Case Studies of Seismic Tomography and Earthquake Location in a Regional Context

William Menke

Lamont-Doherty Earth Observatory of Columbia University

We present three case studies of traveltime tomography applied to regional imaging of the earth's crust and upper mantle. The first case study investigates the imaging of Poisson's ratio using measurements of compressional and shear wave traveltimes. Inversions schemes that jointly fit both types of traveltime data and that selectively damp the compressional to shear wave velocity ratio are shown to be very effective. The second case study focuses on the ability of joint inversions of crustal and Moho-reflected wave traveltimes to image both crustal velocity structure and Moho topography. The ability of this kind of tomography to distinguish Moho topography and lower crustal heterogeneity is shown to be poor. The third case study examines the ability of teleseismic tomography to determine the shape of a mantle plume. A new test for departures from an axial shape is proposed and tested. We also present a case study of earthquake location, that compares locations made with traditional P and S wave traveltimes to those based on differential P wave traveltimes (that is, the double-difference method). The underlying tomography software, "raytrace3d" that is used in this study and which is publicly and freely available is described, and the technical elements of traveltime tomography are fully documented in an appendix.

INTRODUCTION

Since its development in the 1970's, seismic traveltime tomography has had a profound impact upon our understanding of the internal structure of the earth and on the practice of seismology. The literature associated with applications of tomography is huge, and we can make no attempt to review it here. Studies range in scale from a few tens of meters, as in cross-borehole tomography, to the whole earth, and are relevant to issues as diverse as the pattern of mantle convection, the origin of mantle plumes, the formation of continental roots, mantle upwelling and melting at mid-ocean ridges, the maximum depth of subduction, crustal genesis, the magmatic plumbing of volcanoes, the width of fault zones, and the pattern of fractures in the uppermost crust. Furthermore, the desire to image three dimensional earth structures in ever greater detail had driven a tremendous growth in seismic instrumentation, including permanent observatories and short-term deployable land and ocean-bottom seismic stations. This instrumentation permits the deployment of dense arrays that provide high spatial resolution measurements of the seismic wavefield, such as are needed to correctly identify non-first arrival phases and to effectively use natural sources such as earthquakes (Figure. 1). The new instrumentation, as well as its use in large aperture arrays, has had significant impact on other areas of seismology as well (e.g. studies of the earthquake

1

source).

The purposes of this paper are three-fold. The first is to introduce "raytrace3d", a freely available software program for raytracing, seismic traveltime inversion, and earthquake location in three-dimensional isotropic media. The second is to discuss several issues in seismic inversion that are of current interest in regional inversions for the structure of the crust and upper mantle, and earthquake location, using case studies performed with the raytrace3d program. The third is to review the underlying algorithms used in raytracing and similar seismic inversion programs, to make the inner workings of such software more accessible, and a bit less of a "black box".

RAYTRACE3D

There is no single route to seismic tomography. Broadly-speaking, most computer programs that implement tomographic algorithms use measurements of seismic traveltimes to produce estimates of the earth's velocity structure. But as with many other complicated data analysis procedures, many different algorithms and solution strategies can (and are) employed to achieve this overall result. A software developer has perhaps a half-dozen major decision points, each with several alternatives. As a result, the number of variants of tomographic algorithms are legion. We enumerate a few of the major considerations here.

1. Velocity parameterization. While the earth's velocity field varies continuously with position, any practical description of it requires that it be represented by a finite – though perhaps large – number of parameters. Constant-velocity voxels (the three-dimensional analog of pixels), splines and truncated orthogonal expansions (e.g. Fourier series, spherical harmonics) can all be used to this purpose, and all have their advantages and disadvantages. Voxels, for instance, are conceptually simplest, but need to be tiny in size (and hence huge in number) to realistically represent regions of the earth with sharp velocity gradients. Orthogonal expansions are smooth, but are non-local, and can therefore "transport" error to unexpected parts of the model. Splines, which use an interpolation scheme to compute the velocity between a set of control points (or "nodes") are conceptually (and computationally) more complex than voxels, but more localized than orthogonal expansions. For reasons that will be described below, raytrace3d uses linear splines, with nodes forming the vertices of the tetrahedra.

2. Method of computing traveltimes. Two method of computing traveltimes – each with many variants – are commonly used in tomography. Both rely upon an approximation to the wave equation called "ray theory" that is valid only at high frequencies (that is, it omits wave diffraction). The first method, "raytracing" first finds rays – paths along which seismic energy propagates –

connecting a given source and receiver, and then determines the traveltime along each ray. The second, "Eikonal solver" method finds the wavefront (surfaces of equal traveltime) that form concentric shells about the source, and then infers the ray paths from their shape. Current implementations of the Eikonal solver method are faster than raytracing, but suffer from a serious limitation: While more than one ray may connect a given source and receiver, the Eikonal solver can find only the first-arriving of them. In cases where observations of secondary phases are of interest, raytracing may be the preferred method. Raytrace3d uses raytracing, however it does so without explicitly solving the ray equation (which would ordinarily require numerical integration of a differential equation), since the path of a ray inside of a linear spline is exactly an arc of a circle. FAST, another publicly available tomography code, is based on an Eikonal solver [*Zelt, 1998; Zelt and Barton, 1998*].

3. Choice of unknowns. One of the underlying complexities of seismic tomography is that knowledge of the ray paths are required to estimate the unknown velocity structure, but those ray paths depend upon the unknown structure. This problem is usually handled though linearization: the velocity structure is divided into a reference or "background" part, $v_0(x,y,z)$, that is assumed to be known, and a perturbation, $\Delta v(x,y,z)$ that is unknown, but assumed to be small. Only v_0 is used in the raytracing, and the effect of omitting Δv on the raypaths is assumed to be negligible. Once the perturbation, Δv, is found, the whole process can be iterated, with an updated reference velocity.

As far as the tomography is concerned, $v_0(x,y,z)$ and $\Delta v(x,y,z)$ represent two entirely different quantities, and completely different methods can be used to represent them. If, as in the case of raytrace3d, $v_0(x,y,z)$ is represented with splines, then four parameters define each node (its velocity and its position in three dimensions). The velocity perturbation, $\Delta v(x,y,z)$, could also represented with splines, but there is no requirement to do so. Raytrace3d takes a different route, and forms $\Delta v(x,y,z)$ by perturbing the position and/or velocity of the v_0 nodes. The actual model parameters are arbitrary linear combinations of these perturbations. The advantage of such a seemingly complex scheme is described below.

3. Solution method. Most tomography use some variant of the principal of least-squares to select a best-fitting $\Delta v(x,y,z)$. They select a model that minimizes some measure of traveltime error. The known traveltimes and unknown Δv's are connected by a very large system of linear equations. Many techniques are available for solving such a system. Raytrace3d uses an iterative technique based on the biconjugate gradient algorithm [*Press, 1992*].

4. Method of imposing smoothness on the velocity field. While the earth's velocity field probably has

variability at all scales, any experiment makes only a finite number of observations and can therefore resolve features only down to some minimum size. Furthermore, the idea that energy propagates along distinct rays is itself only an approximation. Any real experiment uses waves with finite frequency that average the seismic velocity field on scales on the order of a wavelength. The tomographic solution for the velocity field ought to have a degree of smoothness commensurate with these factors. (The finite wavelength of seismic waves seriously limits resolution. Consider that a 10 Hz seismic wave – about as high of frequency as is used in crustal imaging experiments – has a wavelength of about half a kilometer).

Two strategies can be applied to impose smoothness: First, the velocity parameterization can be designed to exclude short-wavelength features (e.g. by controlling the node spacing in a spline representation). Second, the the solution can be forced to have a prescribed spatial correlation function, by including some measure of the spatial correlation of the solution in the definition of least-square error. Thus the simple definition of the sqared traveltime error, E:

$$E = \Sigma_i^N \, (T_i^{obs} - T_i^{pre})^2 \qquad (1)$$

where T_i^{obs} and T_i^{pre} are the N observed and predicted traveltimes, respectively, can be augmented to:

$$E = \Sigma_i^N \, (T_i^{obs} - T_i^{pre})^2 + \varepsilon^2 \, \Sigma_i^L \, \Sigma_j^L \, C^{-1}_{ij} \, \Delta v_i \, \Delta v_j \qquad (2)$$

Here L is the number of nodes and ε^2 is the so-called "damping parameter". We usually express this damping in terms of a percentage (defined formally in Section A.17), where 0% corresponds to $\varepsilon^2 = 0$ and 100% means that the two terms of Equation 2 are of equal size.

With a proper choice of the matrix, **C**, the second term in Equation 2 becomes a measure of spatial correlation, that is, how slowly the velocity perturbations vary with position. For example, *Abers* [1994] suggests that **C** should be chosen so that C_{ij} exponentially decreases with the geometrical distance between nodes i and j. The solution exactly minimizes the traveltime error when $\varepsilon^2 = 0$, and becomes very smooth as $\varepsilon^2 \to \infty$. Some intermediate value of ε^2 (usually found by trial and error) corresponds to a solution that is both smooth and comes close to minimizing the least-squares traveltime error. Raytrace3d implements only the first method of controlling model smoothness. However, since several nodes can be grouped into the same model parameter, Δv can be made arbitrarily smoother than v_0.

Raytrace3d also provides a limited form of damping, by defining the error as:

$$E = \Sigma_i^N \, (T_i^{obs} - T_i^{pre})^2 + \varepsilon^2 \, \Sigma_i^M \, \Delta m_i^2 \qquad (3)$$

where Δm_i are the M model parameters. A large damping parameter simply tends to drive the Δm's to zero, that is to select a solution that is small. Selective damping of some model parameters over others can be implemented by judicious choice of the coefficients that define the model parameters. Consider, for example, two model parameters, each consisting of the sum of three nodal velocities, such as $\Delta m_1 = 0.2\Delta v_3 + 0.6\Delta v_4 + 0.2\Delta v_5$ and $\Delta m_2 = 2\Delta v_6 + 6\Delta v_7 + 2\Delta v_8$. Each have the same ratio of coefficients (that is, 2:6:2) and thus impose the same smoothness on the three corresponding nodal velocities. But the damping more strongly affects the first set of nodal velocities, since their coefficients are smaller.

5. Finally, the designer of any software must make some practical choices about its user interface, and especially the degree to which it utilizes interactive graphics. Raytrace3d, at the "low end" of this spectrum, is a command-line oriented program that accepts and produces only text files. Visualization must therefore be provided by other software, such as *Wessel and Smith's* [1991] "GMT" Generic Mapping Tools. Raytrace3d is available by anonymous FTP at the URL:

ftp://ftp.ldeo.columbia.edu/pub/menke/raytrace3d.tar.Z

VELOCITY RATIO CASE STUDY

Poisson's ratio, ν, or equivalently, the ratio of compressional to shear wave velocity, α/β, (with $\alpha^2/\beta^2=2(1-\nu)/(1-2\nu)$ has long been understood to provide an important clue about the temperature of the rocks within the earth.   The ideal Poisson solid has a ratio of $\alpha/\beta = \sqrt{3} \approx 1.73$. Most rock-forming minerals have ratios close to this ideal (e.g. quartz, 1.47; plagioclase (An29), 1.83; augite 1.72; diposide, 1.76) [*Christensen, 1982*]. Rocks are observed to have greater variability, probably because of weathering and jointing, but are generally in the 1.1−1.9 range [*Clark, 1996, Table 7.16*]. On the other hand, rocks near their melting point generally have much higher ratios, typically in excess of 1.9, with the limiting value of $\alpha/\beta \rightarrow \infty$ ($\nu=0.5$) reached for fully fluid melts. Measurements of velocity ratio have been used to infer (or rule out) partial melting and magma ponding within the crust and shallow mantle [e.g. *Gebrande et al., 1980; Menke et al., 1996*].

Standard body wave tomography can be used to image both α and β within the earth, as long as compressional and shear wave traveltimes are available. However, forming the ratio of separately-produced compressional wave, α, and shear wave, β, images to produce an image of α/β can be an unstable process. The two traveltime datasets often have somewhat different ray coverage and noise, owing to the difficulty of adequately exciting shear waves with explosive sources and to the difficulty of determining the traveltime of secondary phases such as the S wave. Individually-produced images of compressional and shear

wave velocity are thus subject to different resolution and error, and will have artifacts in different places. Forming a ratio of two such images only further magnifies these problems.

Shaw [1994] puts forward the idea of a simultaneous inversion for compressional velocity, α, and compressional to shear wave velocity ratio, α/β. The idea here is to separately control the damping of the two parameters, so as to select models that favor heterogeneity in α over heterogeneity in α/β (this is sometimes referred to "squeezing" α/β). The resulting α/β image then has the interpretation of the one with smallest variability in ratio that is consistent with the data.

We illustrate this procedure with a synthetic inversion, using "data" calculated from a hypothetical crustal model (the "true model") containing both variability in α and α/β. The true model has α and α/β defined in a $50\times50\times10$ km region containing $11\times11\times11$ nodes. The model contains a broad compressional velocity anomaly, with a maximum amplitude of $\Delta\alpha=-1$ km/s, superimposed on a one-dimensional structure that monotonically increases with depth. The shear wave velocity has a broad, co-located low-velocity anomaly, with $\alpha/\beta=1.75$, also superimposed on a simple one-dimensional structure with $\alpha/\beta=1.83$ in the uppermost layers. A 21×21 array of receivers is placed on the top surface of the model, and 10 compressional and 10 shear sources are placed in a ring around them at 3 km depth (Figure 2, top). Only 30% of the compressional and shear sources are co-located. A total of 4,096 P wave traveltimes and 4,122 shear wave traveltimes are computed for combinations of these sources and receivers. Traveltimes are then perturbed with Gaussian-distributed random noise with a standard deviation of 0.05s for P waves and 0.10s for S waves, to simulate measurement error. The ray coverage (Figure 2, bottom) has some interesting deficiencies typical of many crustal imaging experiments: Relatively few rays sample the lower part of the crust, owing to the source-receiver offsets being limited to less than 25 km, and those that do are all propagating sub-horizontally; All the rays traverse the upper crust, but owing to the depth of the sources, most of them cross it sub-vertically.

We perform both individual and joint inversions (Figure 3). The individual α and β inversions each have 343 model parameters, the velocity of the nodes within the central part of the model (that is, |x| and |y| less than 15 km). The joint inversion has double this number of model parameters, since both α and α/β are being determined at each node. The individual inversions each have a spatially-homogeneous damping of 5%, and converge in three iterations when begun with a one-dimensional starting model. The joint inversion has a spatially-homogenous damping of 5% for α and 50% for α/β, and converges in 6 iterations when begun with a one-dimensional starting model that has homogenous α/β. All inversions reduce the

r.m.s. traveltime error, measured with respect to the starting model, by similar amounts – in the 83-85% range. This degree of noise reduction corresponds to a slightly "overfit" model, as the signal-to-noise ratio of the data is about 3:1 (that is, a 75% improvement would have been sufficient).

All inversions produce compressional and shear wave images that look – at least superficially – like the true model. The position and shape of the low velocity zone is captured remarkably well. A close examination yields some minor differences between the individual-inversion α and β images and their true counterparts. In particular, there are small differences in the shape of the periphery of the low velocity zone, with the inversions tending to have "shoulders" that are not present in the true model. These small differences yield rather large fluctuations in the ratio, α/β, at the shoulders when ratios of the two images are formed. The joint inversion produces a smoother image that correctly captures the low values of α/β at the center of the model.

Interestingly, neither of the α/β images correctly captures the lateral continuity of the high α/β region at the top of the true model. Both image this layer as discontinuous patches that in three dimensions form a ring about the center of the model. This ring results from the interaction of two distinct limitations of the data. First, the receivers are limited to the central part of the model. No traveltime data can constrain the structure at the edges of the model, so the α/β there defaults to the intermediate value assigned to the starting model there. This effect forms the outer boundary of the ring. Second, near the center of the model most rays traverse the shallow crust sub-vertically, interacting with both the upper crust (high α/β) and the low-velocity zone (with low α/β) beneath it. Vertical smearing tends to reduce the contrast in α/β in the center of the model, forming the inner boundary of the ring.

Finally we note that the close correspondence of the inverted and true lower crustal structure is only coincidental. Very few, if any, rays traverse this region, so its properties are unconstrained by the data. The bottom of the low velocity zone is simply being controlled by the starting model, which has no anomalous structure at this depth.

MOHO TOPOGRAPHY CASE STUDY

Many seismic refraction experiments record excellent Moho-reflected phases, such a PmP and SmS (e.g. Figure 1). The traveltimes of these phases are influenced by both the thickness of the crust and by its velocity structure. Thus, for instance, a 1 km downward deflection of Moho increases the traveltime of a normally-incident PmP wave by 0.29s, assuming a lower crustal velocity of 7 km/s. Similarly, a 2.5 km thick low velocity zone (with $\Delta\alpha=-2$

km/s) in the lower crust also increases the PmP traveltime by the same amount. The problem is that these two structures might be given rather different physical interpretations. For instance, if the region being studied was a mid-ocean ridge, then a downward deflection of Moho might be associated with anomalously high crustal production rate, while a low velocity zone might be associated with a partially-molten zone just above Moho. An important issue, then, is whether a tomographic inversion that combines both crustal-turning P waves and Moho-reflected PmP waves can distinguish Moho topography from lower crustal velocity heterogeneity.

We examine this issue with a synthetic inversion, using "data" calculated from a hypothetical crustal model (the "true model") containing both long-wavelength variability in crustal velocity and Moho topography. The true model has velocity and Moho depth defined in a $50\times50\times15$ km region containing $21\times21\times11$ nodes. The model contains a broad compressional velocity anomaly, centered in the mid-crust at the center of the model, with a maximum amplitude of $\Delta\alpha=-1$ km/s. This anomaly is superimposed on a one-dimensional velocity structure that monotonically increases with depth. The Moho has a broad, 2 km deep downward deflection, also centered on the center of the model. A 41×41 grid of receivers are placed on the top surface of the model, and a 21×21 grid of sources are placed at 1.5 km depth (Figure 4, top). About 25% of the possible source-receivers paths are selected on a random basis, yielding a total of 7,370 crustal-turning P wave traveltimes and 6,921 Moho-reflected PmP waves traveltimes. These traveltimes are then perturbed with Gaussian-distributed random noise with a standard deviation of 0.05s. The ray coverage (Figure 4, bottom) is dense, but – as in common with crustal-imaging experiments – has a significant deficiency: Relatively few crustal P waves sample the lower part of the crust, owing to the source-receiver offsets being limited to less than 25 km, and those that do are all propagating sub-horizontally.

We perform joint inversions for crustal velocity structure and Moho depth (Figure 5). Crustal velocity is represented by 1521 model parameter, the velocity of the nodes within the central part of the model (that is, |x| and |y| less than 15 km). The velocity model parameters extend all the way down to Moho. Moho depth is represented by 169 model parameters, the vertical position of each node on the reflecting interface. A suite of inversions are performed, all with velocity damping parameters fixed at 1%, but with the topography damping parameter ranging from 0.1% (that is, squeezing crustal velocity) to 10% (squeezing Moho topography). The inversions required two iterations to converge, when begun with a one-dimensional starting model that has a flat Moho. The two extreme cases of these inversions (Figure 5) both reduce the r.m.s. traveltime error by similar amounts (71-75%), a little less than the 80% expected given the 4:1 signal-to-

8

noise ratio. Unfortunately, the inversions yield rather different estimates of crustal velocity structure and Moho topography.

The velocity-squeezed inversion comes closest to reproducing the true model. It captures both the 2 km depression of Moho and the generally slow crustal velocities in the central part of the model. Some small oscillations of isovelocity surfaces are present in the uppermost crust. These result from an interaction between the high damping and the relatively coarse source ray spacing, which drive the velocity perturbations to small values in the poorly sampled regions of the upper crust that are between the sources. The topography-squeezed inversion fits the traveltime nearly as well as the velocity-squeezed inversion (71% compared to 75%), but has a nearly flat Moho. Anomalously slow PmP traveltimes are produced by a large amplitude low velocity zone in the central part of the lower crust. This zone has a velocity of 5.6 km/s, about 20% less than the 7 km/s "background" velocity at that depth – a value that is well within the range that might be expected by a lower-crustal partial melt zone. This imaging experiment fails to discriminate between the Moho topography and lower-crustal velocity.

The experimental geometry can be improved by increasing the aperture of the source and receiver arrays, so that the lower part of the crust is better-sampled by crustal turning waves. Several practical considerations limit the usefulness of this approach, however: The first is the difficulty of precisely measuring crustal P wave traveltimes at the ranges greater than the distance at which the mantle-turning Pn phase becomes the first arrival. In many crustal models, the crustal P wave turns at a depth of about half the crustal thickness at this "cross-over" range, so the lower half of the crust is usually poorly sampled. The second consideration is the practical difficulty in distinguishing P and PmP at extreme ranges where crustal P arrives just slightly before PmP. The problem is accentuated by the generally low crustal P amplitudes, which are caused by the generally low lower-crustal velocity gradients, and the generally high PmP amplitudes, caused by the large reflection coefficients at near-grazing incidence, both of which conspire to make the onset of crustal P hard to spot.

While traveltime tomography cannot distinguish the model with the depressed Moho from its alternative with slow lower crustal velocities, these two models are by no means seismically indistinguishable. The velocity jump across Moho is much larger in the second case than the first, implying that the *amplitude* of the Moho reflection can be used as an effective discriminant. Other imaging techniques that are more sensitive to wavefield amplitudes, such as the receiver function method, can be used to complement the tomographic results and to resolve the ambiguity.

Since its introduction by *Aki et al.* [1976], teleseismic tomography has been a standard tool for mapping out the velocity structure of the uppermost mantle. This procedure uses waves from distant earthquakes (teleseisms), which are presumed to have simple (e.g. planar) wavefronts as they impinge upon the bottom of a regional model of the earth's crust and upper mantle. Any departure from this simple shape detected by stations on the earth's surface is assumed to be due to velocity heterogeneity within the model. Wave diffraction, which tends to smooth out wavefronts as they propagate (a process often referred to as "wavefront healing") and the fact that the lower mantle has relatively subdued velocity heterogeneities, are often cited in support for the assumption of an initially simple wavefront.

The initial backazimuth and angle of incidence of the teleseismic wavefront can be predicted from the known location of the earthquake and a set of standard tables such as IASPEI 1991 [*Kennett, 1991*]. Such a prediction is imperfect. The IASPEI tables are global averages and omit perturbations in propagation direction caused by lateral heterogeneity. However, the error is usually assumed to be negligible. The initial arrival time of the teleseisms, however, is much more uncertain, owing both to the uncertainty in earthquake origin time and uncertainty in the traveltime of the wave across the earth. As a result, the arrival time is usually considered unknown.

The overall traveltime of the wave through the model cannot be determined without knowledge of its arrival time at the bottom of the model. Teleseismic tomography must rely only upon "relative" traveltime residuals, meaning residuals measured with respect to a best-fitting planar wavefront of known horizontal phase velocity but unknown arrival time. This is a very serious limitation: All vertically-stratified earth models (e.g. where velocity, $v(z)$ varies only with depth, z) have zero residuals in this sense, since they do not perturb the planarity of the wavefront nor change its horizontal phase velocity. From the viewpoint of teleseismic tomography, all vertically-stratified earth models are the same. Furthermore, if a three-dimensional model is viewed as the sum of a vertically-stratified part and small perturbation, $v(x,y,z)=v_0(z)+\Delta v(x,y,z)$, then teleseismic tomography can resolve only Δv and not v_0 (at least if Δv is not so strong as to significantly perturb the raypaths from their shape in the vertically-stratified model, $v_0(z)$). Unfortunately, models with the same Δv but different v_0's can have quite different appearances (Figure 6).

Teleseismic tomography has been used to investigate mantle plumes, such as beneath Iceland [*Tryggvason, 1983; Wolfe et al. 1997; Allen, 2001*] and Yellowstone [*Saltzer and Humphreys, 1997*]. The shape and depth extend of these plumes, and the way in which these

properties are related to geodynamic processes has been a critical issue in these discussion, often engendering considerable controversy [e.g. *Foulger et al., 2000*]. We present here a case study which illustrates techniques for examining the robustness of estimates of plume shape and depth distribution.

The true model has a 30 km thick crust and an upper mantle defined in a 400×400×255 km region containing 41×41×11 nodes. The mantle contains a broad compressional velocity anomaly with maximum amplitude of −3%. The anomaly is circular in cross-section, and uniform with depth (in the sense that Δv is a function only of horizontal position, (x,y)). This anomaly is superimposed on a one-dimensional structure that monotonically increases with depth. A 5×5 grid of receivers are placed on the top surface of the model, and 288 teleseisms (72 azimuths and 4 angles of incidence) are used to compute 5700 traveltimes (Figure 7, top). These traveltimes are then perturbed with Gaussian-distributed random noise with a standard deviation of 0.01s. The ray coverage (Figure 7, bottom) is dense, but – as in common with teleseismic tomography experiments – has a significant deficiency: All of the rays are sub-vertical, with the least vertical ray having an angle of incidence of 32° (that is, the epicentral range of the closest teleseism is about 50°).

We perform two types of inversions for mantle velocity structure (Figure 8).

In the first suite of three inversions, we investigate the ability of the data to resolve the depth distribution of the anomaly, using a depth-squeezing method suggested by *Lerner-Lam and Jordan* [1987]. The velocity is represented by 1089 model parameter, the velocity of the nodes within the central part of the model (that is, |x| and |y| less than 150 km and z beneath the Moho). Three inversions are performed, the first with a spatially uniform damping of 5%, the second with damping that decreases by a factor of 10 with depth and the third with damping that increases by a factor of 10 with depth. The later two inversion squeeze the velocity perturbations deeper and shallower depths in the model, respectively. The models all fit the data equally well, reducing the r.m.s. traveltime residuals by 78% – a much as can be expected given the data's 4:1 signal to noise ratio. All three inversions image an anomaly in the center of the model (Figure 8C–8E), but differ significantly in estimating its depth distribution (Figure 9). The centroid depths of the anomaly are quite different in the three cases, 126 km for the uniformly damped case (which is close to the true value), 137 km for the "squeezed-deep" inversion, and 103 km for the "squeezed-shallow" inversion. Thus this teleseismic inversion is of limited usefulness in studies of the depth distribution of the plume's velocity structure.

The inversion could be improved somewhat by including teleseisms with less steep angles of incidence.

However, such teleseisms turn more shallowly in the mantle and are more influenced by upper mantle heterogeneity outside of the modeled region.

Another aspect of plume shape with geodynamical implications is its departure from "axial shape", meaning the presence of any narrowing or flaring out of the plume with depth. A perfectly axial plume would have a velocity perturbation that was a function of horizontal position only, that is, $\Delta v(x,y,z) \rightarrow \Delta v(x,y)$; its strength does not change with depth. The true plume in the above example has this property. We therefore investigate how to test for departures from axial shape.

We suggest that a "vertical mode" parameterization of velocity is appropriate for such a test. The velocity perturbation along vertical columns of nodes is represented by truncated modal expansion:

$$\Delta v(x,y,z) = \Sigma_{i=0}^{N} a_i(x,y)\, p_i(z) \qquad (4)$$

Here $p_i(z)$ are a set of N functions of depth, z, and $a_i(x,y)$ are corresponding coefficients. They represent the depth behavior of Δv at any given horizontal position, (x, y). We assume that the $p_i(z)$ are ordered by increasing complexity, with $p_0(z)$ being the constant function (that is, $p_0(z)=1$). An axial structure therefore has non-zero a_0, with all higher order coefficients being zero. Many different choices of the $p_i(z)$ are acceptable; we use harmonic functions (e.g. sines and cosines), so that Equation 4 is a truncated Fourier series. We then preferentially damp the higher order coefficients, so as to drive the inversion towards an axial solution. If the resulting solution fits the data and has negligible higher-order coefficients, then we conclude that the data cannot rule out an axial plume (Figure 8F). If higher-order coefficients have significant amplitude, then we conclude that the data require non-axial behavior. In this Case Study, the true plume is perfectly axial, and that axial nature is fully recovered by the inversion.

EARTHQUAKE LOCATION CASE STUDY

While the discussion above emphasizes the role of raytracing and traveltime calculation in seismic tomography, they are also a fundamental part of most earthquake location algorithms. Earthquake location typically relies on some form of Geiger's Method (see Section A.26), in which the location and origin time of an earthquake are determined by matching (usually in the least-squared sense) the observed and predicted arrival times of seismic waves. As we have discussed above, ray theory provides one method of calculating the traveltime, and hence the predicted arrival time of a seismic wave with a specified origin time. In years past, most earthquakes were located in very simple earth models consisting of a stack of homogeneous layers, using software such as

Klein's [1978] HYPOINVERSE program. Raytracing is not explicitly needed in such cases, since traveltimes can be calculated using a simple formula. Rays figure only in this formula's derivation. This simplicity ends as seismologists begin to employ earth models with three-dimensional heterogeneity, where traveltime calculation is a much more difficult problem. The linkage between earthquake location and tomography has also grown because of the desire to locate earthquakes using earth models that are themselves the product of tomography. Compatibility between the two procedures on issues, say, of model representation, is then a practical necessity.

The case study that we present here examines the effect of using a laterally homogeneous model to locate earthquakes, when they actually occurred in a laterally heterogeneous earth. We choose a 50×50×25 km heterogeneous model with two commonly-encountered forms of heterogeneity (Figure 10B): The upper crust, with a mean compressional velocity of 3.1 km/s, thickens from 0.64 km at the left hand side of the model to 1.35 km on the right; and the mid-crust has a high compressional velocity dome with an amplitude of 3%. The homogeneous model (Figure 10D) is a laterally-averaged version of the heterogeneous one. A rectangular, 40×100 grid of earthquakes (Figure 10C) are aligned on the x–z plane at y=1.75, extend from 2 to 10 km depth, and have a regular spacing of 0.2 km. Traveltimes are computed by raytracing from each earthquake to an array of nine receivers (Figure 10A). The traveltime data set consists of a total of 31,277 P wave traveltimes and an equal number of S wave traveltimes (created using the assumption of a constant compressional to shear wave velocity ratio of 1.76). This number of rays corresponds to about 87% of the theoretical total of 36,000, since some of the longest source-receiver distances were excluded from the calculations.

We first compute a set of locations using the laterally homogeneous model, a data set consisting of P and S traveltimes, and starting locations that were based on perturbing their x, y and z coordinates with random numbers generated by a Gaussian distribution with a mean of zero and a standard deviation of 1 km. Eight iteration of Geiger's Method (se Section A.26) are needed to reduce to the traveltime residuals from an initial value of 0.22s to a final value of 0.005s. The general pattern of earthquakes (that is, their lying of a rectangular grid) is well recovered (Figure 10E). On the fine scale, however, systematic mislocation with amplitude of about 0.1 km occurs. The pattern of mislocation is spatially coherent, at least at scales of a few kilometers, but complicated.

There has been much recent interest in performing earthquake locations using only differential traveltimes from nearby earthquakes observed at a common station. The advantages of this technique (which is often called the *double–difference* method) is that differential traveltimes

13

can be measured very accurately using waveform cross-correlation techniques, and they can be predicted more accurately than absolute traveltimes because errors in the velocity model tend to cancel out when traveltimes of neighboring events are subtracted [e.g. *Slunga et al., 1995, Waldhauser and Ellsworth, 2000*]. The disadvantage is that much of the information in the traveltime data that fixes the absolute location of each earthquake has been excluded from the location process. The location process thus emphasizes the spatial relationship of the earthquakes to one another, rather than their absolute position in a fixed reference frame.

We demonstrate the double–difference method by creating a data set of 142,303 differential P traveltimes. Each differential traveltime is for two events with true locations separated by no more than 0.5 km observed on a common station. The earthquakes are located with these data, with the laterally homogeneous model, and with the same randomly–perturbed starting locations as the previous case. Eleven iteration of Geiger's Method (see Section A.27) are needed to reduce the differential traveltime residuals from an initial value of 0.40s to a final value of 0.001s. The general pattern of earthquakes (that is, their lying of a retangular grid) is well recovered, except that large (>1 km) mislocations occur along the edges of the grid (Figure 10F). This problem is due to the fewer data constraining the location of these earthquakes, owing both to their extreme position (that is,, fewer close neighbors) and to our having excluded traveltime data for some of the largest source-receiver offsets. Even so, these mislocations are highly correlated between neighboring earthquakes, so for the most part the relative locations are being reliably recovered. The earthquakes in the central part of the grid – where the data coverage is the best – are located to greater accuracy than the previous case, by a factor of about two (Figure 11). Furthermore, the mislocation errors are mostly vertical, whereas in the previous case the error tended to have a significant horizontal component).

Overall, we form two impressions from this Case Study: that the absolute locations provided by the double-difference method are actually pretty good, and that that improvement in relative location achieved by using the double-difference method is perhaps not as great as one might have expected.

The first impression, good absolute locations, is due to the fact that the double-difference data *do* contain some information about the absolute location of the earthquake. That this is so can be seen by considering two earthquakes, offset by a small vector $\Delta s\mathbf{t}$, where Δs is a small increment of distance and $\mathbf{t}$ is a unit vector, and having origin time difference, Δt_0. The differential arrival time, ΔT, of these earthquakes at a specified station is necessarily bounded by $\Delta t_0 \pm \Delta s/v$, where v is the velocity in the earthquakes' vicinity. These maximum and minimum values occur when the rays are overlapping and leave the earthquake

source with tangent, **t**. Thus, by scanning a large set of data for the extremes of ΔT, one can deduce not only the offset, Δs, but also a raypath – fixed in absolute space – along which both earthquakes must lie. If data for a third neighboring earthquake were available (so long as it wasn't exactly on a line connecting the first two), then a repeated application of this procedure to all three pairs of earthquakes would yield three rays that would converge at the earthquakes' absolute location.

The second impression, that traditional earthquake locations are almost as good as those provided by the double-difference method, reflects the fact that the signal-to-noise ratio for both absolute and differential traveltime data used in this Case Study are similar – and high. In actual applications the ratio for double-difference data would typically be higher than for absolute data, owing to the applicability of high-precision cross-correlation techniques to measuring the former. Proponents of the double-difference method also claim that it better handles cases (such as this one) in which the velocity model used to locate the earthquakes over-simplifies the true model. At best, this effect can account for only very modest improvement here.

CONCLUSIONS

We have presented three case studies of traveltime tomography relevant to the imaging of crust and upper mantle using regional arrays. The first case study investigates the imaging of Poisson's ratio using measurements of compressional and shear wave traveltimes. Inversions schemes that jointly fit both types of traveltime data and that selectively damp the compressional to shear wave velocity ratio are shown to be very effective. The second case study focuses on the ability of joint inversions of crustal and Moho-reflected wave traveltimes to image both crustal velocity structure and Moho topography. The ability of this kind of tomography to distinguish Moho topography and lower crustal heterogeneity is shown to be poor. The third case study examines the ability of teleseismic tomography to determine the shape of a mantle plume. A new test for detecting departures from an axial shape is proposed and tested. It is based on squeezing the higher order coefficients in a modal expansion for vertical velocity profile. Finally, we demonstrate some of the the advantages and limitations of the double-difference earthquake location technique, when it is used to locate earthquakes that have occurred in laterally heterogeneous structures.

REFERENCES

Abers, G.A., Three-dimensional inversion of regional P and S arrival times in the East Aleutians and sources of subduction zone gravity highs., *J. Geophys. Res. 99,* 4395-4412. 1994.

Aki, K. A. Christoffersson and E. Husebye, Three-dimensional seismic structure under the Montana LASA, *Bull. Seism. Soc. Am. 66,* 501-524, 1976.

Allen, R., *The mantle plume beneath Iceland and its interaction with the north Atlantic ridge: a seismologcal investigation (Ph.D. Thesis),* Princeton University, 184 pp, 2001.

Christensen, N.I., Seismic velocities, in Carmichael, R. S., ed., *Handbook of Physical Properties of Rocks*, Vol. II, Boca Raton, Florida, CRC Press, 1-228.

Clark, S.P., *Handbook of Physical Constants, Revised Edition*, Geological Soc. Am., New York. 587 pp.

Foulger, G.R., M.J. Pritchard, B.R. Julian, J.R. Evans, R.M. Allen, G. Nolet, W.J. Morgan, B.H. Bergsson-, P.Erlendsson, S. Jakobsdottir, S. Ragnarsson, R. Stefansson, and K. Vogfjord, The seismic anomaly beneath Iceland extends down to the mantle transition zone and no deeper, *Geophys. J. Int. 142,* F2-F5. 2000.

Gebrande, H., Miller, H. & Einarsson, P., 1980. Seismic structure of Iceland along the RRISP profile, *J. Geophys., 47,* 239-249.

Kennett, B., ed., *IASPEI 1991 Seismological Tables*, Australian School of Earth Science, Canberra, Australia, 167 pp., 1991.

Klein-F., *Hypocenter location program HYPOINVERSE, Open-File Report OF 78-694,* U. S. Geological Survey, 1978.

Lerner-Lam, A.L. And T.H. Jordan, How thick are the continents? *J. Geophys. Res. 92,* 14,007-14,026. 1987.

Menke, W., *Geophysical Data Analysis, Revised Edition, Discrete Inverse Theory*, Academic Press, Orlando, 289 pp, 1989.

Menke, W. and D. Abbott, *Geophysical Theory*, Columbia University Press, 458p,, 1990.

Menke, W., B. Brandsdottir, P. Einarsson and I. Bjarnasson, Reinterpretation of the RRISP-77 Iceland shear wave profiles, *Geophys. J. Int. 126*, 166-172, 1996.

Menke, W., M. West, B. Brandsdottir and D. Sparks, Compressional and shear velocity of the lithosphere in northern Iceland, *Bull. Seism. Soc. Am. 88,* 1561-1571, 1998.

Press, W., *Numerical Recipes in C: The art of Scientific Computing, Second Edition,* Cambridge University Press, Cambridge, 994p, 1992.

Saltzer, R.L. And E.D. Humphreys, Upper mantle P wave velocity structure of the eastern Snake River plain and its relationship to geodynamic models of the region, *J. Geophys. Res. 102,* 11,829-11,841, 1997.

Selby, S.M., *CRC Standard Math Tables, 21-st Edition,*

Chemical Rubber Co., Cleaveland, 714p. 1973.

Shaw, P.R., Age variations of oceanic crust Poisson's ratio; inversion and porosity evelution model, *J. Geophys. Res. 99,* 3057-3066, 1994.

Slunga-R., S.T. Rognvaldsson and B. Bodvarsson, Absolute and relative locations of similar events with application to microearthquakes in southern Iceland, *Geophys. J. Int. 123,* 409-419, 1995.

Spence, W. and S. Alexander, A method of determining the relative location of seismic events, *Earthquake Notes. 39*, 13. 1968.

Tryggvason K., E. Husebye, and R. Stefansson, Seismic image of the hypothesized Icelandic hot spot, *Tectonophysics 100*, 97-118, 1983.

Waldhauser, F. and W. Ellsworth, A double-difference earthquake location algorithm; method and application to the northern Hayward Fault, California*, Bull. Seis. Soc. of Am.. 90*, 1353-1368., 2000.

Wessel, P. and W.H.F. Smith, Free software helps map and display data, *EOS Trans. AGU 72,* 441, 445-446. 1991.

Wolfe, C,J; I.T. Bjarnason, J.C. VanDecar and S.C. Solomon, Seismic structure of the Iceland mantle plume*, Nature 385*, 245-247, 1997.

Zelt, C.A., FAST: 3-D First Arrival Seismic Tomography programs, http://terra.rice.edu/department/faculty/zelt/fast.html, 1998.

Zelt, C.A. and P.J. Barton, 3D seismic refraction tomography: A comparison of two methods applied to data from the Faeroe Basin*, J. Geophys. Res. 103*, 7187-7210,1998.

_____

William Menke, Lamont-Doherty Earth Observatory of Columbia University, Palisades NY 10983 USA.

APPENDIX:
SEISMIC RAYTRACING, TOMOGRAPHY
AND EARTHQUAKE LOCATION IN A
THREE-DIMENSIONAL TETRAHEDRAL MESH

A.1. Choice of Coordinate System.

We use a right-handed Cartesian coordinate system in which position is described by a 3-vector $\mathbf{x}=[x,y,z]^T$, where T indicates transpose. As usual, we identify z as the "vertical-up" direction.

A Cartesian coordinate system is most naturally applied to models of relatively small portions of the earth, say less than 100-200 km in scale, since the earth can be approximated as flat on that scale. It can also be used for somewhat larger models, say up to 1000 km across, by the simple artifice of building the curvature of the earth into the models. But beyond that scale, spherical coordinates

17

are clearly more appropriate.

A.2. The Tetrahedral Mesh.

The material parameters of the seismic model are specified at nodes. Let the total number of nodes be L, with node i being at position, $\mathbf{x}^{(i)}$, and having associated material parameters, such as seismic velocity, $v^{(i)}$. In general, it is possible first to pick a set of nodes and then to organize them into a set of tetrahedra that completely fill – without overlap – the volume of the model. Delaunay tetrahedralization, for instance, is a systematic method for finding the most "compact" set of such tetrahedra. Our own experience, however, is that completely disorganized meshes, where the nodes are at randomly chosen points, are difficult to conceptualize and manipulate. We prefer more organized meshes that have distinct sheets of nodes, at least some of which have a direct correspondence to interfaces within the earth (e.g. sea floor, Moho, etc.).

The Cartesian mesh, in which nodes are regularly spaced in (x, y, z), is the easiest to construct. Each parallelipiped of 8 nodes can easily be divided up into six tetrahedra, in a fashion such that their faces exactly match the faces of adjoining tetrahedral, both within a given parallelepiped and between adjoining ones (Figure 12).

The key property of the Cartesian mesh that allows rapid tetrahedralization is the organization of the nodes by rows, columns and tiers, and not their regular spacing. Irregular meshes which have this underlying organization (that is, the nodes at positions $\mathbf{x}^{(i)(j)(k)}$, where i, j, k are indices) are equally easy to tetrahedralize. We find these "distorted" meshes particularly useful, because they permit the representation of interfaces with varying topography (Figure 13).

A.3. Information Associated With Tetrahedra.

The seismic model consists of both a list of nodes and a list of tetrahedra. Our experience is that at least the following information should to be maintained for each tetrahedron: which node lies at each of its four vertices, the outward-facing unit normals of each of its four faces, the tetrahedron (if any) that adjoins each of its four faces. We note that if a face of a tetrahedron contains vertices $\mathbf{x}^{(j)}$, $\mathbf{x}^{(j)}$, and $\mathbf{x}^{(k)}$, then the vector $\mathbf{n} = (\mathbf{x}^{(j)}-\mathbf{x}^{(i)}) \times (\mathbf{x}^{(k)}-\mathbf{x}^{(i)})$ is normal to the face. The sign of this normal should be chosen to make it point away from the "excluded vertex" (meaning the fourth vertex of the tetrahedron – the one not on this face), so that it is outward-pointing.

Calculations are also facilitated through the maintaining, for each node, of a list of tetrahedra that contain that node as a vertex. In cases where interfaces are of interest, the interfaces can be enumerated. A node's interface (if any) should then be maintained, as should the interface (if any) of each face of each tetrahedron (where a

face is taken to be on the interface if all three of its vertices are on it).

A.4. Interpolation within a Tetrahedron.

Let the four vertices of a tetrahedron be located at position $\{\mathbf{x}^{(i)}, i=1, ... 4\}$ and have corresponding seismic velocity $\{v^{(i)}, i=1, ... 4\}$. Within this tetrahedron, the velocity is assumed to be a linear function of position:

$$v(\mathbf{x}) = a_1 x + a_2 y + a_3 z + a_4 \qquad (5)$$

The 4-vector of coefficients $\mathbf{a}=[a_1, a_2, a_3, a_4]^T$ can be found by solving the 4×4 linear system:

$$v^{(i)} = a_1 x^{(i)} + a_2 y^{(i)} + a_3 z^{(i)} + a_4 \quad ; \quad i=1, ... 4$$
$$\text{or}$$
$$\mathbf{v} = \mathbf{M} \cdot \mathbf{a} \qquad (6)$$

(where $\mathbf{v}$ is a 4-vector of the four nodal velocities and $\mathbf{M}$ is a 4×4 matrix containing node positions) using any standard method (e.g. Gaussian elimination or by simple matrix inversion, $\mathbf{a} = \mathbf{M}^{-1} \cdot \mathbf{v}$). Note that the velocity gradient, $\mathbf{g}$, is just the 3-vector $\mathbf{g}=\nabla v=[a_1, a_2, a_3]^T$.

A.5. Testing Whether a Point is Within a Tetrahedron.

A tetrahedron has four faces, each of which contains three of its four vertices. A face can be labeled by the index of the vertex it excludes (that is, face i does not contain vertex $\mathbf{x}^{(i)}$). A point, $\mathbf{x}$, lies within the tetrahedron if, for all four faces, it lies on the same side of the face as the excluded vertex. The equation $f(\mathbf{x})=0$ that defines each face is needed to perform this test. The standard form of the equation of a plane is $f(\mathbf{x})=Ax+By+Cz+D=0$, where A, B, C, D are constants and where one of these constants is arbitrarily set to unity. (Except when the plane contains the origin, we will adopt the choice D=1). The remaining constants could be computed by solving the 3×3 linear system:

$$A x^{(i)} + B y^{(i)} + Cz^{(i)} = -1 \qquad (7)$$

where $\{\mathbf{x}^{(i)}, i=1,2,3\}$ lie on the plane. However, the equivalent form, $f(\mathbf{x})=(\mathbf{x}-\mathbf{x}^{(i)}) \cdot [(\mathbf{x}^{(j)}-\mathbf{x}^{(i)}) \times (\mathbf{x}^{(k)}-\mathbf{x}^{(i)})] = 0$ (where $\mathbf{x}(i)$, $\mathbf{x}(j)$ and $\mathbf{x}(k)$ lie on the plane) is in practice more useful, since the set of constants $\{A, B, C, D\}$ never need to be computed. Furthermore, this equation has a simple geometrical interpretation: The vectors $(\mathbf{x}^{(j)}-\mathbf{x}^{(i)})$ and $(\mathbf{x}^{(k)}-\mathbf{x}^{(i)})$ both lie in the plane, so their cross product is normal to it. The dot product with $(\mathbf{x}-\mathbf{x}^{(i)})$ is thus proportional to the perpendicular distance of $\mathbf{x}$ from the plane. (It would be exactly distance, if the cross product were normalized to unit length).

Two points $\mathbf{x}$ and $\mathbf{x}^{(i)}$ lie on the same side of a plane if

f(**x**) has the same sign as f(**x**$^{(i)}$), or equivalently if f(**x**)f(**x**$^{(i)}$)>0.

A.6. Finding the Tetrahedron that Contains an Arbitrary Point.

One could, of course, exhaustively test whether the point, **x**, lies in each of the tetrahedra of the model, but more efficient strategies can reduce computational effort. We suggest the following strategy, which is based on the presumption that most calculations will use "neighboring" **x**'s: One should always keep track of the last tetrahedron that was accessed, and test whether **x** is within it. If it isn't, then form a unit vector that points from the center of the tetrahedron (that is, **x**^{c}=(**x**$^{(1)}$+**x**$^{(2)}$+**x**$^{(3)}$+**x**$^{(4)}$)/4) to **x**, compare this vector with the outward-facing normals, **n**, of each of the tetrahedron's four face, select the one in which is most nearly parallel (in the sense of maximizing **n**·(**x**−**x**^{c}), move to the tetrahedron adjoining that face and repeat the test. If it fails, then iterate the process until the tetrahedron is found or the surface of the model is encountered. An exhaustive search may be necessary in this event, since the iterative search can sometimes fail.

A.7. Raytracing in a Linear Medium.

Suppose that a seismic ray starts within a tetrahedron at a point, **x**$_0$, with ray tangent, **t**. Within this tetrahedron, the velocity varies only in one direction, given by the velocity gradient, **g**=∇v, so the ray is confined to the **t**−**g** plane. This geometry is best analyzed in a rotated (x',y',z') coordinate system in which the ray propagates in x'-z' (with z' is parallel to **g** and x' parallel to **t**). The situation is now seen to be equivalent to a ray propagating in a medium with a linear velocity in z'. A ray in such medium is known to have a path that is exactly an arc of a circle (see, for example, Menke and Abbott, 1989) with a radius of curvature R=1/(s|**g**|) (Figure 14). Here s is the phase slowness in the x' direction, which is t_x/v(**x**$_0$). The center of the circle is a distance R from **x** in a direction perpendicular to **t**, which is to say at position (x_c', y_c', z_c') = (x_0'+Rt_z, y_0', z_0'−Rt_x). We note that if the ray were to cross the z' of the center of the circle, its arc would become vertical (with respect to z'), so by Snell's law the velocity at that level must be zero. Hence the velocity at the center of the circle is zero. (This point must, of course, be outside of the tetrahedron, so no physical ray would ever cross that z' level). We will therefore add a coordinate shift that moves the origin of the primed coordinate system to the circle's center.

A.8. Coordinate Rotations and Shifts.

In A.7 above, we need to rotate a coordinate system to (z', y', z') axes parallel to ((**g** × **t**) × **g**, **g** × **t**, **g**) and then

20

shift the origin by (x_c', y_c', z_c'). Let $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ be unit vectors parallel to the primed axes. Then the transformation is $\mathbf{x'}=\mathbf{Sx}-\mathbf{x'}_c$ and the inverse transformation is given by $\mathbf{x}=\mathbf{S^T}(\mathbf{x'}+\mathbf{x'}_c)$ with $\mathbf{S}=[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]^T$.

A.9. Finding the Exit Face.

Suppose that in the primed coordinate system a face has an equation $ax'+by'+cz'+d=0$. The ray lies in the plane of $y=0$, so finding the the intersection point (if any) of the ray with the face is equivalent to finding the intersection point of the line, $ax'+cz'+d=0$ with a circle of radius, R, centered at the origin (Figure 15). (If both a and c are zero, no intersection point exists).

Suppose that we construct a second line that passes through the origin and which is the perpendicular bisector of the first. The bisector has equation, $-cx'+az=0$, and intersection point, $(x',z')=-d(a^2+c^2)^{-1/2}(a,c)$. This intersection point is a distance, $D=|d|(a^2+c^2)^{-1/2}$ away from the origin in the direction, $\mathbf{t}=[-da,-dc]^T|d|^{-1}(a^2+c^2)^{-1/2}$. The polar angle, ψ, from the z' axis to $\mathbf{t}$ can be found using the relation $\mathbf{t}=[\sin(\psi),\cos(\psi)]^T$.

Now suppose that we rotate the coordinate system so that the $\mathbf{t}$ direction becomes the new z'' axis. The problem has mirror symmetry about this axis, with the two intersection points of the line and the circle straddling the z'' axis. The equation of the line is just $z'=D$, where D is nonnegative. The equation of the circle is $z''=R\cos(\phi)$, where ϕ is the polar angle with respect to the z'' axis. One intersection point occurs at angle, $\phi=\cos^{-1}(D/R)$, where ϕ is in the interval $[0,\pi/2]$. The other is at $-\phi$. If $D/R>1$, then the circle and the line do not intersect. In the original primed coordinate system the intersection points are at polar angles $\theta=\psi\pm\phi$.

We accumulate a list of intersection points of the ray with each face of the tetrahedron. If the starting point of the ray is on a face, then this entry point is the solution $\theta=\theta_0$ and it must be removed from the list. Only one of the remaining intersection points is the exit point of the ray – the one with the smallest but positive $\theta-\theta_0$ that lies within the tetrahedron. The ray exit point $\mathbf{x'}=[R\sin(\theta), 0, R\cos(\theta)]^T$ and its tangent $\mathbf{t'}=[R\sin(\theta), 0, -R\cos(\theta)]$ can now be calculated and transformed back into the unprimed coordinate system.

A.10. Traveltime Through a Tetrahedron.

The traveltime is $T=\int_{ray} ds/v$, where s is arclength along the ray. In the rotated and shifted coordinate system, we have $ds=Rd\theta$ and $v=|\mathbf{g}|R\cos(\theta)$ and this integral becomes:

$$T = (1/|\mathbf{g}|) \int_{\theta 0}^{\theta} d\theta \, / \cos(\theta) \qquad (8)$$

Selby [1973, Integral 294] evaluates this indefinite integral:

21

$$\int d\theta \ / \cos(\theta) = \ln |\tan(\theta/2 + \pi/4)| \qquad (9)$$

A.11. Attenuation Through a Tetrahedron.

Suppose that anelastic attenuation of the earth is quantified by a reciprocal quality factor Q^{-1}. The attenuation, $T*$, along the ray is then defined as:

$$T* = \int_{ray} Q^{-1} \ dT$$
$$= \int_{ray} Q^{-1}/v \ ds = (1/|\mathbf{g}|) \ \int_{\theta 0}^{\theta} Q^{-1} \ d\theta/\cos(\theta) \qquad (10)$$

We note that sometimes attenuation data is represented in terms of the "path–averaged reciprocal quality factor" $<1/Q> = T*/T$. Now let the reciprocal quality factor be represented as a linear function of position within the tetrahedron. On the (x', z') plane, it varies as $Q^{-1} = q_1 x' + q_2 0 + q_3 z' + q_4$. Substitution into Equation 10 yields:

$$T* = (q_1 R/|\mathbf{g}|) \int_{\theta 0}^{\theta} \tan(\theta) \ d\theta +$$
$$+ 0 \ q_2 + (q_3 R*/|\mathbf{g}|) \int_{\theta 0}^{\theta} \ d\theta$$
$$+ (q_4/|\mathbf{g}|) \int_{\theta 0}^{\theta} \ d\theta \ / \cos(\theta)$$
$$= \mathbf{q} \cdot \mathbf{h} \qquad (11)$$

Here $\mathbf{h}$ is a 4-vector of the definite integrals. The last integral is the same as Equation 10, above, the second is trivial and the first is (Selby, 1973, Integral 292):

$$\int \tan(\theta) \ d\theta = -\ln |\cos(\theta)| \qquad (12)$$

A.12. Other Ray Integrals.

We mention three other ray integrals that are sometimes of importance:

$$I_1 = \int_{ray} (ax' + bz' + c) \ ds$$
$$= aR^2 \int_{\theta 0}^{\theta} \sin(\theta) \ d\theta$$
$$+ bR^2 \int_{\theta 0}^{\theta} \cos(\theta) \ d\theta$$
$$+ cR \int_{\theta 0}^{\theta} \ d\theta \qquad (13)$$

$$I_2 = \int_{ray} (ax' + bz' + c) \ ds \ / \ v^2$$
$$= (a/|\mathbf{g}|^2) \int_{\theta 0}^{\theta} \sin(\theta)/\cos^2(\theta) \ d\theta$$
$$+ (b/|\mathbf{g}|^2) \int_{\theta 0}^{\theta} 1/\cos(\theta) \ d\theta$$
$$+ [c/(R|\mathbf{g}|^2)] \int_{\theta 0}^{\theta} 1/\cos^2(\theta) \ d\theta \qquad (14)$$

$$I_3 = \int_{ray} ds \ / \ (ax' + bz' + c)$$
$$= \int d\theta \ / \ (a \sin(\theta) + b \cos(\theta) + c/R)$$

The indefinite integrals are given by (Selby, 1973, Integrals 290, 291, 326, 312, 342):

$$\int \sin(\theta) \ d\theta = -\cos(\theta) \qquad (15)$$

$$\int \cos(\theta) \ d\theta = \sin(\theta) \qquad (16)$$

$$\int \sin(\theta)/\cos^2(\theta)\, d\theta = 1/\cos(\theta) \qquad (17)$$

$$\int 1/\cos^2(\theta)\, d\theta = \tan(\theta) \qquad (18)$$

$$\int d\theta / (a\sin(\theta) + b\cos(\theta) + c/R)$$

$$= (1/D)\ln[\ \{a-D+((c/R)-b)\tan(\theta/2)\}$$
$$/\{a+D+((c/R)-b)\tan(\theta/2)\}\ ]$$
$$\textit{if } (c/R)^2<(a^2+b^2) \textit{ and } (c/R)\neq b$$

$$= (2/E)\tan^{-1}[\ \{(a+((c/R)-b)\tan(\theta/2)\} / E\ ]$$
$$\textit{if } (c/R)^2>a^2+b^2$$

$$= (1/(c/R))\ [(c/R)-(a+b)\cos(\theta)-(a-b)\sin(\theta)]/$$
$$[(c/R)-(a-b)\cos(\theta)+(a+b)\sin(\theta)]$$
$$\textit{if } (c/R)^2=a^2+b^2 \textit{ and } (c/R)\neq b \qquad (19)$$

$$\textit{with } D=(a^2+b^2-(c/R)^2)^{1/2} \textit{ and } E=((c/R)^2-a^2-b^2)^{1/2} \quad (20)$$

A.13. Raytracing to a Specific Interface.

The raytracing procedure begins by specifying a starting position and direction for the ray. The tetrahedron inclosing this point is found, and the ray is traced via the above methodology to its exit face. If this face is on the specified interface, then the complete ray has been found. Otherwise the exit point and tangent becomes the entry point and tangent in the adjacent tetrahedron and the process is repeated. We note that having maintained adjacency information greatly improves the efficiency of this process, since no searching for tetrahedra is then required.

The following information needs to be maintained about a raypath: a list of the tetrahedra though which it passes; and for every tetrahedron, the exit face of the ray, the radius of curvature of the ray, the coordinate transformation and origin shift, and the polar angles of the starting and ending point.

We note that the raytracing effort scales linearly with the number of tetrahedra encountered along the ray path, and not with the length of the ray. The use of large tetrahedra, wherever possible, thus improves the efficiency of raytracing. On the other hand, halving the node spacing leads to only a doubling of raytracing effort, even though it increases the number of nodes by a factor of 8.

A.14. Reflecting a Ray off an Interface.

Raypaths containing reflections off of specified interfaces can be calculated though a simple modification of the above procedure. The ray is traced to the reflecting interface, its tangent is modified according to Snell's law, and the raytracing is resumed.

Snell's law requires that the components of the phase slowness parallel to the reflecting interface be equal for

both incident and reflected rays, and that the normal component have opposite signs. Thus one begins by defining a new coordinate system with (x', y', z') axes parallel to $((\mathbf{x}^{(3)}-\mathbf{x}^{(1)}) \times ((\mathbf{x}^{(2)}-\mathbf{x}^{(1)}) \times (\mathbf{x}^{(3)}-\mathbf{x}^{(1)})), \mathbf{x}^{(3)}-\mathbf{x}^{(1)}, (\mathbf{x}^{(2)}-\mathbf{x}^{(1)}) \times (\mathbf{x}^{(3)}-\mathbf{x}^{(1)}))$, where $\mathbf{x}^{(2)}, \mathbf{x}^{(1)}$ and $\mathbf{x}^{(3)}$ are the vertices of the face. The tangent is rotated to this coordinate system, the sign of its z'-component is reversed, and it is rotated back. This vector is the tangent of the reflected ray.

A.15. Point to Point Raytracing.

In many important seismological applications rays need to be traced from a common "source" to a set of "receivers" located either on the surface of the model or on an internal interface. (We note that the sense of source and receivers are wholly interchangeable, because the principle of reciprocity guarantees that the ray from, say, $\mathbf{x}^A$ to $\mathbf{x}^B$ is also the ray from $\mathbf{x}^B$ to $\mathbf{x}^A$).

Another concern arises from the fact that several rays may emerge from a source in different directions, and yet converge upon the same receiver. Identifying all these "multipathed" rays may be important in seismological applications. The search procedure that we outline below has the advantage of addressing both the "several receivers" and "multiple ray" issues simultaneously.

We begin by shooting a large set of rays from the source to the receiver interface. The initial ray tangents are chosen to span all the solid angle around the source, and can conveniently represented as $\mathbf{t}_{ij}$, where the subscripts (i, j) index the polar angles (e.g. azimuth, φ, and angle of incidence, θ) of each ray. The intersection points, $\mathbf{x}_{ij}$, of each ray with the receiver interface (if any) are tabulated. The number of rays needed in this table depends upon the complexity of the underlying model, but our experience is that 10^4 to 10^5 rays are usually sufficient.

If they intersect the receiver interface, then a set of three neighboring rays (that is, with indices (i,j), (i+1,j), (i,j+1) or (i+1,j), (i+1,j+1), (i,j+1) form a triangle on that interface (Figure 16). We now select a single receiver and search for all triangles that enclose it, using a test analogous to the tetrahedron test, above. Associated with each enclosing triangle is a small sliver of solid angle that contains the exact ray tangent that takes a ray from the source to that specific receiver. We use a refinement process, described below, to iteratively improve the estimate of the tangent. Having found one ray connecting the source to the selected receiver, we proceed to the next, and after finishing all that receiver's rays, we begin on the next receiver.

The test of whether a point lies with a triangular patch of an interface should be performed in a coordinate system aligned with that interface (e.g. with the z' axis normal to it). Then only the (x', y') coordinates are relevant. Each side of the triangle is defines by two points, $(x^{(1)'}, y^{(1)'})$ and

$(x^{(2)'}, y^{(2)'})$. The equation of this line is then $f(x',y')=ax'+by'+c=0$, where $a=(y^{(1)'}-y^{(2)'})/D$ and $b=(x^{(2)'}-x^{(1)'})/D$ and $D=x^{(1)'}y^{(2)'}-x^{(2)'}y^{(1)'}$. As with the tetrahedron, a point (x', y') lies within a triangle if it lies on the same side of every side as the excluded vertex. This condition is equivalent to $f(x', y')$ having the same sign as $f(x^{(i)'},y^{(i)'})$, where $(x^{(i)'},y^{(i)'})$ is the vertex excluded from side i, for all three sides.

In our experience, the coordinate transformation mentioned above can be omitted, since in practice most interfaces are sufficiently non-vertical for the untransformed horizontal coordinates (x, y) to be used.

We use a two step refinement procedure. The first step is to trisect the solid angle enclosing the true tangent, by tracing a single new ray that has the mean direction of the original three. This ray subdivides the original triangle in three triangular pieces, only one of which contains the receiver. The trisection process is iterated until some preset limiting precision is reached.

The second step is to use Newton's method to refine the estimate of the tangent. The idea is to view the ray intersection as a function of the take-off angles (θ, φ) and to linearize it about the current estimate of those angles, (θ^0, φ^0), using Taylor's theorem. After denoting the horizontal components of x as $\mathbf{x_H}=[x,y]^T$ and the take-off angles with the 2-vector $\mathbf{a}=[\theta, \varphi]^T$, we have:

$$\mathbf{x_H(t)} \approx \mathbf{x_H(t^0)} + \partial \mathbf{x_H}/\partial \mathbf{a} \cdot (\mathbf{a}-\mathbf{a}^0) \qquad (21)$$

Here $\partial \mathbf{x_H}/\partial \mathbf{a}$ is the matrix of partial derivatives of the ray intersection point with take-off angle. By setting $\mathbf{x_H(t)}=\mathbf{x_H^r}$, where $\mathbf{x^r}$ is the receiver location, we have:

$$\mathbf{a} \approx \mathbf{a}^0 + [\partial \mathbf{x_H}/\partial \mathbf{a}]^{-1} (\mathbf{x_H^r}-\mathbf{x_H(a^0)}) \qquad (22)$$

Here the superscript, -1, signifies the matrix inverse. This equation is iterated until $\mathbf{x_H(t^0)}$ is acceptably close to $\mathbf{x_H^r}$.

Our experience is that the partial derivatives can easily be computed with finite differences (at the expense of tracing two extra rays):

$$\partial \mathbf{x_H}/\partial a_1 = \partial \mathbf{x}/\partial \theta = [\mathbf{x_H}(\theta+\Delta\theta,\varphi) - \mathbf{x_H}(\theta,\varphi)]/\Delta\theta$$
$$\partial \mathbf{x_H}/\partial a_2 = \partial \mathbf{x}/\partial \varphi = [\mathbf{x_H}(\theta,\varphi+\Delta\theta) - \mathbf{x_H}(\theta,\varphi)]/\Delta\varphi \qquad (23)$$

Here $\Delta\theta$ and $\Delta\varphi$ are small increments in take-off angle. We note that methods are also available to calculate these derivatives variationally, using ray integrals, but do not discuss them further here.

The point-to-point raytracing can fail for a variety of reasons. Two rays can fall within the same triangle if the initial search is too coarse, in which case one ray may be missed. If the true ray has a caustic or focus, then the matrix of partial derivatives can become singular and Newton's method will fail. And finally, the receiver may lie within a shadow, in which case no true ray exists.

We note that the derivatives, $\partial \mathbf{x}/\partial \theta$ and $\partial \mathbf{x}/\partial \varphi$ in Equation 23 can be used to find the geometrical spreading function, R, that controls the amplitude of the particle motion of the seismic wave. The idea is that any energy that leaves the source in a small wedge of solid angle, $\Delta \Omega = \sin(\theta) \Delta \theta \Delta \varphi$, remains within the "ray tube" defined by the neighboring three rays, and is spread out over large and larger areas as those rays diverge. The geometrical spreading function, R, relates the perpendicular area, $A_\perp$, of the ray tube to the increment of solid angle, $d\Omega$, that it subtends at the source by $A_\perp = R^2 d\Omega$. Seismic amplitude is proportional to the square root of energy, so it is proportional to $1/R$. The three rays intersect with the interface to form a triangle with sides $d\mathbf{x}_\theta = (\partial \mathbf{x}/\partial \theta) d\theta$ and $d\mathbf{x}_\varphi = (\partial \mathbf{x}/\partial \theta) d\varphi$. The area of this triangle is $A = \frac{1}{2} |d\mathbf{x}_\theta \times d\mathbf{x}_\varphi| = \frac{1}{2} |\partial \mathbf{x}/\partial \theta \times \partial \mathbf{x}/\partial \varphi| \, d\theta d\varphi$. However this triangle is in the plane of the interface, which is not in general perpendicular to the ray. Simple geometry indicates that $A_\perp = A\cos(\psi)$ where ψ is the angle between the ray direction, $\mathbf{t}$, and the interface normal, $\mathbf{n}$ (note $\cos(\psi) = \mathbf{n} \cdot \mathbf{t}$). We thus have:

$$R^2 = A_\perp / d\Omega$$
$$= \frac{1}{2} |\partial \mathbf{x}/\partial \theta \times \partial \mathbf{x}/\partial \varphi| \, (\mathbf{n} \cdot \mathbf{t}) / \sin(\theta) \qquad (24)$$

Here we have used the relationship, $d\Omega = \sin(\theta) d\theta d\varphi$.

A.16. Wavefront to Point Raytracing.

In some applications (e.g. teleseismic tomography), the source is a prescribed wavefront, as contrasted to a point. The goal is to trace a ray, starting at a point on the wavefront and in a direction initially normal to it to a specified receiver. The point-to-point raytracing descried above can easily be modified to accomplish this goal.

The initial search begins by tracing a large set of rays from a grid of points, $\mathbf{x}^{s(i)(j)}$, on the wavefront to corresponding points, $\mathbf{x}^{r(i)(j)}$, on the receiver interface. An search that is exactly analogous to the point-to-point case is now used to identify triangles of $\mathbf{x}^{r(i)(j)}$ that contain a given receiver point, $\mathbf{x}^r$, and to home-in on the exact point on the wavefront that has a ray that connects to the receiver.

The wavefront itself can be specified through its arrival time, $T(x,y,z_0)$, at some reference depth, z_0, at the bottom of the model. The horizontal slowness of a ray leaving the wavefront at a point (x,y,z_0) is then $(s_x, s_y) = (dT/dx, dT/dy)$ and the corresponding ray tangent is $\mathbf{t} = [s_x v(x,y,v_0), \, s_y v(x,y,v_0), \, -\sqrt{(1 - t_x^2 - t_y^2)}]^T$. In many cases the wavefront can be adequately approximated as planar, in which case the arrival time function is $T(x,y,z_0) = s_x x + s_y y + t_0$. Here t_0 represents the arrival time of the wave at the reference position $(0,0,z_0)$.

The arrival time of the wave at the receiver is the sum of the traveltime along the ray and the arrival time of the

wavefront at the point that the ray leaves it.

A.17. Traveltime Tomography.

The key idea in traveltime tomography is that observations of the traveltime of rays through the model contain information about the velocity structure of the model. Traveltime data therefore can be used to solve for estimates of the velocity structure.

While seismic velocity is a continuously varying field, it is being determined by a finite set of discrete parameters, the nodal velocities and positions that describe the model. It is therefore appropriate to view the calculated traveltime, T, along a given ray, i, as a function of these model parameters. Using the M-vector, **m**, as a shorthand for the model parameters, we have:

$$T_i(\mathbf{m}) = T_i^{obs} \qquad (25)$$

where T_i^{obs} are the traveltimes observed in an actual experiment. We now linearize this equation around some initial estimate of the model, $\mathbf{m}_0$, and rearrange:

$$\partial T_i/\partial \mathbf{m} \cdot (\mathbf{m}-\mathbf{m}_0) \approx T_i^{obs} - T_i(\mathbf{m}_0)$$
$$\partial T_i/\partial \mathbf{m} \cdot \Delta \mathbf{m} \approx \Delta T_i$$
$$\mathbf{G} \cdot \Delta \mathbf{m} \approx \Delta \mathbf{T} \qquad (26)$$

Here $\Delta \mathbf{m} = (\mathbf{m}-\mathbf{m}_0)$ represents the perturbation to the reference model, $\mathbf{m}_0$, that leads to a corresponding perturbation in traveltimes sufficient to match the observed traveltime residuals, $\Delta T_i = T_i^{obs} - T_i(\mathbf{m}_0)$. (By "residuals", we mean the deviation of the traveltimes from the prediction of the reference model). If there are N observed traveltimes and M model parameters, then Equation 26 represents a N×M matrix equation for the unknown model perturbations, $\Delta \mathbf{m}$. The N×M matrix is sometime called the "data kernel" and given the symbol, **G**, with $G_{ij}=\partial T_i/\partial m_j$,

Equation 26 is usually ill-conditioned, so that it cannot be solved by simple premultiplication by the matrix inverse, $\mathbf{G}^{-1}$. A common problem is that some parts of the model are completely unsampled by rays, so that no traveltime data constrain the value of certain model parameters. This phenomenon results in a whole column of the matrix, **G**, being zero. One simple approach to the solution of such equations is to use the damped-least squares method (see, for example, Menke, 1989, p. 54), which results in an alternative equation for the unknown $\Delta \mathbf{m}$:

$$[\mathbf{G}^T\mathbf{G} + \varepsilon^2\mathbf{I}] \cdot \Delta \mathbf{m} = \mathbf{G}^T \cdot \Delta \mathbf{T} \qquad (27)$$

$$\Delta \mathbf{m} = [\mathbf{G}^T\mathbf{G} + \varepsilon^2\mathbf{I}]^{-1} \cdot \mathbf{G}^T \cdot \Delta \mathbf{T} \qquad (28)$$

Here **I** is the identity matrix and ε^2 is a "damping parameter". When the damping parameter is zero, the

solution (if it exists) exactly minimizes the squared traveltime error, $E = \Delta\mathbf{T} \cdot \Delta\mathbf{T}$. When ε^2 is very large, model perturbations are all driven toward zero (in the sense that $\Delta\mathbf{m} \cdot \Delta\mathbf{m} \rightarrow 0$). At intermediate values of the damping parameter the solution has the desired property of being very nearly the least-squares solution but having the unconstrained model parameter perturbations driven close to zero. Some experimentation is needed to find an optimum value of the damping parameter. Our approach is to express damping as a percentage, that is, in terms of the ratio of ε^2 to the largest element of $\mathbf{G}^T\mathbf{G}$, in which case optimal values are typically in the range of 0.1 to 10 percent.

In practical problems, the number of traveltime observations, N, the number of nodes, L, and the number of model parameters, M, are large (e.g. both of order 10^4). The data kernel, $\mathbf{G}$, is thus a very large matrix (e.g. of order $N \times M = 10^8$ elements). However, in cases where each model parameter affects only a few neighboring nodes, most of these elements are zero, since a row of $\mathbf{G}$ corresponds to a single traveltime observation and ray, and a single ray will interact with only a small subsets of nodes in the model. For instance, suppose that a single model parameter influences J neighboring nodes, and the total number of model parameters is about the same as the number of nodes. Then a single ray crossing the model, which interacts with about $L^{1/3}$ nodes, will interact with about $JM^{1/3}$ model parameters. $\mathbf{G}$ will have on order $JN^{1/3}M$ non-zero elements. Taking J=10, as might be typical in an actual application, the above example gives 2×10^6 non-zero elements, or just 2% of the total number of elements. The data kernel is thus a "sparse" matrix. The $M \times M$ matrix, $\mathbf{G}^T\mathbf{G}$, that appears in Equation 27 is also sparse, although in our experience it is not usually as sparse as $\mathbf{G}$ (e.g. 90% sparse).

Since both $\mathbf{G}$ and $\mathbf{G}^T\mathbf{G}$ are typically very large, efficient storage techniques that explicitly recognize the sparseness and store only non-zero elements are usually a necessity. The most common strategy is to organize the non-zero elements as linked lists of elements (rather than as an array of elements) (Press, 1992). Special effort must also be taken to optimize the storage strategy for computing the dot product of the sparse matrix and its transform with an arbitrary vector, as these operations are required in many parts of the inversion process.

Care must be taken not to destroy the sparseness of Equation 27 during its solution. For this reason, explicit computation of the matrix inverse, as in Equation 28, is usually not advisable, since the inverse of a sparse matrix is, in general, not sparse. Methods that involve row and column transformations, such as Gaussian elimination and Householder rotation, also destroy the sparseness. On the other hand, many iterative methods for solving a linear system (like Equation 27) are available that require only the sparse matrix, and possibly its transform, to be dotted

with vectors. (Note that the matrix, $[\mathbf{G}^T\mathbf{G} + \varepsilon^2\mathbf{I}]$, in Equation 27 is symmetric, so that it is its own transform). Our own experience is that the biconjugate gradient iterative method (Press, 1992) is particularly suited to solving Equation 27.

We note that these iterative methods can solve the least-squares Equation 27 without ever calculating the matrix, $[\mathbf{G}^T\mathbf{G} + \varepsilon^2\mathbf{I}]$, since these methods only require the multiplication of this matrix with an arbitrary vector, $\mathbf{u}$. We simply note that $[\mathbf{G}^T\mathbf{G} + \varepsilon^2\mathbf{I}]\mathbf{u} = \mathbf{G}^T[\mathbf{Gu}] + \varepsilon^2\mathbf{u}$, and perform the multiplication $\mathbf{Gu}$ first. Explicit calculation of the matrix, $\mathbf{G}^T\mathbf{G}$, should be avoided whenever possible, both because it tends to be less sparse than $\mathbf{G}$ (and hence requires more storage) and because it requires significant computation.

Since Equation 27 is derived by linearization, the estimated model $\mathbf{m}=\mathbf{m}_0+\Delta\mathbf{m}$ is only an approximation to the best-fitting model. Iteration of the equation may therefore be necessary. Each iteration requires a major computational effort, since a new set of rays need to be traced though the updated model.

A.18. Model Parameters.

While the model parameters need to control that nodal velocities and positions, they need not have a one-to-one correspondence to them. In our experience, considerable flexibility is gained by defining them to be linear combinations of either nodal velocities or nodal positions. We thus define velocity model parameters:

$$\Delta v^{(i)} = \Sigma_{j=1}^{M} w_{ij}\,\Delta\mathbf{m}_j\,; \quad i=1,\,\ldots,\,L \qquad (29)$$

and position model parameters:

$$\Delta\mathbf{x}^{(i)} = \Sigma_{j=1}^{M}\mathbf{W}_{ij}\,\Delta\mathbf{m}_j\,; \quad i=1,\,\ldots,\,L \qquad (30)$$

Here L is the number of nodes and w_{ij} and $\mathbf{W}_{ij}$ are weights. Note that each $\mathbf{W}_{ij}$ is a 3-vector that represents both a weight and the direction in which the node is displaced.

In the case of velocity model parameters, increasing the i-th model parameter changes the velocities of a group of nodes – those for which the weights, w_{ij}, are nonzero – by an amount specified by the weight. In the case of position model parameters, increasing the i-th model parameter changes the position of a group of nodes – those for which the weights, $\mathbf{W}_{ij}$, are nonzero – by an amount and direction specified by the weight.

Defining both position and velocity model parameters that effect the same node should be avoided, because they are to a large degree redundant. Similar perturbations in the velocity structure can be achieved by either changing the velocity of a node or by moving it to a new position. This redundancy only adds to the ill-conditioned nature of the tomography. Our practice is to use velocity model

29

parameters exclusively, except where we have data from reflected phases that specifically sample the position of an interface within the model.

Model parameters that are the weighed sums of nodal velocities (or positions) are useful because they allow one to decouple the spatial scale of the model (which may need to be fine in order to represent small-scale features such as surface topography) from the spatial scale of the inversion (which may need to be coarse if only few data are available). The weighting functions therefore control the spatial correlation length of the inversion.

The shortest correlation is achieved when M=L and $w_{ij}=\delta_{ij}$, where δ_{ij} is the Kronecker delta function. Here every model parameter controls the velocity of exactly one node. Choosing M=L and w_{ij} to decrease with the distance between nodes, $w_{ij}=\exp(-r^2/c^2)$ with $r=|\mathbf{x}^{(i)}-\mathbf{x}^{(j)}|$ and c is a constant imposes smoothness with a length scale of c. For models that are organized into a series of K sub-horizontal interfaces, the choice M=K and $w_{ij}=1$ if node i is on interface j and zero otherwise permits one to find the best-fitting "stratified" solution. Our practice is to always begin the inversion process with stratified model parameters, so as to first match the overall – and typically strong – increase of velocity with depth that is present in most places.

We normally choose the model parameters to have equal overall weight, in the sense that $\Sigma_{i=1}^{L}w_{ij}=1$ for all model parameters. We note however that different weights can be used to selectively damp different parts of the model. Since the damping tends to drive $\Delta\mathbf{m}\cdot\Delta\mathbf{m}$ to zero, giving a specific model parameter a larger weight tends to allow the corresponding velocity perturbations to be larger than average. This strategy can be used to find solutions that force velocity perturbations to be in particular places, which can be useful for hypothesis testing. For instance, assigning large weights to "shallow" model parameters ensures that if shallow and deep perturbations trade off with one another, the inversion will recover a model that favors the shallow ones. Thus any deep perturbations that are present are likely to be "real", in the sense of being required by the data.

A.19. Partial Derivatives of Traveltime with Nodal Velocity.

We present two methods of calculating the derivative of traveltime with velocity-type model parameters.

Method 1 uses the chain rule to construct the desired quantity, $\partial T_i/\partial m_j$, from several simpler partial derivatives. We start from the statement that the traveltime is an integral of the velocity field along the ray, $T=\int_{ray} ds/v$. Let us now write the velocity in terms of the velocity of a reference velocity and a perturbation, $v=v^0+\Delta v$. The traveltime integral becomes:

$$T = \int_{ray} ds/v = \int_{ray} ds/v_0 - \int_{ray} \Delta v/v_0^2 \, ds \qquad (31)$$

$$\Delta T = -\int_{ray} \Delta v/v_0^2 \, ds \qquad (32)$$

The integration path is most properly the perturbed raypath (that is, along whatever ray corresponds to velocity, v). We will use Fermat's principle, however, to justify performing the integral along the unperturbed ray path (that is, the ray path corresponding to velocity, v_0). Fermat's principle roughly states that the traveltime integral is insensitive to small errors in the raypath.

Within a tetrahedron, both the velocity field and its perturbation are linear functions of position. Suppose that we work in a rotated coordinate system defined in A.10. Then along the circular unperturbed raypath these linear functions are:

$$v_0 = R \, |g| \, z' = R \, |g| \cos(\theta) \qquad (33)$$

$$\begin{aligned}
\Delta v &= \Delta \mathbf{a} \cdot [x', y', z', 1]^T \\
&= \Delta a_1 x' + 0 + \Delta a_3 z' + \Delta a_4 \\
&= R\sin(\theta)\Delta a_1 + 0\Delta a_2 + R\cos(\theta)\Delta a_3 + \Delta a_4 \qquad (34)
\end{aligned}$$

Here $\Delta \mathbf{a}$ is a 4-vector of coefficients. Substituting Equation 33 into Equation 32 yields:

$$\begin{aligned}
\Delta T &= \\
-\int_{\theta 0}^{\theta} (\Delta a_1 R\sin(\theta) &+ \Delta a_3 R\cos(\theta) + \Delta a_4) / (R^2 \, |g|^2 \cos^2(\theta)) \, Rd\theta \\
&= \Delta a_1 \, (-1/|g|^2) \int_{\theta 0}^{\theta} \sin(\theta)/\cos^2(\theta) \, d\theta + \Delta a_2 \, 0 \\
&\quad + \Delta a_3 \, (-1/|g|^2) \int_{\theta 0}^{\theta} 1/\cos(\theta) \, d\theta \\
&\quad + \Delta a_4 \, (-1/(R|g|^2)) \int_{\theta 0}^{\theta} 1/\cos^2(\theta) \, d\theta \\
&= \mathbf{f} \cdot \Delta \mathbf{a} \qquad (35)
\end{aligned}$$

The three integrals can all be performed analytically, and are given by Eqns. 17, 9 and 18, respectively. The resulting formula is thus of the form $\Delta T = \mathbf{f} \cdot \Delta \mathbf{a}$, where f is a known 4-vector involving these integrals. Since in general, $\Delta T = \partial T/\partial \mathbf{a} \cdot \Delta \mathbf{a}$, the 4-vector, $\mathbf{f}$, can be identified as the partial derivative, $\partial T/\partial \mathbf{a}$:

$$\partial T/\partial \mathbf{a} = \mathbf{f} \qquad (36)$$

The coefficients, $\mathbf{a}$, of the linear velocity function are related to the four nodal velocities, $\mathbf{v}$, of the tetrahedron though the linear Equation 6, $\mathbf{v} = \mathbf{M} \cdot \mathbf{a}$, where $\mathbf{M}$ is a 4×4 matrix containing nodal positions (but note that the positions are in the transformed coordinate system). The solution of this equation is $\mathbf{a} = \mathbf{M}^{-1} \cdot \mathbf{v}$. Differentiation yields:

$$\partial \mathbf{a}/\partial \mathbf{v} = \mathbf{M}^{-1} \qquad (37)$$

The nodal velocities perturbations are related to perturbations in the velocity model parameters though Equation 29, $\Delta v^{(i)} = \Sigma_{j=1}^{M} w_{ij} \Delta \mathbf{m}_j$. Hence we conclude that:

$$\partial\mathbf{v}/\partial\mathbf{m} = \mathbf{w}^* \qquad (38)$$

where $\mathbf{w}^*$ is a 4×M matrix whose elements are the weights that connect the model parameters with the four nodal velocities corresponding to this tetrahedron. The chain rule now allows us to combine Eqns. 36, 37 and 38:

$$\partial\mathrm{T}/\partial\mathbf{m} = \partial\mathrm{T}/\partial\mathbf{a} \cdot \partial\mathbf{a}/\partial\mathbf{v} \cdot \partial\mathbf{v}/\partial\mathbf{m} = \mathbf{f} \cdot \mathbf{M}^{-1} \cdot \mathbf{w} \qquad (39)$$

This resulting partial derivative matrix represents the contribution from one tetrahedron, and must therefore be summed over all tetrahedra though which the ray passes.

Method 2 uses finite differences to calculate $\partial\mathrm{T}/\partial\mathbf{m}$ from the formula:

$$\partial\mathrm{T}/\partial m_i = [\mathrm{T}(m^0_1, \dots, m^0_i + \Delta m, \dots, m^0_M) - \mathrm{T}(\mathbf{m}^0)] / \Delta m_i \quad (40)$$

Here Δm_i is a small increment that is added to the i-th model parameter. The calculation of the perturbed traveltime is performed simply by using Equation 29 to perturb the nodal velocities, calculating the linear formula using Equation 6, and integrating along the raypath using Equation 19.

This method is less efficient than the first, but is algebraically simpler.

A.20. Teleseismic Tomography.

One limitation of teleseismic tomography (that is, where the sources are wavefronts entering from the bottom of the model) is that the arrival time of that wavefront at some reference depth deep in the model is rarely known to any useful accuracy. Uncertainties in the origin time of the earthquake and in global earth structure are usually much larger than the effect of heterogeneities in the model. On the other hand, the error is common to all stations that observe the arrival time of the same teleseism.

We can explicitly modify Equation 26, $\mathbf{G} \cdot \Delta\mathbf{m} = \Delta\mathbf{T}$, to remove the effect of the origin time uncertainty. We begin by considering just the rows of $\mathbf{G}$ associated with a particular teleseism, and adding a term, e, that represents the error in that teleseism's origin time:

$$\Sigma_{j=1}^{M} G_{ij} \Delta m_j = \Delta T_i + e \qquad (41)$$

We now sum over those i's – say there are N' of them – that correspond to observations of a single teleseism:

$$\Sigma_i \Sigma_{j=1}^{M} G_{ij} \Delta m_j = \Sigma_i \Delta T_i + N' e \qquad (42)$$

We then divide Equation 42 by N' and subtract it from Equation 41, to yield:

$$\Sigma_{j=1}^{M} G_{ij} \Delta m_j - (1/N') \Sigma_i \Sigma_{j=1}^{M} G_{ij} \Delta m_j = \Delta T_i - (1/N') \Sigma_i \Delta T_i$$

$$\Sigma_{j=1}{}^{M} [G_{ij} - (1/N') \Sigma_i G_{ij}] \Delta m_j = \Delta T^R_i$$
$$\Sigma_{j=1}{}^{M} G^R_{ij} \Delta m_j = \Delta T^R_i \qquad (43)$$

The right hand side of this equation has the interpretation of "relative" traveltime error, ΔT^R. The left hand side contains a correspondinly modified form of the data kernel, $\mathbf{G}^R$. This procedure for elimination of origin time error has two undesirable effects. First, and most importantly, information about the laterally-averaged velocity structure is lost. Only lateral heterogeneities in velocity structure can be detected. Second, the new data kernel, $\mathbf{G}^R$, is much less sparse than the unmodified kernel, $\mathbf{G}$, in Equation 26. This loss of sparseness leads to increased storage requirements and computational time.

The loss of sparseness can be avoided through working directly with Equation 41, rather than Equation 43. The error parameter, e, is treated as an unknown. This formulation adds one unknown parameter for each earthquake source. These error parameters, $\mathbf{e}$, must now be determined along with the unknown velocity model parameters, $\Delta \mathbf{m}$. Differential weighting is then used to favor solutions with a large e_i's over solutions with large velocity heterogeneity. With the proper choice of weighting, one achieves a solution that is very close to the exact solution of 43.

A.21. Partial Derivatives of Traveltime with Nodal Position.

We present two method of calculating the derivative of traveltime with position-type model parameters.

Method 1 is analogous to the first method described for velocity-type derivatives (Equation 31).

$$\partial T/\partial \mathbf{m} = \partial T/\partial \mathbf{a} \cdot \partial \mathbf{a}/\partial \mathbf{X'} \cdot \partial \mathbf{X'}/\partial \mathbf{m} \qquad (44)$$

A change in model parameters, $\Delta \mathbf{m}$, causes the position, $\mathbf{x'}^{(i)}$, (in the primed coordinate system) of each of the four nodes of a particular tetrahedron to displace a distance, $\Delta \mathbf{x'}^{(1)}$, as specified by Equation 30. Let us now group the four node positions into a single 12-vector, $\mathbf{X'} = [x'^{(1)}, y'^{(1)}, z'^{(1)}, \ldots z'^{(4)})]^T$ and note that the equation for relating node positions, $\mathbf{X'}$, to the coefficients, $\mathbf{a}$, is Equation 6, $\mathbf{a} = [\mathbf{M}(\mathbf{X'})]^{-1} \cdot \mathbf{v}$. Its derivative can be calculated using the rule:

$$(\mathbf{M} + \Delta \mathbf{M})^{-1} \approx \mathbf{M}^{-1} - \mathbf{M}^{-1} \cdot \Delta \mathbf{M} \cdot \mathbf{M}^{-1}$$

(see, for example, Menke and Abbott, 1990, p. 435). Then

$$\partial \mathbf{a}/\partial X'_i = -\mathbf{M}^{-1} \cdot \partial \mathbf{M}/\partial X_i \cdot \mathbf{M}^{-1} \qquad (45)$$

As each node coordinate, X'_i, appears in only one element of $\mathbf{M}$, and since it appears as itself (as contrasted to a complicated function), $\partial \mathbf{M}/\partial X_i$ is a 4×4 matrix with one

33

element equal to unity and all the rest zero.

In the unprimed coordinate system, the derivative, $\partial\mathbf{M}/\partial X_i$, is analogous to Equation 38, $\partial\mathbf{X}/\partial\mathbf{m}=\mathbf{W}^*$, where $\mathbf{W}^*$ is a 12×M matrix of the components of the weights, $\mathbf{W}_{ij}$, from Equation 30 that correspond to the four nodes of the tetrahedron under consideration. As the primed and unprimed coordinates are related through the transformation, $\mathbf{S}$, the derivative, $\partial\mathbf{X'}/\partial\mathbf{X}$, just contains the elements of $\mathbf{S}$. Finally, we can use the chain rule to write:

$$\partial T/\partial\mathbf{m} = \partial T/\partial\mathbf{a} \cdot \partial\mathbf{a}/\partial\mathbf{X'} \cdot \partial\mathbf{X'}/\partial\mathbf{X} \cdot \partial\mathbf{X}/\partial\mathbf{m} \qquad (46)$$

As before, the contribution of all tetrahedra along a ray must be summed. An important issue concerns the integration path in the ray integral in the $\partial T/\partial\mathbf{a}$ part of the derivative. Equation 35 indicates that it should be performed along the unperturbed raypath. However, a perturbation in the node positions changes the shape of the tetrahedron, and thus shortens or lengthens the raypath in each tetrahedron. The derivative for any given tetrahedron thus contains an error associated with this change in ray length.   On the other hand, the errors for adjoining tetrahedra are of opposite sign. Any perturbation of the position of a face increase the length of the ray in one tetrahedron but decreases it in the adjoining tetrahedron by an equal amount. Since the velocity is continuous across the face, the traveltime perturbations exactly cancel. Integration along the unperturbed raypath is thus acceptable.

Exceptions to this error cancellation occur when the perturbation moves the receiver interface or a reflector interface, since there is then no compensating adjacent tetrahedron. In practice, receiver interface perturbations are rarely a concern, since in most experiments receivers are placed at known positions. The reflector interface issue is more interesting, and is discussed in the next section.

Method 2 simply uses a combination of raytracing and finite differences to calculate the derivative. The idea is to perturb a model parameter, m_i, (and hence the node positions) and trace a new ray from the source to the receiver interface, using the take-off angles of the unperturbed ray. The traveltime is computed along this perturbed ray. However, the perturbed ray will not in general pass through the receiver point, $\mathbf{x}^r$, but instead intersect the receiver interface at some other point, $\mathbf{x}$. The traveltime thus need to be corrected by adding a term $\mathbf{t} \cdot (\mathbf{x}-\mathbf{x}^r)/v(\mathbf{x}^r)$, where $\mathbf{t}$ is the ray tangent at $\mathbf{x}$. This term represents propagation of a planar wavefront from $\mathbf{x}$ to $\mathbf{x}^r$. The traveltime is calculated along this ray and Equation 39 is used to estimate $\partial T/\partial m_i$.

This method is less efficient than the first but is algebraically simpler. In general, this method requires that one extra ray be traced for each traveltime-model-parameter combination, a total of N×M in all. In practice, however, the number of extra rays is much less, since a

given traveltime is typically associated with a ray that is influenced by only a few model parameters. In order to gain this efficiency, one must test whether $\partial T/\partial m_i$ is zero before proceeding with a full calculation. Lists of nodes affected by a particular model parameter, and tetrahedra affected by a particular node, are very useful for this purpose.

A.22. Partial Derivatives of Reflected Wave Traveltime with Nodal Position.

The method of calculating $\partial T/\partial \mathbf{m}$ that we describe here accounts for the fact that the ray must satisfy Snell's law on the reflector interface. Thus as the position of the interface is perturbed, the point of reflection moves in a compensatory way.

In the following derivation we will rely heavily upon both the chain rule (e.g. as used in Sections A.19 and A.21) and on the fact that, for any n-vectors $\mathbf{x}$ and $\mathbf{y}$, where we can consider one to be a function of the other, the partial derivative $\partial \mathbf{x}/\partial \mathbf{y}$ is the matrix inverse of the partial derivative $\partial \mathbf{y}/\partial \mathbf{x}$.

We assume that the unperturbed ray leaves the source, $\mathbf{x}^s$, with tangent, $\mathbf{t}^{s0}$, reflects from the interface and arrives at the receiver, $\mathbf{x}^r$, with tangent, $\mathbf{t}^{r0}$. We divide the ray into two parts, ray A from source to reflection point and ray B from receiver to reflection point (Figure 17). Note that we have reversed the sense of propagation of ray B. It now leaves the receiver with tangent, $-\mathbf{t}^{r0}$, and propagates back towards the reflecting interface. Let us denote the takeoff angles associated with $\mathbf{t}^{s0}$ and $-\mathbf{t}^{r0}$ as ψ^{A0} and ψ^{B0}, their traveltimes as T^{A0} and T^{B0}, their intersections points with the reflecting interface as $\mathbf{x}^{A0}$ and $\mathbf{x}^{B0}$, and their tangents at those points as $\mathbf{t}^{A0}$ and $\mathbf{t}^{B0}$, respectively. Since rays A and B constitute a true reflected ray, $\mathbf{x}^{A0}=\mathbf{x}^{B0}$ and their tangents $\mathbf{t}^{A0}$ and $\mathbf{t}^{B0}$ satisfy Snell's law (that is, in a coordinate system normal to the reflecting interface, their interface-parallel components are equal). Let these components be denoted by the 2-vectors, $\mathbf{t}_H^{A0}$ and $\mathbf{t}_H^{B0}$, and the corresponding components of position be denoted, $\mathbf{x}_H^{A0}$ and $\mathbf{x}_H^{B0}$. Then Snell's law implies $\mathbf{t}_H^{A0}=-\mathbf{t}_H^{B0}$ and $\mathbf{x}_H^{A0}=\mathbf{x}_H^{B0}$.

We now perturb a single model parameter i by Δm_i (thus causing a change in the position of the reflecting interface) and retrace rays A and B. In general neither the conditions $\mathbf{x}_H^A=\mathbf{x}_H^B$ nor $\mathbf{t}_H^A=-\mathbf{t}_H^B$ will be satisfied. We must therefore perturb the take-off angles, ψ^A and ψ^B, so that these conditions are once again satisfied.

We begin by calculating, by finite differences (e.g. Eqn 23), the derivatives $\partial \mathbf{x}_H^A/\partial \psi^A$, $\partial \mathbf{t}_H^A/\partial \psi^A$, $\partial T^A/\partial \psi^A$, $\partial \mathbf{x}_H^B/\partial \psi^B$, $\partial \mathbf{t}_H^B/\partial \psi^B$, and $\partial \mathbf{T}^B/\partial \psi^B$. Note that two additional rays must be traced from both the source point and the receiver point to the reflecting interface. We then use the chain rule to eliminate the take-off angles from these derivatives, that is, to find the derivatives of tangent and

traveltime with reflector position:

$$\partial \mathbf{t}_H{}^A / \partial \mathbf{x}_H{}^A = \partial \mathbf{t}_H{}^A / \partial \psi^A \, [\partial \mathbf{x}_H{}^A / \partial \psi^A]^{-1} \qquad (47)$$

$$\partial T^A / \partial \mathbf{x}_H{}^A = \partial T^A / \partial \psi^A \, [\partial \mathbf{x}_H{}^A / \partial \psi^A]^{-1} \qquad (48)$$

$$\partial \mathbf{t}_H{}^B / \partial \mathbf{x}_H{}^B = \partial \mathbf{t}_H{}^B / \partial \psi^B \, [\partial \mathbf{x}_H{}^B / \partial \psi^B]^{-1} \qquad (49)$$

$$\partial T^B / \partial \mathbf{x}_H{}^B = \partial T^B / \partial \psi^B \, [\partial \mathbf{x}_H{}^B / \partial \psi^B]^{-1} \qquad (50)$$

The matrix inversion can be computed analytically, as all the above matrices are 2×2. We use these derivatives to write first order equations for variation of the tangents and the traveltime with the position, $\mathbf{x}_H$, of the reflection point:

$$\mathbf{t}_H{}^A(\mathbf{x_H}) = \mathbf{t}_H{}^A(\mathbf{x}_H{}^A) + \partial \mathbf{t}_H{}^A / \partial \mathbf{x}_H{}^A \cdot (\mathbf{x}_H - \mathbf{x}_H{}^A) \qquad (51)$$

$$\mathbf{t}_H{}^B(\mathbf{x_H}) = \mathbf{t}_H{}^B(\mathbf{x}_H{}^B) + \partial \mathbf{t}_H{}^B / \partial \mathbf{x}_H{}^B \cdot (\mathbf{x}_H - \mathbf{x}_H{}^B) \qquad (52)$$

$$T^A(\mathbf{x_H}) = T^A(\mathbf{x}_H{}^A) + \partial T^A / \partial \mathbf{x}_H{}^A \cdot (\mathbf{x}_H - \mathbf{x}_H{}^A) \qquad (53)$$

$$T^B(\mathbf{x_H}) = T^B(\mathbf{x}_H{}^B) + \partial T^B / \partial \mathbf{x}_H{}^B \cdot (\mathbf{x}_H - \mathbf{x}_H{}^B) \qquad (54)$$

By inserting Eqns. 51 and 52 into Snell's law ($\mathbf{t}_H{}^A = -\mathbf{t}_H{}^B$) we achieve a 2×2 matrix equation for the position of the bounce point, $\mathbf{x}_H$:

$$[\partial \mathbf{t}_H{}^B / \partial \mathbf{x}_H{}^B + \partial \mathbf{t}_H{}^A / \partial \mathbf{x}_H{}^A] \cdot \mathbf{x_H} = $$
$$- \mathbf{t}_H{}^A(\mathbf{x}_H{}^A) - \mathbf{t}_H{}^B(\mathbf{x}_H{}^B)$$
$$+ \partial \mathbf{t}_H{}^B / \partial \mathbf{x}_H{}^B \cdot \mathbf{x}_H{}^B + \partial \mathbf{t}_H{}^A / \partial \mathbf{x}_H{}^A \cdot \mathbf{x}_H{}^A \qquad (55)$$

The solution of this simple 2×2 equation is inserted into Eqns. 53 and 54 to yields an estimate of the traveltimes along the true perturbed ray (that is, one that satisfies Snell's law). The finite difference derivative is then:

$$\partial T / \partial m_i = [\ (T^A + T^B) - (T^{A0} + T^{B0})\ ] / \Delta m_i \qquad (56)$$

A.23. Inversion for Both Compressional and Shear Velocity.

The methodology that we have established above will work for either compressional or shear velocity, taken individually Sometimes, however, a joint inversion for both compressional velocity, α, and shear velocity, β, is preferable, because it provides a opportunity to assess the degree to which their variations track one another. Empirically, one often finds that $\beta = r\alpha$, where the proportionality constant, r, is in the range 0.5-0.6. Identifying regions in which r is significantly different from some background level may be important, because such variations may be indications of strong changes in lithology or temperature.

We begin by defining model parameters for compressional velocity, α, and shear:compressional

velocity ratio, r, that are analogous to Equation 29:

$$\Delta\alpha^{(i)} = \Sigma_{j=1}^{M\alpha} w^{\alpha}_{ij}\, \Delta\mathbf{m}^{\alpha}_{j}\,; \quad i=1, L \qquad (57)$$

$$\Delta r^{(i)} = \Sigma_{j=1}^{Ma} w^{r}_{ij}\, \Delta\mathbf{m}^{r}_{j}\,; \quad i=1, L \qquad (58)$$

Here $\Delta\alpha^{(i)}$ and $\Delta r^{(i)}$ are the perturbations of the nodal compressional velocity and ratio, respectively, and $\Delta\mathbf{m}^{\alpha}_{j}$ and $\Delta\mathbf{m}^{r}_{j}$ are their corresponding model parameters. The total number of model parameters is $M=M^{\alpha}+M^{r}$. In general, the number of velocity and ratio model parameters may be different. In many practical instances the choice, $M^{\alpha}\!>\!>\!M^{r}$, may be reasonable (that is, the ratio is assumed to vary much more slowly with position than does compressional velocity).

Let us assume that there are N^{P} P-wave traveltime data, $\mathbf{T}^{P}$, and N^{S} S-wave traveltime data, $\mathbf{T}^{S}$, with $N^{P}+N^{S}=N$. The data kernel, $\mathbf{G}$, is an $N\times M$ matrix of partial derivatives. Following the approach of A.19, we calculate these derivatives by summing the contributions of each tetrahedron through which the ray P or S rays passes (the two rays in general have different paths). For each tetrahedron we identify for nodal compressional velocities, $\boldsymbol{\alpha}=[\alpha^{(1)},\ \alpha^{(2)},\ \alpha^{(3)},\ \alpha^{(4)}]^{T}$, four nodal shear velocities, $\boldsymbol{\beta}=[\beta^{(1)},\ \beta^{(2)},\ \beta^{(3)},\ \beta^{(4)}]^{T}$, and four nodal ratios, $\mathbf{r}=[r^{1)},\ r^{(2)},\ r^{(3)},\ r^{(4)}]^{T}$. The requirement that both $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ vary linearly within each tetrahedron implies that we must interpolate $\beta(\mathbf{x})$, not $r(\mathbf{x})$, within the tetrahedron. We denote that coefficients of the linear compressional velocity and shear velocity functions as $\mathbf{a}$ and $\mathbf{b}$, respectively (that is, $\alpha=\mathbf{a}\cdot[x', y', z', 1]^{T}$ and $\beta=\mathbf{b}\cdot[x', y', z', 1]^{T}$. The partial derivatives are then:

$$\partial T^{P}/\partial\mathbf{m}^{\alpha} = \Sigma^{\text{tetrahedra}}\ \partial T^{P}/\partial\mathbf{a}\cdot\partial\mathbf{a}/\partial\boldsymbol{\alpha}\cdot\partial\boldsymbol{\alpha}/\partial\mathbf{m}^{\alpha} \qquad (59)$$

$$\partial T^{P}/\partial\mathbf{m}^{r} = 0 \qquad (60)$$

$$\partial T^{S}/\partial\mathbf{m}^{\alpha} = \Sigma^{\text{tetrahedra}}\ \partial T^{S}/\partial\mathbf{b}\cdot\partial\mathbf{b}/\partial\boldsymbol{\beta}\cdot d\boldsymbol{\beta}/d\boldsymbol{\alpha}\cdot\partial\boldsymbol{\alpha}/\partial\mathbf{m}^{\alpha} \quad (61)$$

$$\partial T^{S}/\partial\mathbf{m}^{r} = \Sigma^{\text{tetrahedra}}\ \partial T^{S}/\partial\mathbf{b}\cdot\partial\mathbf{b}/\partial\boldsymbol{\beta}\cdot\partial\boldsymbol{\beta}/\partial\mathbf{r}\cdot\partial\mathbf{r}/\partial\mathbf{m}^{r} \qquad (62)$$

The derivative of nodal shear velocity with nodal compressional velocity is just the 4×4 matrix, $\partial\boldsymbol{\beta}/\partial\boldsymbol{\alpha}=\mathbf{diag}(r^{1)},\ r^{(2)},\ r^{(3)},\ r^{(4)})$. Similarly, the derivative of nodal shear velocity with nodal ratio is the 4×4 matrix, $\partial\boldsymbol{\beta}/\partial\mathbf{r}=\mathbf{diag}(\alpha^{(1)},\ \alpha^{(2)},\ \alpha^{(3)},\ \alpha^{(4)})$. The other component derivatives are just the compressional and shear velocity versions of the derivatives given in in Section A.19.

One sensible arrangement of $\mathbf{G}\cdot\Delta\mathbf{m}=\Delta\mathbf{T}$ is to organize the traveltime perturbations, $\Delta\mathbf{T}$, with the P-wave traveltime perturbations, $\Delta\mathbf{T}^{P}$, on top and the S-wave traveltime perturbations, $\Delta\mathbf{T}^{S}$, on the bottom. Similarly, the the model parameter perturbations, $\Delta\mathbf{m}$, can be organized with the compressional velocity model

perturbations, $\Delta\mathbf{m}^{\alpha}$, on top and the ratio model perturbations, $\Delta\mathbf{m}^{r}$, on the bottom. The data kernel then consists of four blocks of derivatives, with $\partial T^{P}/\partial\mathbf{m}^{\alpha}$ in the upper left, zeroes in the upper right, $\partial T^{S}/\partial\mathbf{m}^{\alpha}$ in the lower left and $\partial T^{S}/\partial\mathbf{m}^{r}$ in the lower right.

A larger damping is usually applied to ratio model parameters than to the compressional velocity model parameters, in order to select a solution preferentially biased toward velocity heterogeneity (as contrasted to ratio heterogeneity). This damping can be achieved by making the overall magnitude of the ratio weights, w^{r}_{ij}, smaller than the magnitude of the velocity weights, w^{α}_{ij}.

A.24. Time-dependent or "4D" Inversions.

Suppose that we perform two experiments, one at time, t_1, and the other at some later time t_2. We might want to determine whether the underlying velocity models differ significantly. We write:

$$
\begin{aligned}
v(\mathbf{x},t_2) &= v(\mathbf{x},t_1) + (dv/dt)\,\Delta t \\
&= [\,1 + \Delta t\,(dv/dt)\,/\,v(\mathbf{x},t_1)\,]\ \ v(\mathbf{x},t_1) \\
&= r(\mathbf{x})\,v(\mathbf{x},t_1) \qquad\qquad (63)
\end{aligned}
$$

with $\Delta t = t_2 - t_1$. The quantity $(r-1)$ represents the fractional change in velocity between times t_1 and t_2. Equation 63 has the same form as the the compressional-shear velocity relationship, $\beta = r\alpha$, of Section A.23, so the corresponding problem can be solved in the same way.

A.25. Inversion for Attenuation.

We begin by defining attenuation model parameters, $\mathbf{m}$, that control the nodal reciprocal quality factors, $Q^{-1(i)}$, at the nodes of the model:

$$
Q^{-1(i)} = \Sigma_{j=1}^{M}\,w_{ij}\,m_j \qquad\qquad (64)
$$

Note that the model is defined in terms of the reciprocal quality factors themselves, rather than in terms of perturbations about a reference model. As we will show below, the attenuation inversion is completely linear, and will not require linearization about a starting model. Within each tetrahedron, the reciprocal quality factor is taken to vary linearly, so that Q^{-1} is give by the usual interpolation formula, $Q^{-1} = \mathbf{q} \cdot [x', y', z', 1]^{T}$, where $\mathbf{q}$ is a 4-vector of coefficients. These coefficients can be found by solving the usual matrix equation, $\mathbf{a} = \mathbf{M} \cdot \mathbf{q}$, where $\mathbf{a} = [Q^{-1(1)},\ Q^{-1(2)},\ Q^{-1(3)},\ Q^{-1(4)}]^{T}$ is a 4-vector of nodal reciprocal quality factor values and $\mathbf{M}$ is a 4×4 matrix of node coordinates in the primed coordinate system of the same form as in Equation 5. Inserting the linear formula into the defining equation for attenuation, $T^* = \int_{ray} Q^{-1}ds/v$, and integrating using the formulas of Section A.11 yields $T^* = \mathbf{q} \cdot \mathbf{h}$, where $\mathbf{h}$ is a 4-vector of trigonometric functions

given by Equations. 11 and 12. The partial derivative of attenuation with model parameters is then:

$$\partial T^*/\partial \mathbf{m} = \Sigma^{\text{tetrahedra}} \; \partial T^*/\partial \mathbf{q} \cdot \partial \mathbf{q}/\partial \mathbf{a} \cdot \partial \mathbf{a}/\partial \mathbf{m}$$
$$= \Sigma^{\text{tetrahedra}} \; \mathbf{h} \cdot \mathbf{M}^{-1} \cdot \mathbf{w} \qquad (65)$$

Here $\mathbf{w}$ is a 4×M matrix of those weights, w_{ij}, that relate the model parameters to the reciprocal quality factors at the 4 nodes of the tetrahedron under consideration. The equation relating attenuation to model parameters is $\mathbf{G} \cdot \mathbf{m} = \mathbf{T}^*$, where $G_{ij} = \partial T^*_i/\partial m_j$. It can be solved via damped least-squares, as given by Equation 28.

A.26. Earthquake Location (Geiger's Method).

Suppose that we have observations of the arrival time, τ_i, of seismic waves from a source at $\mathbf{x}^s$ to a set of N receivers at $x^{r(i)}$. The problem is to use the arrival times to deduce to source location, $\mathbf{x}^s$, and source origin time, t^0. We begin by noting that the arrival time is the origin time plus the traveltime:

$$\tau_i(\mathbf{x}^s) = T_i(\mathbf{x}^s, \mathbf{x}^{r(i)}) + t_0 \qquad (66)$$

where $T_i(\mathbf{x}^s, \mathbf{x}^{r(i)})$ is the traveltime, and then linearize this equation about an initial estimate of the source location, $\mathbf{x}^{s0}$:

$$\tau_i(\mathbf{x}^s) - T(\mathbf{x}^{s0}, \mathbf{x} r^{(i)}) \approx [\partial T_i/\partial \mathbf{x}^s] \cdot \Delta \mathbf{x} + t_0 \qquad (67)$$

Here $\Delta \mathbf{x} = (\mathbf{x}^s - \mathbf{x}^{s0})$ is a 3-vector that represents a small correction to the $\mathbf{x}^{s0}$ that improves the fit to the observed traveltime data. To first order, only the ray-parallel component of a perturbation in source location changes the traveltime, so that $\partial T_i/\partial \mathbf{x}^s = \mathbf{t}(\mathbf{x}^{s0})/v(\mathbf{x}^{s0})$, where $\mathbf{t}$ is the direction of the ray as it leaves the initial estimate of the source, and v is the velocity at that point.

Equation 67 constitutes a N×4 matrix equation for the 4 unknowns, $\Delta \mathbf{x}$ and t_0. It can be solved by damped least-squares, as in Equation 28. However, one must be careful to control the relative weighting of $\Delta \mathbf{x}$ and t_0, which (after all) have different units of measurement. One reasonable approach is to rescale time to distance, using a "typical" velocity. Equation 67 involves a first-order approximation, so several iterative applications of it are generally necessary (with new rays being traced at each step) before the earthquake can be considered adequately located.

A problem of ill-conditioning sometimes arises when only P-wave traveltimes are available (as contrasted to both P and S wave data), and when the source-receiver geometry is such that all observed rays leave the source within a narrow range of directions (that is, with a narrow range about some mean direction, $\mathbf{t}_m$). Origin time then strongly trades off with position, in the sense that a source location perturbation of $\mathbf{t}_m \Delta s$, where Δs is a small

increment of distance, can very nearly compensate an origin time perturbation of $-\Delta s/v(\mathbf{x}^s)$. Damped least-squares can be used to suppress the effects of the ill-conditioning, or more advanced techniques such as singular-value decomposition can be used, instead.

A.27. Earthquake Location Using Differential Traveltimes.

The time difference between phase arrivals from several earthquakes observed on a common station can often be determined more accurately than the individual arrival times, especially if waveform cross-correlation techniques are used to calculate time differences. Similarly, the time difference between phase arrivals to a common station can often be predicted more accurately than the traveltimes of individual phases, especially if the earthquake hypocenters are spatially close to one another. The errors in predicted traveltimes due to a poorly known velocity model are usually highly correlated between the phases, and tend to cancel out when time differences are computed. Many authors [e.g. *Spence and Alexander, 1968; Slunga et al., 1995; Waldhauser and Ellsworth, 2000*] have pointed out that these two effects imply that relative location of a group of earthquakes can often be determined more accurately than the absolute location of individual members of the group.

Suppose that we have a total of J sources, each with location, $\mathbf{x}^s_j$, and origin time, t_{0j}, j=1, ... J. There are a total of 4J unknown source parameters (the three components of position and origin time). Suppose that there are K receivers, $\mathbf{x}^r_k$, k=1, ... K. And suppose that there are M arrival times, τ_m, m=1, ... M each from from a source at $\mathbf{x}^s_{j(m)}$ to a receiver at $\mathbf{x}^r_{k(m)}$, m=1, ... M. Finally, we suppose that the actual data, d_n, are not the arrival times themselves, but rather N linear combinations of these arrival times:

$$d_n = \Sigma_{m=1}^{M} W_{nm} \tau_m , \quad n=1, ... N \qquad (68)$$

The each row of the matrix, $\mathbf{W}$, specifies the coefficients for one datum, d_n. Several special cases are worth noting: A row that is all zero except for a single element with a value of unity would represent an individual traveltime measurement. A row that is all zero except for two elements, one +1 and the other −1, would represent a differential traveltime measurement. Finally, we note that weight of an individual measurement, d_n, in the earthquake location process can be specified by multiplying the corresponding row of $\mathbf{W}$ by an appropriate constant.

The fundamental earthquake location equation (Equation 67) can now be extended to this case by multiplying by the weight matrix and summing:

$$d_n - \Sigma_{m=1}^{M} W_{nm} T(\mathbf{x}^{s0}_{j(m)} , \mathbf{x}^r_{k(m)}) \approx$$
$$\Sigma_{m=1}^{M} W_{nm} \Sigma_{p=1}^{3} [\partial T_m/\partial x^s_{j(m)p}]\Delta x_{j(m)p} + \Sigma_{m=1}^{M} W_{nm} t_{0j(m)} \quad (69)$$

Here $\mathbf{x}^{s0}_j$ represents the initial guess of the location of the j-th source, and $\Delta\mathbf{x_j}$ is an improvement to it ($\Delta x_{j\,p}$ is its p-th component). These are a set of N equations for the 4J unknowns (the three components of the improvement $\Delta\mathbf{x}$ and origin times, t_0, for each of the J events). The locations of all the J events are coupled. They must all be solved for simultaneously.

Normally, the number of earthquakes greatly exceeds the number of stations. Raytracing from the stations to the earthquakes, rather than *vice versa*, is thus most efficient Unfortunately, the point–to–point raytracing procedure discussed above in Section A.15 is not well–suited to this problem, because it is rather artificial to think about the earthquake hypocenters lying on a single interface, and because that interface would have to be redefined from iteration to iteration (as the hypocentral locations are updated). We solve this problem by generalizing the two dimensional traveltime table described in Section A.15 to a fully three-dimensional one. Prior to the start of the earthquake location process, we shoot a large set of rays from each station (say with takeoff angles θ and φ), and build a table of the positions, $\mathbf{x}(\phi,\varphi,T)$ along each ray, sampled at equally spaced traveltimes, T. We then view the ray tube formed by three neighboring rays as being composed of a set of tetrahedra whose vertices are at $\mathbf{x}(\phi,\varphi,T)$. An estimate of the takeoff angles and traveltime of a ray connecting a given station to a hypocenter at $\mathbf{x}$ can then be estimated by selecting the ray tube tetrahedron (or tetrahedra, if there are multiple arrivals) that contains $\mathbf{x}$ and linearly interpolating using the values at the vertices. If greater accuracy is required, then Newton's method Equation 21 can then be used to iteratively refine this estimate.

We note two issues related to this traveltime calculation strategy. First, a traveltime table must be maintained for each station, and the size of this table is necessarily rather large. Second, finding which ray tube tetrahedron (or tetrahedra, if there are multiple arrivals) contains a given point, $\mathbf{x}$, is a non-trivial process, especially if one wants to avoid the time-consuming process of scanning the entire table. The problem is that the topology of the traveltime surfaces can be very complicated, so that search algorithms such as described in Section A.6 are not applicable. Our solution is based on maintaining a list of all model tetrahedra touched by a given ray tube. One first finds the model tetrahedron containing $\mathbf{x}$, and then exhaustively searches within only those ray tubes that intersect that model tetrahedron.

FIGURE CAPTIONS

Figure 1. Record section of vertical-component (bottom) and horizontal-component (top) seismograms of regional earthquake

observed on a linear array of seismometers. Note that clear crustal-turning waves (P and S), Moho-reflected waves (PmP and SmS) and free-surface reflections (PP and PPP) are present. This example is from northern Iceland (Menke et al. 1998), where the crust is mafic and 25-30 km thick.

Figure 2. Top: Hypothetical source-receiver geometry for α/β Case Study. Map view of receivers (circles), P-wave sources (triangles) and S-wave sources (stars). Bottom: Vertical cross-section through center of three-dimensional model, showing typical raypaths. Note deepest ray turns at a depth of about 6 km.

Figure 3. Tomographic images for α/β Case Study. Row A: true model. Row B: starting model. Row C. Results of separate P velocity, α, and S velocity, β, inversions. Row D: Results of joint inversion for α and α/β in which the inversion has been selectively damped to favor variability in α over variability in β. Models C and D fit the traveltime data to within its noise.

Figure 4. Top: Hypothetical source-receiver geometry for Moho Topography Case Study. Map view of receivers (circles) and sources (triangles). Bottom: Vertical cross-section through center of three-dimensional model, showing typical raypaths.

Figure 5. Tomographic images for Moho Topography Case Study. A: True model. B: Starting model. C. Results of an inversion that has been selectively damped to favor variability in Moho topography over variability in crustal velocity. D: Results of inversion that has been selectively damped to favor variability in crustal velocity over variability in Moho topography. Models C and D fit the traveltime data to within its noise.

Figure 6. Six hypothetcial upper-mantle models. These six models have the velocity $v(x,y,z)=v_0(z)+\Delta v(x,y,z)$. Each has a different v_0, but all have the same Δv, implying that they are indistinguishable using teleseismic tomography. The upper-left model has the anomaly superimposed on the constant background, $v_0(z)=8$ km/s, so the anomaly, $\Delta v(x,y,z)$, is countoured.

Figure 7. Top: Hypothetical source-receiver geometry for Plume Shape Case Study. Map view of receivers (circles) and teleseismic sources (arrows). Bottom: Vertical cross-section through center of three-dimensional model, showing typical raypaths.

Figure 8. Tomographic images for Plume Shape Case Study. A: True model. B: Starting model. C: Results of an inversion with depth-constant damping. D: Results of "squeezed-deep" inversion that has been selectively damped to favor deep variability in velocity. E: Results of a "squeezed-shallow" inversion that has been selectively damped to favor shallow variability in velocity. F: Modal inversion. Models C through F all fit the traveltime data to within its noise.

Figure 9. Depth profiles of the velocity anomaly from the center of the three nodal inversions for plume structure. A). Case of depth-constant damping. B) "Squeezed-deep" inversion. C) "Squeezed-shallow" inversion.

Figure 10. Earthquake Location Case Study. A) Top view of model showing stations (squares) and earthquakes (dots). B) True model (shown with 0.5 km/s contours) has significant lateral heterogeneity in upper and mid crust. C) True locations of earthquakes. D) Model used in location (shown with 0.5 km/s contours) is similar to the true model, but laterally homogeneous. E) Mislocation vectors when data are P and S wave traveltimes. F) Mislocation vectors when data are differential P times. See text for further discussion.

Figure 11. Enlargement of central area of earthquake location case study, showing mislocation vectors for A) P and S traveltime data, XZ plane. B) P and S traveltime data, YZ plane. C) Differential P traveltime data, ZX plane. D) Differential P traveltime data, XY plane.

Figure 12. Four neighboring nodes (circles) of a Cartesian mesh form a parallepiped that can be subdivided into six tetrahedra.

Figure 13. A "warped" Cartesian mesh. Two sub-horizontal interfaces have been shaded.

Figure 14. The ray path in the tetrahedron is an arc of a circle. The center of this circle is outside the tetrahedron at the point where the velocity, v, is zero.

Figure 15. Finding the intersection of a circle and a line. A) The intersection points (black circles) are positioned symmetrically about the perpendicular bisector (dashed). B) The intersection points can be easily calculated in a coordinate system chosen so the z" axis is parallel to the perpendicular bisector.

Figure 16. Procedure for point-to-point interface. A) A suite of rays are shot from the source to an interface containing the receiver. The three rays (bold curves) forming a triangle enclosing the receiver are selected. B) The triangle is recursively trisected into smaller triangles (shaded) by tracing a sequence of rays (labeled 2, 3, 4), each one of which divides a triangle into three smaller pieces.

Figure 17. Diagram for computing the partial derivative of reflected wave traveltime with reflected interface position. A) Unperturbed case. B) Perturbed case. See text for further discussion.

Figure 1. Record section of vertical-component (bottom) and horizontal-component (top) seismograms of regional earthquake observed on a linear array of seismometers. Note that clear crustal-turning waves (P and S), Moho-reflected waves (PmP and SmS) and free-surface reflections (PP and PPP) are present. This example is from northern Iceland (Menke et al. 1998), where the crust is mafic and 25-30 km thick.

Figure 2. Top: Hypothetical source-receiver geometry for α/β Case Study. Map view of receivers (circles), P-wave sources (triangles) and S-wave sources (stars). Bottom: Vertical cross-section through center of three-dimensional model, showing typical raypaths. Note deepest ray turns at a depth of about 6 km.

Figure 3. Tomographic images for α/β Case Study. Row A: true model. Row B: starting model. Row C. Results of separate P velocity, α, and S velocity, β, inversions. Row D: Results of joint inversion for α and α/β in which the inversion has been selectively damped to favor variability in α over variability in β. Models C and D fit the traveltime data to within its noise.

Figure 4. Top: Hypothetical source-receiver geometry for Moho Topography Case Study. Map view of receivers (circles) and sources (triangles). Bottom: Vertical cross-section through center of three-dimensional model, showing typical raypaths.

Figure 5. Tomographic images for Moho Topography Case Study. A: True model. B: Starting model. C. Results of an inversion that has been selectively damped to favor variability in Moho topography over variability in crustal velocity. D: Results of inversion that has been selectively damped to favor variability in crustal velocity over variability in Moho topography. Models C and D fit the traveltime data to within its noise.

Figure 6. Six hypothetcial upper-mantle models. These six models have the velocity $v(x,y,z)=v_0(z)+\Delta v(x,y,z)$. Each has a different v_0, but all have the same Δv, implying that they are indistinguishable using teleseismic tomography. The upper-left model has the anomaly superimposed on the constant background, $v_0(z)=8$ km/s, so the anomaly, $\Delta v(x,y,z)$, is countoured.

Figure 7. Top: Hypothetical source-receiver geometry for Plume Shape Case Study. Map view of receivers (circles) and teleseismic sources (arrows). Bottom: Vertical cross-section through center of three-dimensional model, showing typical raypaths.

Figure 8. Tomographic images for Plume Shape Case Study. A: True model. B: Starting model. C: Results of an inversion with depth-constant damping. D: Results of "squeezed-deep" inversion that has been selectively damped to favor deep variability in velocity. E: Results of a "squeezed-shallow" inversion that has been selectively damped to favor shallow variability in velocity. F: Modal inversion. Models C through F all fit the traveltime data to within its noise.

Figure 9. Depth profiles of the velocity anomaly from the center of the three nodal inversions for plume structure. A). Case of depth-constant damping. B) "Squeezed-deep" inversion. C) "Squeezed-shallow" inversion.

Figure 10. Earthquake Location Case Study. A) Top view of model showing stations (squares) and earthquakes (dots). B) True model (shown with 0.5 km/s contours) has significant lateral heterogeneity in upper and mid crust. C) True locations of earthquakes. D) Model used in location (shown with 0.5 km/s contours) is similar to the true model, but laterally homogeneous. E) Mislocation vectors when data are P and S wave traveltimes. F) Mislocation vectors when data are differential P times. See text for further discussion.

Figure 11. Enlargement of central area of earthquake location case study, showing mislocation vectors for A) P and S traveltime data, XZ plane. B) P and S traveltime data, YZ plane. C) Differential P traveltime data, ZX plane. D) Differential P traveltime data, XY plane.

Figure 12. Four neighboring nodes (circles) of a Cartesian mesh form a parallepiped that can be subdivided into six tetrahedra.

Figure 13. A "warped" Cartesian mesh. Two sub-horizontal interfaces have been shaded.

Figure 14. The ray path in the tetrahedron is an arc of a circle. The center of this circle is outside the tetrahedron at the point where the velocity, v, is zero.

Figure 15. Finding the intersection of a circle and a line. A) The intersection points (black circles) are positioned symmetrically about the perpendicular bisector (dashed). B) The intersection points can be easily calculated in a coordinate system chosen so the z" axis is parallel to the perpendicular bisector.

Figure 16. Procedure for point-to-point interface. A) A suite of rays are shot from the source to an interface containing the receiver. The three rays (bold curves) forming a triangle enclosing the receiver are selected. B) The triangle is recursively trisected into smaller triangles (shaded) by tracing a sequence of rays (labeled 2, 3, 4), each one of which divides a triangle into three smaller pieces.

Figure 17. Diagram for computing the partial derivative of reflected wave traveltime with reflected interface position. A) Unperturbed case. B) Perturbed case. See text for further discussion.

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

TOMOGRAPHY AND EARTHQUAKE LOCATION

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE

MENKE