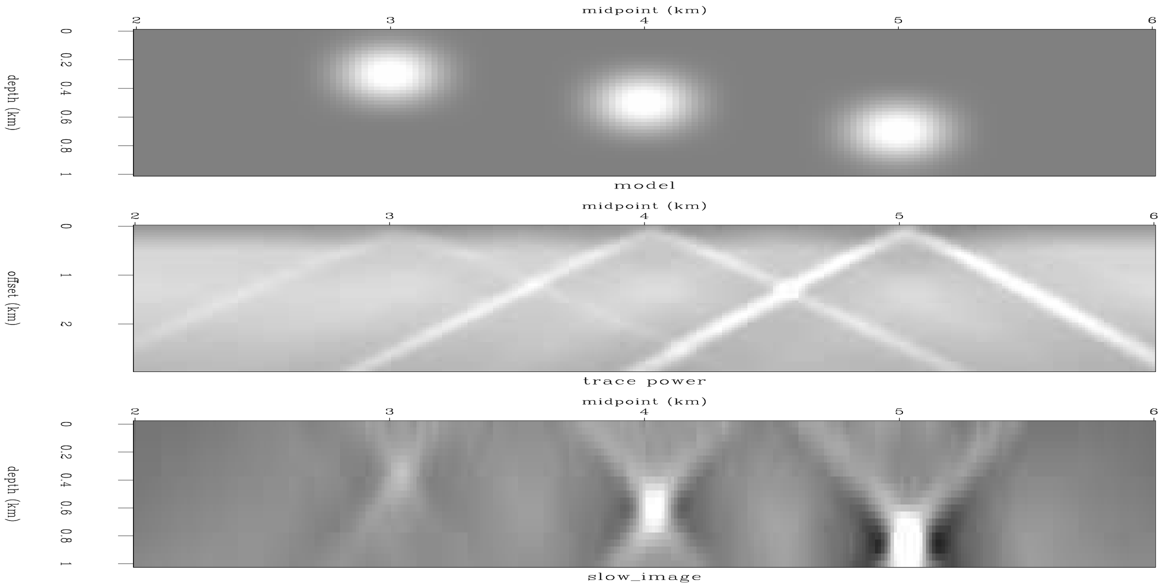


Figure 8: Kjartansson style tomography for slow anomalies. dimitri-kjimage.slow [ER]

appear larger than the actual anomalies because the backprojection of a fast anomaly with straight rays tends to spread into a larger region. The tomographic images presented in this section confirm the presence of near-surface anomalies in the data and the synthetics. The synthetics confirm that kinematic focusing can cause the amplitude anomalies observed in the Grand Isle data. Although Kjartansson's method can detect and locate anomalies, it does not provide a means of estimating the near-surface velocity. In order to do that, I have developed the constrained tomographic method of the next section.

TRAVELTIME TOMOGRAPHY

Wave-equation datuming has been demonstrated to be effective for downward continuing marine data through the water layer and performing layer replacement (Yilmaz and Lucas, 1986; Berryhill, 1986). In that application, the datuming velocity is well known. In my application it is unknown. In order to find the datuming velocity, I backproject residual traveltimes picked from deep non-hyperbolic events. The goal of the method is to improve imaging by removing distortions from events. Except for the near-surface anomalies, the velocity structure of the Grand Isle data is pretty simple, so that normal moveout correction and stacking provide an adequate imaging method. Stack power is maximized when events have hyperbolic trajectories. This means that after NMO correction, events should be flat. A measure of this flatness is to crosscorrelate a pilot trace t_{pilot} with the data after NMO. The maximum lag of this crosscorrelation gives a time shift $\triangle t_k$ which is the residual traveltime to be backprojected (k is a

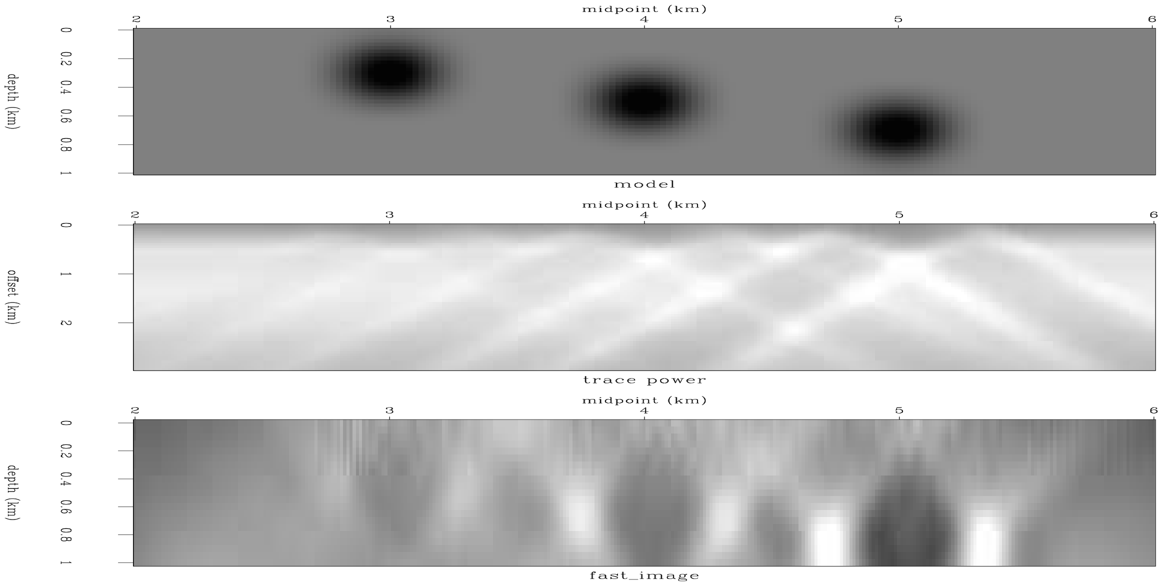


Figure 9: Kjartansson style tomography for fast anomalies. $\boxed{\text{dimitri-kjimage.fast}}$ [ER]

trace index). The objective function to be *minimized* is then

$$Q(\hat{m}) \;=\; \sum_k (\triangle t_k)^2 \;=\; \sum_k \left(t_{pilot} - t_k(\hat{m})\right)^2. \tag{1}$$

The residual time $\triangle t_k$ is directly related to residual slowness $\triangle m$ in the near-surface model $\hat{m}$. The traveltime calculation can be linearized by assuming straight rays and writing the traveltime as

$$\mathbf{t} \;=\; \mathbf{M}\hat{\mathbf{m}}, \tag{2}$$

where $\mathbf{M}$ is the matrix of path lengths which multiply the slowness vector $\hat{\mathbf{m}}$. Rewriting equation (1) as

$$Q(\hat{m}) \;=\; \mathbf{r}\cdot\mathbf{r} \;=\; \left(\mathbf{t}_{pilot} - \mathbf{M}\hat{\mathbf{m}}\right)^T \left(\mathbf{t}_{pilot} - \mathbf{M}\hat{\mathbf{m}}\right). \tag{3}$$

Defining $\hat{\mathbf{m}} = \hat{\mathbf{m}}_i + \triangle\mathbf{m}$, and taking the gradient with respect to the slowness vector, results in the regression:

$$\mathbf{0} \;\approx\; \triangle\mathbf{t} - \mathbf{M}\triangle\mathbf{m} \tag{4}$$

The near-surface slowness model is updated with $\triangle\mathbf{m}$. Including an estimate of the inverse covariance matrix (Menke, 1989) changes the regression to:

$$\begin{aligned} \mathbf{0} &\approx \triangle\mathbf{t} - \mathbf{M}\triangle\mathbf{m} \\ \mathbf{0} &\approx \mathbf{A}\triangle\mathbf{m}. \end{aligned} \tag{5}$$

If the inverse covariance matrix is unknown (as is usually the case), it can be estimated by considering $\mathbf{A}$ to be a prediction error filter with coefficients to be determined (Claerbout, 1994). The model and the prediction error filter can then be written as

$$\begin{aligned} \mathbf{m} &= \mathbf{m}_i + \triangle\mathbf{m} \\ \mathbf{A} &= \mathbf{A}_i + \triangle\mathbf{A}, \end{aligned} \tag{6}$$

where the subscript i indicates initial values which are to be updated. The product of the filter and the model is

$$\begin{aligned} \mathbf{Am} &= (\mathbf{A}_i + \triangle\mathbf{A})(\mathbf{m}_i + \triangle\mathbf{m}) \\ &= \mathbf{A}_i\mathbf{m}_i + \mathbf{A}_i\triangle\mathbf{m} + \triangle\mathbf{A}\mathbf{m}_i + \triangle\mathbf{A}\triangle\mathbf{m}. \end{aligned} \tag{7}$$

The regression equation (4) then becomes:

$$\begin{aligned} \mathbf{0} &\approx \triangle\mathbf{t} - \mathbf{M}\mathbf{m}_i - \mathbf{M}\triangle\mathbf{m} \\ \mathbf{0} &\approx \mathbf{A}_i\mathbf{m}_i + \mathbf{A}_i\triangle\mathbf{m} + \triangle\mathbf{A}\mathbf{m}_i + \triangle\mathbf{A}\triangle\mathbf{m}. \end{aligned} \tag{8}$$

Dropping the nonlinear term and rearranging, the final regression can be written as:

$$\begin{aligned} \triangle\mathbf{t} - \mathbf{M}\mathbf{m}_i &\approx \mathbf{M}\triangle\mathbf{m} \\ -\mathbf{A}_i\mathbf{m}_i &\approx \mathbf{A}_i\triangle\mathbf{m} + \triangle\mathbf{A}\mathbf{m}_i. \end{aligned} \tag{9}$$

The data space and model space residuals are defined as:

$$\begin{aligned} \mathbf{r}_d &= \triangle\mathbf{t} - \mathbf{M}\mathbf{m}_i \\ \mathbf{r}_m &= -\mathbf{A}_i\mathbf{m}_i. \end{aligned} \tag{10}$$

Packing the unknown filter coefficients, $\triangle\mathbf{A}$ into the abstract vector $\triangle\mathbf{a}$ and putting the elements of the slowness model $\mathbf{m}_i$ into a convolutional matrix $\bar{\mathbf{m}}$, the last convolution of equation (9) can express as $\bar{\mathbf{m}}\triangle\mathbf{a}$. The regression of equation (9) can then be written in matrix form as

$$\begin{bmatrix} \mathbf{r}_d \\ \mathbf{r}_m \end{bmatrix} = \begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{A}_i & \bar{\mathbf{m}} \end{bmatrix} \begin{bmatrix} \triangle\mathbf{m} \\ \triangle\mathbf{a} \end{bmatrix}. \tag{11}$$

This system of equations is solved using the method of conjugate gradients. The first step in the solution is to initialize the model by calculating

$$\mathbf{m}_i = \mathbf{M}^T\mathbf{t}, \tag{12}$$

and making an initial guess of the filter

$$\mathbf{A}_i = \partial_x^2 + \partial_y^2. \tag{13}$$

The regression equation (11) can then be used to calculate the gradient:

$$\begin{bmatrix} \triangle\mathbf{m} \\ \triangle\mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{M}^T & \mathbf{A}_i^T \\ \mathbf{0} & \bar{\mathbf{m}}^T \end{bmatrix} \begin{bmatrix} \mathbf{r}_d \\ \mathbf{r}_m \end{bmatrix}, \tag{14}$$

and the conjugate gradient:

$$\begin{bmatrix} \triangle\mathbf{r}_d \\ \triangle\mathbf{r}_m \end{bmatrix} = \begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{A}_i & \bar{\mathbf{m}} \end{bmatrix} \begin{bmatrix} \triangle\mathbf{m} \\ \triangle\mathbf{a} \end{bmatrix}. \tag{15}$$

In summary, the conjugate gradient solution begins by initializing the model and filter with equations (12) and (13). The residuals are initialized using equation (10) and iterations begin by calculating the gradient [equation (14)] and the conjugate gradient [equation (15)]. In practice, the algorithm is allowed to run for a while using the Laplacian filter of equation (13) before the prediction error filter coefficients are found.

SYNTHETIC EXAMPLES

The synthetic data of Figures 6 and 7 are used to test the tomography algorithm of the previous section. Figure 10 is a plot of the residual traveltimes which represent the deviation of the curves in Figures 6 and 7 from hyperbolas. The effects of the slow anomalies are confined to narrow trajectories and the effects of the fast anomalies are spread over wider trajectories. The resulting tomograms will be used to calculate traveltimes maps for the Kirchhoff

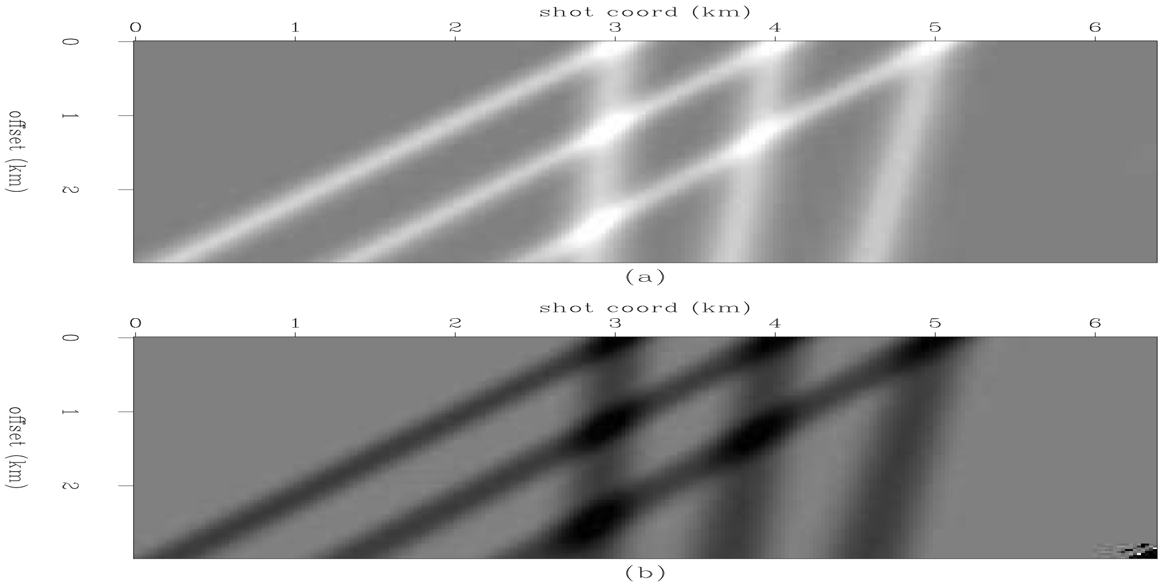


Figure 10: Residual traveltimes picked from (a) the synthetics in Figure 6 and (b) the synthetics in Figure 7. dimitri-picktt [ER]

datuming code. In order to compare the tomographic results, I present the travel time map generated by calculating finite-difference traveltimes in the near-surface velocity model depicted in Figure 5. Figure 11 is a display of the residual traveltime from each subsurface location at 1 km depth (horizontal axis) to all surface locations (vertical axis). These traveltimes can be considered to be the goal of the tomographic inversion since they are the desired datuming traveltimes. The tomographic image obtained with the simplest backprojection algorithm (equation (4)) is compared to the original depth model in Figure 12. Although this image is considerably better than the Kjartansson style tomogram of Figure 8, it exhibits undesired artifacts such as vertical elongation and “X” patterns. This slowness image is used to calculate the finite-difference traveltime map displayed in Figure 13. This map does not match the traveltime map of Figure 11 very well. Figures 14 and 16 are a comparison of the original depth model to several tomographic images. The slow images (Figure 14b to 14d) look smaller than the fast images (Figure 16b to 16d) because the residuals are backprojected along straight lines. The amplitudes of the various tomograms are compared in Figures 15 and 17. Using the prediction error formulation of equation (9) (frame `c` in Figures 14 through 17) is a significant improvement over the simpler formulation of equation (4). (frame `b` in Figures 14 through 17): The amplitudes in frame `c` look smoother and the images are better located. The

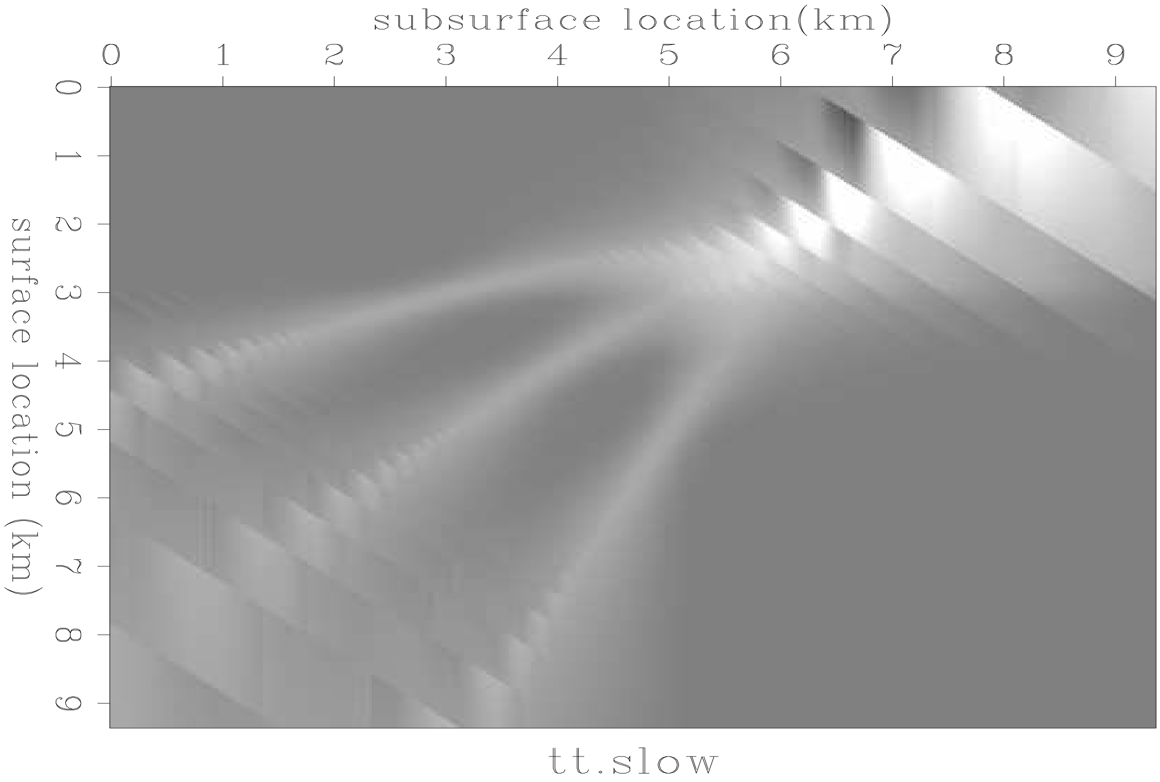


Figure 11: Residual traveltime map corresponding to the slowness model depicted in Figure 5 using slow near-surface velocity anomalies. Fast anomalies result in a similar map with reversed polarity. dimitri-resid.tt.slow [ER]

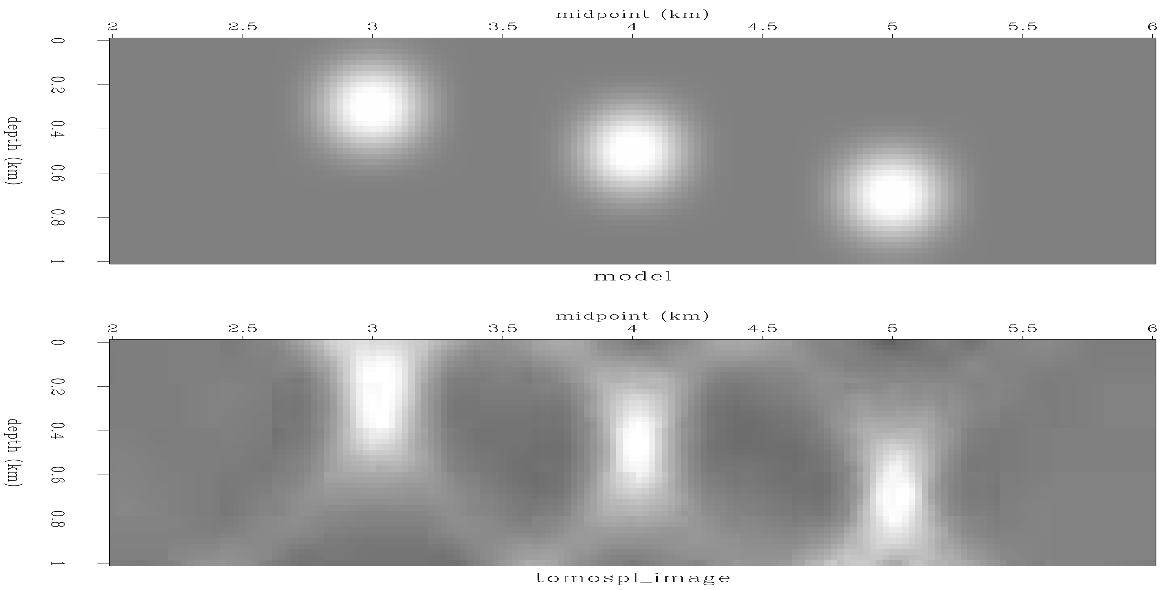


Figure 12: Comparison of the original slowness model (top) to the tomographic image (bottom), reconstructed without any estimate of inverse covariance. dimitri-tomo.tomospl.slow [ER]

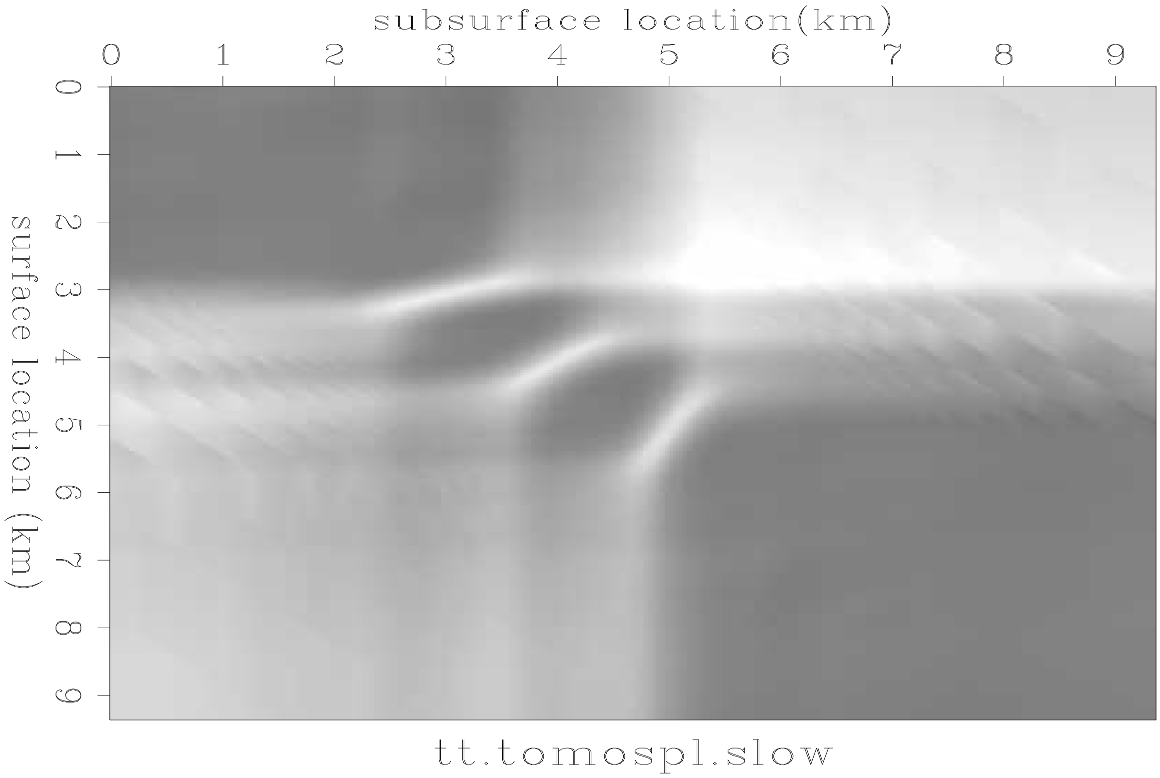


Figure 13: Residual traveltime map based on the slowness model of Figure 12. This map does not match Figure 11 very well. dimitri-resid.tt.tomospl.slow [ER]

reconstruction artifacts are also ameliorated. By using the knowledge that straight ray reconstruction broadens fast anomalies and compresses slow anomalies, the information from the images generated by equation (9) in frame c of Figures 14 through 17 can be used to add additional constraints to the inversion. The result of doing this is displayed in frame d of Figures 14 through 17. These images match the original depth models (frame a of Figures 14 through 17) pretty well. The most significant improvement is in the amplitudes and in the reduction of reconstruction artifacts. A closeup of the inversion result using the prediction error filter to

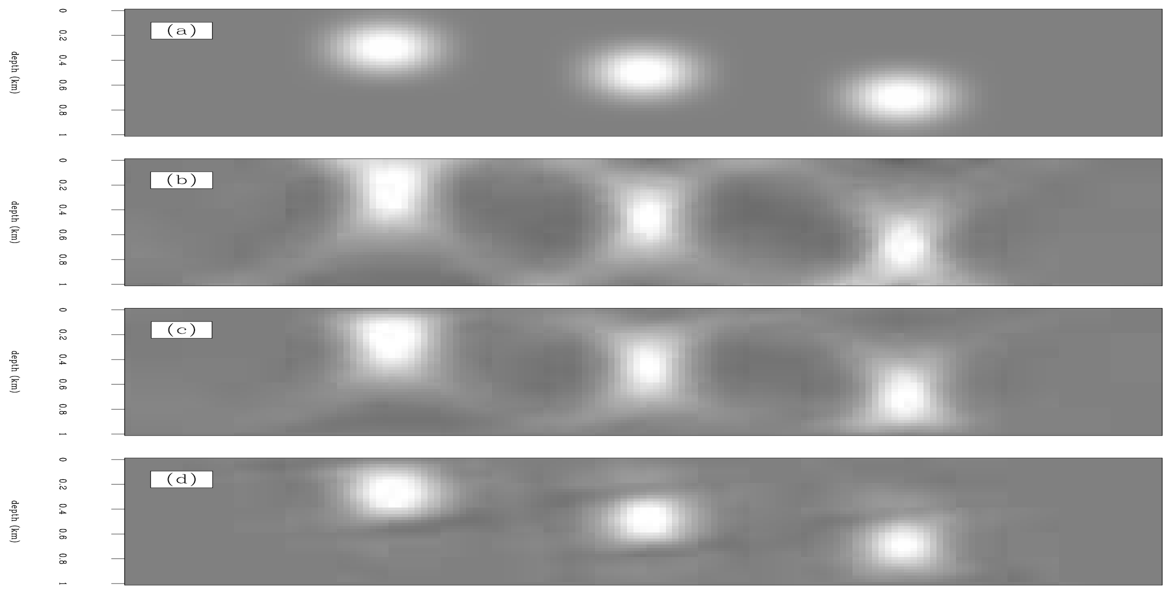


Figure 14: Depth model and tomograms for slow near-surface anomalies. The vertical axis is depth and the horizontal axis is midpoint. (a) Original model with slow near-surface anomalies. (b) Tomographic inversion using equation (4). (c) Tomographic inversion using equation (9). (d) Tomographic inversion using equation (9) and additional model constraints. dimitri-comptomo.slow [ER]

estimate inverse covariance as implied by equation (9) and model constraints is displayed in Figure 18. This is the same image as Figure 14d. The corresponding traveltime map, shown in Figure 19, matches the desired traveltime map of Figure 11 almost exactly.

LAYER REPLACEMENT

The traveltimes of Figure 19 and a similar set of traveltimes for the fast near-surface anomalies are used to generate kinematic Green's functions for Kirchhoff wave-equation datuming. The result of layer replacement for the slow and fast models is shown in Figures 20 and 21. The events now exhibit hyperbolic moveout.

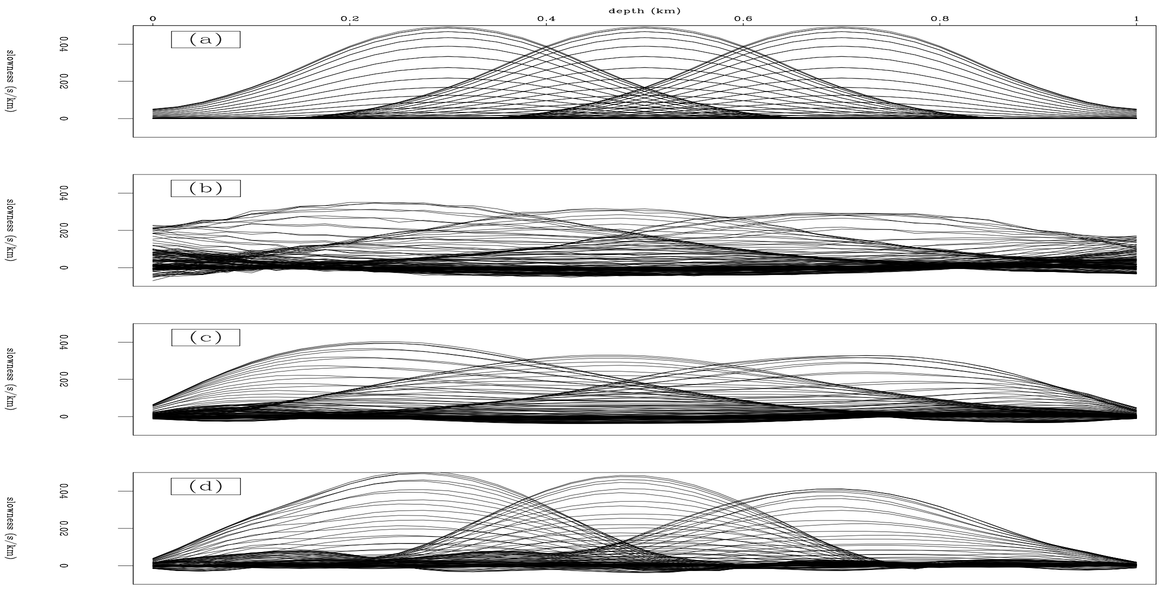


Figure 15: Residual slowness amplitudes for slow near-surface anomalies. The vertical axis is residual slowness and the horizontal axis is depth. (a) Original model with slow near-surface anomalies. (b) Tomographic inversion using equation (4). (c) Tomographic inversion using equation (9). (d) Tomographic inversion using equation (9) and additional model constraints. dimitri-compamp.slow [ER]

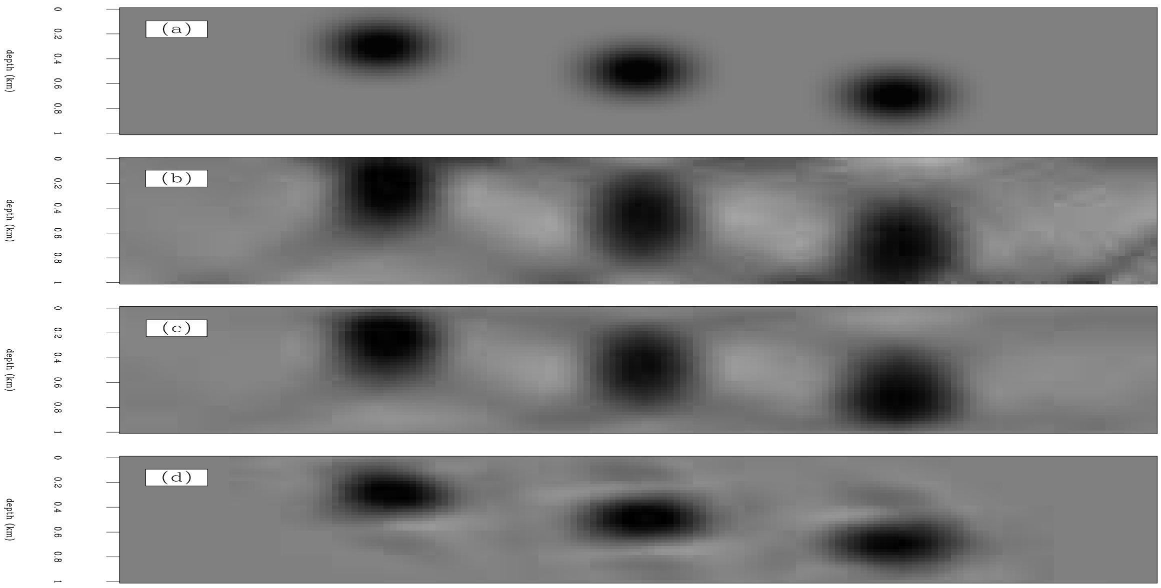


Figure 16: Depth model and tomograms for fast near-surface anomalies. The vertical axis is depth and the horizontal axis is midpoint. (a) Original model with fast near-surface anomalies. (b) Tomographic inversion using equation (4). (c) Tomographic inversion using equation (9). (d) Tomographic inversion using equation (9) and additional model constraints. dimitri-comptomo.fast [ER]

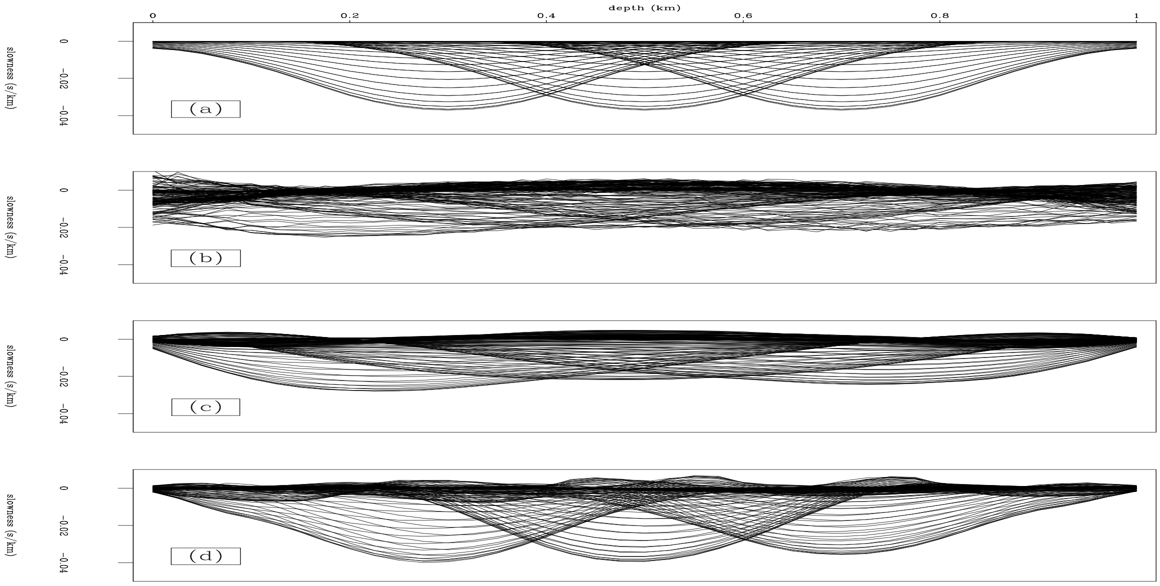


Figure 17: Residual slowness amplitudes for fast near-surface anomalies. The vertical axis is residual slowness and the horizontal axis is depth. (a) Original model with fast near-surface anomalies. (b) Tomographic inversion using equation (4). (c) Tomographic inversion using equation (9). (d) Tomographic inversion using equation (9) and additional model constraints. dimitri-compamp.fast [ER]

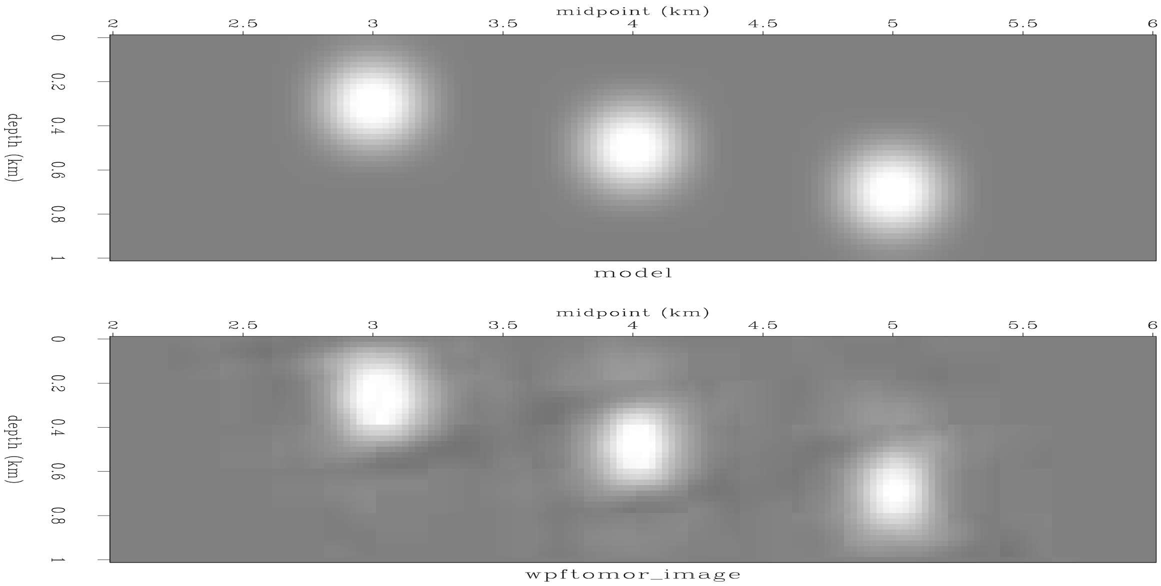


Figure 18: Comparison of the original slowness model (top) to the tomographic image (bottom), reconstructed by using weights and PEF to estimate inverse covariance. dimitri-tomo.wpftomor.slow [ER]

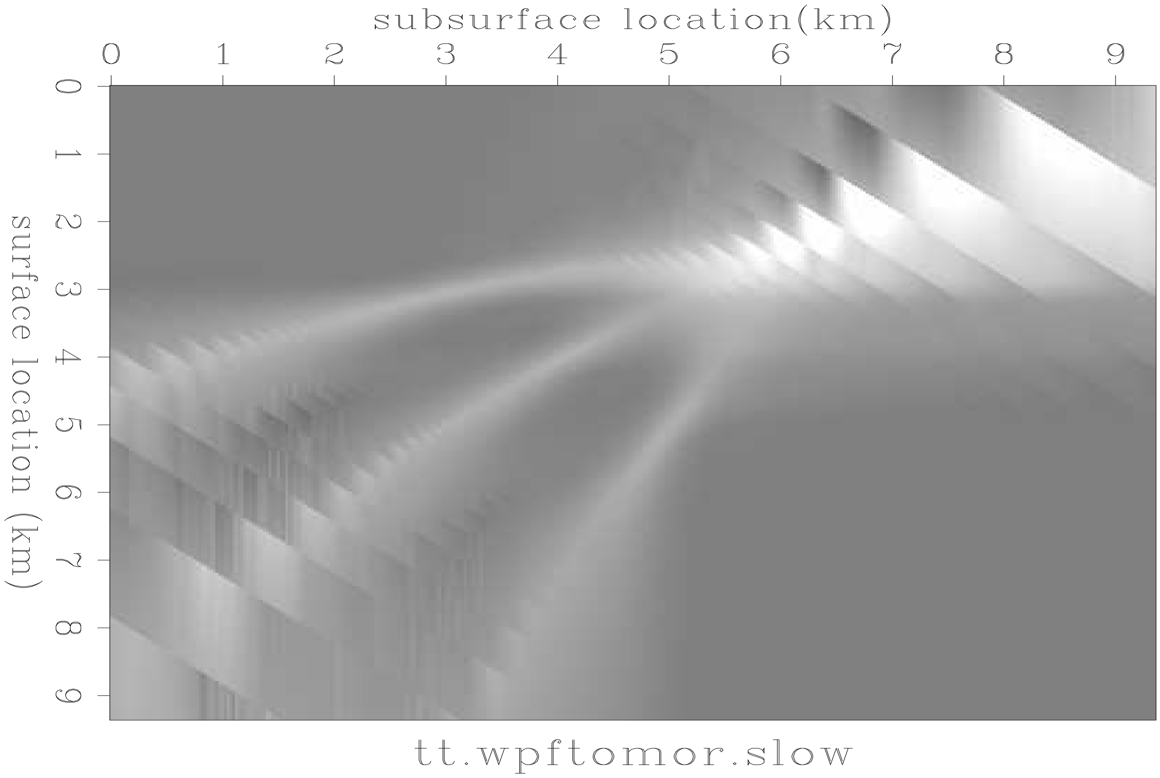


Figure 19: Residual traveltime map based on the slowness model of Figure 18 and used for layer replacement in Figure 20. This map compares favorably with Figure 11 dimitri-resid.tt.wpftomor.slow [ER]

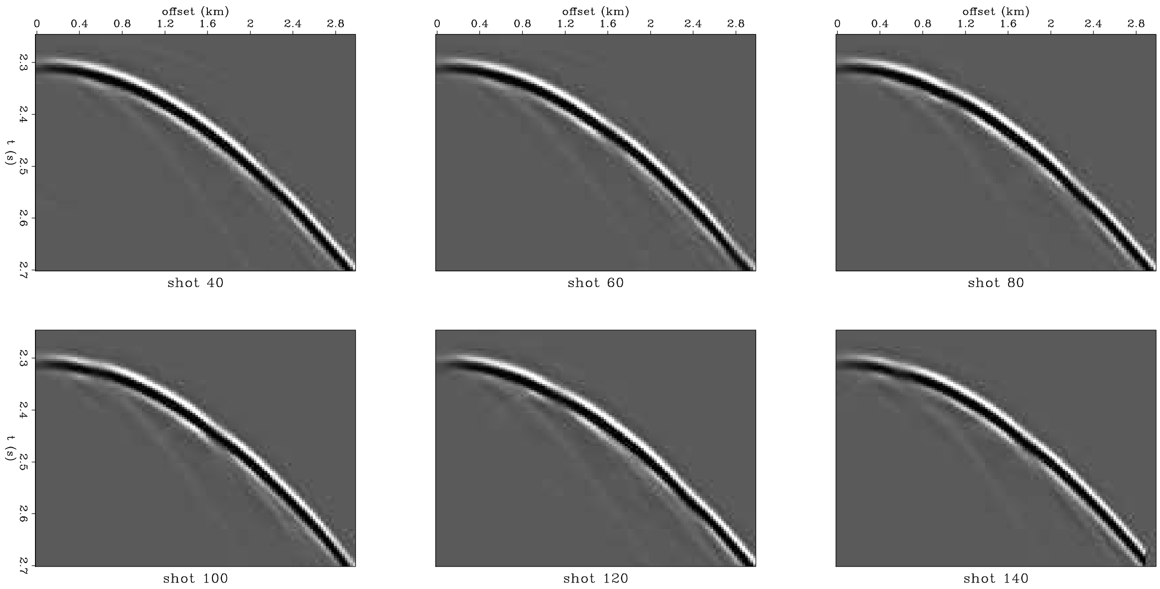


Figure 20: Layer replacement with slow near-surface anomalies. dimitri-rep.wpftomor.slow [ER]

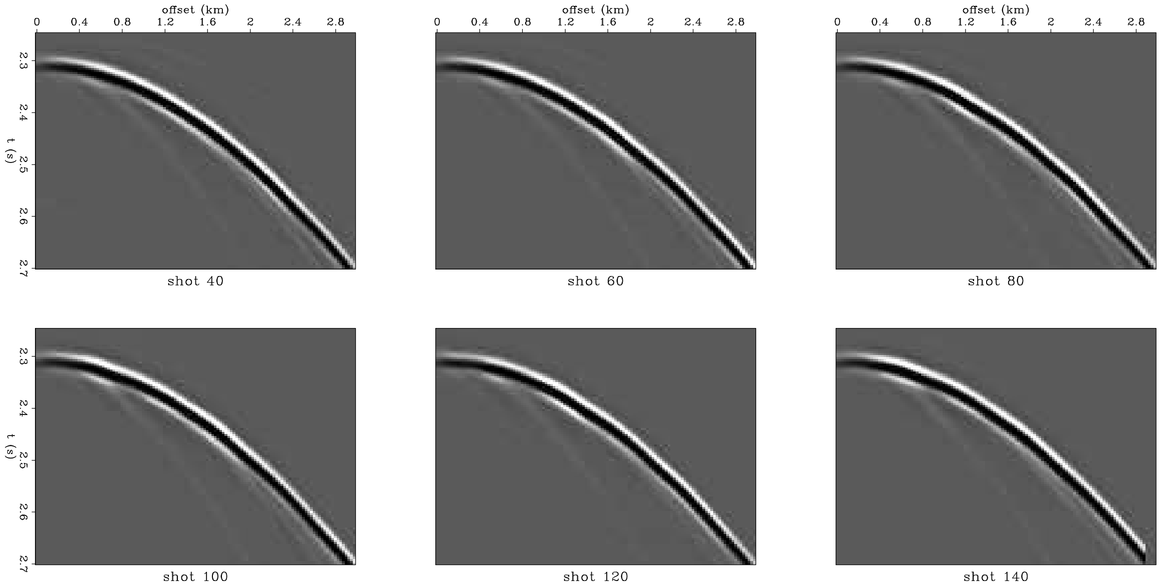


Figure 21: Layer replacement with fast near-surface anomalies. dimitri-rep.wpftomor.fast [ER]

DISCUSSION AND CONCLUSIONS

I backproject residual traveltimes picked from deep non-hyperbolic events to determine the near-surface velocity for datuming and layer replacement. The best image is obtained by constraining the solution and by solving for both the slowness and the coefficients of the prediction error filter which estimates the inverse covariance matrix. Once the near-surface velocity model is found, wave-equation datuming is used to perform layer replacement and remove unwanted distortions. The tomographic solution may be non-unique; however, this is not a problem because the ultimate goal is to remove distortions from deeper events. So it does not matter if the velocity used for datuming is an accurate representation of the of the actual velocity. It only needs to provide an adequate representation of the kinematics. I am not interested in imaging the near surface, just in removing its deleterious effects. The method presented here works well on synthetics. It may not work as well on real data, so a more sophisticated velocity estimation method, such as the one outlined in the Appendix, may be required.

REFERENCES

Berryhill, J. R., 1979, Wave equation datuming: Geophysics, **44**, 1329–1344.

Berryhill, J. R., 1984, Wave equation datuming before stack (short note): Geophysics, **49**, 2064–2067.

Berryhill, J. R., 1986, Submarine canyons - velocity replacement by wave equation datuming before stack: Geophysics, **51**, 1572–1579.

Berryman, J. G., 1991, Lecture notes on nonlinear inversion and tomography: Lawrence Livermore National Laboratory, Report No. UCRL-LR-105358-Rev.1.

Bevc, D., 1993, Data parallel wave equation datuming: SEP–**77**, 131–140.

Bevc, D., 1994, Near-surface velocity estimation and layer replacement: SEP–**80**, 361-380.

Biondi, B., 1990, Velocity analysis using beam stack: Ph.D. thesis, Stanford University.

Claerbout, J., 1993, Reflection tomography: Kjartansson revisited: SEP–**79**, 59–67.

Claerbout, J., 1994, Applications of three dimensional filtering: manuscript in preparation.

Etgen, J., 1990, Residual prestack migration and interval velocity estimation: Ph.D. thesis, Stanford University.

Fowler, P., 1988, Seismic velocity estimation using prestack time migration: Ph.D. thesis, Stanford University.

Kjartansson, E., 1979, Attenuation of Seismic Waves in Rocks and Applications in Energy Exploration: Ph.D. thesis, Stanford University.

Menke, W., 1989, Geophysical data analysis: Discrete inverse theory: Academic Press, Inc.

Popovici, A. M., 1991, Finite difference travel time maps: SEP–**70** 245–256.

Toldi, J. L., 1985, Velocity analysis without picking: Ph.D. thesis, Stanford University.

van Trier, J., and Symes, W. W., 1991, Upwind finite-difference calculation of traveltimes: Geophysics, **56**, 812–821.

Yilmaz, O., and Lucas, D., 1986, Prestack layer replacement: Geophysics, **51**, 1355–1369.

APPENDIX A

An alternative to traveltime tomography is to perform a wavefield inversion by formulating the problem in terms of maximizing a non-quadratic objective function. This formulation is more computationally intensive, but it does not require crosscorrelation and picking and may be less sensitive to noise. Similar approaches are employed by other authors for different velocity estimation settings (Toldi, 1985; Fowler, 1988; Biondi, 1990; Etgen, 1990)

Maximizing stack power

In formulating the objective function, I assume that I can get a good estimate of the overall background velocity structure, minus the near-surface anomalies (i.e., approximate stacking trajectory). After wave-equation datuming with the best estimate of near-surface velocity, the stack power is maximized. Formally, this can be expressed as maximizing the following non-linear objective function:

$$Q_p(\hat{m}) \;=\; \sum_t \sum_{x_s} \left(\sum_{x_g} \mathrm{Nmo}\, P(x_s, x_g, t(x_s, \xi_s, x_g, \xi_g, \hat{m})) \right)^2 . \tag{A-1}$$

In this formulation, the data have been downward continued from the surface having shot and group coordinates (ξ_s, ξ_g), to a datum below the near-surface anomalies having shot and group coordinates (x_s, x_g). I express the Kirchhoff downward continuation (without the cluttering detail of amplitude and phase terms) as a summation over shot and group coordinates:

$$P(x_s, x_g, t) \;=\; \sum_{\xi_s} \sum_{\xi_g} P(\xi_s, \xi_g, \tau + tt(x_s, \xi_s, \hat{m}) + tt(x_g, \xi_g, \hat{m})) \tag{A-2}$$

In this equation, each input trace is time shifted by an amount that depends on the traveltime tt between input and output locations and summed into the output location (Berryhill, 1979; Berryhill, 1984). $Q_p = (P(t(\hat{m})))$ is a complicated function of traveltime and slowness. The gradient of the objective function with respect to the model parameters is calculated by applying the chain rule:

$$\begin{aligned} \nabla_{\hat{m}} Q_p \;&=\; \left.\tfrac{\partial Q_p}{\partial m}\right|_{m=\hat{m}} \;= \\ 2 \textstyle\sum_t \sum_{x_s} &\left[\left(\textstyle\sum_{x_g} \mathrm{Nmo}\, P(x_s, x_g, t(m)) \right) \left(\textstyle\sum_{x_g} \mathrm{Nmo} \tfrac{\partial P}{\partial t}(x_s, x_g, t) \tfrac{\partial t}{\partial m}(x_s, \xi_s, x_g, \xi_g) \right) \right] \end{aligned} \tag{A-3}$$

The derivative of the downward continued data is evaluated numerically using finite differences. The derivative of time with respect to the model parameters can be most easily evaluated by linearizing the traveltime relationship, as in equation (2) of the main text. The non-linear inversion implied by equation (A-3) then proceeds by (1) evaluating the gradient $\nabla_{\hat{m}} Q_p$ and (2) using it to update the slowness model $\hat{m} = \hat{m} + \eta \nabla_{\hat{m}} Q_p$. The step size η is determined so that it maximizes Q_p. Iterations are continued until convergence.

Maximizing semblance

Instead of maximizing stack power, which favors large amplitudes, a better alternative might be to maximize semblance,

$$S(x_s,t) = \frac{\left(\sum_{x_g} P(x_s,x_g,t)\right)^2}{\sum_{x_g} P^2(x_s,x_g,t)},$$

so that the objective function becomes

$$Q_s(\hat{m}) = \sum_t \sum_{x_s} S(x_s, t(x_s,\xi_s,x_g,\xi_g,\hat{m})). \tag{A-4}$$

The gradient of the objective function with respect to the model parameters is then

$$\left.\frac{\partial Q_s}{\partial m}\right|_{m=\hat{m}} = \left.\frac{\partial S}{\partial t}\right|_{t=t(\hat{m})} \left.\frac{\partial t}{\partial m}\right|_{m=\hat{m}}. \tag{A-5}$$

The time derivative of semblance is evaluated numerically and the derivative of time with respect to the model parameters can be evaluated by linearizing the traveltime relationship. The inversion proceeds as before.