A Simultaneous Tomographic Inversion of ISC Travel Time Residuals for Mantle P Velocity, Source Mislocations, and Station Corrections

By

R. Jay Pulliam¹, D.W. Vasco²,³, and Lane R. Johnson²,³

Technical Report No. 361
July 1992

¹Department of Statistics, University of California, Berkeley
²Department of Geology and Geophysics, University of California, Berkeley
³Center for Computational Seismology, Lawrence Berkeley Laboratory, Berkeley, California

Department of Statistics
University of California
Berkeley, California 94720
A Simultaneous, Tomographic Inversion of ISC Travel Time Residuals for Mantle P Velocity, Source Mislocations, and Station Corrections

R. JAY PULLIAM$^{1,3}$, D.W. VASCO$^{2,3}$, AND LANE R. JOHNSON$^{2,3}$

$^1$ Department of Statistics, University of California, Berkeley
$^2$ Department of Geology and Geophysics, University of California, Berkeley
$^3$ Center for Computational Seismology, Lawrence Berkeley Laboratory, Berkeley, California

ABSTRACT

We use ISC P-arrival data (1964-1987) and the LSQR algorithm to solve for a three-dimensional P-velocity model, source mislocations, and station corrections simultaneously. The model mantle is parameterized by approximately equal-area blocks: $5^\circ \times 5^\circ$ and generally 200 km in depth. More than 3 million rays from 46,000 shallow events satisfying selection criteria are averaged according to $2^\circ \times 2^\circ \times 10$ km deep bins to construct nearly 726,000 summary rays for the inversion. The LSQR algorithm is a least-squares approach which minimizes the $l^2$ norm of the residuals. The data are weighted by the inverse variance of travel time residuals as a function of epicentral distance; model parameters are weighted by a measure of the quality of sampling in each model block, or voxel. A roughness penalty is included in the inversion algorithm.

Due to the averaging procedure employed to construct summary events, source location corrections are generally small. Source location corrections in subduction zones generally move the source toward the positive velocity anomaly. Sources in continental regions have the smallest, nearly insignificant, corrections. Station corrections range from -0.94 to +1.07 sec, with no obvious correlation to tectonics or elevation of the station region.

Resolution and co-variance are evaluated by approximate methods. Resolution is estimated by the inversion of a synthetic, checkerboard test pattern and the calculation of point spread functions for selected voxels. Covariance is estimated by averaging results from inversions of realistic errors and by a jackknife procedure.

We present our three-dimensional velocity model in conjunction with the resolution estimates produced by our checkerboard test. Normalized checkerboard output values, ranging from 0 to 1, are used to modify each voxel's red-blue velocity value from full color saturation, indicating good resolution, to white, which indicates no resolution. The velocity model shows a fast anomaly in the lower mantle beneath the Tonga-New Hebrides subduction zone to a depth of 1670 km and another fast anomaly beneath the Japanese Island arc and eastern Asia reaching nearly to the core-mantle boundary. Continuity between these anomalies and shallower fast anomalies is not clear. A fast anomaly extending from 670 km to 2070 km depth appears beneath the eastern United States, Caribbean Sea, and Central South America. In addition, a number of slow anomalies associated with hotspots extend through the upper mantle but are extinguished in the lower mantle by our resolution weighting. Mid-ocean ridges are associated with moderately slow anomalies in the top 400 km of our model. The transition zone shows large $l=1, 2$, and 3 spherical harmonic components. Diminished heterogeneity in the lower mantle, reported by other authors, is confirmed by our study.
1. INTRODUCTION

Attempts to image the three-dimensional seismic heterogeneity of Earth's mantle differ in their approaches to parametrizing the model mantle and in their formulation and solution of the constraining equations. Popular model parametrizations include regionalization of the crust and mantle based on geographical association with surface tectonic processes, spherical harmonic series expansion of the anomalous velocity field, cubic splines, and division into a number of non-overlapping blocks. No one parametrization has been demonstrated to be clearly superior to the others. The bias inherent in a tectonic regionalization makes such a scheme inappropriate for studies of the lower mantle, uneven and incomplete ray coverage makes it useful for studies of the upper mantle. Spherical harmonic expansions and cubic splines require fewer terms to describe a model to the same level of detail as a block parametrization, but do not offer the blocks' geometrical simplicity. With independent block parameters, one may examine the ray sampling, resolution, and covariance of a geographical location more easily.

In addition to differences in model parametrization, studies differ in their construction of the system of equations to be solved and the numerical methods employed to solve them. Early efforts to map the three-dimensional velocity structure of Earth, all of which use a block parametrization, were limited in their structural detail by numerical methods that calculate the explicit inverse of the coefficient matrix. Those methods allow the formal calculation of covariance and resolution matrices to evaluate the reliability of the model, but severely restrict the number of parameters available to describe the model due to limitations of existing computers. Increased detail became possible when iterative, approximate techniques were employed to solve the system of constraining equations. The cost of this increased detail was that resolution and covariance could not be calculated formally and presented alongside the mantle model. Several means to approximate resolution and covariance were developed in conjunction with SIRT backprojections and other methods. For example, some workers calculate a "point spread function", a column of the non-symmetric resolution matrix, and invert model error distributions to investigate the propagation of errors in the data through the algorithm to the solution. Others perform an inversion for a set of anomalies, each of which extends beyond the bounds of a single voxel, distributed throughout the model as a means to evaluate resolution.

While global studies performed to date vary in their methods for obtaining model solutions and in their approaches to parametrizing the model, the data involved in each study of P-velocity are nearly identical.
Since 1964, the International Seismological Center has collected seismic arrival times from around the world. They employ these times in a sophisticated procedure in which times are associated into "events" and the events are located with P-arrival times and the one-dimensional, Jeffreys-Bullen (J-B) travel time tables and a standard least-squares technique [Adams et al., 1982]. For the purposes of seismic imaging and accurate location of events, the geographical distribution of sources (earthquakes and large explosions) and seismographic stations, that so far are located almost exclusively on continents, is unfortunate. With the oceans nearly empty of stations, large portions of Earth, particularly in the southern hemisphere, are poorly sampled by recorded seismic energy, and event locations are inadequately constrained geometrically. Also, using the one-dimensional J-B model, which has known deficiencies in its representation of the spherically averaged structure of Earth in addition to its inability to account for lateral velocity variations, produces inaccurate event locations. The problems surrounding these inaccuracies in source locations form a central concern of this thesis.

In this study we use ISC P-arrival data from January 1964 through January 1987 and the LSQR algorithm to solve for a three-dimensional P-velocity model of the mantle, source mislocations, and station corrections. Reliability of our model is checked by approximate means, and the models produced by Dziewonski [1984], Clayton and Comer [1983; Hager and Clayton, 1989], and Inoue et al. [1990] provide valuable comparisons. Our study differs from Clayton and Comer's [1983; Hager and Clayton, 1989] in that we use the LSQR algorithm instead of the SIRT algorithm to solve the constraining equations, and that we solve for source and station terms simultaneously. In contrast to Dziewonski [1984] who parametrizes the model mantle with spherical harmonics, and solves the equations with the generalized inverse, we use a block parametrization and the LSQR algorithm. Inoue et al.'s [1990] procedure severely downweights outliers, while ours does not. The consequences of disregarding outliers are discussed more fully later in this paper. Also, we solve for source and station terms differently than Inoue et al. [1990], use summary rays formed from the entire data set instead of using subsets consisting of actual rays, and perform the model smoothing differently. A companion study to this one, undertaken by Pulliam et al. [1992] with the same data set, considers the case in which the $l^1$ norm of the travel time residuals is minimized, rather than the $l^2$ norm. The former case more closely approximates Inoue et al.'s [1990] treatment of outlying residuals than the $l^2$ minimization performed here.

2. Method
2.1 Model Parametrization

The starting model used in this study is a one-dimensional, spherically-symmetric P-velocity model modified from *Jeffreys* [1960]. Modifications to the Jeffreys velocity model were necessary to obtain a model consistent with the J-B travel-time tables [*Jeffreys and Bullen*, 1940]. These modifications are small but important because they remove a systematically slow trend for the mid-mantle from the model published by Jeffreys and make the model more consistent with the tables, that are used by the the ISC to find source locations. We use the modified J-B model rather than PREM [*Dziewonski and Anderson*, 1981] or the Herrin model [*Herrin et al.*, 1968], because in tests it provides equally accurate locations for sources with known hypocenters, and because using a model consistent with the J-B tables, that are used by the ISC to locate the events, eliminates the need to relocate all events in a new velocity model.

Our three-dimensional model is parametrized with approximately equal-area volume elements (voxels), 5° by 5° laterally and 14 shells, generally 200 km thick, for a total of 22,876 voxels (see figure 1 and table 1). The exceptions to the 200 km thick shells occur in the upper mantle, in order to place a radial boundary at the 670 km discontinuity, resulting in a 270 km thick shell, and above the core-mantle boundary, where the lowermost shell is 228 km thick.

We write the vector of travel time residuals for a single event as the sum of three terms:

\[ \delta t_j = t_j^{\text{obs}} - t_j^{\text{calc}} = \delta t_j^{\text{hypo}} + \delta t_j^{\text{model}} + \delta t_j^{\text{station}}, \]  

where the length of all vectors, \( m_j \), is equal to the number of arrival times reported for the \( j \)th event. The hypocentral term itself has four contributions: so

\[ \delta t_j^{\text{hypo}} = \frac{\partial t_i}{\partial h_1} \delta h_{1j} + \frac{\partial t_i}{\partial h_2} \delta h_{2j} + \frac{\partial t_i}{\partial h_3} \delta h_{3j} + \frac{\partial t_i}{\partial h_4} \delta h_{4j}, \]  

where

\( t_i = \) the travel time for the \( i \)th ray,

\( \delta h_{1j} = \) origin time correction for the \( j \)th event,

\( \delta h_{2j} = \) latitude correction for the \( j \)th event,

\( \delta h_{3j} = \) longitude correction for the \( j \)th event,

\( \delta h_{4j} = \) depth correction for the \( j \)th event,
\[ i = 1,2,...,m_j; \ j = 1,2,...,n_e, \]

\[ m_j = \text{the number of arrival times for the } j^{\text{th}} \text{ event}, \]

\[ n_e = \text{the number of events in the data set}. \]

In matrix notation,

\[ \delta t_j^{\text{hypo}} = H_j \delta h_j, \tag{3} \]

where

\[ (H_{ji})_j = \frac{\partial t_i}{\partial h_j} = \text{matrix of source mislocation partial derivatives}, \]

\[ l = 1,2,3,4, \]

\[ \delta h_j = \text{vector of hypocenter corrections for event } j. \]

The "model" term in equation (1), \( \delta t_j^{\text{model}} \), represents the deviation of travel times predicted by our starting velocity model, \( \bar{c}(r) \) from travel times through Earth's actual velocity structure, \( c(r) \), for the \( j^{\text{th}} \) event. For single travel time residual, \( \delta t_j^{\text{model}} \), we assume that our starting velocity model is within a few percent of the true structure and seek to reconcile the discrepancy by solving for the perturbation term,

\[ \delta t_i^{\text{model}} = \int_{\bar{s}_i} \frac{ds}{c(r)} - \int_{\bar{s}_i} \frac{ds}{\bar{c}(r)} \]

\[ = \int_{\bar{s}_i} \left( \frac{1}{c(r)} - \frac{1}{\bar{c}(r)} \right) ds \]

\[ = -\int_{\bar{s}_i} \frac{\delta c(r)}{\bar{c}^2(r)} ds, \]

where \( \bar{s}_i \) is the path of the \( i^{\text{th}} \) ray through the starting velocity model, \( \bar{c}(r) \).

We represent the velocity perturbations as a finite linear combination of "basis" functions,

\[ \delta c(r) = \sum_{k=1}^{n} \gamma_k f_k(r), \tag{5} \]

and choose a set of local basis functions in which the medium under investigation is divided into non-overlapping cells, or voxels. Following Nolet [1987], let
\[ f_k(r) = \begin{cases} 1 & \text{if } r \text{ is in cell } k \\ 0 & \text{elsewhere} \end{cases} \]  \hspace{1cm} (6)

Our choice of a local basis is arbitrary in many respects. We prefer the block parametrization to a series representation because it allows a more accurate assessment of ray sampling of Earth and the resulting coefficient matrix is quite sparse. This sparseness may be exploited to solve the matrix problem efficiently. Fewer terms are required to describe the model to the same level of resolution with the global spherical harmonic basis (fewer by up to an order of magnitude), but the coefficient matrix in the spherical harmonic case is dense.

An expression for the travel time perturbations in terms of velocity perturbation basis functions results:

\[ \delta t_{\text{model}}^{\text{model}} = \sum_{k=1}^{n} \gamma_k \frac{f_k(r)}{c(r)^2} ds = \sum_{k=1}^{n} A_{ik} \gamma_k, \]  \hspace{1cm} (7)

where

\[ A_{ik} = -\int_{S_i} \frac{f_k(r)}{c(r)^2} ds. \]

In matrix form,

\[ \delta t_{\text{model}}^{\text{model}} = A \gamma. \]  \hspace{1cm} (8)

The "station" term of equation (1) includes errors in observed travel time residuals resulting from incorrect observations, such as instrument errors and systematic phase mispicks or misidentifications at a particular station, or from inaccuracies in the starting model near the station that occur on a scale too small to be resolved by the 3-D model parametrization. We express these contributions to the travel time residual as

\[ \delta t_{\text{station}} = S \mu, \]  \hspace{1cm} (9)

where

\[ S_{ik} = \begin{cases} 1 & \text{if } k = \text{station number} \\ 0 & \text{if } k \neq \text{station number}, \end{cases} \]

\[ \mu_k = \text{the station correction for the } k^{\text{th}} \text{ station.} \]
Substituting equations (3), (8), and (9) into equation (1), we find the problem we wish to solve is

$$\delta t = A\gamma + H\delta h + S\delta \mu,$$

where

- $\delta t \in R^{M \times 1}$ = vector of travel time residuals,
- $A \in R^{M \times n_e}$ = matrix of ray segments in voxels,
- $\gamma \in R^{n_e \times 1}$ = vector of coefficients in the expansion of perturbations to the starting model,
- $H \in R^{M \times 4n_e}$ = matrix of partial derivatives for all events,
- $\delta h \in R^{4n_e \times 1}$ = vector of perturbations to the hypocenters,
- $S \in R^{M \times n_p}$ = matrix of partial derivatives for stations,
- $\delta \mu \in R^{n_p \times 1}$ = vector of station corrections,

- $M$ = number of data (reported arrivals),
- $n_e$ = number of events,
- $n_p$ = number of model blocks,
- $n_s$ = number of reporting stations.

### 2.2 Simultaneous vs. Progressive Inversion

At this point we may combine matrices into a single, partitioned matrix and solve for all parameters simultaneously, or take advantage of the problem’s natural separation into three distinct classes of parameters and solve for each class progressively. Each approach has its appeal. Simultaneous inversion is simpler conceptually and requires fewer computational operations, but deals with a much larger matrix ($O(M \times n_p + 4n_e + n_s)$), so it demands more core memory, even in row-active implementations. Further, due to the different nature of the parameters to be estimated and their differing scales, results are very sensitive to the scaling applied to the coefficient matrix. Progressive inversion, on the other hand, enables us to exploit the natural separation of the matrix problem to solve for each set of parameters in a step-wise fashion, reducing demands for core memory and eliminating the need for careful scaling of parameter
classes. This approach follows Pavlis and Booker [1980], Spencer and Gubbins [1980], Jordan and Sverdrup [1981], Kennett and Williamson [1988] and O'Connell and Johnson [1991], among others, and allows a more detailed analysis of resolution and uncertainty in the determination of mislocation terms than would be practical otherwise. Unfortunately, the projections involved for each source mislocation matrix tend to fill in elements in originally sparse coefficient matrices, which increases the computation time required to solve the complete problem and increases disk-access time for row-active algorithms. This last point is critical for global-scale problems involving large data sets and large numbers of model parameters. Still, each method is feasible for problems involving matrices of the order $10^6 \times 10^4$. Preliminary results of a progressive inversion algorithm have been presented [Pulliam and Johnson, 1989b], as well as synthetic tests of the algorithm [Pulliam and Johnson, 1989a].

2.3 Simultaneous Inversion

Here we choose to combine the three coefficient matrices and solve for all parameters simultaneously, i.e.

$$\begin{bmatrix} \gamma \\ A | H | S \end{bmatrix} \begin{bmatrix} \delta h \\ \delta \mu \end{bmatrix} = \delta t,$$

(11)

or

$$Gx = \delta t,$$

(12)

where

$$G = \begin{bmatrix} A | H | S \end{bmatrix}$$

and

$$x = \begin{bmatrix} \gamma \\ \delta h \\ \delta \mu \end{bmatrix}$$

H and S are first scaled so that each row has the same euclidean norm as the same row of A.

Equation (12) presents us with a classical linear inverse problem. Typically, the $M \times N$ coefficient matrix, G, will have many more data than parameters ($M \gg N$) and, given that errors are contained in the data, the equations will almost surely be inconsistent. As a criterion for minimizing the misfit of parameters to data we choose the euclidean ($l^2$) norm, resulting in the least squares problem:
\[ \text{Min} \| Gx - \delta t \|^2 = \text{Min} (Gx - \delta t)^T (Gx - \delta t). \] 

(13)

This is not a simple choice. Strictly, least-squares is most appropriate for problems involving a Gaussian distribution of errors. When applied to such a distribution, least-squares produces the maximum likelihood solution to the linear matrix equation. However, residuals contained in ISC travel time data are not clearly Gaussian [Buland, 1984; Pulliam et al., 1992]. More observations are found in the distribution's tails than one would expect in a Gaussian distribution. We may transform the entire matrix problem to one involving a Gaussian distribution of residuals by applying a set of weights to rows of the problem (12). These weights may be obtained through uniform variance reduction analysis of the original vector of residuals, \( \delta t \) [Jeffreys, 1939; Buland, 1986]. Alternatively, we can remove the bulk of the blunders and gross random errors by truncating the distribution. We choose to truncate the summary residual distribution (figure 7) at \( \pm 7 \) seconds. This truncation value assures us we will not discard too many reliable observations and, as a test of whether we are keeping too many unreliable observations, we apply the uniform variance reduction method to our truncated set of residuals. After inverting the sets of modified and unmodified summary residuals, we find differences in velocities for individual voxels on the order of 0.01%. Apparently the outliers are sufficiently few in number, relative to the central portion of the distribution, that the influence they exert on the final model is minor.

Analysis of a set of travel time residuals to which corrections for a three-dimensional model are applied reveals that their distribution may be more similar to a two-sided exponential than a Gaussian. If this is true, minimizing the \( l^1 \) norm would be more appropriate than \( l^2 \) minimization. Since the \( l^1 \) norm is less sensitive to outliers in the distribution of residuals, gross errors in the dataset -- due to mis-identification of phases, mis-readings of time codes, faulty instruments, and source mislocation, for example -- would be less likely to propagate through the inversion to the model. On the other hand, some of the outlying residuals in the tails of the distribution constitute real and significant data, indicating relatively large velocity differences between the real earth and our starting model. Raytracing from "calibration" events, sources with known locations, produce travel time residuals amounting to 5 seconds or more at some stations. Providing the residuals resulting from the more extreme velocity anomalies in the real earth do not violate the assumptions under which we linearized the originally nonlinear travel time problem, these data are the ones we wish to emphasize in the inversion, not the extreme errors or the minor deviations clustered around zero. Both \( l^2 \) and \( l^1 \)-minimization approaches warrant our attention in order to compare resulting models. The \( l^1 \)-minimization is pursued in a companion study by Pulliam et al. [1992] using the same dataset.
From here on our development parallels the development of the inverse problem in Pulliam et al., [1992]. Ultimately we arrive at the modified set of equations

\[
\begin{bmatrix}
\mathbf{b} \\
\mathbf{0}
\end{bmatrix} = \begin{bmatrix}
\mathbf{G} \\
\mathbf{\lambda B}
\end{bmatrix} \mathbf{x},
\]

(14)

where

\[
\mathbf{G} = \mathbf{W}_d \mathbf{G} \mathbf{W}_x,
\]

\[
\mathbf{b} = \mathbf{W}_d \mathbf{b}_t,
\]

\[
\mathbf{x} = \mathbf{W}_x \mathbf{x}
\]

subject to the minimization of \( ||\mathbf{x}||^2 \). Our estimate of \( \mathbf{W}_d \) comes from the standard errors of ISC travel time residuals as a function of epicentral distance. Assuming the data are independent, we form the diagonal matrix

\[
\mathbf{W}_d = \begin{bmatrix}
\frac{1}{\sigma_1} & 0 & \ldots & 0 \\
0 & \frac{1}{\sigma_2} & \ldots & 0 \\
0 & 0 & \ldots & 1 \\
0 & 0 & \ldots & \frac{1}{\sigma_M}
\end{bmatrix}
\]

\( M = \) number of rays in the data set,

\( \sigma_i = \) standard errors of ISC travel time residuals,

\( i = 1, 2, \ldots, M. \)

Our estimate of \( \mathbf{W}_x \) is

\[
\mathbf{W}_{xx} = \frac{\sum_i l_i}{n_k \sqrt{\kappa}}.
\]

where
\begin{align*}
l_i &= \text{the length of the } i^{th} \text{ ray segment in voxel } k, \\
v_k &= \text{the volume of the } k^{th} \text{ voxel}, \\
n_k &= \text{the number of ray segments in the } k^{th} \text{ voxel.}
\end{align*}

We solve these equations with the LSQR algorithm.

3. Data

3.1 Data Selection

The data inverted in this study were obtained from the catalog of the International Seismological Centre (ISC) for the period January 1964 through January 1987 (frontispiece). To avoid contamination of our mantle phases by Earth’s core we limit the range of our coverage to epicentral distances between 0° and 96°. The confusing scatter caused by refractions from the 400 km and 670 km discontinuities at about 15° to 25° is dealt with in the inversion process by weighting each summary ray by the inverse of the standard deviation of travel time residuals as a function of delta. A single event must have a minimum of twenty-five reporting stations, and source depths as reported by the ISC must be greater than 0 km and less than 70 km. We discard all events located by the ISC at Earth’s surface, but retain events located at the other default depths. All observations are corrected for ellipticity by integration along the raypath. Travel time residuals are formed by subtracting the time calculated by tracing rays through the spherically-symmetric starting velocity model from the observed time corrected for ellipticity. Raytracing is performed by a shooting method involving the direct numerical integration of the eikonal equations with an integration scheme that checks the local error at each integration step. Lengths of ray segments in voxels are found by integrating distance along the curved raypath and finding the intersections of rays with voxel boundaries. Rays associated with residuals greater than seven seconds are discarded. Approximately 3.02 million rays satisfy these criteria. Figures 2 and 3 show the locations of the selected events and seismographic stations, respectively.

We examined histograms of travel time residuals associated with events assigned by the ISC to five different default depths, 0, 5, 10, 15, and 33 km, and compared the residual distributions for these events to the residual distribution of remaining events. Due to a problem with our FORTRAN subroutine, ISC records in which the source depth was left blank defaulted to zero source depth. Although the residual distribution shows a clear bimodal pattern, we were unable to distinguish reliably between true, zero-depth locations and blank-depth locations after the data had been extracted from the ISC master set. For this
study we discarded all events with source depth equal to zero. However, histograms for the four remaining
default depths are indistinguishable from the histogram for all remaining events, so we cannot justify culling
events with source depth = 5, 10, 15, or 33 km. A histogram of the culled data set is shown in figure 4
along with the first four moments of the travel time residual distribution.

3.2 Summary Rays

Summary rays were formed as composites of rays that sample nearly the same portion of Earth. Bins con-
 sist of 2° by 2° voxels divided into seven depth intervals of 10 km each, for a total of 71,904 bins. Residu-
 als of rays emanating from and ending in the same two bins are averaged into a single, summary ray.
These bins are quite small, even compared to our model voxels. The result is a set of summary rays in
which most are composed of very few actual rays, typically two or three. We apply no minimum cutoff, so
nearly half of the resulting rays consist of just a single ray, not a composite at all. By constructing sum-
mary rays we seek to reduce the redundancy of the data set, in order to mitigate the effects of nonuniform
sampling of Earth on our final model, and remove variations in travel times due to heterogeneity on a scale
smaller than our velocity model parametrization. At the same time we would like to preserve as much of
the original variation of residuals as possible. We wish to allow the inversion algorithm to reconcile the
discrepancies in travel times, rather than remove these discrepancies in a pre-processing step. This approach
allows us to evaluate the performance of our algorithm using test cases that better represent the true case of
inconsistent and erratic travel times in Earth. Figures 5 and 6 show the locations of summary events and
summary stations.

Our final data set consists of a total of 725,993 summary rays emanating from 5,986 summary events. 
Figure 7 (compare to figure 4) shows the mean, variance, and skewness of the data are reduced significantly
by constructing summary rays. However, the statistics of the two distributions are not directly comparable
because we truncate the summary residuals at ±7 seconds. Compared to the distribution of actual residuals,
the distribution of summary residuals is slightly more like a Gaussian distribution, as indicated by the sizes
of its tails, and slightly less like a two-sided exponential, as indicated by a comparison of cumulative distri-
butions, than is the distribution of actual residuals. Ultimately, the residuals remaining after three-
dimensional structure is accounted for will resemble a two-sided exponential distribution more closely than a
Gaussian.

3.3 Ray Coverage
Figures 8a-f show the distribution of ray segment lengths that make up our coefficient matrix, $A$. These values consist of column sums, indicating the total sampling of each individual voxel by the data set used in this study. In the absence of a weighting matrix, $W_x$, that balances the column norms, results of an inversion would be expected to follow this pattern quite closely. Figure 8a shows the clear demarcation of plate boundary source regions that are well-sampled. Asia, North America, Europe, and Australia are also well-sampled. In contrast, other regions tend to be quite poorly-sampled. There are also large oceanic areas that are completely unsampled by our data set. These voxels do not enter into the inversion. The next depth layer, figure 8b, shows a broadening of the well-sampled regions and a slight reduction of the unsampled oceanic areas. At 400-670 km and 670-870 (figures 8c and 8d) these trends continue, and by the mid-mantle (figure 8e) virtually all voxels are sampled. In general, sampling becomes more homogeneous with depth, and at the bottom of the mantle (figure 8f) the sampling is much more uniform than in the first layer. However, in absolute numbers the sums of ray segments in voxels decrease with depth, even as more voxels are sampled in each layer. Table 1 details the average number of hits for sampled voxels in each layer along with the the average sum of ray segments in a voxel at a given depth and the number of voxels sampled in each depth interval. These averages include only voxels that have non-zero sampling. The trends in Table 1 show that while homogeneity of sampling increases with depth, voxels tend to be less frequently and less heavily traversed by recorded seismic rays. Note in all six figures the strong bias toward the northern hemisphere and toward continents. As we will discover in our treatment of resolution and covariance, the uneven mantle sampling translates directly into uneven constraints and resolution for our final model.

3.4 Stochastic vs. Deterministic Analysis

Gudmundsson et al. [1990] show that there exists a minimum level of stochastic noise in the ISC data set below which we cannot expect to resolve structure and which casts doubt upon the reliability of schemes such as ours to resolve the apparently small velocity anomalies in the lower mantle. This noise might arise from reading or instrument errors or very small-scale structure in the upper mantle that causes multi-pathing, or a breakdown of the ray approximation in general. However, Gudmundsson et al.'s [1990] analysis indicates the level at which the behavior of travel time residuals becomes non-systematic is well below the starting level of our data. For the scale of our model blocks, $5^\circ \times 5^\circ$, their figures show maximum extrapolated variances of somewhat greater than $2\; \text{sec}^2$ at the distance range contaminated by reflections from discontinuities, $15^\circ$ to $25^\circ$, and averaging slightly less than $1\; \text{sec}^2$ outside this range. In contrast, our residuals
average variance is 3.1 sec$^2$ for the raw travel time residuals, which is reduced to 2.6 sec$^2$ after the formation of summary rays and 2.1 sec$^2$ after inversion. Our variance values are not consistent with those of Gudmundsson et al. [1990] because they choose to truncate the residual distribution at 4 sec, whereas we truncate at 7 sec. Clearly, even if half the original variance in travel time residuals cannot be accounted for by our inversion method, the half that can be accounted for is substantial and significant. Gudmundsson et al. [1990] note as well that the signal to random noise ratio in the teleseismic ISC P-wave data is about S/N=2. Their other results elegantly confirm previous indications that upper mantle structure is significantly more heterogeneous than mid- and lower-mantle structure. These results do not stand in the way of imaging lower mantle structure, where that structure is comparable in amplitude to upper mantle heterogeneity, with the portions of the travel time residuals that do vary systematically.

4. Resolution and Uncertainty

Equally important to producing a seismic velocity model is a thorough investigation of the reliability of that model. We need to evaluate the "resolution," the image of an input model as seen through the "filter" of the inverse method, and the "uncertainty," the errors contained in our output resulting from errors in the input data that propagate through the inversion. For a discrete problem of the form (14) in which the data contain Gaussian errors, the estimate of uncertainty takes the form of an a posteriori covariance matrix [Tarantola, 1987],

$$C_m = (\tilde{G}^T C_d^{-1} \tilde{G} + B^T C_m^{-1} B)^{-1}. \tag{15}$$

The resolution matrix is then

$$R = C_m \tilde{G}^T C_d^{-1} \tilde{G}. \tag{16}$$

However, due to the large numbers of data and model parameters required to image Earth's interior to a useful level of detail, formal calculation of covariance and resolution matrices has been beyond our computational capacity. Calculation of resolution and covariance matrices in tomographic inversions have been necessarily approximate and incomplete, and a number of methods have been developed to evaluate a tomographic model's reliability.

4.1 Approximating Resolution

Humphreys and Clayton [1988] explore the resolution of an inversion by means of a synthetic test in which a velocity perturbation is introduced to one or more voxels in a region of interest. Using Fermat's
principle, which holds that travel times calculated through the three-dimensional Earth are insensitive to changes in raypath, one may calculate the travel time residuals that would be produced by the synthetic anomalies without tracing rays in the 3-D model. Synthetic data is constructed with the same raypaths as the original data set and the same model parametrization. The synthetic data is inverted and the voxels adjacent to the perturbed voxels examined for smeared and "ghost" images that are artifacts of the inversion. The result may be thought of as the response of the algorithm to an impulse introduced to the system, and forms one column of the resolution matrix, which is non-symmetric. Humphreys and Clayton [1988] call this vector the "point spread function," distinguishing it from the "resolving kernel" which is the corresponding row of the resolution matrix. One limitation of this approach is that the anomalies are introduced in the span of the parametrization, and the resulting estimate of "resolution" may therefore be misleadingly high.

Figures 9-14 show point spread functions for six voxels. A 5% velocity anomaly is introduced at (17°S, 178°W, 600 km depth) beneath the Tonga subduction zone (figure 9a), and an inversion returns a value of 3.7% for the anomalous voxel (figure 9b). Note the relatively small leakage to adjacent voxels, indicating the region is well-resolved by our method. The maximum smeared anomalies occur in voxels directly above and below the perturbed voxel, at 0.7% and 0.6%, respectively. No other smeared anomaly exceeds 0.2%. A second impulse of 3%, shown in figure 10a, was placed beneath the New Hebrides subduction complex at (16°S, 166°W, 900 km depth). The inversion returns a value of 2.1% and neighboring voxels returned 0.4%, 0.1%, 0.2%, and 0.1% in the same layer (figure 10b). Some streaking along raypaths to the south and north appears at low amplitudes. Both images of anomalies placed in subduction zones (figures 9b and 10b) display the broadening and smearing with depth noted by Spakman et al. [1989] in simulated tests with a much finer model grid. Although the amplitudes of the smeared anomalies are quite small, the smearing is indeed systematic and apparently oriented along the dip of the subduction zone features we will discover later in our model. The voxels above and below return 0.2% and 0.6%, respectively. The voxel two layers below the introduced anomaly returns 0.3%.

The third and fourth tests examine negative anomalies. The first, shown in figure 11a, is introduced to the first layer beneath the Hawaiian islands at (21.4°N, 158°W, 70 km depth). The -3% anomaly is returned as -2.1%, and the maximum spurious perturbation, -0.66%, occurs in the voxel immediately below and to the west of the perturbed voxel (figure 11b). Note the smearing in this case into voxels to the west and east, which lie along common raypaths to sources in the northwest Pacific and to North America, respectively. This smearing results from poor geometrical constraint of the voxels beneath Hawaii; most arriving rays
travel along parallel paths. An anomaly beneath Iceland produces a similar smearing pattern to the east and west. A -4% anomaly at (65°N, 18°W, 70 km depth) (figure 12a) returns as -2.5% with a smeared value in the voxel just below reaching -1.5% and the voxel two layers below reaching -0.3% (figure 12b). Figure 13a shows a 3% anomaly introduced beneath central South America at (0°, 65°W, 800 km depth). A value of 1.8% is returned by our inversion, with smeared values above reaching 0.4% and below reaching 0.3% (figure 13b). A systematic smearing feature dips to the east, possibly along rays emanating from subduction zone events, but at very low amplitudes. A 4% velocity anomaly is introduced at (40°N, 85°W, 800 km depth) beneath the eastern North America (figure 14a), and an inversion returns a value of 2.5% for the anomalous voxel (figure 14b). Here the leakage to adjacent pixels displays an interesting pattern. As we will see later, a significant anomaly appears in our model through the Carribbean and into central South America that corresponds to the smeared velocities shown here. As in other tests, the largest values appear in voxels directly above and below the perturbed voxel, both at 0.5%. No other value exceeds 0.2%. While the amplitudes of smeared values are small, the systematic error suggests we would not have strong control on the spatial extent of an actual anomaly and warns us to be cautious in the interpretation of results in this region. In addition to the values returned to perturbed voxels and their immediate neighbors, the test inversion produces several "ghost" images away from introduced anomalies. The largest of these amounts to 0.1%, or one-tenth of the smallest value returned by the inversion for an introduced anomaly.

*Inoue et al.* [1990] show a way to approximate the resolving kernel for one model parameter, as well as the corresponding row of the covariance matrix. The idea is to use the LSQR algorithm to solve

\[ \mathbf{G}^T \mathbf{G} \mathbf{x}_j = e_j, \]

where \( e_j \) is a vector whose \( j^{th} \) element is 1, while all other elements are 0. After \( \mathbf{x}_j \) is found, the \( j^{th} \) row vector, \( \mathbf{y}_j \) of \( \mathbf{R} \) is given by

\[ \mathbf{y}_j = \mathbf{x}_j \mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G}. \]

The drawback of calculating a single row or column of the resolution matrix is that each interesting feature must be examined individually with a separate inversion. The result is a visual representation of resolution presented in a number of figures that must be compared simultaneously to the model produced with real data.

A similar, though more complete, approach is to introduce a full model, so that a value is specified for every voxel, and invert the synthetic data generated through this model. *Inoue et al.* [1990] advocate a
checkerboard pattern in which adjacent voxels alternate between two extreme values. Let the model checkerboard pattern be

\[ m_{cb} = \beta \sum_j c_j e_j, \]

where \( c_j = 1 \) or \(-1\) and \( \beta \) is a scaling factor. Calculate the synthetic travel times from real sources to the real stations included in the data set (e.g., using the data kernel \( G \)), so

\[ d_{cb} = G m_{cb}. \]

Use the LSQR algorithm to solve for an estimate of \( m_{cb} \):

\[ \hat{m}_{cb} = \beta R \sum_j c_j e_j = \beta \sum_j c_j e_j, \]

that is simply a superposition of point spread functions. The estimated model image may then be compared to the starting model to identify regions with poorly recovered values. Spakman and Nolet [1987] use a harmonic function instead of a checkerboard pattern. Both test patterns share the advantage that a more complete sense of resolution for the model may be presented with just a few figures. Unfortunately, they also share the disadvantage that separate inversions must be performed for various input models with different wavelengths and amplitudes. Both patterns are parametric in the sense that we assume before inversion that we have some idea of the spatial scale and amplitudes of interesting features of the real earth. Impulse tests assume we know the location and amplitude of interesting features. Both methods require the same computation time as generating the best-fitting solution itself.

Figure 15a shows our input checkerboard model. Values alternate between +0.3 km/s and -0.3 km/s for adjacent voxels. Figure 15b shows the results for the top layer, 0-200 km. Well-resolved regions correlate strongly with good ray coverage (figure 8a), though continents are generally imaged more clearly than mid-ocean ridges. This discrepancy is probably due to better geometrical constraints on continental voxels. At mid-ocean ridges, rays connecting to seismographic stations propagate more vertically than horizontally, thus rays sampling ridges tend to be parallel. Continental voxels are generally sampled by a more complex set of criss-crossing rays that provide stronger constraints on the continental voxel’s velocity. Checkerboard results for subsequent layers are incorporated into maps of our velocity model, which are shown in the next section.
Nolet and Snieder [1990] suggest a less time-consuming means of producing a resolving kernel with a reduced basis, produced by the LSQR algorithm. It is common for coefficient matrices in tomographic problems to be numerically singular, so the matrix is rank-deficient and the range space may be spanned by a basis that is considerably smaller in dimension than the original matrix. Each iteration of the LSQR algorithm produces a single basis vector of the range space. One must decide when a sufficient number of basis vectors has been produced to represent the solution to the desired degree of accuracy. In the absence of the singular value spectrum, this decision presents a serious problem. Scales [1987] offers a way to obtain the singular values from the tri-diagonal matrix also produced by the LSQR algorithm, but he points out that numerical round-off errors can produce artifact entries in the set of singular values. We performed tests on a real, cross-hole tomographic problem in which the results of Scales' method were compared to singular values obtained via SVD and confirmed the deleterious effect of these errors, which do not allow the singular value spectrum to be produced reliably.

Vasco [1991] presents an extremal bound approach to evaluating resolution and uncertainty in a tomographic inversion. Instead of finding a single model that is "best-fitting" in some sense, he finds properties of the range of models that are consistent with the data. While this method is much different in its approach, it shares the computational drawbacks of methods mentioned previously. Each parameter must be considered individually and the computation time required for each block is comparable to the time required to find the entire "best-fitting" model. A subset of the model parameters could be examined, but calculating bounds for every block is not feasible.

4.2 Approximating Covariance

Attempting to evaluate covariance, we examine the distribution of travel time residuals and find that it more closely resembles a two-sided exponential distribution of deviates than a Gaussian, though it falls somewhere between the two [Pulliam et al., 1992]. To investigate the propagation of these errors through the inversion procedure to the final model, we replace the vector of travel time residuals with a vector of synthetic residuals that are distributed, in the first case, as a two-sided exponential and, in the second case, as a Gaussian. In each case, the variance of the random distribution after truncation is adjusted to the same level as that of the actual data. We perform 25 inversions with different residual vectors for each case and find corresponding "model" vectors, \( e^n \), where \( n = 1, 25 \). An estimate of the model covariance may be obtained as
\[ c_{ij} = \frac{1}{N} \sum_{n=1}^{N} e_i^ne_j^n \]

where

\[ e_i = \text{value for voxel } i \text{ produced by inversion of errors.} \]

Figures 16a and 16b show the covariance estimates for our model's top layer, 0-200 km, for the Gaussian and exponential cases, respectively. Results for both types of distribution share the same general patterns. Locations of large and small errors are quite similar; only the amplitudes of the errors vary significantly. This difference follows immediately from the main difference between Gaussian and exponential distributions: the exponential distribution has much longer tails than does the Gaussian. Given our inversion procedure, which minimizes the \( l^2 \) residual norm and therefore weights large deviates more heavily than small ones, the longer-tailed distribution naturally translates into larger "model" values. The most striking feature of these covariance estimates is their strong correlation with ray coverage. Areas sampled by many rays have large errors associated with them, while sparsely-sampled areas tend to have smaller errors. This reflects the tendency of the LSQR algorithm to image anomalies where they are well-constrained, but to bias model values toward zero where there is inadequate coverage. This tendency is desirable when dealing with real data and a real model; regions of large uncertainty receive small values rather than large values. However, the same tendency renders the values produced by this method poor estimates of model covariance. In fact, these figures do indicate where random errors in our data are mapped in our model, but with the noise maps alone we cannot assess the reliability of individual model parameters. Perhaps the maps do serve as reliable estimates of model uncertainty in regions that are constrained reasonably well, but distinguishing regions that have reliable estimates from regions which must be dismissed from consideration is not possible.

A second approach to estimating model covariance, the jackknife, is outlined by Efron [1982] and applied to seismic data by Lees and Crosson [1989]. Here we need not assume a particular distribution for the errors, since model variability is assessed directly from the variability of the data. Unfortunately, our algorithm still comes into play and we must be again be wary of the influence of uneven ray coverage and artifacts of our parametrization. To form the jackknife estimate we perform \( N \) inversions of the real data, leaving out a subset of the data without replacement. For each inversion we produce a model

\[ \mathbf{m}^n = N \mathbf{m} - (N-1)\mathbf{m}^n \]
\[ \hat{m} = \frac{1}{N} \sum_{n=1}^{N} m_i^n \]

which has the variance

\[ c_{ul} = \frac{1}{N(N-1)} \sum_{n=1}^{k} (m_i^n - \hat{m}_i)^2 \]

where

\[ N = \text{number of data subsets.} \]

Figure 15 shows a map of jackknife standard errors for the 0-200 km depth layer. The general tendency of the jackknife estimate is also to place anomalies in heavily-sampled regions, but some features distinguish the jackknife estimate of variance from the model error distributions estimates. The jackknife estimates show a weaker correlation with the number of rays sampling a voxel and the sum of ray lengths in a voxel than do the model error distributions' estimates, and shows more variability due to the geometrical distribution of rays sampling a given voxel. If leaving out a few rays produces a velocity estimate for a given voxel much different from previous estimates, the voxel is poorly constrained. However, the known effect of our algorithm again causes us to distrust the results of the jackknife procedure. Poorly constrained voxels are biased towards zero, exactly the opposite tendency we desire for an investigation of covariance.

Keeping in mind our algorithm's tendency, we should simply disregard results in poorly sampled regions and concentrate on portions of Earth that are well-sampled. Distinguishing well- from poorly-sampled regions is easily done with a block model parametrization. We adopt this approach in our presentation of model maps in the next section.

5. RESULTS AND DISCUSSION

5.1 Source Mislocations and Station Corrections

The averaging procedure employed to construct summary rays should reduce both the mislocation of events and the station errors, at least in the cases where several actual stations are averaged into one summary station, and so we expect to see only general features of source mislocations and station errors in our results. These terms are included for the sake of a realistic formulation of the problem, and for stability in the inversion. When an inversion is performed without source mislocation or station terms, variance is reduced by 18%. Solving for velocity and source terms, but not for station terms, the total variance of
summary travel time residuals is reduced by 19%. Solving for all velocity, source, and station terms reduces residual variance by 22%. The amount of variance explained by our model depends critically on the weights given to the roughness penalty and damping coefficient; further reduction may be achieved by relaxing the minimum norm criterion and at the expense of a reasonably smooth and physically plausible model. However, the relative increase in variance reduction as velocity, source mislocation, and station terms are progressively included is a consistent feature of the inversions. While not orthogonal, the three parameter classes clearly account for distinct parts of the travel time residuals. Leaving out one set of terms from the inversion leaves a corresponding portion of the residual variance unexplained.

Due to the averaging procedure employed to construct summary events, source location corrections are generally small. Source location corrections in subduction zones generally move the source toward the positive velocity anomaly. Sources in continental regions have the smallest, nearly insignificant, corrections. Station corrections range from -0.94 to +1.07 sec, with no obvious correlation to tectonics or elevation of the station region.

5.2 Velocity Model

Figures 18a-c illustrate the technique we use to show maps of individual layers of our P velocity model for the 670-870 km depth range. For each layer we take the results from our checkerboard test (figure 18a), find the absolute value for each pixel, and normalize so the checkerboard output represents the portion of the recovered input value. The checkerboard results now range from 0 to 1. For the layer velocity map (figure 18b) we remove the mean from each layer and show velocity in gradations of red to blue. Red indicates slower-than-average velocity and blue marks a faster-than-average region. Next we apply the normalized checkerboard results, for a given layer, to that layer’s velocities (figure 18c). Colors range from full saturation to white as each pixel’s velocity is modified by its checkerboard resolution value. Full recovery of the input checkerboard value is indicated by full color saturation and no recovery produces zero color saturation, in which case the pixel is white. Areas not sampled by our data set are left black. Figures 19a-f show six additional depth layers of model ISC5_LSQR.

5.3 Correlations with Surface Tectonics

5.3.1 Rift zones

A significant feature of model ISC5_LSQR is slow anomalies associated with mid-ocean spreading centers (figure 19a). The mid-Atlantic rift, tracing the middle of the Atlantic Ocean, appears quite clearly, although
not continuously. Anomalies reach -1.5%, but average closer to -0.25%. Resolution for parts of the rift is poor but anomalies that do appear are consistently slow, as expected. Resolution is particularly poor for the discontinuous parts of the anomaly in the South Atlantic. The rift extending to the South Sandwich subduction zone appears faintly, as do the Chile and East Pacific rifts. These areas, along with most of the southern hemisphere, suffer from poor resolution in our checkerboard tests.

A reasonably continuous, slow anomaly emerges from the Red Sea along the Carlsberg and mid-Indian rifts and diverges into two anomalies where the mid-Indian rift splits into the Southeast and Southwest Indian rifts. The anomaly associated with the Southeast Indian rift extends through the southern Indian Ocean and south of Australia, after which it disappears. The checkerboard tests indicate this anomaly is poorly resolved. The anomaly marking the Southwest Indian rift continues around the tip of South Africa to join the mid-Atlantic anomaly. The perturbations associated with Indian Ocean spreading centers reach -1%, but again average closer to -0.25%. The East African rift shows quite clearly in our model, with anomalies reaching -2.0%. The only major rift zone that finds no reliable expression in our model is the ridge that extends across the southern Pacific. The ISC catalog contains very few events located along this ridge (see figure 2), which results in poor resolution.

Many of the rift zone anomalies that appear in the 0-200 km range of model ISC5 LSQR also appear in the 200-400 km layer (figure 19b), including the mid-Atlantic, East African, and mid-Indian rifts in particular. All of these are diminished in their magnitudes, spatial extent, and continuity. Many anomalies associated with rifts do not appear at all, despite generally increased resolution in the second layer. By the third layer, 400-670 km depth, only the anomalies associated with the Carlsberg, mid-Indian, and East African rifts are clearly visible (figure 19c). In the fourth layer (figure 18c) the strong correlation between rift zones and slow anomalies is gone, although diffuse, slow anomalies beneath the mid-Indian and East African rifts persist.

5.3.2 Subduction zones

The backarc basins in the western Pacific are clearly marked by slow anomalies that average around -1.0% (figure 19a). At 200-400 km the pattern is still clear (figure 19b), but in several cases, such as beneath the Aleutians and the Japanese island arc, the slow anomaly has been pinched out by an adjacent fast anomaly. This may be due to the dominance of the subducted slab over the excess volatiles released by slabs in mantle material below 200 km. However, in the Tonga-Kermadec and New Hebrides subduction
zones the fast anomaly shows the opposite effect, having been partially displaced by a slow anomaly. Still, the Tonga anomaly is fast and clearly continuous through the 400-670 km layer (figure 19c). Another fast anomaly, much larger in lateral extent, appears in the 670-870 km layer and extends to more than 1670 km depth, but the continuity between this deep anomaly and the shallower anomaly associated with the subduction beneath Tonga is questionable. The shallower fast anomaly, 0-670 km, is consistent with the results of the regional study of Zhou [1990], though our model grid is not able to show the finer detail of the slow backarc basin directly above the dipping slab. The large fast anomaly to the west of the Tonga trench does not appear in Zhou's cross-sections. Beneath Japan and the Kurile Islands another fast anomaly protrudes into the lower mantle at a much shallower angle than the Tonga feature. Again the continuity between a fast upper mantle anomaly dipping to the northeast, and a broadened, diffuse, but similarly fast anomaly below 670 km depth is not clear. The deeper anomaly appears to be continuous all the way to the lowermost layer in the mantle beneath northeastern Asia. The shallower features of our model correspond to anomalies shown by Zhou and Clayton [1990] but the deeper anomalies of model ISC5_LSQR fall in regions not included in their cross-sections. A fast anomaly beneath the Andes is abruptly pinched out in the second layer by slow anomalies. The fast anomaly reappears in the third layer below South America and continues into the 670-870 km layer. The Bering Sea appears as a slow anomaly, consistent with other back-arc basins around the Pacific. In the second layer a fast anomaly extends along the Aleutian trench and displaces the southernmost extension of the slow, Bering Sea anomaly. In the 400-670 km and 670-870 km layers the fast anomaly becomes progressively more dominant, but diminishes in the 870-1070 km depth interval. A significant fast anomaly appears in the 400-870 km depth range beneath the Mariana subduction zone, and a broad, fast region occurs in the 670-870 km layer underneath the Philippines.

In general, the broadening and flattening to sub-horizontal noted by Zhou [1990] and Zhou and Clayton [1990] also appear in our model at the 670-870 km layer (figure 18). Van der Hilst and Spakman [1989] demonstrate that inaccuracies of the J-B model in the upper mantle and the geometry of ray coverage provided by the P phase can lead to a similar sort of flattening and horizontal extension purely as artifacts of the imaging procedure. They caution that the interpretation of such images must be approached with caution, though a more accurate one-dimensional starting model can help limit these flattening artifacts. Our starting model differs from the starting model used by van der Hilst and Spakman [1989] in that we do not include discontinuities in the upper mantle, but is similar in that we do not include a low-velocity zone in the uppermost mantle either. Without discontinuities our one-dimensional model will generate raypaths that
appear to sample the transition zone well, while in fact the travel time residuals associated with those rays actually correspond to rays that were refracted at the discontinuities. One effect may be the inaccurate mapping of anomalies to locations in the transition zone, and perhaps to locations below the 670 km discontinuity. The extent of this mis-mapping is unclear. A second problem is associated with the apparent existence in some regions of the real earth of a low-velocity zone, which is not present in our starting model. A low-velocity zone in Earth's mantle would produce a shadow zone at the surface in which few rays would emerge and those that did emerge would have small amplitude. Readers of seismograms might overlook the actual first arrival and instead pick a larger-amplitude arrival that was refracted at the 400 km discontinuity or turned upward by a steep gradient below the low-velocity zone. Unless we can re-identify the picks supplied to the ISC and discard or use appropriately the late arrivals (which we cannot with a J-B model) we will introduce a set of systematically slow residuals which will propagate through the inversion to produce slow anomalies in our model. The distribution of travel time residuals as a function of epicentral distance contains an anomalously sparse section clustered about zero seconds between 10° and 17° but indeed shows a systematic trend toward slow (positive) residuals. Rays emerging at these distances follow paths that bottom between 100 and 250 km in the J-B model. Rays are most sensitive to velocity perturbations near their sources, receivers, and bottoming points, so unexpected slow anomalies found in these depths in models produced with J-B starting model should be considered suspect. If low-velocity zones were distributed according to some pattern, preferentially under continents rather than oceans, for example, rather than appearing consistently worldwide, the effect could be even more difficult to uncover. Rather than appearing as a slow mean to the three-dimensional model, which may be removed and used to update the one-dimensional model, the effect would be specific to particular regions. This may provide an alternative explanation for images that suggest a detached slab feature due to a lack of continuity between fast features above and below about 200 km depth. We believe our results are relatively, though not entirely, uncontaminated by this effect since the inverse weighting by residuals' standard errors applied to rows of the constraining matrix equation reduces the influence of these data.

5.3.3 Continental shields

Fast anomalies show quite clearly in continental shield regions (figure 19a). Alaska, Canada, Greenland, Fennoscandia, Siberia, and northern Australia all show fast anomalies on the order of +1.5%. All of these fast anomalies persist through the 200-400 km layer (figure 19b), but appear broken and discontinuous in
the transition zone. Northern Africa and eastern South America are poorly resolved.

5.3.4 Hotspots

A number of hotspots correlate well visually with strong slow anomalies in our model. The Azores, Cape Verde, Canary Islands, Afar/Ethiopia, Lake Victoria/East Africa, Comores Islands, Kerguelen, Christmas Island, Tasmania, Caroline Islands, Hawaii, Galapagos, Vema Seamount, and Mt. Erebus hotspots all appear as isolated, negative velocity perturbations (figure 19a). In addition, the Yellowstone and Raton, New Mexico hotspots appear subsumed into the general slow anomaly covering the western United States. The Mehetia/Society Islands/Tahiti hotspot appears as one member of a complex set of four hotspots, including the MacDonald Seamount, and Marquesas Islands and Pitcairn Island/Gambier Islands hotspots. Between these four hotspots, marked by two diffuse slow anomalies, lies an apparent change to a fast anomaly. The Jan Mayen and Iceland hotspots appear as parts of the mid-Atlantic rift, as does the St. Helena hotspot just to the west of Africa. It is interesting to compare the St. Helena hotspot anomaly to the Ascension hotspot, located about 8° closer to Africa, which finds no expression in our model. St. Helena is located near the rift and is thus illuminated by seismic events associated with the formation of new oceanic crust. In all, about half the set of hotspots compiled by Richards et al. [1988] are marked by slow anomalies in at least the first layer of our model.

None of these anomalies changes sign through the first three layers. Below 670 km none clearly changes, but the Yellowstone and Iceland anomalies appear displaced. In the 870-1070 km range, several more hotspot anomalies are displaced from their surface locations, but each anomaly persists in some form nearby. Beneath Hawaii a slow anomaly, reaching -1.5%, extends deep into the mantle, trailing off to the northwest with depth. Both Pulliam et al. [1992] and Inoue et al. [1990] show this area to have a weak fast perturbation, contrary to our expectation. This difference can probably be explained by the differences in the way we treat the data. Inoue et al. [1990], though they use an $l^2$ residual norm minimization, severely downweight outliers. Pulliam et al. [1992] minimize an $l^1$ norm of the residuals. Experience with calibration events that have known locations shows that some of the most extreme, slow travel times are recorded at the Hawaiian stations. Downweighting these extreme residuals in the inversion procedure causes the algorithm to overlook the anomalies that give rise to these slow travel times.

5.4 Transition Zone
In the transition zone, 400-670 km, the pattern of anomalies changes completely (figure 19c). The correlations between anomalous velocities and surface tectonics observed in the top two layers do not exist here. Shield regions are not generally fast and backarc basins are not generally slow, though the region extending northward and westward from New Zealand, a complex subduction zone, is quite slow. A striking slow feature, amounting to -2%, appears under southern Eurasia and India. The fast feature beneath Tonga is not clearly continuous through this layer, though neighboring voxels show a similar anomaly to the west in the next lower layer.

Several hotspot anomalies including those at Hawaii, Kerguelen, Iceland, Lake Victoria, Yellowstone, Raton, Afar, Mt. Erebus, Galapagos, Canary Islands, and Mehetia persist through this layer. In the 670-870 km layer, the features at Tonga spread laterally to the west. The slab-related anomaly under Japan has migrated a similar distance westward but without similar lateral extension.

Overall, our topmost layer shows a similar pattern to Inoue et al.'s [1990] 78-148 km layer, though our layer is not so heavily smoothed. Our 400-670 km layer is also quite similar to their 478-629 km layer in regions for which we have ray coverage. Though they do not show their hitcount map for this layer, the long slow feature extending northeast-southwest across the Pacific in their model may be an artifact of their smoothing procedure. Our data set shows the central Pacific to be largely unsampled in this depth range.

5.5 Midmantle

Confirming results of previous studies, our model shows diminished amplitudes of velocity anomalies in the mid-mantle. There is no obvious large-scale radial continuity throughout the mid-mantle. On a smaller scale, several hotspot anomalies persist. Most striking are fast anomalies beneath eastern North America, the Caribbean, and central South America, and the features, mentioned earlier, beneath Tonga and Japan/eastern Asia. The fast anomaly beneath eastern North America and the Caribbean appears in the same location as a large S-velocity anomaly reported by Grand [1987] and the anomaly’s apparent continuation beneath South America also appears in Grand’s recent results (personal communication) (see figures 18c and 20b). Similar features for P-velocity appear in the inversions performed by van der Hilst [1990]. The fast feature beneath Tonga broadens and continues to dip to the west to a depth of 1670 km. Beneath Japan and eastern Asia the fast anomaly is diffuse but extends all the way to the core-mantle boundary.

5.6 Lower mantle
The lowermost layers, 2470-2670 km (figure 19f) and 2670 km-CMB, show a significantly different pattern from the mid-mantle. In general, our results for the 2470-2670 km layer agree with both Dziewonski's L02.56 model and Hager and Clayton's [1989] smoothed version of Clayton and Comer's [1983] model. When expanded in spherical harmonics and recombined using only $1 \leq l \leq 6$, our model also shows a large slow anomaly over southern Africa, though this anomaly is displaced relative to L02.56 and in a way that is more consistent with Hager and Clayton's result. Other slow anomalies appear beneath the southern Pacific Ocean, beneath the Bering Sea and toward the North Pole, in the northern Atlantic Ocean, and beneath Papua New Guinea. Fast anomalies appear beneath Asia, South America, and north of New Zealand. Some oscillation appears to occur between the lowermost layer and the 2470-2670 km layer just above. In the South Pacific, a large-amplitude slow anomaly just above the CMB trades off with a fast anomaly above it. A ring of slow material surrounds the fast anomaly in the second-to-bottom layer. The checkerboard tests show reasonable resolution in this area, but such oscillations may be just the sort of problem checkerboard tests cannot reveal. Results for the bottom layer, $D''$, are suspect because our restriction to rays with epicentral distance less than 96° results in poor coverage of this layer. This restriction is intended to avoid contamination of our data set by arrivals diffracted at the core. Reduced resolution in the bottom layer is the price we pay for avoiding this contamination.

5.7 Continuity of Features

Figure 20a shows several of the features described above in a set of cross-sections through model ISC5_LSQR. A constant-latitude slice at 24°S through the Tonga-New Hebrides subduction complex shows the associated fast anomaly dipping to the west. A fast continental root appears further to the west at this latitude under northern Australia. To the east the Pitcairn Island/Gambier Islands hotspot is associated with a slow anomaly. Note the pinching out of the dipping Tonga anomaly between the 200-400 km and the 400-670 km layers. The appearance of such a broad and deep fast anomaly to the west of the subduction zone, extending to 1670 km depth in the dip direction of the downgoing slab, offers tantalizing circumstantial evidence for slab penetration into the lower mantle. Further evidence is shown in the other constant-latitude cross-section of figure 20a, at 52°N, where a fast anomaly appears under the Kuriles and dips to the northwest. This anomaly is much broader, extending continuously through the Japanese Island arc and under eastern Asia (see cross-section at 124°E), but is less distinct than the Tonga anomaly. A slight fast anomaly beneath the Aleutians is also shown.
A constant-longitude slice at 84°W, under eastern North America, the Caribbean, and western South America shows a distinct fast anomaly beginning at about 1070 km depth and extending to 2070 km. Under North America and the Caribbean, the anomaly is consistent in both size and location with an S-velocity anomaly reported by Grand [1987]. The fast anomaly veers to the east south of the Caribbean (see figure 20b). Above the fast anomaly in the Caribbean and through Central America the model is slow. A constant-depth section shows the 2470-2670 km layer with longer-wavelength anomalies than the upper mantle. A slow anomaly appears under the western Pacific and a region beneath the eastern Pacific and North America is fast.

Figure 20b shows the long fast anomaly beneath eastern North America, the Caribbean, and South America in a depth section at 670-870 km. A constant-longitude slice at 88°W shows that this depth constitutes the top of the anomaly at this longitude, though the fast continental shield in the northern United States and Canada appears at the top of the section. To the east of the fast anomaly in the northern hemisphere lies a broad slow anomaly under the Atlantic Ocean. A slice at 29°S latitude shows the fast anomaly under central South America extends continuously to the surface at the Chilean subduction zone. A section at 64°N shows fast shields beneath Canada, Greenland, and Fennoscandia and a slow anomaly beneath Iceland. A broad slow region appears at 32°E under the East African rift zone.

5.8 Spherical Harmonic Expansion

Surface spherical harmonic series expansions to degree 15 were calculated by integration around the globe for each coefficient, rather than by fitting coefficients to model values by least-squares. In this way, coefficients are independent of each other and coefficient values are independent of the point of truncation of the harmonic series. That is, coefficients do not change if the series expansion is calculated a second time with a different number of terms. The associated Legendre polynomials are fully normalized, i.e.,

\[
p_l^m(\theta) = \left[ (2-\delta_{l,0})(2l+1)(l-m)!(l+m)! \right]^{1/2} P_l^m(\cos \theta).
\]

Figure 21 shows the total power in the series expansion for each layer plotted as a function of depth. The anomalously low power in the 200-400 km layer probably is due to the fact that rays that bottom in this layer, which emerge at the epicentral distance range 15° ≤ Δ ≤ 20°, have the largest variance of all the travel time residuals. These rays are the most sensitive to velocity perturbations in the 200-400 km layer, but in our inversion their influence on the final model is (severely) downweighted by the inverse of the
residuals' standard errors. The first layer (0-200 km) and the transition zone (400-670 km) have the highest power, indicating the greatest heterogeneity in our model occurs at these depths. Again, because our starting model does not contain discontinuities at 400 and 670 km depth we are probably mapping more power into the transition zone than is justified. Power decreases in the mid-mantle and increases again as we approach the core. The small increase in power at the 1270-1470 km depth layer is the result of an unusually large \( l = 1 \) component. The drop in power from the 2470-2670 km layer to the lowermost layer, 2670-CMB, is probably due to our poor ray coverage. Figure 22 shows the power in series expansions of each layer as a function of angular degree. The top two layers appear relatively devoid of power at the lower degrees, despite the strong concentration of ray coverage in \( l = 1-5 \) patterns. In the 400-670 km layer \( l = 1, 2, 3, \) and 6 dominate. The large \( l = 2 \) component confirms previous reports, but to our knowledge, no other study has shown the equally prominent \( l = 1 \) and 3 components. In the mid-mantle power is more or less evenly distributed across the harmonic terms. The exception is in the layer 1270-1470 km, where the \( l = 1 \) harmonic is strong and the \( l = 2 \) and 3 components rise above the higher-degree harmonics. These components clearly are responsible for the high power total of this layer.

Since sign information is not included in power calculations, figure 22 does not show how the distribution patterns for all layers combine constructively or destructively to form a pattern for the whole mantle. Figure 23 shows the power in the spherical harmonic expansions averaged through the whole mantle and through the upper and lower mantle separately. The averaging is performed on the the individual harmonic coefficients, weighted at each layer by the square of the layer mid-point's radius, which normalizes the power in each layer to the layer's surface area. For the upper mantle the power spectrum shows a dominant \( l = 6 \) component, along with prominent \( l = 2, 5, 12, \) and 13 terms. The \( l = 1 \) power for the upper mantle is low, simply reflecting the results in figure 22 which show the small \( l = 1 \) components in the first two layers. However, in the lower mantle the same component is unexpectedly low, given the large values in several of the individual layers. This may mean the \( l = 1 \) component is poorly resolved in the lower mantle and trades off between layers in our model. In general, high power in harmonic degrees of the upper mantle coincides with high power in the same degrees of the lower mantle, though relative amplitudes once again suggest that heterogeneity is concentrated in the upper mantle. Degrees 2, 5, and 12 dominate the expansion averaged through the whole mantle. Surprisingly, degree 6 has the lowest average power even though its power dominates the top layers and contributes significantly to the lower mantle's total. From the relatively low power of the whole-mantle integration compared to the separate lower- and upper-mantle
averaged series, it is clear that either one part of the real mantle is compensating for anomalies in the other part or our inversion scheme is trading off power between the lower and upper portions of our model mantle.

6. CONCLUSIONS

We present a three-dimensional P-velocity model for the Earth's mantle found by inverting ISC travel time data for the time period January 1964 - January 1987. Our inversion minimizes the $l^2$ norm of the travel time residuals, by means of the conjugate gradient variant LSQR algorithm. Model maps show values only for sampled voxels, with a weighting scheme based on our estimate of model resolution. This allows the most realistic presentation of what is known about mantle velocity structure.

Model values and patterns for the top two layers correlate well visually with surface tectonics. Backarc basins, rift zones, and some known hotspots all find expression as slow anomalies in our model while continental shields and some subduction zones are marked by unusually fast velocities. Comparison of model ISC5_LSQR to the model of Inoue et al. [1990] reveals great visual similarities in patterns of fast and slow velocities. Overall, the mid- and lower mantle show distinctly less heterogeneity than the upper mantle. Fast anomalies appear in both the mid- and lower mantle beneath the Tonga subduction zone, eastern North America, the Caribbean, central South America, and Japan/eastern Asia. The lower mantle correlates well at low orders with the models reported by Dziewonski [1984], Morelli and Dziewonski [1985, 1986], and Clayton and Comer [1983; Hager and Clayton, 1989], though we cannot be confident of our results for the lowermost layer (D').

Seismological studies of lateral heterogeneity in the mantle differ in their general approaches and in specific decisions made along the way. Our model parametrization differs significantly from those employed by Dziewonski [1984] and Inoue et al. [1990], and differs slightly from Clayton and Comer [1983; Hager and Clayton, 1989]. We use summary rays while Dziewonski [1984] and Inoue et al. [1990] do not. We weight rows of the matrix problem by the inverse standard errors of the travel time residuals as a function of delta, a measure of data quality, while Inoue et al. [1990] weight preferentially according to the size of the residual. We weight columns of the coefficient matrix by a measure of the quality of each voxel's sampling, while Clayton and Comer [1983; Hager and Clayton, 1989] weight by the number of hits, and Inoue et al. [1990] do not weight columns at all. We use the LSQR algorithm to solve the constraining matrix equation while Clayton and Comer [1983; Hager and Clayton, 1989] and Dziewonski [1984] do not.
While each of these differences has consequences that may be traced to differences in our final models, our results are quite similar in general. All these studies depend on traditional ray theory and Fermat's principle in the construction of the tomographic equations and, perhaps most importantly, each study makes use of the same set of global travel time data.

Acknowledgments. This research was supported by grants EAR-8720879 and DMS-8810192 of the National Science Foundation and by the Director, Office of Energy Research, Division of Basic Energy Sciences, Engineering, and Geosciences, of the U.S. Department of Energy under contract DE-AC03-76SF00098. During a portion of this work, RJP was supported by an Associated Western Universities/Department of Energy Laboratory Graduate Research Fellowship. Computations were carried out at the Center for Computational Seismology of Lawrence Berkeley Laboratory and at the Statistical Computing Facility, Department of Statistics, UC Berkeley. The authors would like to thank Philip Stark, Barbara Romanowicz, and Joel Ita for helpful discussions and software. Wes Bethel for his help with the figures, Guust Nolet for a preprint, and Andrea Morelli and Hiroshi Inoue for complete versions of their models.

7. REFERENCES


Clayton, R.W., and R.P. Comer, A tomographic analysis of mantle heterogeneities from body wave travel times, EOS Transactions AGU, 64, 776, 1983.

Dziewonski, A.M., Mapping the lower mantle, Determination of lateral heterogeneity in P velocity up to degree and order 6, J. Geophys. Res., 89, 5929-5952, 198


List of Figures

Table 1 Details of the model parametrization and the sampling provided by our data set. Include are the average number of hits for sampled voxels in each layer along with the average sum of ray segments in a voxel at a given depth and the number of voxels sampled in each depth interval. These averages include only voxels which have non-zero sampling.

Figure 1 The model mantle is parametrized as voxels, $5^\circ \times 5^\circ$ at the equator and generally 200 km thick, for a total of 22,876 model parameters. Voxels in a given layer have approximately equal surface area.

Figure 2 Locations of sources used in this study. The data set consists of about 46,000 shallow events located by the ISC for the time period January 1964 - January 1987.

Figure 3 Locations of seismographic stations reporting to the ISC in January 1987.

Figure 4 Histogram of the 3.02 million travel time residuals associated with the events included in this study. Also shown are the first four moments of the distribution.

Figure 5 Locations of summary sources. The averaging procedure, based on a 2$^\circ \times 2^\circ$ grid, reduces the number of sources to about 6,000.

Figure 6 Locations of summary stations. The number of stations is reduced to 979 by the ray averaging procedure.

Figure 7 Histogram of the summary residuals with the first four moments of the distribution.

Figure 8 The distribution of column sums of our coefficient matrix, $A$, indicating the total sampling of each voxel by our summary data set. Shown in each panel is a different depth layer: (a) 0-200 km, (b) 200-400 km, (c) 400-670 km, (d) 670-870 km, (e) 1270-1470 km, and (f) 2470-2670 km.

Figure 9 (a) A 5% velocity anomaly introduced at (17$^\circ$S, 178$^\circ$W, 600 km depth), beneath the Tonga subduction zone and (b) the resulting point spread function. Cross-sections are at 20$^\circ$S, 174$^\circ$W, 6$^\circ$E, and 1700 km depth.

Figure 10 (a) A 3% anomaly placed beneath the New Hebrides subduction zone at (16$^\circ$S, 166$^\circ$E, 900 km depth) and (b) the resulting point spread function. Cross-sections are at 20$^\circ$S, 168$^\circ$E, 12$^\circ$W, and 2470 km depth.

Figure 11 (a) A -3% anomaly introduced to the first layer beneath the Hawaiian islands at (21.4$^\circ$N, 158$^\circ$W, 70 km depth) and (b) the resulting point spread function. Cross-sections are at 20$^\circ$N, 156$^\circ$W, 24$^\circ$E, and 1700 km depth.
Figure 12 (a) A -4% anomaly introduced to the first layer beneath Iceland at (65°N, 18°W, 70 km depth) and (b) the resulting point spread function. Cross-sections are at 64°N and 1300 km depth.

Figure 13 (a) A 3% anomaly introduced to the fourth layer beneath central South America at (0°, 65°W, 870 km depth) and (b) the resulting point spread function. Cross-sections are at 0°, 62°W, 118°E, and 1400 km depth.

Figure 14 (a) A 4% anomaly placed beneath eastern North America at (40°N, 85°W, 870 km depth) and (b) the resulting point spread function. Cross-sections are at 42°N, 85°W, 870 km depth.

Figure 15 Map of the top layer, 0-200 km, of the resolution test using a synthetic "checkerboard" model.
(a) Input values alternate between ±0.3 km/s. (b) Output of the test inversion.

Figure 16 Model standard error estimates for the 0-200 km layer using (a) a Gaussian distribution of deviates in place of the real data as input to the inversion and (b) a two-sided exponential distribution as input.

Figure 17 Model standard error estimates for the 0-200 km layer obtained with a jackknife procedure.

Figure 18 A set of layer maps for the 670-870 km depth range illustrating the technique used to plot velocity maps. The checkerboard resolution test results (a) are used to modify the velocity values for each block (b) based on the percentage of input value recovered by the test inversion. Final values (c) range from red (slow) to blue (fast) and from full color saturation, indicating full recovery of the input checkerboard value, to white, indicating no recovery.

Figure 19 (a-f) Six depth layers of model ISC5_LSQR: (a) 0-200 km, (b) 200-400 km, (c) 400-670 km, (d) 1270-1470 km, and (e) 1470-1670 km, and (f) 2470-2670 km. Each layer's mean has been removed. Velocity perturbations grade from red (slow) to blue (fast). In addition, color values are modified from full saturation, indicating the voxel is well-resolved as determined by the checkerboard test, to white, which indicates no recovery of the checkerboard value.

Figure 20 Cross-sections of model ISC5_LSQR. Shown are two constant-latitude slices at 24°S and 52°N, constant-longitude slices at 124°E and 84°W, and a constant-depth section showing the 2470-2670 km layer.

Figure 21 Cross-sections of model ISC5_LSQR. Shown are two constant-latitude slices at 24°S and 64°N, two constant-longitude slices at 34°E and 98°W, and a constant-depth section at 670-870 km.
Figure 22 Power contained in surface spherical harmonic series expansions of model ISC5_LSQR for each depth layer.

Figure 23 Power in the spherical harmonic expansions for each depth interval as a function of angular degree. All values are normalized to the maximum value appearing in the figure. Numbers on the right refer to the maximum power for each layer.

Figure 24 Power in spherical harmonic series generated by averaging ISC5_LSQR layer expansions through the whole mantle and through the upper and lower mantle separately.
Model parameterization
ISC source locations

January 1964 – January 1987
MEAN = -0.21E+00
VAR = 0.54E+01
SKEW = -0.13E-01
KURT = 0.63E+01
Summary stations
fig. 15
Fig. 19 (cont'd)

(d)

(e)

(f)

checkerboard resolution

velocity

-0.5° 0° 0.5°
Fig. 2: A graph showing the power (10**(-3) km**2/axis**2) against depth (km). The graph exhibits a peak at approximately 600 km depth with a gradual decrease and increase as depth increases, stabilizing around 2400 km.
Model isc5s_12 averaged

Angular degree (l)

maximum power

0.41E-04

0.17E-04

0.92E-05

upper mantle

lower mantle

whole mantle