JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 101, NO. B10, PAGES 22,245–22,272, OCTOBER 10, 1996

Global mantle shear velocity model developed using nonlinear asymptotic coupling theory

Xiang-Dong Li[1] and Barbara Romanowicz

Seismographic Station and Department of Geology and Geophysics
University of California, Berkeley

Abstract. We present a three-dimensional shear velocity model of the whole mantle developed using SH waveform data. The model is expressed horizontally in terms of spherical harmonics up to degree 12 , and vertically in terms of Legendre polynomials up to degrees 5 and 7 in the upper and lower mantle, respectively. What distinguishes this model from other tomographic models published to date is (1) the theoretical normal mode-based wave propagation approach, where we include across branch mode coupling terms in order to model the body wave sensitivity to structure along the path more accurately; (2) the wave-packet weighting scheme which allows to balance contributions from high-amplitude and low-amplitude phases, increasing the resolution in some parts of the mantle. We also relax the constraints on the Moho depth, which is allowed to vary in the inversion, thus absorbing some uncertainties in crustal structure.The resulting model is generally in good agreement with other recent global mantle S velocity models and with some regional models. The rms profile with depth has more power than other models in the upper mantle/lower mantle transition region and the zone of increased power and low degree structure near the base of the mantle is confined to the last 500 km in depth. This model provides a particularly good fit to the non-hydrostatic geoid through harmonic degree 12 (79% variance reduction), as well as good fits to observed splitting functions of S velocity sensitive mantle modes, indicating that both large-scale and small-scale features are really well constrained.

Introduction

The first global tomographic models of the Earth's mantle were derived separately for the upper and the lower mantle using different data sets, primarily long-period surface wave phase velocity measurements for the upper mantle [*Nataf et al.*, 1986; *Montagner and Tanimoto*, 1991] and short-period P wave travel times, as collected by the International Seismological Centre (ISC), for the lower mantle [*Dziewonski et al.*, 1977; *Dziewonski*, 1984; *Inoue et al.*, 1989]. Waveform modeling was first introduced for the purpose of modeling the upper mantle [*Woodhouse and Dziewonski*, 1984] and the corresponding modeling technique was based on normal mode summation, with the reconstruction of the phase of surface waves primarily in mind. The first attempts to apply this technique to the whole mantle were made by *Woodhouse and Dziewonski* [1987] and *Tanimoto* [1990]. A review of global mantle tomography is given by *Romanowicz* [1991].

While much interest remains in the study of the upper mantle using surface wave techniques [e.g., *Zhang and Tanimoto*, 1993; *Ekstrom et al.*, 1995; *Trampert and Woodhouse*, 1995], the accumulation of high-quality digital data from the expanding international global seismic network [e.g., Integrated Research Institutions for Seismology (IRIS), Geoscope, Federation of Digital Seismic Networks (FDSN)], has made it possible to consolidate the waveform approach in whole mantle tomographic inversions and to develop techniques for the complementary measurement of long-period body waves [e.g., *Woodward and Masters*, 1991a, b]. Current whole mantle tomographic models thus make use of a mix of techniques applied to digital waveform data, trying to jointly exploit a combination of free oscillation spectral measurements, and the phase information in surface waves and body waves, in order to achieve better depth and spatial resolution in the mantle [e.g., *Woodward et al.*, 1993; *Su et al.*, 1994; *Johnson et al.*, 1994].

The waveform inversion techniques have built upon the methodology developed more than a decade ago by *Woodhouse and Dziewonski* [1984], who assumed that a seismogram is only sensitive to the horizontally averaged structure along the great-circle path between the source and the receiver. The validity of this approximation, which we will refer to as the "path average approximation" (PAVA), has been verified both theo-

[1]Now at Molecular Dynamics, Sunnyvale, California

Paper number 96JB01306.
0148-0227/96/96JB-01306$09.00

retically and numerically for the case of long period surface waves [*Mochizuki*, 1986; *Park*, 1987; *Romanowicz*, 1987; *Li and Romanowicz*, 1995]. It has also been applied to body waves, [*Woodhouse and Dziewonski*, 1987; *Tanimoto*, 1990], but it is much less valid in this case, since it fails to represent the preferential sensitivity of body waves to structure along the ray path. In order to remedy this shortcoming, long-period absolute and differential travel time measurements are made and interpreted in the framework of geometrical ray theory [*Woodward and Masters*, 1991a,b; *Su and Dziewonski*, 1991]. The main disadvantage of this ray approach is that the usable portion of seismograms is limited to well isolated phases, which severely limits sampling of the mantle. More recently, the PAVA and travel time approach have been used simultaneously in the hope that they would compensate each other's drawbacks [*Dziewonski and Woodward*, 1992; *Woodward et al.*, 1993; *Su et al.*, 1994].

Accounting accurately for the effects of lateral heterogeneity on body waves is a computationally intensive and rather prohibitive task, although the theoretical tools for this exist at the present time within the framework of normal mode theory [e.g., *Lognonne and Romanowicz*, 1990; *Geller and Hara*, 1993]. Recently, *Li and Tanimoto* [1993] demonstrated that the ray character of long-period body waves can be brought out by a normal mode theoretical approach if cross-branch modal coupling is considered in an asymptotic fashion. The original version of the theory presented by *Li and Tanimoto* [1993] follows the approach developed by *Romanowicz* [1987] and involves the complete linearization of phase perturbation terms, an approximation not valid beyond very short times after the event. In a recent contribution [*Li and Romanowicz*, 1995] (hereafter referred to as paper 1), we have introduced a nonlinear method to overcome this deficiency. The method is referred to as the "nonlinear asymptotic coupling theory" (NACT). We also provided an efficient algorithm for the calculation of partial derivatives with respect to model parameters. In paper 1, we focused on the comparison of the NACT with the more traditional PAVA. We found that, in particular, for a given data set, NACT yields better resolution in the lower mantle. The two models obtained using PAVA and NACT respectively differed increasingly at decreasing spatial wavelengths and model NACT performed better in predicting the observed geoid.

In the present study we apply the NACT inversion technique to a larger data set of SH seismograms. Since the formulation of NACT has been presented in paper 1, we shall not repeat it here. In what follows, we give detailed descriptions on data selection, weighting scheme, and some modeling considerations. We also present the results of formal resolution tests, which are informative in interpreting the inversion results. We assess the validity of our model by examining how successful it is in predicting independent seismic and geophysical observations. We finally compare this model with other existing whole mantle models and discuss its geophysical implications.

Data Selection and Weighting Scheme

In this study, two kinds of waveform data are considered, both extracted from transverse component traces, obtained by appropriately combining the two horizontal component records for each source-receiver pair. The first kind consists of 9368 SH accelerograms (which, in contrast to displacement seismograms, emphasizes higher frequency components) of body waves, low-pass filtered with a cutoff frequency of 1/32 Hz and a corner frequency of 1/37 Hz. The second consists of 7849 SH accelerograms of mantle Love waves of the first and second orbits (G_1 and G_2), low-pass filtered with a cutoff frequency of 1/80 Hz and a corner frequency of 1/100 Hz. The high-frequency content of the data is somewhat larger than in some recent global waveform inversions, which typically low-pass filter with a cutoff frequency lower than 1/40 Hz for body waves[*Tanimoto*, 1990; *Su et al.*, 1994]. The advantage of using shorter-period waveform data is that they have higher resolution due to their shorter wavelengths and that smaller earthquakes, which have higher corner frequencies, can be used. On the other hand, shorter period waveform data are more difficult to model, as pointed out by *Woodhouse and Dziewonski* [1984]. For example, the data variance reductions are 3% for the body waves and −12% (variance increased) for the mantle waves when comparing waveform synthetics obtained using the preliminary reference Earth model (PREM) [*Dziewonski and Anderson*, 1981] as the Earth model and Harvard centroid moment tensor (CMT) solutions [*Dziewonski et al.*, 1981] for the source parameters to our observed waveforms. In contrast, *Dziewonski et al.* [1993] reported that with the same Earth and source models, the variance reduction for their body wave data set with a cutoff frequency of 1/45 Hz was 28% and that for their mantle wave data set with a cutoff frequency of 1/135 Hz was 30%.

Seismograms recorded near the epicenters or their antipodal points ($\Delta < 15^o$ or $\Delta > 165^o$) are not selected since the asymptotic coupling theory breaks down near $\Delta = 0$ or $\Delta = 180^o$ [*Li and Tanimoto*, 1993].

The data were recorded at 85 seismographic stations of of IRIS, Geoscope, Chinese Digital Seismic Network (CDSN), and RSTN networks. The geographic distribution of these stations is shown in Figure 1a. The seismograms used in this study are from 716 earthquakes occurring between 1977 and 1992. The epicenter distribution of these events is shown in Figures 1b and 1c. The scalar moments of the events span 2 orders of magnitudes, from 5×10^{17} to 5×10^{19} N m. We avoid using very large events to prevent possible complications from their long source-time durations, which make the point source assumption invalid.

We manually select windows in the time domain (Figure 2) in order to use only those portions of data which are associated with major energy arrivals. Obviously, this yields a big saving in computation of mode coupling. Another reason for windowing is that it gives us flexibility on how to weigh different phases in the inversion. For example, if we consider the whole wave

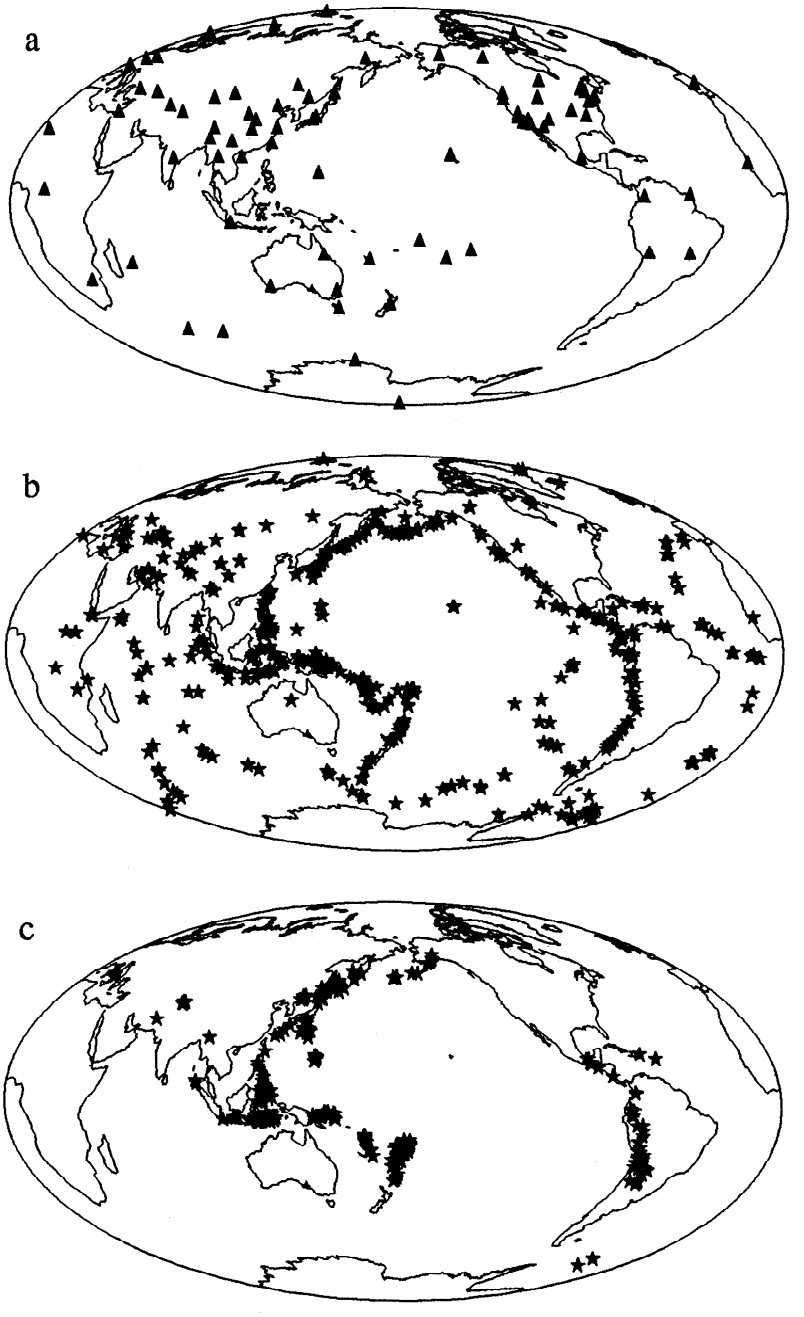

Figure 1. (a) Distribution of seismographic stations whose recordings are used in this study. (b) Distributions of epicenters of shallow events (depth<100 km). (c) Distributions of epicenters of deep events (depth > 100 km).

train in Figure 2a as a block, the S_{dif} phase will be overwhelmed by large-amplitude phases (e.g., SS and S_4) in the inversion, considering that waveforms are matched according to the l_2 norm. However, information carried by the S_{dif} phase is very valuable, because of its strong sensitivity to the lowermost part of the mantle, which is sampled by fewer waves. The windowing scheme enables us to assign weights to different phases with more flexibility.

Generally speaking, the data covariance matrix $\mathbf{C}_d$ used in the inversion (see next section) has both diagonal and off-diagonal non trivial elements. While the diagonal elements of the matrix represent the error in a specific datum, off-diagonal elements quantify the independence of the error between the two corresponding data. For example, unmodeled small-scale structure of the Earth causes highly correlated errors if the data are associated with waves having similar paths. Although there are ways to assign values to off-diagonal elements assuming some error models for the data, the inversion of the matrix $\mathbf{C}_d$ is impractical for a very large data set (the number of data points involved in this study is of the order of a million). Here we take a pragmatic approach. We simply set the off-diagonal elements of the

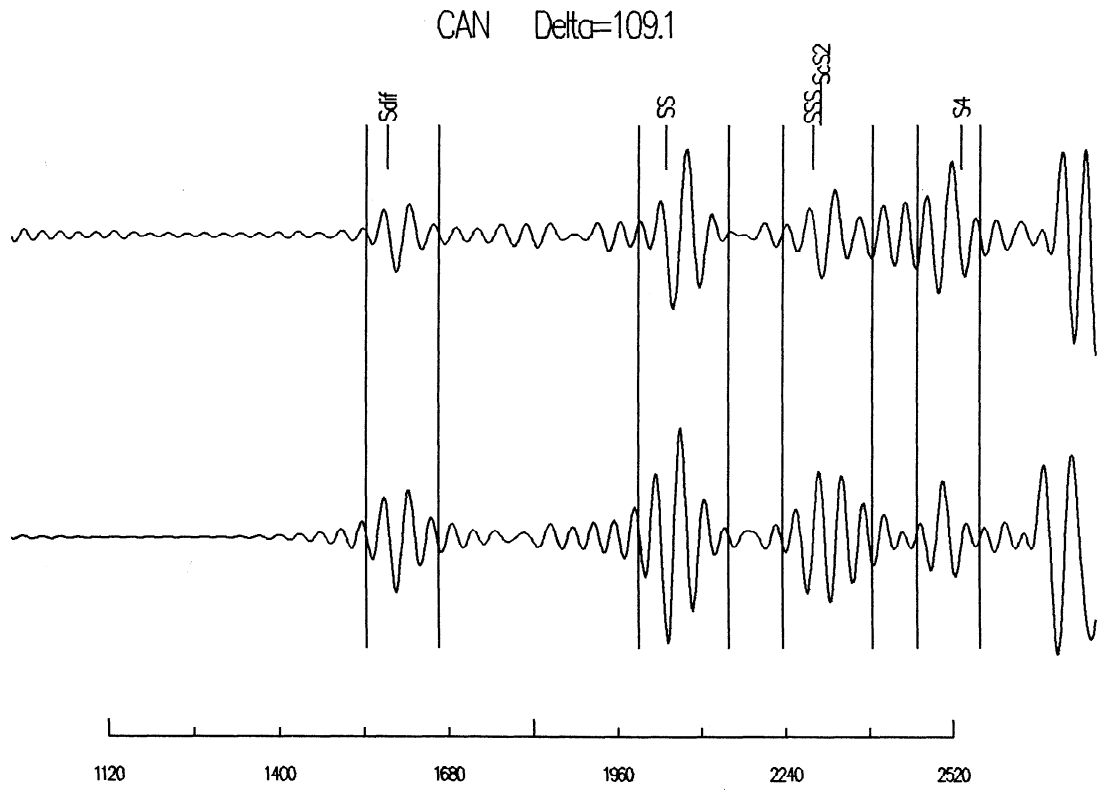

Figure 2a. Low-pass filtered ($f < 1/32$ Hz) *SH* body wave seismogram of January 19,1988, Chile earthquake recorded at Geoscope station CAN. In both Figures 2a and 2b the observed seismogram is plotted on the top trace; and the synthetic seismogram calculated from the spherical reference model PREM [*Dziewonski and Anderson*, 1981] is plotted on the bottom trace for reference. Only the data within the indicated windows are used in the inversions. The time scales are in seconds.

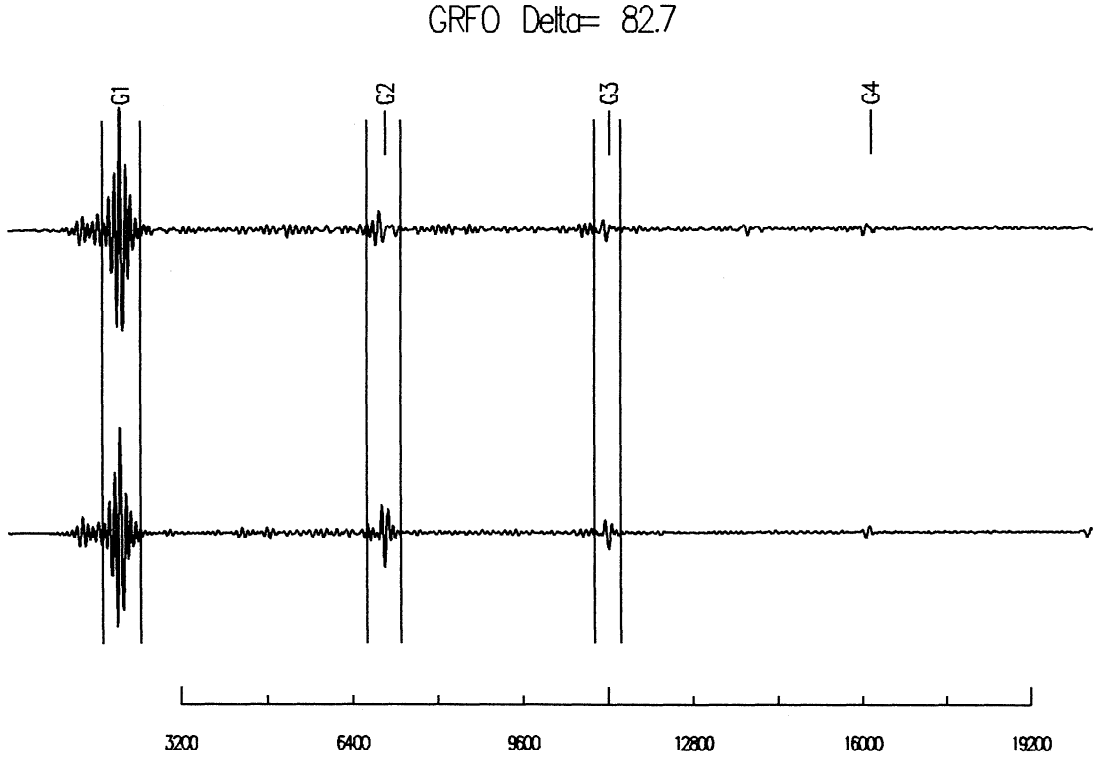

Figure 2b. Low-pass filtered ($f < 1/80$ Hz) *SH* mantle-wave seismogram of 04/22/1987 Japan earthquake recorded at GDSN station GRFO.

data covariance matrix to zero, but we weigh down data if their errors are significantly correlated with errors in the remaining data set.

We introduce a weighting function w for each datum such that the corresponding diagonal element of the inversed data covariance matrix $\mathbf{C}_d^{-1}$ is proportional to w^2. We assume that data belonging to the same "wave packet", which consists of data within a time window such as those shown in Figure 2, have the same weight w. Three factors are taken into account in constructing w:

$$w = w_e w_n w_r, \tag{1}$$

where w_e characterizes the error in the data, w_n measures the redundancy of the data points within a wave packet, and w_r quantifies the redundancy of the whole wave packet with respect to the other wave packets which sample the same regions of the mantle. Evaluation of these factors requires assumptions, which, to a certain extent, can be ad hoc. We explain how we assign these factors in Appendix A.

Inverse Problem for Earth Structure and Source Parameters

Although the main goal of this study is to determine global three-dimensional mantle structure, it is important to address the issue of earthquake source parameter estimation. The two problems cannot be completely separated. Formally, we may write for the seismogram of the sth earthquake recorded at the rth receiver:

$$u_{rs}(t) = f(t; \mathbf{X}_E, \mathbf{X}_s; \mathbf{r}_r), \tag{2}$$

where t is the time, $\mathbf{X}_E$ represents 3-D structure of the Earth, $\mathbf{X}_s$ describes the source parameters of earthquake s, and $\mathbf{r}_r$ is the (known) location of the receiver. Functional f represents the theoretical formalism used to calculate the synthetic seismogram. Generally speaking, both Earth structure $\mathbf{X}_E$ and source parameters $\mathbf{X}_s$ are unknown and need to be determined from the equation. Although it is possible to simultaneously invert for both, we choose an iterative scheme, in which we alternatively solve for either structure parameters $\mathbf{X}_E$ or source parameters $\mathbf{X}_s$, keeping the other set of parameters fixed at the values obtained from the previous iteration.

Inversion for Earth Structure

In this step, we keep fixed the source parameters $\mathbf{X}_s$ of all the earthquakes in our data set. We use the Harvard CMT solutions [*Dziewonski et al.*, 1981] as initial source parameters. Our data set is a collection of discrete time series, $\mathbf{d}$, representing the used portions of seismograms corresponding to different sources and receivers. We seek a (finite dimensional) set of model parameters, $\mathbf{x}$, which represents the 3-D Earth model. Under the assumption that both $\mathbf{d}$ and $\mathbf{x}$ have Gaussian distributions with covariance matrices $\mathbf{C}_d$ and $\mathbf{C}_x$, respectively, a stochastic solution $\mathbf{x}$ [e.g., *Jackson* , 1979; *Tarantola and Valette*, 1982] is one which minimizes the objective function

$$\begin{aligned}\Phi(\mathbf{x}) &= [\mathbf{f}(\mathbf{x}) - \mathbf{d}]^T \mathbf{C}_d^{-1} [\mathbf{f}(\mathbf{x}) - \mathbf{d}] \\ &+ (\mathbf{x} - \mathbf{x}_0)^T \mathbf{C}_x^{-1} (\mathbf{x} - \mathbf{x}_0)\end{aligned} \tag{3}$$

where $\mathbf{x}_0$ is the a priori value of $\mathbf{x}$ and represents our knowledge on the Earth before the inversion. When the forward theory $\mathbf{f}(\mathbf{x})$ is a nonlinear function, the minimum of $\Phi(\mathbf{x})$ is found by the iterative application of the recursion

$$\begin{aligned}\mathbf{x}_{i+1} &= \mathbf{x}_i + (\mathbf{A}_i^T \mathbf{C}_d^{-1} \mathbf{A}_i + \mathbf{C}_x^{-1})^{-1} [\mathbf{A}_i^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{f}(\mathbf{x}_i)) \\ &- \mathbf{C}_x^{-1} (\mathbf{x}_i - \mathbf{x}_0)]\end{aligned} \tag{4}$$

where $\mathbf{A}_i$ is the matrix of partial derivatives

$$\mathbf{A}_i = \left[\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}\right]_{\mathbf{x}=\mathbf{x}_i}, \tag{5}$$

In this study we use the nonlinear asymptotic coupling theory of paper 1 to calculate synthetic seismograms $\mathbf{f}(\mathbf{x})$. In paper 1, an algorithm for calculating the partial derivative matrix was also presented. The evaluation of the data covariance matrix $\mathbf{C}_d$ was given in the previous section. In the next section we will discuss the problems associated with the model covariance matrix $\mathbf{C}_x$ and the a priori model $\mathbf{x}_0$.

Inversion for Earthquake Source Parameters

Under the assumption that earthquake sources can be represented as points in both time and space, we use the concept of centroid moment tensor (CMT) [*Dziewonski et al.*, 1981] to model earthquake sources. For each earthquake, there are 10 parameters describing the CMT solutions, which are the six independent components of the moment tensor [*Dziewonski et al.*, 1981], the centroid origin time, and the three centroid coordinates of the hypocenter. Assuming a 3-D Earth model, the inversion problem for the CMT solution can be solved using a formalism similar to the one presented in the previous subsection. The essential difference is that partial derivatives are calculated with respect to the CMT source parameters rather than with respect to structure parameters. In Appendix B, we give some useful expressions for partial derivatives with respect to the CMT source parameters using NACT.

Modeling Considerations

We invert for models which represent 3-D deviations from a reference model. Since only SH components are used in this study, we model lateral heterogeneity in terms of mantle shear velocity perturbation from the reference model. The contribution of the perturbation in compressional velocity and density is small and hence ignored, a common practice in global waveform modeling [e.g, *Woodhouse and Dziewonski* 1984; *Tanimoto*, 1990]. In order to account for strong lateral heterogeneity in the crust, we also introduce undulations of the free

surface and of the Moho depth. We do not attempt in the present study to invert for anelastic structure, which would require taking into account focusing effects and therefore off great circle plane propagation [e.g., *Romanowicz*, 1987, 1990, 1995]. The anelastic structure assumed is that of PREM [*Dziewonski and Anderson*, 1981] and is kept fixed. We are therefore not taking into account all the information contained in the wave amplitudes and this is a shortcoming common to all elastic tomographic mantle models up to now.

Ellipticity Corrections

In this study the spherically symmetric Earth model PREM [*Dziewonski and Anderson*, 1981] plus the associated hydrostatic ellipticity model, calculated using Clairaut's equation [*Jeffreys*, 1970], are used as the reference model. To correctly account for the effect of the hydrostatic ellipticity is a delicate and important matter, in particular because of the geodynamical importance of non hydrostatic ellipticity in the Earth's mantle, as discussed further in the "results and discussion" section. First, we have found that for the seismic data as used in this study ($T \sim 40$ s), it is very important to convert receiver and source latitudes from geographic (as reported in catalogs) to geocentric (as required by the theory). The error due to the lack of such a conversion can be as large as the sum of all other ellipticity effects. Another important factor is the precision in calculating the effect on normal modes due to hydrostatic ellipticity. Standard expressions in the case of normal modes without coupling were given by *Woodhouse and Dahlen* [1978]. *Woodhouse and Dziewonski* [1984] pointed out that there is rather delicate cancellation, in calculating the splitting effect of the ellipticity in the case of isolated multiplets (no coupling), between the contribution of internal structure and that of boundaries, such as the surface. This is also true for the case when coupling between modes is considered, as we do here. In Appendix C we extend the technique of *Woodhouse and Dziewonski* [1984] for the calculation of the ellipticity effect on normal modes, to the case of multiplet coupling.

Crustal Corrections

The importance of the effect of lateral heterogeneity in the crust on modeling mantle velocity anomalies has been pointed out previously [e.g., *Woodhouse and Dziewonski*, 1984; *Montagner and Tanimoto*, 1991]. In developing their model M84C, *Woodhouse and Dziewonski* [1984] introduced an a priori crustal model to correct for the crustal effect. Such an approach has been followed by numerous authors [e.g., *Zhang and Tanimoto*, 1993; *Su et al.*, 1994]. A problem with this approach is that the uncertainty in the a priori crustal model can bias the inversion results for velocity anomalies at the top of mantle and the extent of the bias is unknown. This could potentially affect the answers to some geophysically very important questions.

In this study, we attempt to investigate the problem by a simultaneous inversion for the Moho undulation and S-velocity structure in the mantle. Such an inversion allows us to perform resolution analyses to see to what extent we can distinguish the crustal structure from the mantle velocity perturbation. We start with an a priori crustal thickness model in which we take a topography-bathymetry model (ETOPO5, 1992) as the perturbation to the ocean floor of the PREM model [*Dziewonski and Anderson*, 1981]. The a priori model of Moho undulation is constructed the same way as by *Woodhouse and Dziewonski* [1984], and represents the difference in Moho depth between the continental and oceanic regions. In the inversion, we allow further perturbations in the Moho undulation as required by the data while keeping the surface topography (the ocean floor of PREM) fixed as in the a priori starting model.

From the point of view of stochastic inverse formalism, the crustal corrections by *Woodhouse and Dziewonski* [1984], and some more recent studies are equivalent to assuming that the *a priori* crustal model has zero variance (precisely known), while in this study we relax this assumption.

In order to obtain a realistic Moho model, one needs data of higher frequency than those used in this study. Reflection data from the Moho could also be crucial. The purpose of inverting for Moho here is to reduce the bias in our mantle model due to the error in the starting a priori crustal model: the inversion result of Moho undulations should also incorporate, to a large extent, the effects of strong lateral velocity variations within the crust, which we do not consider explicitly.

Model Parameterization

We use spherical harmonics, up to angular degree 12, as basis functions to represent lateral variations of structure. This is a natural choice for the NACT approach (paper 1).

For the radial dependence of the perturbation in S velocity, we choose to use Legendre polynomials [*Dziewonski*, 1984; *Woodhouse and Dziewonski*, 1984]. We parameterize the upper and lower mantle separately; that is, we use two independent sets of Legendre polynomials in the upper and lower mantle. This gives us flexibility to model the upper and lower mantle with different resolutions, which is convenient since we have better constraints on the upper mantle than on the lower mantle structure due to surface wave data and depth phases of body waves. We truncate the Legendre expansion at degree 5 in the upper mantle and at degree 7 in the lower mantle.

In recent studies some authors have used different polynomial sets. For example, *Su et al.* [1994] used Chebyshev polynomials. Under the restriction that only the diagonal elements of the model covariance matrix (see (3)) are allowed to be non zero (simple damping of model parameters), the two sets of polynomials could have advantages and disadvantages relative to each other. However, if we specify our a priori constraints using all the necessary elements of the model covariance matrix, there is no difference in the inversion results using one or another set of polynomials (see Appendix D for a proof).

In summary, we expand the relative perturbation in S velocity in each shell (either the upper mantle or the lower mantle) by

$$\delta v_S(r,\theta,\phi)/v_S(r) = \sum_{q=0}^{q_{\max}} \sum_{s=0}^{12} \sum_{t=0}^{s} [a_{st}^q \cos t\phi + b_{st}^q \sin t\phi] p_s^t(\theta) f_q[x(r)], \quad (6)$$

where $p_s^t(\theta)\{\cos t\phi, \sin t\phi\}$ are spherical harmonics as defined in *Stacey* [1977], $v_S(r)$ is the background S velocity evaluated for PREM [*Dziewonski and Anderson*, 1981], $q_{\max} = 5$ for the upper mantle and $q_{\max} = 7$ for the lower mantle, $\{a_{st}^q, b_{st}^q\}$ are the coefficients to be determined in the inversion, and $x(r)$ is reduced, normalized radius [*Dziewonski*, 1984; *Woodhouse and Dziewonski*, 1984]:

$$x = (2r - r_{670} - r_{moh})/(r_{moh} - r_{670}) \quad (7)$$

for the upper mantle and

$$x = (2r - r_{cmb} - r_{670})/(r_{670} - r_{cmb}) \quad (8)$$

for the lower mantle, with r_{moh}, r_{670}, and r_{cmb} being the radii of the Moho, 670-km discontinuity, and core-mantle boundary, respectively. For the perturbation in the surface and Moho radii we simply have

$$\delta r = \sum_{s=0}^{12} \sum_{t=0}^{s} [a_{st} \cos t\phi + b_{st} \sin t\phi] p_s^t(\theta), \quad (9)$$

where δr can be either perturbation in surface radius, δr_{sur}, or in Moho radius δr_{moh}. Thus the complete model parameters are, in the notation used in (3),

$$\mathbf{x} = (\{a_{st}^q, b_{st}^q\}_u, \{a_{st}^q, b_{st}^q\}_l, \{a_{st}, b_{st}\}_s, \{a_{st}, b_{st}\}_m) \quad (10)$$

where the first set of parameters, $\{\}_u$, is for the upper mantle, the second, $\{\}_l$, is for the lower mantle, and the third, $\{\}_s$, and the fourth, $\{\}_m$, are for the surface and Moho, respectively. Altogether there are 2704 model parameters, among which 169 parameters are for the surface undulation and are kept fixed in the inversion.

Model Covariance Matrix

The model covariance matrix $\mathbf{C}_x$ (Cf. (3)) represents our a priori knowledge [*Tarantola and Valette*, 1982] of the Earth structure $\mathbf{x}$. Since seismic data for the given distribution of the sources and receivers do not sample the mantle perfectly, inverse problems for mantle structure are typically mix determined. There is often a range of solutions which can explain the data equally well. A certain specification of $\mathbf{C}_x$ helps to single out, from among them, a particular solution which possesses certain characteristics. We seek a solution which can explain the data well and, at the same time, is small and smooth.

We specify the model covariance matrix $\mathbf{C}_x$ such that the value of $\mathbf{x}^T \mathbf{C}_x^{-1} \mathbf{x}$, often referred to as the penalty function, takes the form

$$\begin{aligned}
&\mathbf{x}^T \mathbf{C}_x^{-1} \mathbf{x} \\
&= \iint \left[\eta_1 m^2 + \eta_2 \left(\frac{\partial m}{\partial r}\right)^2 + \eta_3 \left(\frac{\partial^2 m}{\partial r^2}\right)^2 + \eta_4 |\nabla_1 m|^2 \right] dr d\Omega \\
&+ \int \left[\eta_5 (m|_{r=r_+} - m|_{r=r_-})^2 + \eta_6 \left(\frac{\partial m}{\partial r}|_{r=r_+} - \frac{\partial m}{\partial r}|_{r=r_-}\right)^2 \right. \\
&\quad \left. + \eta_7 \left(\frac{\partial^2 m}{\partial r^2}|_{r=r_+} - \frac{\partial^2 m}{\partial r^2}|_{r=r_-}\right)^2 \right] d\Omega \\
&+ \int \left[\eta_8 (\delta r_{moh})^2 + \eta_9 |\nabla_1 \delta r_{moh}|^2 \right] d\Omega, \quad (11)
\end{aligned}$$

where $m \equiv \delta v_S / v_S$ is the S velocity model in the mantle, the radial intergration $\int dr$ is from the core-mantle boundary to the Moho, the surface intergration $\int d\Omega$ is over the surface of the unit sphere, ∇_1 denotes the surface gradient operator, r_+ is the radius just above the 670-km discontinuity, r_- is the radius just below the 670-km discontinuity, In equation (11), $\eta_1, \eta_2, \ldots \eta_9$ are empirical damping parameters, which are adjusted by trial and error to achieve reasonable size and smoothness of the model rms perturbations as a function of depth. Substituting (6) and (9) into the right side of (11), the elements of $\mathbf{C}_x^{-1}$ are readily obtainable by identifying corresponding terms on both sides of the equation. The η_1 term on the right side of (11) corresponds to the requirement of smallness. The η_2 and η_3 terms constrain the solution to be smooth radially. The η_4 term imposes horizontal smoothness. We introduce the η_5, η_6, and η_7 terms to reduce the discontinuity in the perturbation model across the upper-lower mantle boundary. Lateral heterogeneity is not necessarily continuous across the boundary. Given the long wavelengths of the data used in this study, however, we think any sharp discontinuity could just be an artifact due to our particular separate parameterization. In order to study whether or not lateral heterogeneity has a sharp discontinuity across the boundary, higher frequency data, signals reflected from the 670-km discontinuity in particular, are needed [e.g., *Shearer and Masters*, 1992]. Finally, the η_8 and η_9 terms make the perturbation in Moho depth smaller and smoother.

Formal Resolution Analysis

In the resolution analyses, we attempt to investigate the effect on the inversion results of the a priori model covariance matrix, which is introduced to reduce the effect of errors in the data at the cost of resolution loss. Since the forward theory used in this study, NACT, is nonlinear in terms of the relationship between the model and data, the inverse problem has to be solved iteratively. As for formal resolution, however, we choose to perform conventional linear analysis. In such an analysis we assume that we can write

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0) + \mathbf{A}\Delta\mathbf{x}, \quad (12)$$

where $\mathbf{A}$ is the partial derivative matrix and $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}_0$. Thus the minimum of the objective function (3) is given by

$$\Delta \mathbf{x} = (\mathbf{A}^T \mathbf{C}_d^{-1} \mathbf{A} + \mathbf{C}_x^{-1})^{-1} \mathbf{A}^T \mathbf{C}_d^{-1} [\mathbf{d} - \mathbf{f}(\mathbf{x}_0)]. \quad (13)$$

In the resolution analyses, we calculate synthetic data for a given input model $\Delta \mathbf{x}_{\text{in}}$:

$$\mathbf{d} - \mathbf{f}(\mathbf{x}_0) = \mathbf{A} \Delta \mathbf{x}_{\text{in}}. \quad (14)$$

Thus the output model is obtained by inversion,

$$\Delta \mathbf{x}_{\text{out}} = \mathbf{G} \Delta \mathbf{x}_{\text{in}}, \quad (15)$$

where $\mathbf{G}$ is the resolution matrix [*Aki and Richards*, 1980] and is given by

$$\mathbf{G} = (\mathbf{A}^T \mathbf{C}_d^{-1} \mathbf{A} + \mathbf{C}_x^{-1})^{-1} \mathbf{A}^T \mathbf{C}_d^{-1} \mathbf{A}. \quad (16)$$

Note that the resolution matrix becomes the identity matrix when no a priori constraints are applied ($\mathbf{C}_x^{-1} = 0$). In addition to the forward theory used to calculate the partial derivative matrix $\mathbf{A}$, the resolution matrix depends upon two factors: (1) the data set, determined by the source-receiver distribution, data selection and weighting scheme $\mathbf{C}_d$; and (2) the model covariance matrix $\mathbf{C}_x$. In all the experiments presented in this section, we keep these two factors the same as used in the real inversion.

The purpose of the first resolution experiment is 1) to see whether the uncertainty in the Moho undulation will significantly affect the inversion results in the mantle; and 2) to assess how much of a realistic input mantle model can be recovered in the inversion. In constructing the input model, we let the *S*velocity structure in the mantle be the same as that of model S12/WM13 [*Su et al.*, 1994], reparametrized using Legendre polynomials in depth. We introduce a checkerboard Moho undulation with a peak amplitude of ±15 km to represent the uncertainty in the a priori Moho model. The input Moho model consists of a spherical harmonic angular degree 5 and azimuthal order 1 sine term, which has very weak correlation with the input mantle lateral heterogeneity (see the solid curve in Figure 3). We correlate the output mantle model with the pattern of the input Moho model and the correlation coefficient, as a function of depth, is shown as the dashed curve in Figure 3. Although a slightly increased correlation coefficient can be seen in some regions, in particular the top 300 km of the mantle, the effect of unmodeled Moho depth variation on the mantle result is small, especially considering that a large input Moho model is used.

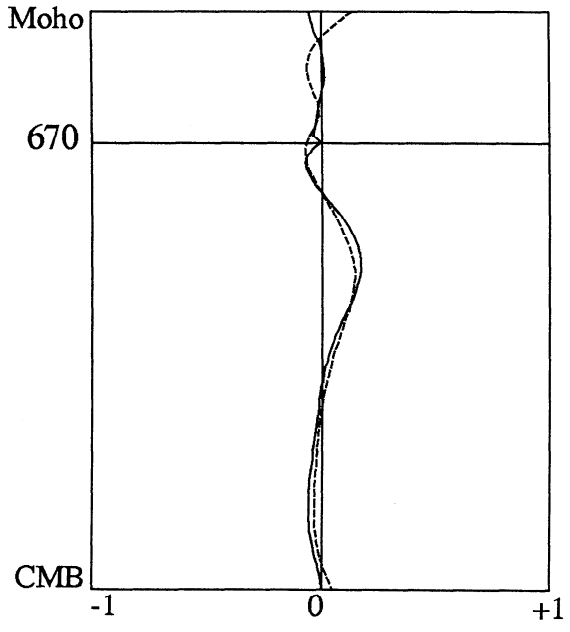

Figure 3. Resolution test on contamination of mantle tomographic result due to the uncertainty in Moho depth. A large Moho undulation (with ±15 km peak value) model is introduced to represent the uncertainty. The correlation coefficient between lateral heterogeneity in the mantle and the lateral variations in the input Moho model is given as a function of depth. The solid curve is for the input mantle model, and the dashed curve is for the output mantle model obtained after inversion.

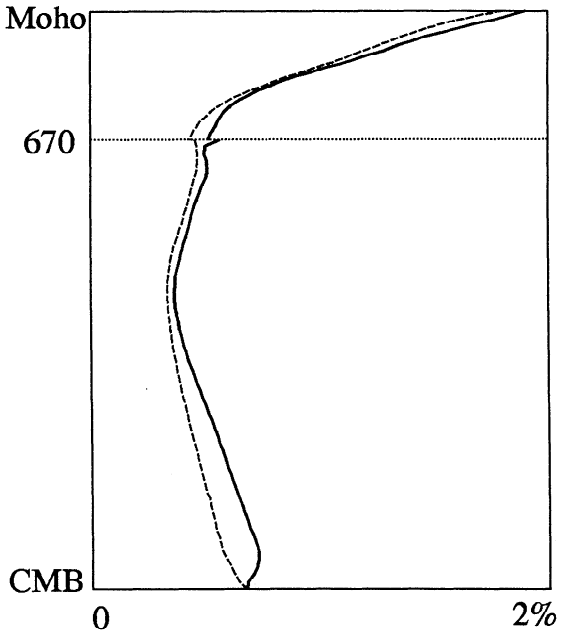

Figure 4. The rms amplitude of the input (solid curve) and output (dashed curve) models in the resolution test described in text.

In Figure 4 we show the comparison of the rms lateral heterogeneity in the input and output models. In general, there is a loss of amplitude due to the damping in the inversion throughout the mantle, with most of the loss occurring at the bottom of the lower mantle. Note also that the extent of amplitude loss depends upon the smoothness of the input model. The smoother the input model, the smaller the loss. In order to further demonstrate this, we compare the power spectra of the input and output models in Figure 5. In the lower mantle, some higher-degree components lose more power than lower-degree components. For example, in the depth range 1700-2300 km, approximately 25% power is lost for degree 2 and 50% is lost for degree 12.

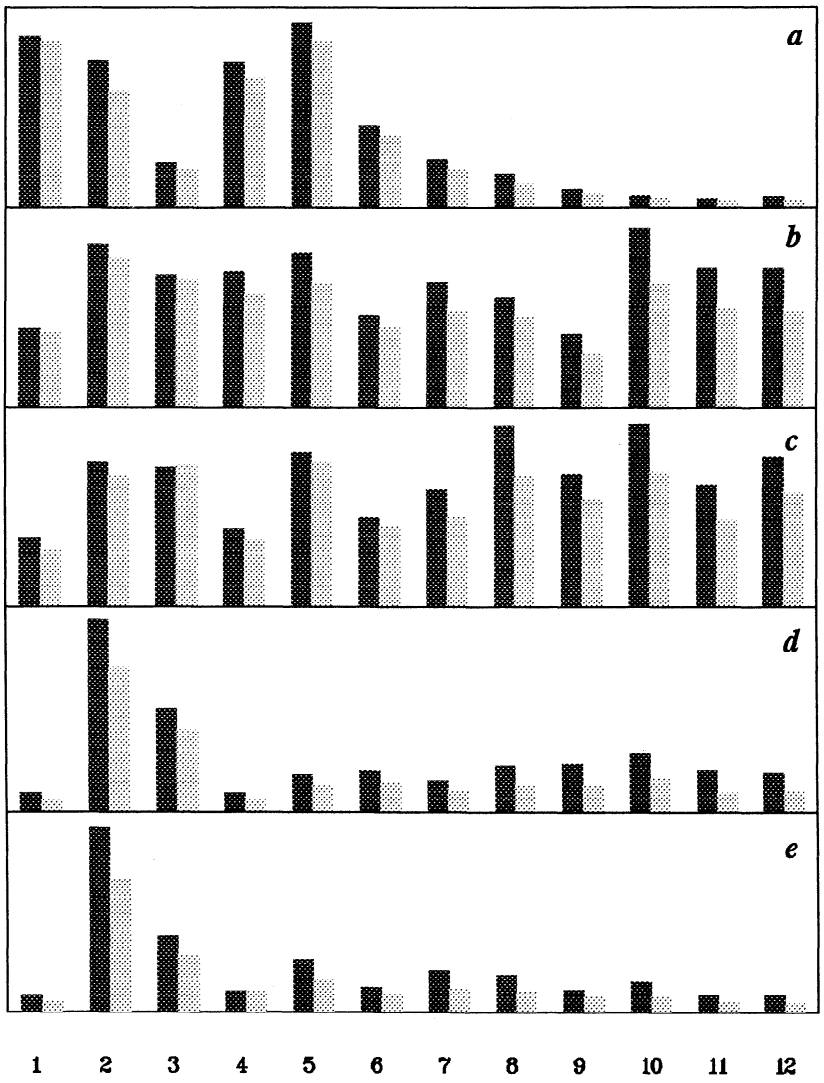

Figure 5. Power spectra, as a function of angular degree, of the input (darker-shaded bars) and output (lighter-shaded bars) models in the resolution test described in Section 5. The input mantle model is S12/WM13 of [*Su et al.* , 1994]. Vertically averaged power spectra are given for the following depth ranges: (a) 0–500 km; (b) 500–1100 km; (c) 1100-1700 km; (d) 1700-2300 km; and (e) 2300–2900 km. Each panel is normalized by its maximum value.

In Figure 6 we show the result of a "checkerboard test" for the whole mantle. The input model has a degree 12 and order 6 horizontal pattern (degree 12 is the highest degree in our spherical harmonic expansion). Radially, it has one zero crossing in the upper mantle and four zero crossings in the lower mantle. An equatorial vertical cross section of this model is shown in Figure 6a. The recovered model is shown in Figure 6b. The pattern of the model is preserved satisfactorily in the inversion. The amplitude is also recovered well except in some regions around 2000 km depth.

In paper 1, we used a smaller data set to demonstrate that NACT does better than the conventional PAVA in resolving a degree 8 checkerboard pattern. In Figure 6c we show the corresponding result, obtained using PAVA, a degree 12 checkerboard pattern, and the current data set. The only difference between Figures 6b and 6c is that different theories are used in calculating the partial derivative matrix. Clearly, the NACT still has better resolution power than the PAVA for the larger data set used in this study. One may argue that the shortcoming of a less accurate theory can become insignificant if enough data are available. Supposing this is true, we still emphasize the importance of the theory because a data set which is sufficient for one problem will become insufficient for more complicated problems. Finally, we note that the comparison between Figures 6b and 6c does not give the error due to the use of the less accurate PAVA, since the resolution matrix depends on the interaction between the partial derivative matrix $\mathbf{A}$ and the model covariance matrix $\mathbf{C}_x$, not directly on the error in the theory used in constructing $\mathbf{A}$.

We note, in particular, that the resolution we obtain

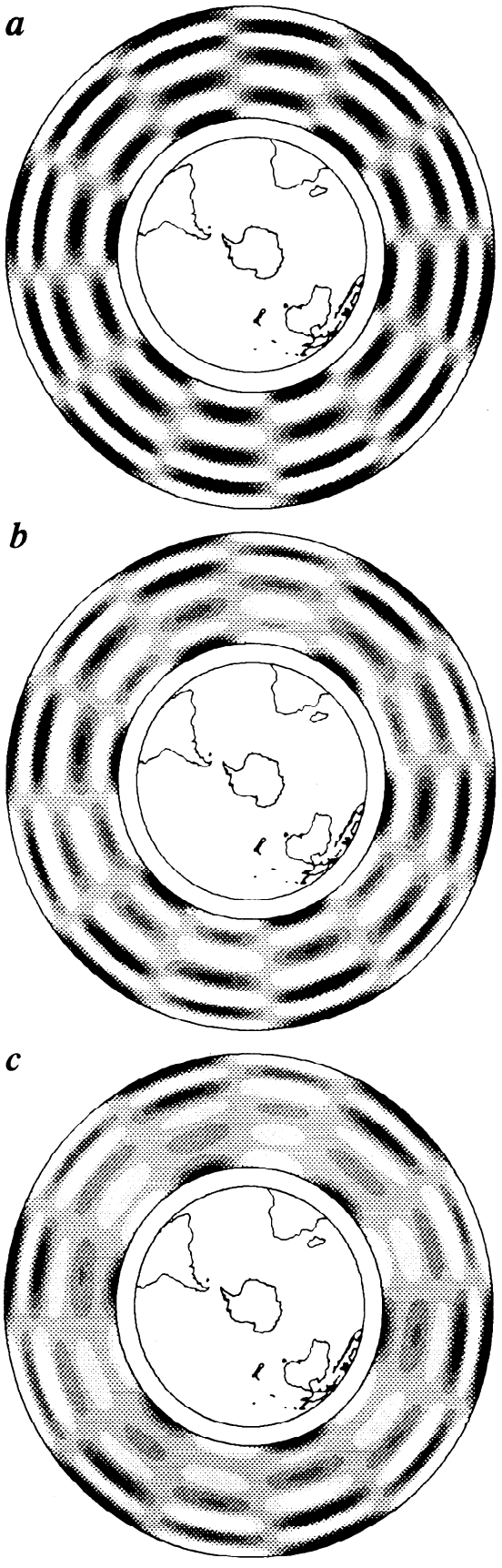

Figure 6. Vertical cross sections showing the results of a checkerboard test for the whole mantle. (a) Input model; (b) output model after inversion using NACT; (c) output model after inversion using PAVA.

in the lowermost mantle using the PAVA approximation (Figure 6c) appears better than that shown by *Su et al.* [1994, Figure 13], although they have a larger waveform data set and, in addition, travel time data. We attribute the difference mainly to the fact that we window individual body wave packets, which allows us to give appropriate weight to weaker phases in the seismograms, such as S_{dif}, improving resolution in the lowermost mantle (see Figure 2a).

Results and Discussion

The final model, SAW12D (S velocity, asymptotic theory, waveform data, degree 12 model of version D), is tabulated in Table A1[1]. As mentioned earlier, the surface topographic correction is taken from the ETOPO5 data set and is fixed in the inversion. The Moho undulation and mantle S velocity perturbations are the result of the inversion.

Of the order of 3-4 iterations are required for the model to converge, when starting from a spherically symmetric model such as PREM, and one iteration is sufficient to invert for perturbations in source parameters. We note that of the order of 35% of the total variance reduction is achieved by allowing adjustments in source parameters, indicating that this step is significant. In this paper we will not examine the perturbations in source locations and mechanisms obtained, postponing this until a later contribution, in which we will show results of inverting both SV and SH sensitive waves, to constrain all elements of the seismic moment.

The final variance reductions obtained with model SAW12D are 40% for the body waves and 52% for the surface wave data. The variance reductions using a recently published 3-D S velocity model S12/WM13 [*Su et al.*, 1994] are −6% and 46% for body waves and surface waves, respectively. The poor fit to the body waves using S12/WM13 comes mainly from the discrepancy in small-scale features between the two models. Unlike the path average approximation, which desensitizes the data to higher-degree structure, NACT predicts very localized sensitivity kernels [*Li and Tanimoto*, 1993; paper 1]. In Figure 7 we compare the sensitivity kernels of an SS phase predicted using PAVA and NACT. In order to compute a synthetic seismogram that matches the data, NACT requires an accurate local velocity model along the ray path, whereas PAVA needs only the horizontally averaged velocity structure, between the source and receiver, to be correct. Indeed, if we use PAVA the variance reduction to the body wave data from S12/WM13 is increased to 17%. Of the order of 3-4 iterations are required at most for the model to converge, when starting from a spherically symmetric model such

[1] Supporting data table is available on diskette or via ANonymous FTP from kosmos.agu.org directory APEND (Username = anonymous, Password = guest). Diskette may be ordered from AMerican Geophysical Union, 2000 Florida Avenue, N.W., Washington, DC 20009 or by phone at 800-966-2481; $15.00. Payment must accompany order.

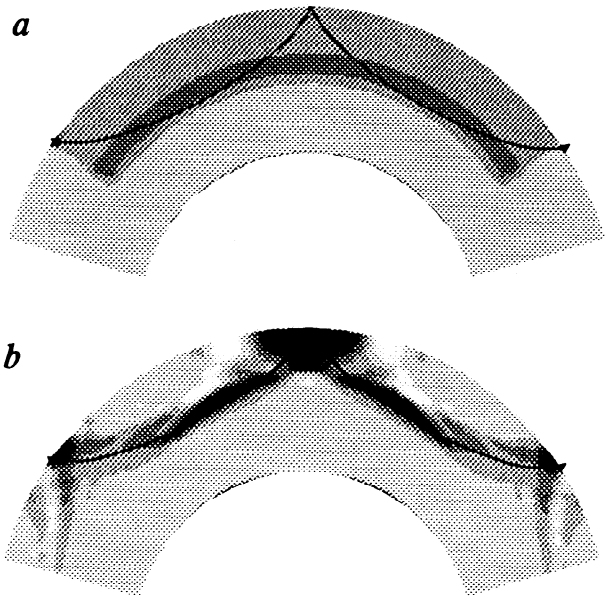

Figure 7. Sensitivity kernels for an *SS* phase, using two different theoretical approaches: (a) PAVA and (b) NACT. Black and white shades indicate positive and negative values of kernels, respectively. The background gray shade indicates zero sensitivity. The star and triangle indicate the locations of the source and receiver, respectively. The thin curve represents the ray predicted using the geometrical ray theory.

as PREM, and that one iteration is sufficient to invert for perturbations in source parameters. We note that of the order of 35% of the total variance reduction is achieved by allowing for adjustments in source parameters, indicating that this step is important. In this paper we will not examine the perturbations in source locations and mechanisms obtained, postponing this until a later contribution, in which we will show results of inverting both *SV* and *SH* sensitive waves, and which will allow us to constrain all elements of the seismic moment.

Prediction of Independent Seismological Measurements

A powerful test to assess the quality of a model is to determine how well it can predict independent seismic data. Using model SAW12D and the geometrical ray theory, we have calculated synthetic differential travel times for the *S-SS* and *S-ScS* data set of *Woodward and Masters* [1991a, b]. The variance reduction is 27% for the *S-SS* data and 25% for the *S-ScS* data. These numbers are rather low, compared with the variance reductions given by model S12/WM13 (62% and 66%, respectively), which was developed using the travel time measurements. The question now is whether the travel time data, as interpreted by the geometrical ray theory, require a significantly different model. In order to answer this question, we performed a simultaneous inversion of both data sets. The resulting model, SAWT12D, reduces the data variance by 66% and 62% for the *S-SS* and *S-ScS* data sets, respectively. In order to assess the change in the model caused by addition of traveltime data, we plot the correlation coefficients, as a function of depth, between SAWT12D and SAW12D and between SAWT12D and S12/WM13 in Figure 8a. SAWT12D is perfectly correlated with SAW12D in the upper mantle, and the correlation in the lower mantle is still good, compared to the correlation with S12/WM13. In Figure 8b we compare the profiles with depth of rms lateral heterogeneity of these three models. The change introduced by the addition of the travel time data is relatively small and confined to the lower mantle, particularly between 1500 and 2500 km depth.

Another kind of independent seismological measurements comes from free-oscillation data. For well-isolated multiplets, we may retrieve splitting function coeffi-

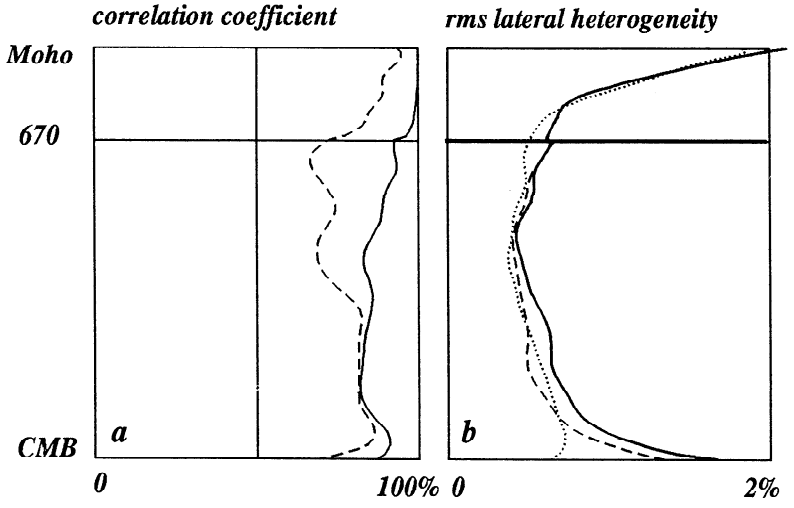

Figure 8. (a) Correlation coefficients as a function of depth between models SAW12D and SAWT12D (solid curve) and between models SAW12D and S12/WM13 (dashed curve); b) comparison of rms velocity variations as a function of depth for models SAWT12D (solid curve), SAW12D (dashed curve), and S12/WM13 (dotted curve).

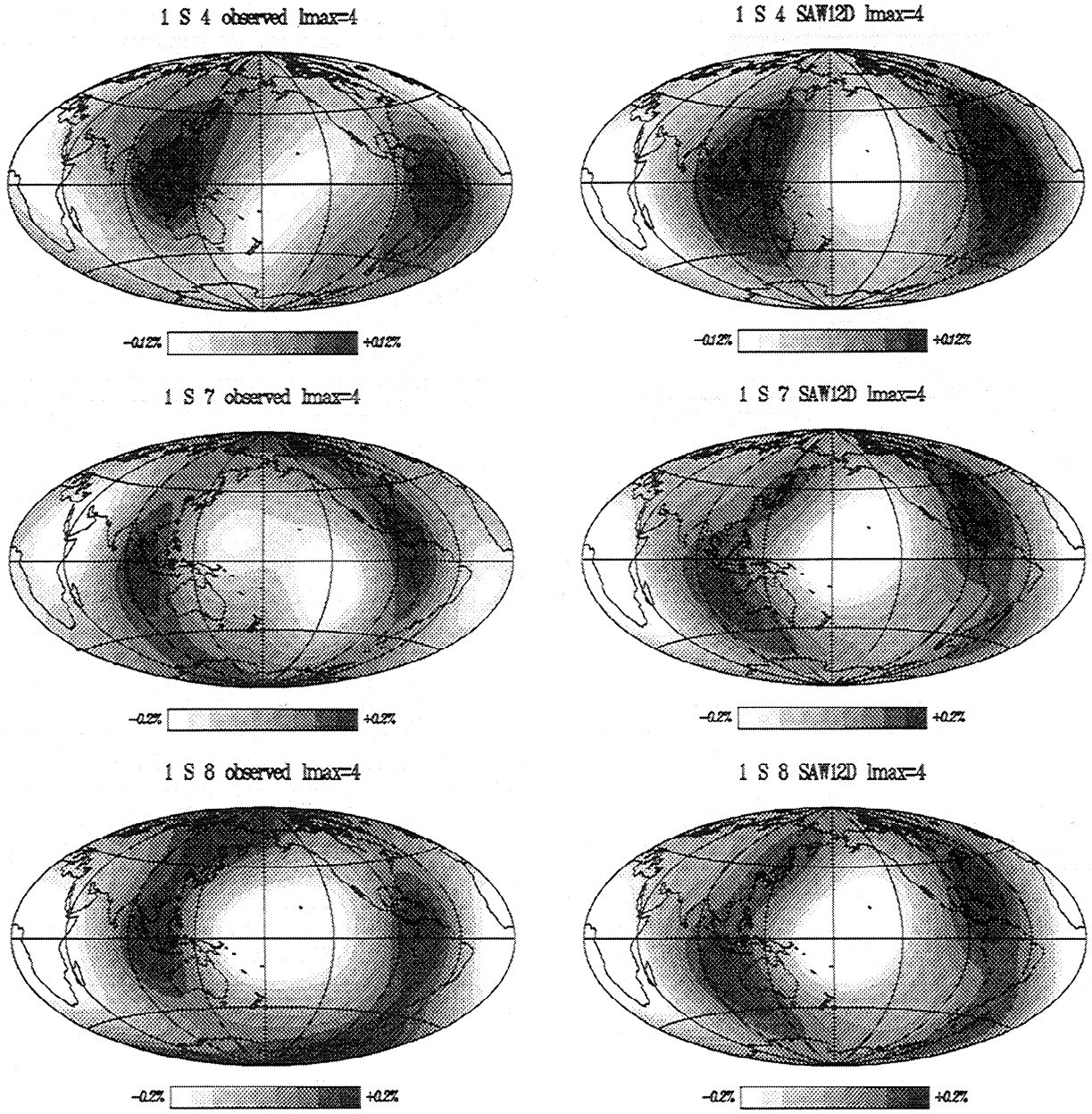

Figure 9. Comparison of splitting functions as (left) observed and (right) predicted using SAW12D. The splitting functions are truncated at spherical harmonic degree 4.

cients from long-period seismograms [*Woodhouse and Giardini*, 1985; *Ritzwoller et al.*, 1986]. Splitting functions impose constraints on large-scale structure of even spherical-harmonic degrees. In Figure 9 we show a comparison between splitting functions retrieved from the observation and those predicted using SAW12D. These splitting functions belong to modes which are primarily sensitive to S velocity structure in the mantle. The observed splitting functions are retrieved from seismic spectra of June 9, 1994 Bolivia and October 4, 1994, Kuril Islands events. Although no modal data have been used in the inversion, SAW12D can predict these splitting functions well.

Spatial Distribution of Structural Heterogeneities

The rms fluctuations as a function of depth of model SAW12D is shown as the thick solid curve in Figure 10, compared with other models. SAW12D agrees very well with recently published S12/WM13 of *Su et al.* [1994] (the dashed curve in Figure 10) in the top 400 km of the mantle. The level of lateral heterogeneity of the former is somewhat higher around the 670 km-discontinuity. The largest difference between these two models is in the D'' region, where SAW12D has significantly stronger lateral heterogeneity, in better agree-

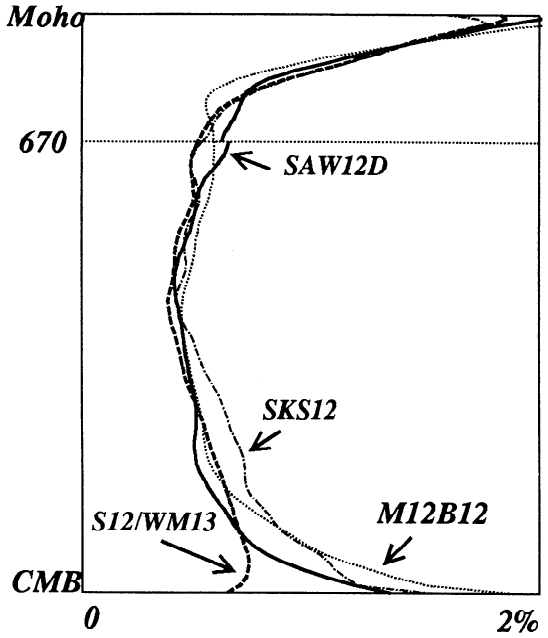

Figure 10. The rms velocity variations as a function of depth for models: SAW12D (solid curve), S12/WM13 (dashed curve), M12B12 (dotted curve), and SKS12 (dashed-dotted curve).

ment with some recent mantle models [*Johnson et al.*, 1994; *Liu et al.*, 1994] (the dotted and dashed-dotted curves in Figure 10). The spherically symmetric part of model SAW12D differs from PREM by less than 0.2% in most of the mantle, except at the top of the upper mantle, where it is faster by 0.7%.

In Figure 11 we show the starting and final Moho undulation models. The starting model is the same as used by *Woodhouse and Dziewonski* [1984] and represents the difference between oceanic and continental regions. The final model, which is damped toward the starting model in the inversion, has a smaller amplitude in general. Some tectonic features are seen in the final model. For example, shields have deep roots, and younger ocean floor near mid-ocean ridges is generally thinner than older oceanic crust. Some small-scale features are suspicious and will need to be corrected in the future, for example, by adding shorter-period surface wave data. As discussed in subsection on modeling considerations/crustal corrections", the goal of inverting for Moho depth is primarily to minimize contamination of the mantle part of the model by potentially erroneous crustal structure.

In Plate 1, we present maps of model SAW12D at various depths. The S velocity anomaly at a depth of 150 km follows the major tectonic features as in previous tomographic upper mantle models. Mid-ocean ridges are associated with slow anomalies. The slowest anomalies are located under Baja California, the triple junction of Pacific-Nazca-Antarctic plates, and the East African Rift. Back-arc regions of some major subduction trenches are also slow. The most striking fast anomaly is underneath the Canadian Shield. The other tectonically stable areas, such as West Australia, Antarctic, Siberia, Russian Platform, Brazil, West Africa Craton, and central west Africa (Zaire) are fast. Older ocean floor near subduction zones is faster than young ocean floor near mid-ocean ridges.

At a depth of 250 km, the pattern of lateral heterogeneity remains similar to that at 150 km depth, while the amplitude decays. The dominance of high velocities beneath shields is still clear, supporting the hypothesis of the continental tectosphere [*Jordan*, 1975, 1978], whereas the strongest slow features are associated with the East Pacific and Antarctica rises. This is in agreement with other upper mantle studies [*Montagner and Tanimoto*, 1991; *Su et al.*, 1994; *Trampert and Woodhouse*, 1995].

Between 250 km and 350 km, the distribution has started to change. The shield signature has disappeared in North America, and an elongated zone of slow anomalies is emerging in the central Pacific, going through Hawaii. At a depth of 450 km the shield signature has disappeared practically everywhere, except in west Africa, which is still fast, indicating that

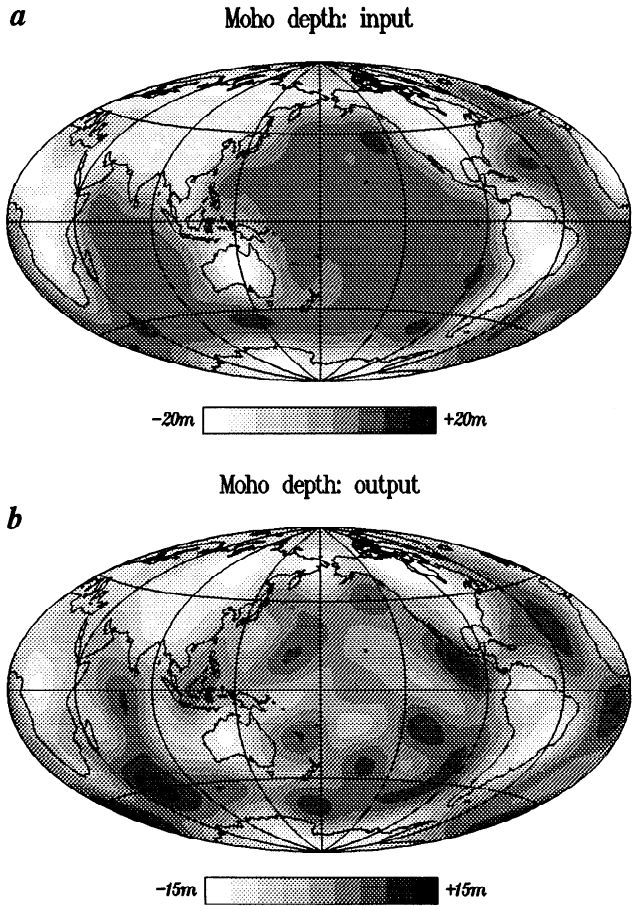

Figure 11. (a) The a priori starting model and (b) the final model of perturbations in Moho radius. The darker shades indicate shallower depth.

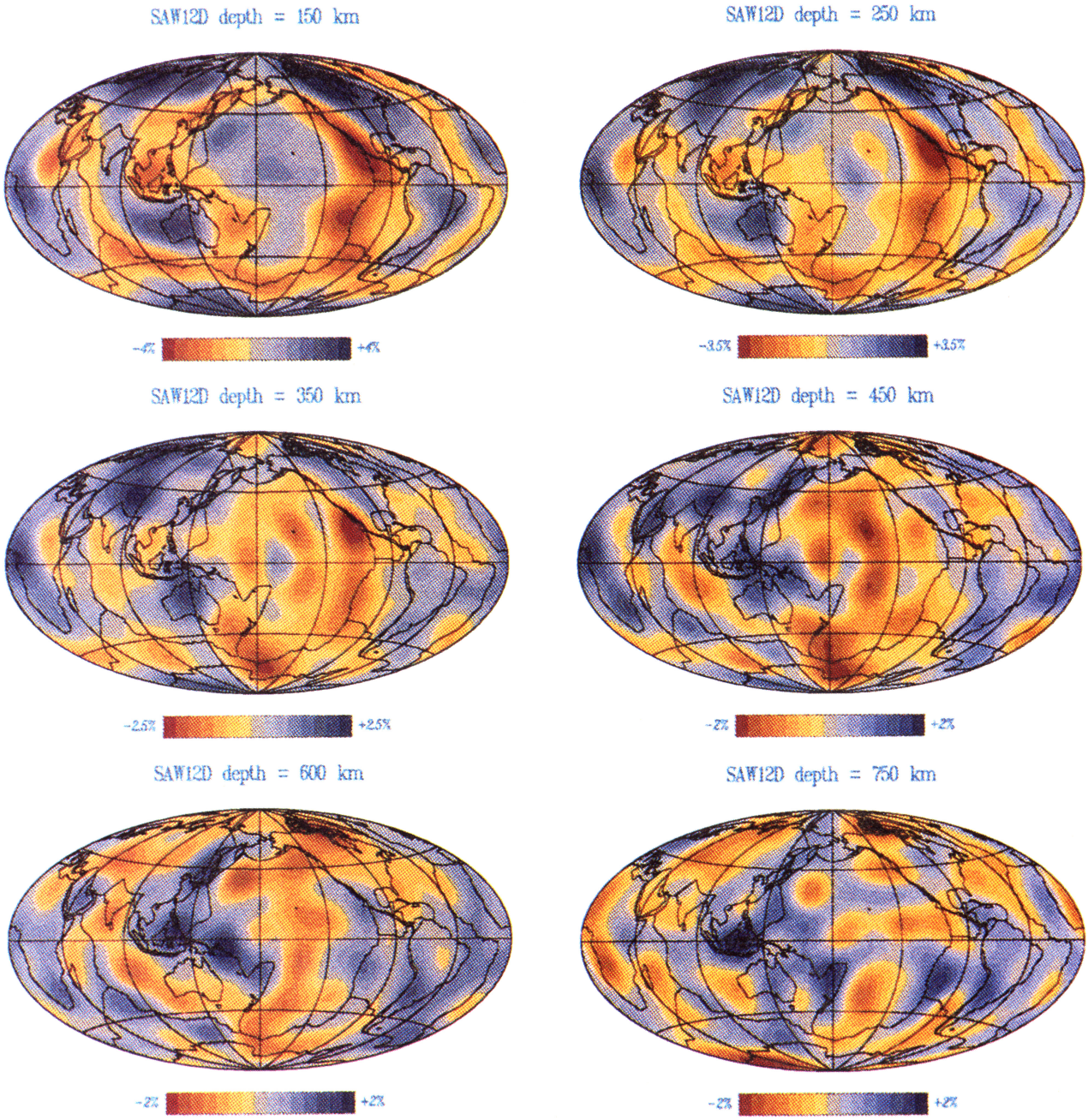

Plate 1. Relative S velocity perturbations of model SAW12D shown at various depths. The average shift with respect to the PREM model (C_{00} coefficient) at each depth has been removed before plotting.

the maximum depth of continental roots [*Jordan* 1975, 1978] is shallower than 450 km. A striking feature is the fast anomalies associated with western Pacific subduction. There is some indication of faster than average anomalies in western South America, likely associated with the subduction there. Faster than average velocities are also seen under Tibet, in agreement with other global models [*Su et al.*, 1994]. The slowest regions have now definitely moved away from the ridge, and in the Pacific, they now appear to follow a north-south trend parallel to the East Pacific Rise, but centered on Hawaii and the Pacific-Antarctic-Australian triple junction. This gradual offset of slow anomalies to the west becomes apparent at around 300 km depth and is well established at 450 km depth. At 600 km depth, higher than average velocities have replaced low ones under the East Pacific Rise. The shift to the west of the center of low velocities in the Pacific, between the uppermost mantle (0-250 km) and the transition zone (400-650 km) has been reported previously and detected in the lowest harmonic degrees of upper mantle structure [e.g., *Montagner and Romanowicz*, 1993].

In Plate 1, we also plot the S velocity anomalies above (600 km depth) and below (750 km depth) the 670-km

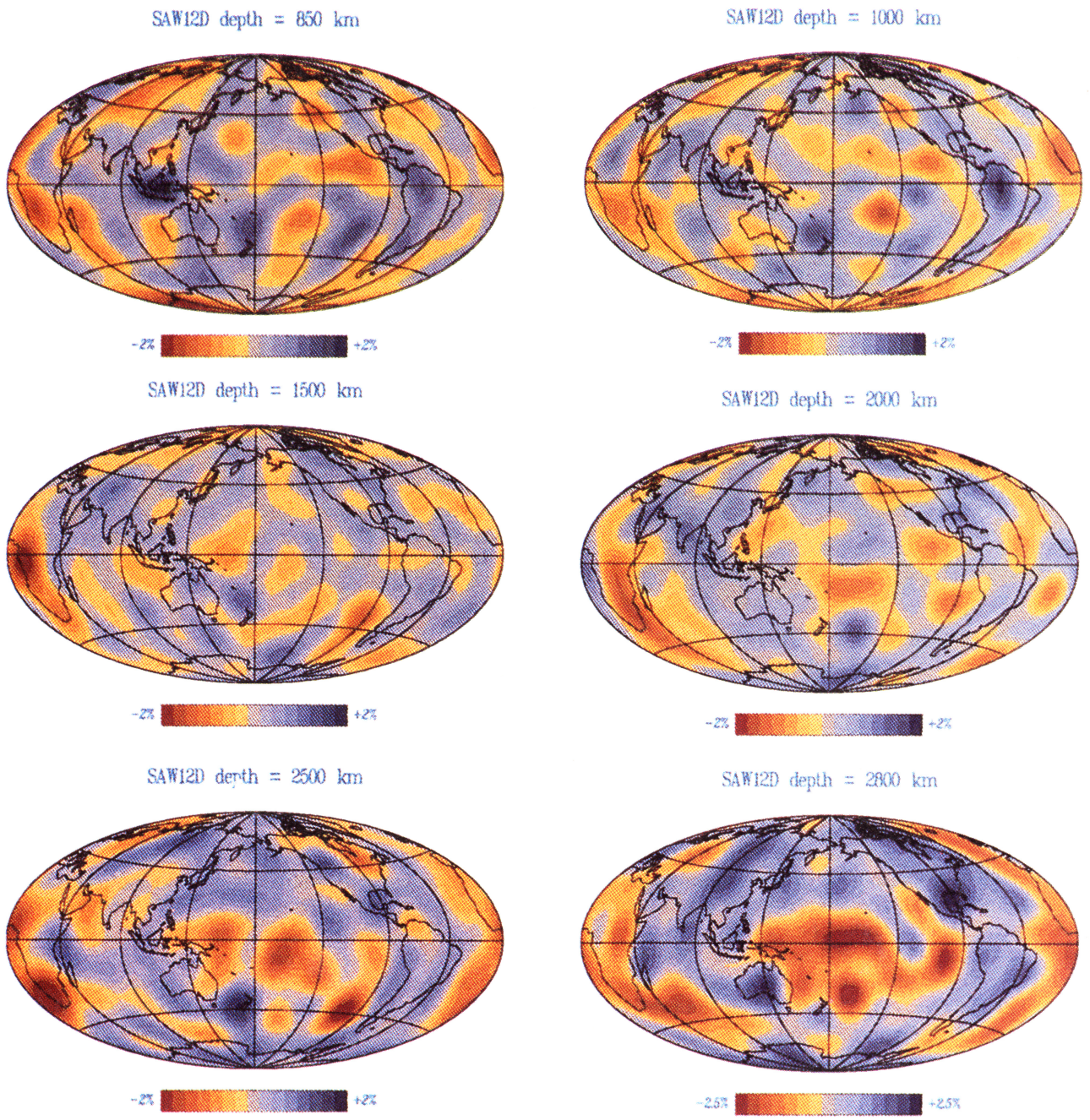

Plate 1. (continued)

discontinuity. At a depth of 600 km, the strongest positive anomalies are the fast features associated with the western Pacific subductions. Fast anomalies associated with the South American subduction zone can also be seen, though not as strongly. We see slab related fast velocity anomalies on both sides of the discontinuity, especially beneath Fiji, Tonga, Indonesia, and South America. However, the fast anomalies in the west Pacific fade out at 1000 km depth. Although the limited vertical resolution of the model does not allow us to provide definitive evidence either for or against the slab penetration across the 670-km discontinuity, the areas where the fast anomalies subsist at 750 km (Fiji-Tonga and Indonesia) are in good agreement with results of regional studies [*VanDerHilst*, 1993, 1995]. In Figure 12, we compare the vertical stack of model SAW12D between 580 and 725 km depths to the slab model of *Ricard et al.* [1993]. Note that the latter model is based simply upon Cenozoic and Mesozoic plate motion and thus does not take into account lateral heterogeneity caused by any other sources, in particular potential hot rising currents. This comparison is, however, useful in that it provides a framework for, on the one hand, interpreting the fast anomalies in our modeling and, on the other hand, validating some features of the slab model. Except for the strong fast feature near the East Pacific Rise, the fastest anomalies correlate well with the slab model in their general trend. We note that this

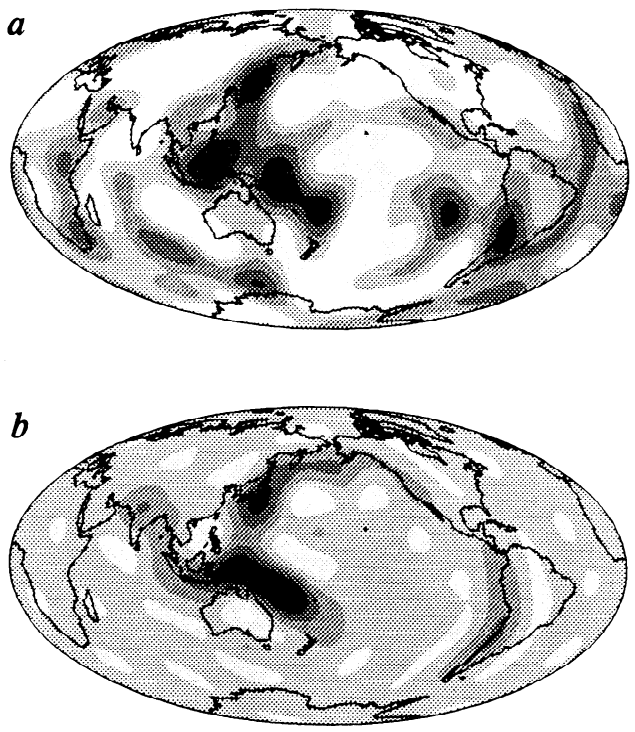

Figure 12. Comparison of (a) SAW12D S velocity structure, stacked between 580 and 725 km depths, and (b) the slab model of *Ricard et al.* [1993] in the same depth range.

intriguing fast feature near the East Pacific Rise is also present, in this depth range, in the more regional model of *Grand* [1994]. We also note that the comparison with the slab model reveals that the region of slow anomalies in the central Pacific cannot be explained simply by a truncation of the slab model at low harmonics. This is also the case for slow anomalies under Eurasia and Africa. The slow velocity "corridor" in the central Pacific starts breaking apart at 750 km depth (Plate 1), indicating that it may originate near the base of the upper mantle. At a depth of 850 km, the distribution of heterogeneity is similar to that at 750 km, with weaker amplitudes.

At a depth of 1000 km, the large-scale high- and low-velocity patterns have been replaced by small-scale anomalies that are responsible for the "white" character of the heterogeneity spectrum throughout the mid mantle. We note remaining blobs of high velocities under Fiji-Tonga and South America, the latter in agreement with *Grand*'s [1994] observation of the lost Farallon plate. At a depth of 1500 km, the pattern of lateral heterogeneity is characterized by the lack of dominant large-scale structure, in agreement with the results of *Su et al.* [1994]. The maximum velocity anomaly is the slow feature underneath central west Africa, which is still striking at 2000 km depth. Starting at 2500 km depth, we see a fast Pacific rim with slow central Pacific and south Africa, a pattern in the lowermost mantle which has been reported in practically every global mantle model published so far. Unlike S12/WM13 and some other models which have a dominant spherical harmonic degree 2 component, however, SAW12D is still rich in higher degrees at this depth.

In the D'' region, represented here at a depth of 2800 km, as in other models [*Su et al.*, 1994; *Johnson et al.*, 1994], long wavelength features become dominant again: fast Pacific rim, slow central Pacific and South Africa. The details of the pattern in D'' needs to be validated through comparison with other, more specific studies of this part of the mantle. The fast velocity anomaly under the Caribbean, not as apparent in other global models, is in agreement with *Grand*'s [1994] regional model. Also, the fine structure in the central Pacific, with a minimum of velocity located around 30° south seems to correlate well with preliminary results on the geographical thickness pattern of a thin very low velocity zone at the base of D'' (E. Garnero, personal communication, 1995). Although we use *ScS* phases, we cannot resolve here any possible trade-off between lateral variations in D" and undulations of the core-mantle boundary. For these, both reflected and transmitted waves are needed, as well as shorter-period measurements.

Wavelengths of Lateral Heterogeneity

The spectral character of mantle heterogeneity provides important constraints for geodynamicists. Numerical simulations of mantle convection [e.g., *Tackley et al.*, 1993; *Bunge et al.*, 1995] do not give realistic geographic locations of lateral heterogeneity. Instead, they attempt to predict realistic spectral characteristics of lateral heterogeneity. The scale of lateral heterogeneity in global tomographic models is also the subject of current debate [e.g., *Snieder et al.*, 1991; *Su and Dziewonski*, 1991, 1992; *Zhang and Tanimoto*, 1991].

In Figure 13 we show vertically averaged power spectra of SAW12D in each of 10 layers throughout the mantle. In Figure 13a, each panel is normalized to its own maximum value. In FIgure 13b, normalization is the same for all panels and clearly shows the much larger power of lateral heterogeneity at the top and the bottom of the mantle.

In the top 300 km of the mantle, lateral heterogeneity is dominated by spherical harmonic degrees 2,4,5,6. Beyond degree 6, the power decays with increasing degree. The strong degree 1 component present in S12/WM13 at this depth range [see Figure 6a *Su et al.*, 1994] is absent here.

Between 300 and 500 km depths, the spectrum is dominated by degree 1. Although the radially integrated degree 1 component through the mantle is smaller for SAW12D than that for S12/WM13 due to much decreased total power in this depth range, it is still particularly interesting to know where this anomaly is compensated in the Earth so that the total contribution to the geoid is zero, as required by the choice of the geographic coordinate system. In Figure 14 we show the degree 1 component of S velocity anomaly at 400 km depth. It presents roughly an ocean-continent distri-

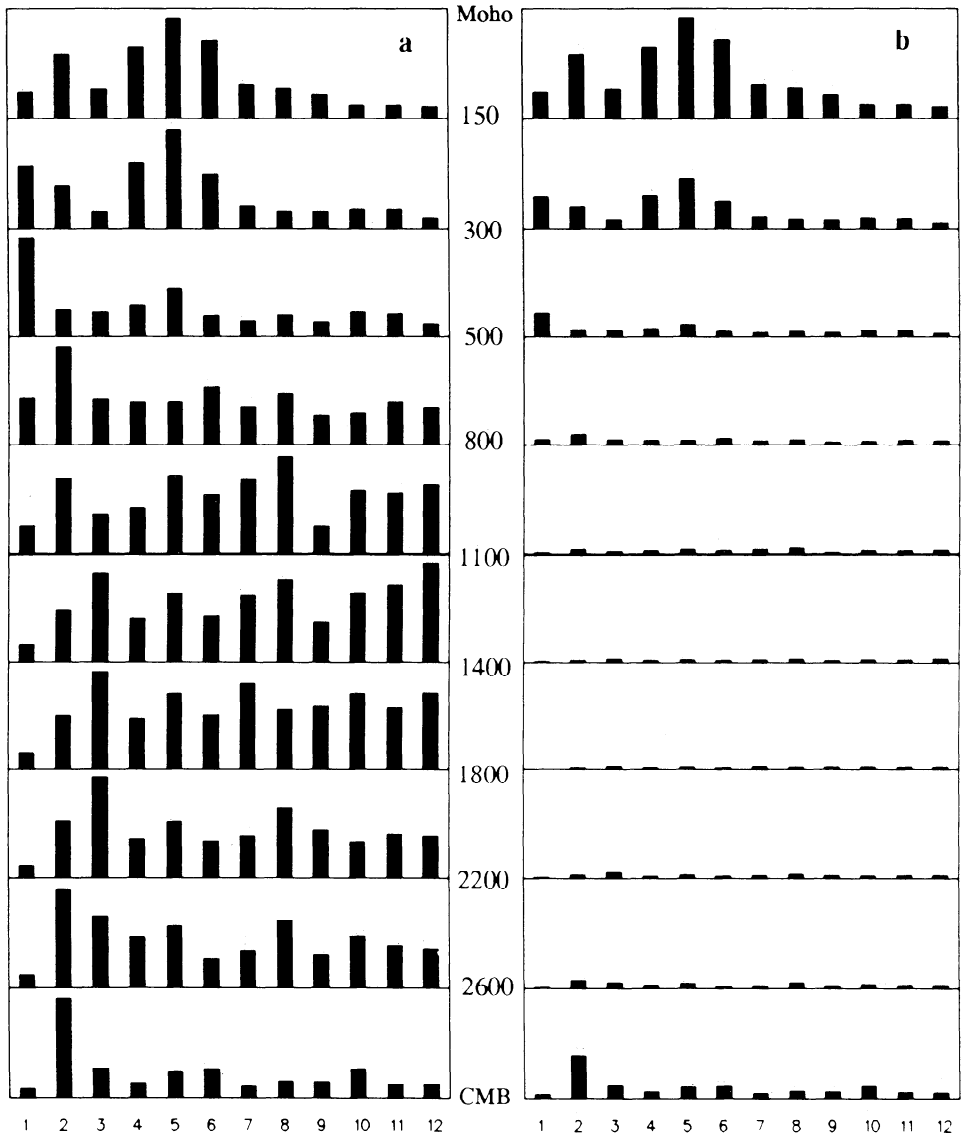

Figure 13. Power spectra of model SAW12D as a function of spherical harmonic degree. The spectra are vertically averaged over each layer, for each degree. (a) Spectra in each layer are normalized to their maximum value. (b) The normalization is the same for all layers.

bution. This suggests that the degree 1 component of density anomaly is not completely compensated within the crust and needs to be compensated by deeper structure in the upper mantle.

In the upper-lower mantle transition zone (500-800 km depth), lateral heterogeneity is dominated by degree 2. Among all the degree 2 components, the largest term is of order 0 (C_{20}). This C_{20} component in the transition zone, which is stronger than in some previous global models [e.g., *Dziewonski et al.*, 1993; *Su et al.*, 1994], accounts for, to a very large extent, the observed excess ellipticity of the non hydrostatic geoid, as will be explained in subsection on "excess ellipticity of geoid interpreted in a dynamic Earth".

In the lower mantle, below 1000 km, the spectrum is rather white until we reach the D'' region. In model S12/WM13 of *Su et al.* [1994], in contrast, the spectrum of lateral heterogeneity is white only down to about 2000 km depth, with the bottom 1000 km of the mantle characterized by a dominant degree 2 pattern. In SAW12D, the degree 2 pattern emerges strongly only at the very bottom of the mantle (see the bottom panel of Figure 13b). This reddening of the spectrum in the D'' region supports the hypothesis that the D'' region is, at least partly, a thermal boundary layer [*Jarvis and Peltier*, 1986]. The dominance of degree 2 in the depth range 500-800 km, although not as strong, may also suggest a thermal boundary layer component in this depth range.

It is interesting to point out that some of these spectral characteristics are in good agreement with the result from a recent study of numerical simulations of mantle convection [*Tackley et al.*, 1993], in which the upper mantle has a peak at degree 6 in the spectrum

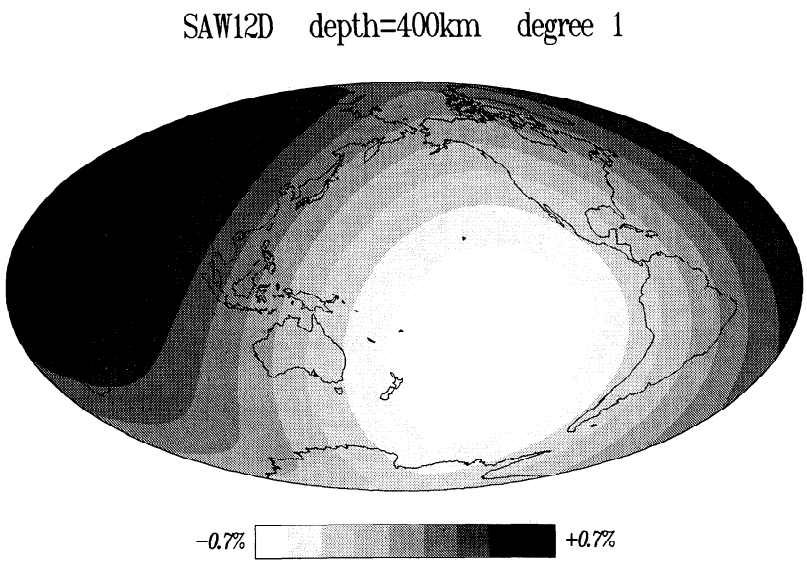

Figure 14. The distribution of degree 1 S velocity anomaly of SAW12D at 400 km depth.

of lateral heterogeneity, whereas the bottom 400 km of the mantle is dominated by degrees 2 and 3.

Mantle Lateral Heterogeneity and Geoid

Any seismic tomographic model has finite resolution and is subject to uncertainties. Generally speaking, we can resolve larger-scale (lower-degree) features better than smaller-scale (higher-degree) ones. In this study we attempt to resolve lateral heterogeneity up to spherical harmonic degree 12. It is important to assess whether high-degree components of the model are representative of the real Earth. One of the independent tests is to see how well we can predict the non hydrostatic geoid on the Earth's surface.

The relationship between seismic mantle lateral heterogeneity and the non-hydrostatic geoid has been studied by many authors [e.g., *Hager et al.*, 1985; *Forte et al.* 1993; *Corrieu et al.*, 1994]. The basic ideas involved in these studies are [*Richards and Hager*, 1984]: (1) the velocity anomalies from global tomographic models represent density anomalies in the mantle; and (2) the density anomalies in the mantle drive viscous flow and, in particular, introduce dynamic topography at the boundaries, which together with the internal density anomalies determines the geoid on the surface. Thus the prediction of the geoid depends not only on the seismic tomographic result but also on the viscosity structure in the mantle and the conversion factor between density and seismic anomalies.

As an approximation, the mantle viscosity η and the conversion factor c are assumed to be laterally homogeneous in the mantle. Namely,

$$\eta = \eta(r) \tag{17}$$

and

$$\delta\rho(r,\theta,\phi)/\rho(r) = c(r)\delta v_S(r,\theta,\phi)/v_S(r), \tag{18}$$

where $\rho(r)$ and $v_S(r)$ are the value of density and S velocity of the reference spherical model (PREM of *Dziewonski and Anderson* [1981]), respectively. Although numerous researchers [e.g., *Hager and Richards*, 1989; *Ricard and Vigny*, 1989; *Mitrovica and Peltier*, 1992; *Forte et al.*, 1993; *Corrieu et al.*, 1994] have inferred the profiles of $\eta(r)$ and $c(r)$ using various data and techniques, the uncertainty in these parameters is still very large.

In our experiment, we adopt a very simple model of $\eta(r)$ and $c(r)$. We use a three-layered viscosity model: constant viscosity η_{li} in a 100-km-thick lithosphere, constant viscosity η_{up} in the remaining upper mantle, and constant viscosity η_{lo} in the lower mantle. Since the synthetic geoid is independent of the absolute viscosity, we take $\eta_{up} = 1$. The lithosphere and the lower mantle are assumed to have higher viscosity: $\eta_{li} = \eta_{lo} = 30$. The conversion factor is assumed to increase from 0.3 in the whole upper mantle to 0.4 in the lower mantle. This profile of viscosity and conversion factor, also presented as dashed lines in Figure 15a, is within the range of those reported in the literature. With such a simple model of $\eta(r)$ and $c(r)$ and seismic model SAW12D, the synthetic geoid does not correlate very well with the observed one in the lowest degrees, implying that the overall model of viscosity and conversion factor needs to be adjusted (see the next subsection). For the higher degrees (6–12), for which the synthetic depends mainly upon the upper mantle structure, the synthetic geoid correlates with the observed one very well. In Figure 16, the darker shaded bars represent the significance level of the correlation coefficients, for each degree from 6 to 12, between synthetic and observed geoid. The high correlation indicates that the small-scale features of model SAW12D represent the real Earth structure in the upper mantle well. In Figure 16, we also show (lighter shaded bars) the correlation between the observed geoid and the a priori surface undulation model used in the inversion. The predicted geoid correlates with the observed

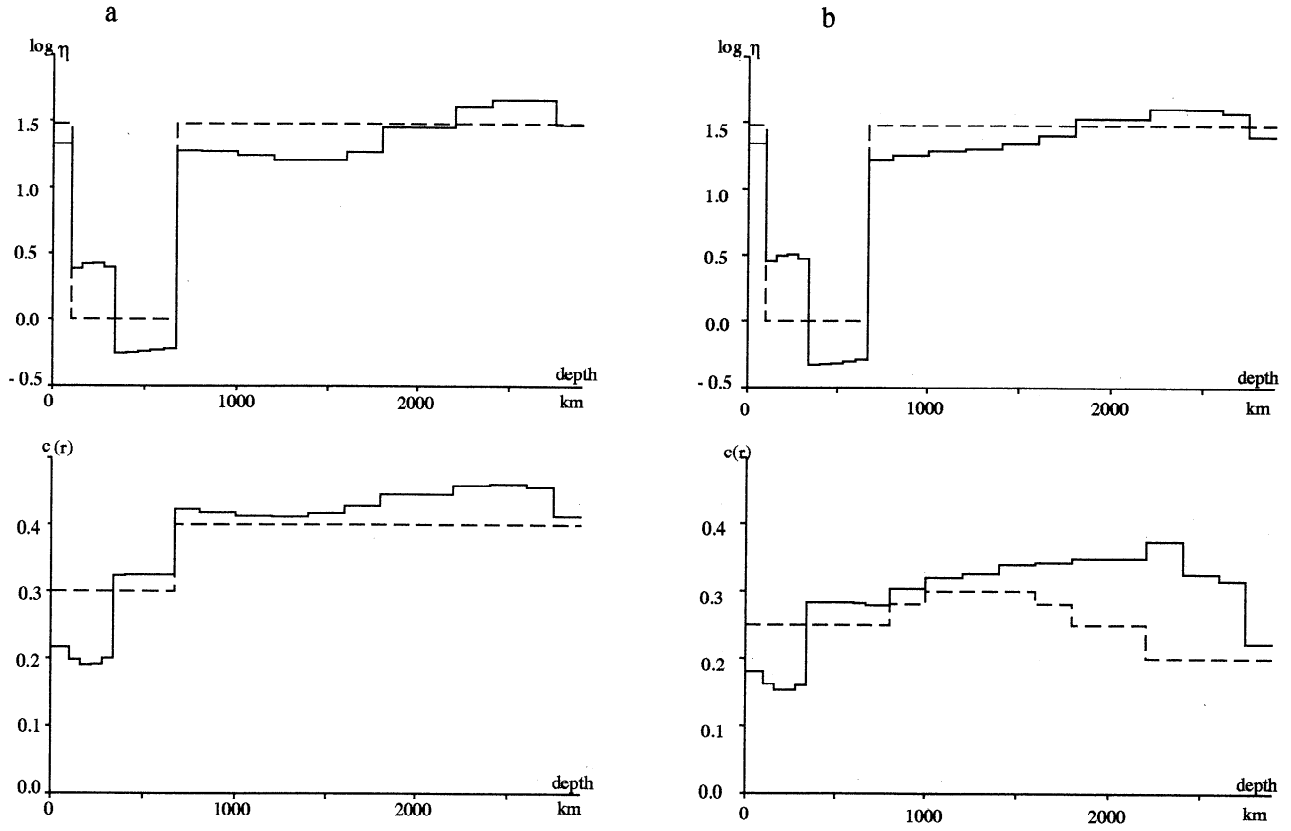

Figure 15. (top) Viscosity and (bottom) density conversion coefficient $c(r)$ profiles as a function of depth. Dashed lines represent starting models, and solid lines are the results after optimizing the fit to the observed geoid, using SAW12D. Two different starting models for $c(r)$ are tested: (a) similar to predictions from mineral physics experiments, and (b) similar to *Karato*'s [1993] predictions.

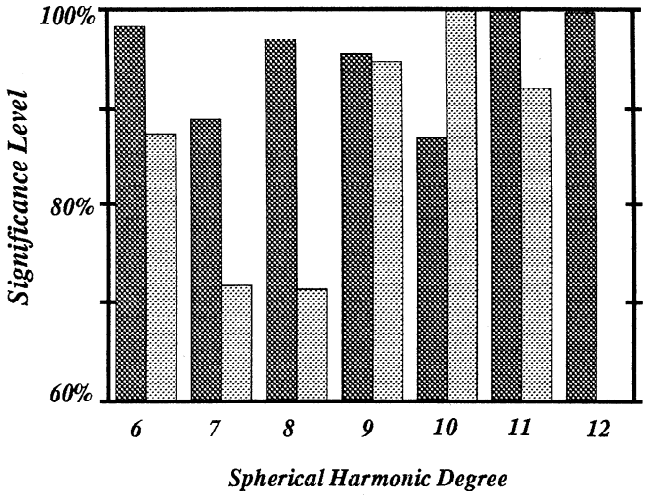

Figure 16. Correlation between the observed and predicted geoid for spherical harmonic degrees 6–12 (darker shaded bars). The prediction is done using the viscosity and density model as indicated by dotted lines in Figure 15a. The lighter shaded bars indicate correlation between the observed geoid and topography model. Correlation bars with a significance level lower than 60% are not shown.

one systematically better than does the surface undulation, and this is also true for the Moho model. This indicates that the geoid-related signal in SAW12D does not come merely from the crustal corrections.

Excess Ellipticity of Geoid Interpreted in a Dynamic Earth

As mentioned above, many authors have attempted to infer viscosity and velocity-density conversion factor in the mantle by optimizing the fit of the synthetic geoid to the observation. In particular, they found [e.g, *Forte*, 1989; *Ricard and Vigny*, 1989; *Forte et al.*, 1993] that it was difficult to simultaneously fit the degree 2 order 0 non hydrostatic geoid (excess geoid) and its other components, using previous seismic tomographic models. The source of the excess ellipticity is, in particular, of geodynamic interest: if the attempts to account for the excess ellipticity in terms of mantle heterogeneity should fail, other sources must be sought to explain it.

Here we attempt to predict the surface geoid from SAW12D by optimizing radial viscosity structure $\eta(r)$ and velocity–density conversion factor $c(r)$. Among others, *Corrieu et al.* [1994] have developed a procedure to directly invert for $\eta(r)$ and $c(r)$. We use this technique and a starting model of $\eta(r)$ and $c(r)$ as shown by

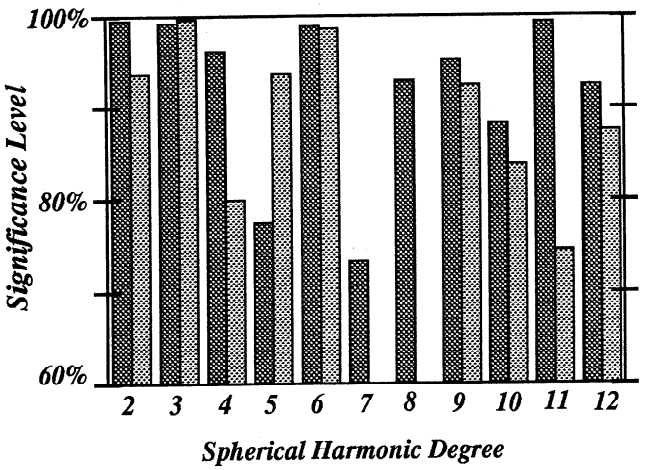

Figure 17. Correlation between the predicted and observed geoid. The darker shaded bars are for SAW12D and the lighter shaded bars are for S12/WM13. Correlation bars with a significance level lower than 60% are not shown.

the dotted lines in Figure 15a. After inversion for $\eta(r)$ and $c(r)$, the variance reduction for the observed geoid (degree 2 to degree 12) reaches 79%. Figure 17 gives degree-by-degree correlation between the predicted and observed geoid. In order to see how well we can predict the nonhydrostatic ellipticity of the geoid, Figure 18 shows the residual geoid for each azimuthal order of degrees 2 and 3, which are the dominant degrees in the geoid and account for 86% of the total power between degree 2 and 12. As a comparison, we repeat this experiment with the same starting model of viscosity and velocity-density conversion factor, replacing SAW12D with model S12/WM13 [*Su et al.*, 1994]. Here, the final variance reduction for the geoid after the optimizing $\eta(r)$ and $c(r)$ is 66% (the same experiment with model M12B12 [*Johnson et al.*, 1994] yields a variance reduction of 67%). The degree-by-degree correlation between observed and predicted geoid using S12/WM13 is shown as the lighter shaded bars in Figure 17. The residual geoid of degrees 2 and 3 for S12/WM13 is also shown in Figure 18. It is clear from Figure 18 that the degree 2 and order 0 term of the geoid (C_{20}), which is not well explained by S12/WM13 and other previous models, can be explained reasonably well using SAW12D. Figure 19 shows the C_{20} S velocity profile, as a function of depth, for model SAW12D, as compared to S12/WM13. Because the geoid Green's function at degree 2 is primarily sensitive to structure in the transition zone, we infer that the better fit to the geoid degree 2 structure is explained by the difference in relative values of the degree 2 coefficients and, in particular, C_{20}, in the depth range 500-900 km. Figure 20 shows a comparison of the stacked degree 2 velocity structure, in this depth range, for models SAW12D and S12/WM13. The white polar caps indicate that SAW12D has relatively more amplitude in C_{20} than S12/WM13, and this is largely what contributes to the geoid fit.

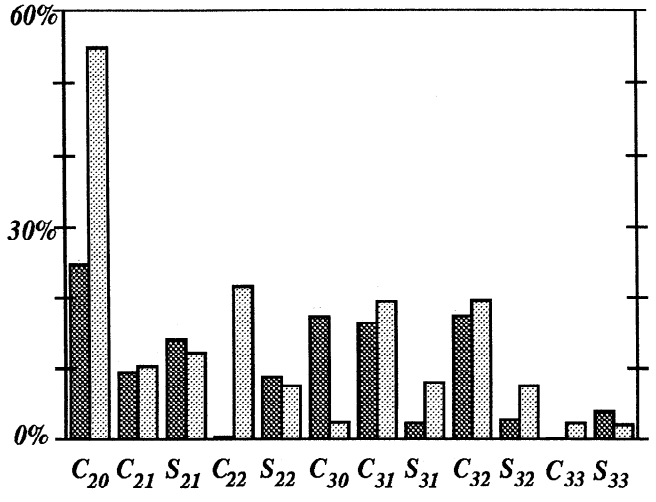

Figure 18. Residual geoid of degrees 2 and 3 after fit by predictions using SAW12D (darker shaded bars) and S12/WM13 (lighter shaded bars), respectively, using the starting viscosity and density conversion model shown by dotted lines in Figure 15a . The residuals are normalized by the observed C_{20} term.

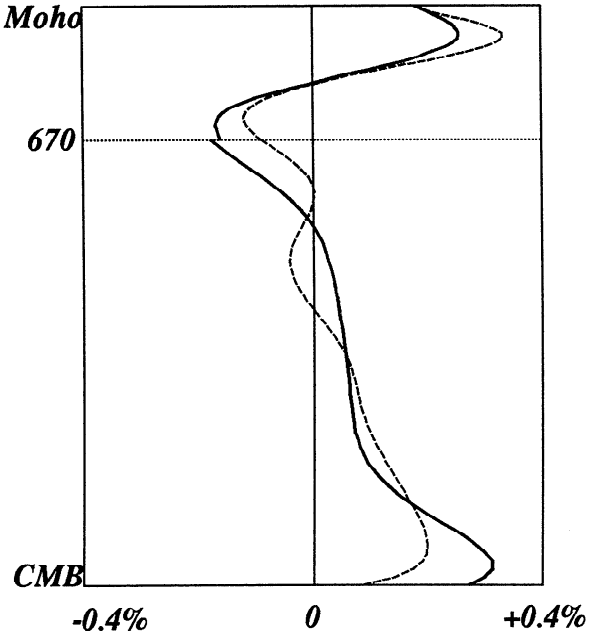

Figure 19. Depth profile of degree 2 and order 0 S velocity anomalies. The solid curve is for SAW12D, and the dashed curve is for S12/WM13.

We note that the inverse problem for viscosity and velocity-density conversion factor, using seismic tomographic models as constraints, is very non unique. The primary goal of the experiment described here is simply to demonstrate that the excess ellipticity and other terms of the surface geoid can be explained in terms of mantle lateral heterogeneity if appropriate $\eta(r)$ and $c(r)$ are used. The resulting radial profiles of viscosity and density conversion factor $c(r)$ can therefore only be discussed with caution. Having said that, we note nev-

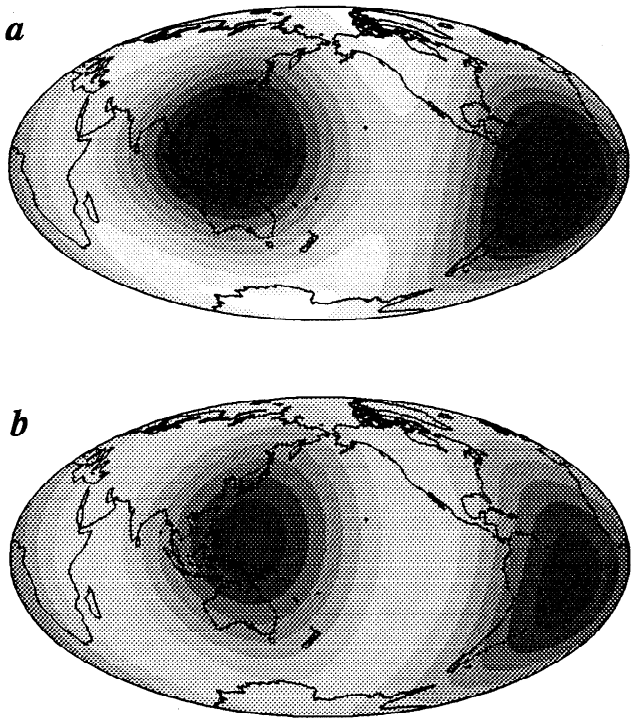

Figure 20. Degree 2 S velocity anomalies stacked between 500 and 900 km depths for models (a) SAW12D and (b) S12/WM13.

ertheless that the inverted viscosity profile (Figure 15a) exhibits a low-viscosity zone at the base of the upper mantle, in agreement with some recent geodynamical studies [e.g., *Forte et al.*, 1993; *Mitrovica*, 1995]. Also, the inverted density conversion factor remains high in the lower half of the lower mantle, contrary to the profile proposed by *Karato* [1993], which has a maximum around a depth of 1400 km. We have verified that this is not simply an artifact imposed by our starting solution: when using a model similar to Karato's as a starting model, the inversion produces a $c(r)$ profile which tends to flatten out the mid mantle maximum (see Figure 15b). Again, because of damping issues in the tomographic inversion, this observation is only tentative at this point.

Another set of observables for which SAW12D provides encouraging fits are plate velocities: 88% variance reduction for the poloidal part, compared to 66% for S12/WM13 (C. Lithgow-Bertelloni, personal communication,1995).

Conclusions

We have presented a new 3-D S velocity model of the Earth's whole mantle, which has been derived using SH waveforms. This model is generally in good agreement with other recent global mantle models, and there are indications, from comparison with more regional studies and independent geophysical observables, that its small-scale structure is somewhat better constrained. Fits to the geoid and to plate motions at the 80% level are particularly encouraging. Other whole mantle models are being derived using a combination of data, in particular travel time measurements. The reason why we can get away with only waveform data is due to the combination of two components of our approach which differ from that used in previous studies: (1) windowing of body wave energy packets, which allows us to weight down the highest-amplitude phases with respect to lower-amplitude ones that contribute to sampling of structure in an important way (e.g., S_{diff} in the lower mantle); and (2) the inclusion of cross-branch coupling terms in the theoretical formalism, which provides a more realistic description of the sensitivity of body waves to structure along the path.

While the computation time involved in our approach is somewhat larger than for the classical PAVA approximation, it is still within reasonable limits on current generation fast computer workstations. One of the advantages of this approach is that it avoids mixing of heterogeneous data sets, while maintaining good resolution power. The next step is to extend this approach to the P-SV case, which should increase the resolution further, and is the subject of our current investigation. It is also straightforward to extend this waveform modeling approach to include focusing and attenuation information contained in the amplitudes and obtain better constraints on the mantle anelastic structure [e.g., *Durek et al.*, 1993; *Romanowicz*, 1995].

Appendix A: Scheme to Evaluate Weighting Factors

In the following, we describe how we evaluate in practice each of the three factors w_e, w_n, and w_r in constructing the weighting w, defined for each wave packet as $w = w_e w_n w_r$ (see section on "data selection and weighting scheme" in main text).

First, w_e is introduced to characterize the error in the data. Since the absolute values of many systematic errors in waveform data are, to first order, scaled to the amplitude of the wave packet, we simply put $w_e = 1/\bar{d}$, where $\bar{d}$ is the rms amplitude of the wave packet. Errors falling into this category include those due to the uncertainty in the source parameters, unmodeled effects of off-great-circle focusing and defocusing, attenuation, and high-degree structure, etc. If an estimate of random observational noise were available, a more sophisticated formula could be employed. However, this may not be very crucial for this study since the data set has been carefully examined and wave packets with high noise level are excluded.

Factor w_n measures the redundancy among the data within a wave packet. Two extreme cases are that for case a, each datum provides completely independent information; and for case b, all the data can be represented by one of them and the rest of them are totally redundant. Denoting the number of data points of the wave packet by n, it is appropriate to set $w_n = 1$ for case a and $w_n = n^{-1}$ for case b. We here simply take the mid way of the two extremes and put $w_n = n^{-1/2}$.

Finally, we use w_r to quantify the redundancy of the whole wave packet with respect to wave packets sampling similar ray paths. We introduce a "correlation factor" c_{ij} for any given wave packet pair i and j. When two wave packets sample the same ray path, we set $c_{ij} = 1$ (thus $c_{ii} \equiv 1$). When two wave packets have "very different" ray paths, we let $c_{ij} = 0$. We give a recipe below to assign the correlation factor c_{ij}. A measurement of redundancy of wave packet i is $\sum_j c_{ij}$, where the summation is over all the wave packets (including i). However, $\sum_j c_{ij}$ may be an overestimate of the redundancy, since two wave packets with very similar paths but for different earthquakes may have independent errors. Thus we choose $w_r = (\sum_j c_{ij})^{-1/2}$ (note that $\sum_j c_{ij} \geq 1$). In evaluating c_{ij}, we take two factors into account: $c_{ij} = p_{ij} g_{ij}$. p_{ij} is either 1 or 0: if the two packets, i and j, correspond to the same seismic phase (say, both are S phases), $p_{ij} = 1$; otherwise, $p_{ij} = 0$ (e.g. one is S_{dif} and the other is SS phase). Factor g_{ij} $(0 \leq g_{ij} \leq 1)$ measures the geometrical relationship between the corresponding source-receiver pairs (i.e., it is defined for each path pair). Again $g_{ii} \equiv 1$ and $g_{ij} = 0$ unless the two paths have some redundancy:

$$g_{ij} = B(\frac{h(s_i, s_j)}{h_0}) B(\frac{h(r_i, r_j)}{h_0}) B(\frac{v(s_i, s_j)}{v_0}) \times B(\frac{v(r_i, r_j)}{v_0}) + B(\frac{h(s_i, r_j)}{h_0}) B(\frac{h(r_i, s_j)}{h_0}) \times B(\frac{v(s_i, r_j)}{v_0}) B(\frac{v(r_i, s_j)}{v_0}), \quad \text{(A1)}$$

where h_0 and v_0 are desired resolution lengths in horizontal and vertical directions, respectively (we set $h_0 \sim$ 2000 km and $v_0 \sim 150$ km in this study); $h(p_i, p_j)$ is the horizontal distance between point p_i and point p_j, with p being either a source s or receiver r (e.g., $h(s_i, r_j)$ is the horizontal distance between the source of ith seismogram and the receiver of the j seismogram); vertical distances $v(p_i, p_j)$ are defined similarly $(v(r_i, r_j) \equiv 0$, since all the receivers are on the surface); and the "cosine bell" function B is defined as

$$B(x) = \begin{cases} 0 & , \text{if } |x| > 1 \\ \cos(\frac{\pi}{2} x) & , \text{if } |x| \leq 1. \end{cases} \quad \text{(A2)}$$

In Figure A1, we give a 2-D illustration (for the case of $v(p_i, p_j) = 0$) on how the redundancy index g_{ij} works. The first term on the right side of (A1) is nonzero for case 3 in Figure 21; and the second term is nonzero for case 4. Usually, at least one of the two terms is zero.

It is interesting to point out that g_{ij} is also very useful in selecting data. For a candidate seismogram i, the value of $\sum_j c_{ij}$, where the summation is over all existing seismograms j, gives an indication on how valuable this candidate seismogram is in terms of providing independent information.

Appendix B: Partial Derivatives With Respect to Earthquake Location

Preliminaries

Let us start with equation (1) of paper 1, which gives an expression of an acceleration seismogram $u(t)$ based upon first-order perturbation theory of normal modes,

$$u(t) = \mathbf{R} \exp(i\mathbf{\Omega} t)\mathbf{S}, \quad \text{(B1)}$$

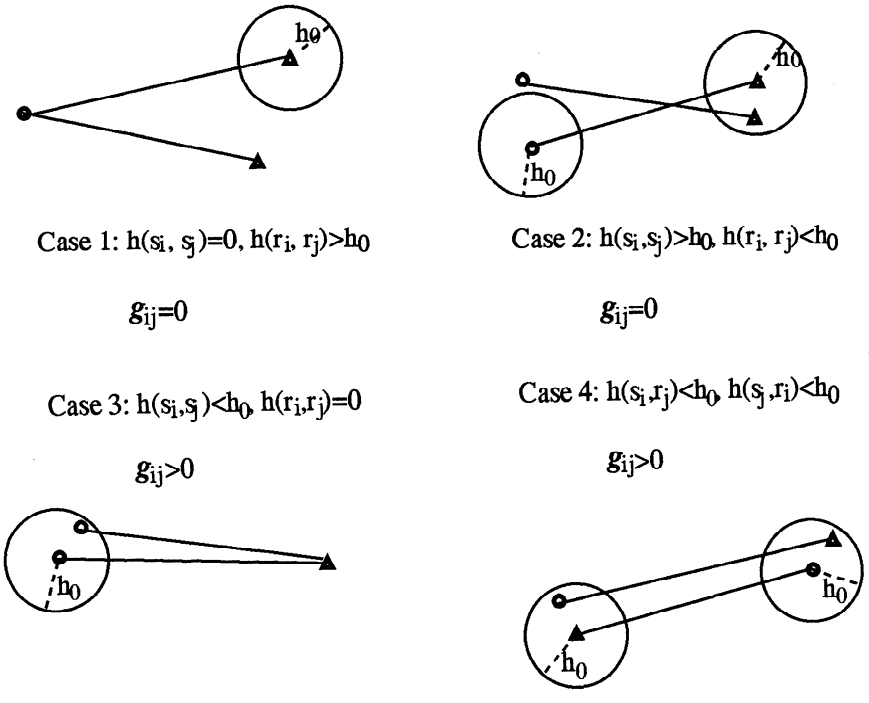

Figure A1. Two-dimensional schematic illustration on how redundancy index g_{ij} works. In each case, the small circles represent the sources of the events, and triangles are the receivers.

where the real part of the right side of the equation is understood, $\mathbf{R}$ and $\mathbf{S}$ are the receiver and source vectors, respectively, and $\mathbf{\Omega}$ represent the effect of the Earth structure.

The elements R_k^m of $\mathbf{R}$ represent receiver function of the (k,m)th singlet and are given by

$$R_k^m = \mathbf{v} \cdot \mathbf{u}_k^m, \tag{B2}$$

where $\mathbf{v}$ is the "instrument vector" [*Woodhouse and Girnius*, 1982] and $\mathbf{u}_k^m$ are the eigenfunctions of singlet (k,m) , which take the form

$$\begin{aligned}\mathbf{u}_k^m &= U(r_r)\hat{Y}_l^{0m}(\theta_r,\phi_r)\mathbf{e}_r + V(r_r)\nabla_1\hat{Y}_l^{0m}(\theta_r,\phi_r) \\ &- W(r_r)\mathbf{e}_r \times \nabla_1\hat{Y}_l^{0m}(\theta_r,\phi_r).\end{aligned} \tag{B3}$$

In (B3) U, V, W are known functions for a given reference spherically symmetric Earth model, (r_r, θ_r, ϕ_r) are spherical polar coordinates of the receiver, $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi)$ denote unit vectors in the coordinate directions, $\nabla_1 = \mathbf{e}_\theta\partial_\theta + cosec\theta\mathbf{e}_\phi\partial_\phi$ is the surface gradient operator. $\hat{Y}_l^{Nm}$ are fully normalized generalized spherical harmonics and related to those, Y_l^{Nm}, defined by *Phinney and Burridge* [1973] through $\hat{Y}_l^{Nm} = \sqrt{(2l+1)/4\pi}Y_l^{Nm}$.

Using the complex basis $\{\mathbf{e}_-, \mathbf{e}_0, \mathbf{e}_+\}$ of *Phinney and Burridge* [1973],

$$\mathbf{e}_- = \frac{1}{\sqrt{2}}(\mathbf{e}_\theta - i\mathbf{e}_\phi) \tag{B4}$$

$$\mathbf{e}_0 = \mathbf{e}_r \tag{B5}$$

$$\mathbf{e}_+ = \frac{1}{\sqrt{2}}(-\mathbf{e}_\theta - i\mathbf{e}_\phi) \tag{B6}$$

we may rewrite (B3)

$$\begin{aligned}\mathbf{u}_k^m &= U\hat{Y}_l^{0m}\mathbf{e}_0 + \Omega_0^l(V - iW)\hat{Y}_l^{-1m}\mathbf{e}_- \\ &+ \Omega_0^l(V + iW)\hat{Y}_l^{1m}\mathbf{e}_+,\end{aligned} \tag{B7}$$

where $\Omega_N^l = [\frac{1}{2}(l+N)(l-N+1)]^{1/2}$. Then we obtain equation (8) of *Woodhouse and Girnius* [1982] from (B2)

$$R_k^m = \sum_{N=-1}^{1} R_{kN}\hat{Y}_l^{Nm}, \tag{B8}$$

where R_{kN} are given in Table 1 of *Woodhouse and Girnius* [1982].

Now we look at the elements, S_k^m, of the source vector $\mathbf{S}$ in (B1). According to *Woodhouse and Girnius* [1982] we can write

$$S_k^m = \mathbf{E}_k^{m*} : \mathbf{M}, \tag{B9}$$

where $\mathbf{M}$ is the moment tensor of the earthquake and $\mathbf{E}_k^m$ is the strain tensor in the (k,m)th singlet. Note that (B9) has a sign difference from equation (3) of *Woodhouse and Girnius* [1982], because we use acceleration instead of displacement, which also requires the eigenvector $\mathbf{u}_k^m$ subject to the normalization

$$\int_V \rho_0 \mathbf{u}_k^{m*} \cdot \mathbf{u}_{k'}^{m'} dv = \delta_{kk'}\delta_{mm'} \tag{B10}$$

rather than equation (5) of *Woodhouse and Girnius* [1982].

In order to evaluate the strain tensor $\mathbf{E}_k^m \equiv \frac{1}{2}[\nabla\mathbf{u}_k^m + (\nabla\mathbf{u}_k^m)^T]$, we apply gradient operator ∇ on both sides of (B7)

$$\begin{aligned}\nabla\mathbf{u}_k^m &= \frac{1}{r}\Omega_{-1}^l\Omega_0^l(V - iW)\hat{Y}_l^{-2m}\mathbf{e}_-\mathbf{e}_- \\ &+ [\frac{1}{r}(U-V+iW)\mathbf{e}_-\mathbf{e}_0 + (\dot{V}-i\dot{W})\mathbf{e}_0\mathbf{e}_-]\Omega_0^l\hat{Y}_l^{-1m} \\ &+ \{\dot{U}\mathbf{e}_0\mathbf{e}_0 + \frac{1}{r}[-U + \Omega_{+1}^l\Omega_0^l(V - iW)]\mathbf{e}_+\mathbf{e}_- \\ &+ \frac{1}{r}[-U + \Omega_{+1}^l\Omega_0^l(V + iW)]\mathbf{e}_-\mathbf{e}_+\}\hat{Y}_l^{0m} \\ &+ [\frac{1}{r}(U-V-iW)\mathbf{e}_+\mathbf{e}_0 + (\dot{V}+i\dot{W})\mathbf{e}_0\mathbf{e}_+]\Omega_0^l\hat{Y}_l^{+1m} \\ &+ \frac{1}{r}\Omega_{-1}^l\Omega_0^l(V + iW)\hat{Y}_l^{+2m}\mathbf{e}_+\mathbf{e}_+,\end{aligned} \tag{B11}$$

where an overdot denotes differentiation with respect to r.

$$\begin{aligned}\mathbf{E} &= \frac{1}{r}\Omega_0^l\Omega_2^l(V + iW)^*\mathbf{e}_-\mathbf{e}_-\hat{Y}_l^{-2m} \\ &+ \frac{1}{2}\Omega_0^l(X + iZ)^*(\mathbf{e}_0\mathbf{e}_- + \mathbf{e}_-\mathbf{e}_0)\hat{Y}_l^{-1m} \\ &+ [\dot{U}\mathbf{e}_0\mathbf{e}_0 - \frac{F}{2}(\mathbf{e}_+\mathbf{e}_- + \mathbf{e}_-\mathbf{e}_+)]\hat{Y}_l^{0m} \\ &+ \frac{1}{2}\Omega_0^l(X - iZ)^*(\mathbf{e}_0\mathbf{e}_+ + \mathbf{e}_+\mathbf{e}_0)\hat{Y}_l^{1m} \\ &+ \frac{1}{r}\Omega_0^l\Omega_2^l(V - iW)^*\mathbf{e}_+\mathbf{e}_+\hat{Y}_l^{2m},\end{aligned} \tag{B12}$$

where

$$\begin{aligned}X &= \dot{V} + \frac{1}{r}(U - V) \\ Z &= \dot{W} - \frac{1}{r}W \\ F &= \frac{1}{r}[2U - l(l+1)V].\end{aligned} \tag{B13}$$

We may also write the (symmetric) moment tensor

$$\begin{aligned}\mathbf{M} &= M_{rr}\mathbf{e}_r\mathbf{e}_r + M_{\theta\theta}\mathbf{e}_\theta\mathbf{e}_\theta + M_{\phi\phi}\mathbf{e}_\phi\mathbf{e}_\phi \\ &+ M_{r\theta}(\mathbf{e}_r\mathbf{e}_\theta + \mathbf{e}_\theta\mathbf{e}_r) + M_{r\phi}(\mathbf{e}_r\mathbf{e}_\phi + \mathbf{e}_\phi\mathbf{e}_r) \\ &+ M_{\theta\phi}(\mathbf{e}_\theta\mathbf{e}_\phi + \mathbf{e}_\phi\mathbf{e}_\theta)\end{aligned} \tag{B14}$$

in terms of complex basis $\{\mathbf{e}_-, \mathbf{e}_0, \mathbf{e}_+\}$,

$$\begin{aligned}\mathbf{M} &= \frac{1}{2}(M_{\theta\theta} - M_{\phi\phi} + 2iM_{\theta\phi})\mathbf{e}_-\mathbf{e}_- \\ &+ \frac{1}{\sqrt{2}}(M_{r\theta} + iM_{r\phi})(\mathbf{e}_0\mathbf{e}_- + \mathbf{e}_-\mathbf{e}_0) \\ &+ M_{rr}\mathbf{e}_0\mathbf{e}_0 - \frac{1}{2}(M_{\theta\theta} + M_{\phi\phi})(\mathbf{e}_+\mathbf{e}_- + \mathbf{e}_-\mathbf{e}_+) \\ &+ \frac{1}{\sqrt{2}}(-M_{r\theta} + iM_{r\phi})(\mathbf{e}_0\mathbf{e}_+ + \mathbf{e}_+\mathbf{e}_0) \\ &+ \frac{1}{2}(M_{\theta\theta} - M_{\phi\phi} - 2iM_{\theta\phi})\mathbf{e}_+\mathbf{e}_+.\end{aligned} \tag{B15}$$

Substituting (B12) and (B15) into (B9), we obtain

equation (9) of *Woodhouse and Girnius* [1982]:

$$S_k^m = \sum_{N=-2}^{2} S_{kN} Y_l^{Nm*}, \tag{B16}$$

where S_{kN} are given in Table 1 of *Woodhouse and Girnius* [1982] with a sign difference as mentioned above.

Derivatives With Respect to Earthquake Location

Since the calculation of derivatives of seismograms with respect to r_r is relatively trivial, we shall only give expressions of $\nabla_1 u$, where the surface gradient operator ∇_1 acts on the source coordinates. Applying ∇_1 on both sides of (B1), we have

$$\nabla_1 u = \mathbf{R} \exp(i\Omega t) \nabla_1 \mathbf{S}. \tag{B17}$$

The element of $\nabla_1 \mathbf{S}$ is given, using (B9),

$$\nabla_1 S_k^m = (\nabla_1 \mathbf{E})^* : \mathbf{M}. \tag{B18}$$

Using

$$\begin{aligned}
r\nabla_1 \mathbf{E} &= \frac{1}{r}\Omega_0^l \Omega_2^l \Omega_{-2}^l (V+iW)^* \mathbf{e}_-\mathbf{e}_-\mathbf{e}_- \hat{Y}_l^{-3m} \\
&+ \Omega_0^l \Omega_2^l [\frac{1}{2}(X+iZ) - \frac{1}{r}(V+iW)]^* \\
&\quad \times \mathbf{e}_-(\mathbf{e}_-\mathbf{e}_0 + \mathbf{e}_0\mathbf{e}_-)\hat{Y}_l^{-2m} \\
&+ \Omega_0^l [\frac{1}{r}\Omega_2^l \Omega_2^l (V+iW) - (X+iZ)]^* \mathbf{e}_+\mathbf{e}_-\mathbf{e}_- \hat{Y}_l^{-1m} \\
&+ \Omega_0^l [\dot{U} - (X+iZ)]^* \mathbf{e}_-\mathbf{e}_0\mathbf{e}_0 \hat{Y}_l^{-1m} \\
&- \frac{1}{2}\Omega_0^l [F + (X+iZ)]^* \mathbf{e}_-(\mathbf{e}_-\mathbf{e}_+ + \mathbf{e}_+\mathbf{e}_-)\hat{Y}_l^{-1m} \\
&+ \frac{1}{2}[\Omega_0^l \Omega_1^l (X+iZ) + F - 2\dot{U}]^* \mathbf{e}_+(\mathbf{e}_-\mathbf{e}_0 + \mathbf{e}_0\mathbf{e}_-)\hat{Y}_l^{0m} \\
&+ \frac{1}{2}[\Omega_0^l \Omega_1^l (X-iZ) + F - 2\dot{U}]^* \mathbf{e}_-(\mathbf{e}_+\mathbf{e}_0 + \mathbf{e}_0\mathbf{e}_+)\hat{Y}_l^{0m} \\
&- \frac{1}{2}\Omega_0^l [F + (X-iZ)]^* \mathbf{e}_+(\mathbf{e}_-\mathbf{e}_+ + \mathbf{e}_+\mathbf{e}_-)\hat{Y}_l^{+1m} \\
&+ \Omega_0^l [\dot{U} - (X-iZ)]^* \mathbf{e}_+\mathbf{e}_0\mathbf{e}_0 \hat{Y}_l^{+1m} \\
&+ \Omega_0^l [\frac{1}{r}\Omega_2^l \Omega_2^l (V-iW) - (X-iZ)]^* \mathbf{e}_-\mathbf{e}_+\mathbf{e}_+ \hat{Y}_l^{+1m} \\
&+ \Omega_0^l \Omega_2^l [\frac{1}{2}(X-iZ) - \frac{1}{r}(V-iW)]^* \times \\
&\quad \mathbf{e}_+(\mathbf{e}_+\mathbf{e}_0 + \mathbf{e}_0\mathbf{e}_+)\hat{Y}_l^{+2m} \\
&+ \frac{1}{r}\Omega_0^l \Omega_2^l \Omega_{-2}^l (V-iW)^* \mathbf{e}_+\mathbf{e}_+\mathbf{e}_+ \hat{Y}_l^{+3m},
\end{aligned} \tag{B19}$$

we obtain

$$\nabla_1 S_k^m = \sum_{N=-3}^{3} \mathbf{D}_{kN} Y_l^{Nm*}, \tag{B20}$$

where

$$\begin{aligned}
\mathbf{D}_{k\pm3} &= -\frac{k_3}{r^2}(V \mp iW)(M_{\theta\theta} - M_{\phi\phi} \mp 2iM_{\theta\phi}) \\
&\quad \times(\pm\mathbf{e}_\theta - i\mathbf{e}_\phi)
\end{aligned} \tag{B21}$$

$$\begin{aligned}
\mathbf{D}_{k\pm2} &= \frac{k_2}{r}[(X \mp iZ) - \frac{2}{r}(V \mp iW)](\pm M_{r\theta} - iM_{r\phi}) \\
&\quad \times(\pm\mathbf{e}_\theta - i\mathbf{e}_\phi)
\end{aligned} \tag{B22}$$

$$\begin{aligned}
\mathbf{D}_{k\pm1} &= -\frac{k_1}{2r}[\frac{L_1}{r}(V \mp iW) - (X \mp iZ)] \\
&\quad \times(M_{\theta\theta} - M_{\phi\phi} \mp 2iM_{\theta\phi})(\mp\mathbf{e}_\theta - i\mathbf{e}_\phi) \\
&\quad - \frac{k_1}{r}\left\{[\dot{U} - (X \mp iZ)]M_{rr} + \frac{1}{2}[F + (X \mp iZ)] \times \right. \\
&\quad \left. (M_{\theta\theta} + M_{\phi\phi})\right\}(\pm\mathbf{e}_\theta - i\mathbf{e}_\phi)
\end{aligned} \tag{B23}$$

$$\begin{aligned}
\mathbf{D}_{k0} &= \frac{k_0}{2r}[L_0(X+iZ) + F - 2\dot{U}](-M_{r\theta} - iM_{r\phi}) \\
&\quad \times(\mathbf{e}_\theta - i\mathbf{e}_\phi) + \frac{k_0}{2r}[L_0(X-iZ) + F - 2\dot{U}] \\
&\quad \times(M_{r\theta} - iM_{r\phi})(-\mathbf{e}_\theta - i\mathbf{e}_\phi)
\end{aligned} \tag{B24}$$

with $L_1 = (l+2)(l-1)/2$, $L_0 = l(l+1)/2$ and k_n being defined by *Woodhouse and Girnius* [1982]:

$$k_n = \frac{1}{2^n}\left[\frac{2l+1}{4\pi}\frac{(l+n)!}{(l-n)!}\right]^{1/2}. \tag{B25}$$

The formal similarity between (B16) and (B20) suggests a simple way to obtaining the NACT expressions of $\nabla_1 u$. Namely, $\nabla_1 u$ takes the same form of equation (10) of paper 1 with the following substitutions. On the left side of equation $u \longrightarrow \nabla_1 u$; and on the right side of equation: $A_k \longrightarrow \mathbf{A}_k$, $T_{kk'}^{(1)} \longrightarrow \mathbf{T}_{kk'}^{(1)}$, and $T_{kk'}^{(2)} \longrightarrow \mathbf{T}_{kk'}^{(2)}$, where $T_{kk'}^{(1)}$ and $T_{kk'}^{(2)}$ are defined in equations (A3) and (A4) of paper 1 and

$$\mathbf{A}_k = \sum_m R_k^m \nabla_1 S_k^m \tag{B26}$$

$$\mathbf{T}_{kk'}^{(1)} = \sum_{N=-1}^{1}\sum_{M=-3}^{3} i^{N+M} R_{kN} \mathbf{D}_{k'M} \cos\left(\frac{N+M}{2}\pi\right) \tag{B27}$$

$$\mathbf{T}_{kk'}^{(2)} = \sum_{N=-1}^{1}\sum_{M=-3}^{3} i^{N+M} R_{kN} \mathbf{D}_{k'M} \sin\left(\frac{N+M}{2}\pi\right). \tag{B28}$$

It is important to notice that (B27) and (B28) are valid in the great circle coordinates [*Li and Tanimoto*, 1993]. Finally, we note that although $\mathbf{A}_k = \nabla_1 A_k$ is true, the equations $\mathbf{T}_{kk'}^{(1)} = \nabla_1 T_{kk'}^{(1)}$ and $\mathbf{T}_{kk'}^{(2)} = \nabla_1 T_{kk'}^{(2)}$ do not hold in general.

Appendix C: Correction for Hydrostatic Ellipticity

In principle, the lateral heterogeneity due to hydrostatic ellipticity can be treated as general lateral heterogeneity of angular degree 2 and azimuthal order 0, and its contribution to a seismogram can be calculated in the same way as nonhydrostatic lateral heterogeneity. For the particular reference spherical Earth model used in this study (PREM of *Dziewonski and Ander-*

son [1981]), however, special treatments are needed for the ellipticity effect. *Woodhouse* [1980] gave a recipe for calculating the coupling effect due to the perturbations in Voigt bulk modulus, shear modulus, and density. PREM, however, is an anisotropic (transversely isotropic) model with five independent elastic parameters A, C, L, N, F [*Love*, 1927]. Consequently, the hydrostatic ellipticity causes lateral heterogeneity in these parameters. Thus we need expressions for calculating coupling effects due to the perturbations in A, C, L, N, F, as well as in density.

Generalizing formulae of *Woodhouse* [1980] to the transversely isotropic case and using his notation, we can write for $Z^{kk'}_{mm'}$, as used in equation (20) of *Li and Tanimoto* [1993],

$$Z^{kk'}_{mm'} = \sum_{st}[(2l+1)(2l'+1)(2s+1)/(4\pi)]^{1/2} \times (-1)^{m'}\begin{pmatrix} l' & s & l \\ -m' & t & m \end{pmatrix} \times \left\{\int_0^a [\delta z_{st}]\, r^2 dr - \sum_d r_d^2 h_d^{st}\,[z_s]^+_-\right\}, \quad \text{(C1)}$$

where

$$\delta z_{st} = \delta A_{st}\bar{A}_s + \delta C_{st}\bar{C}_s + \delta L_{st}\bar{L}_s + \delta N_{st}\bar{N}_s + \delta F_{st}\bar{F}_s + \delta\rho_{st}R_s^{(2)} \quad \text{(C2)}$$

and

$$z_s = A\tilde{A}_s + C\tilde{C}_s + L\tilde{L}_s + N\tilde{N}_s + F\tilde{F}_s + \rho_0 R_s^{(1)} \quad \text{(C3)}$$

where the effect of the Earth rotation (terms associated with the rotation rate Ω in equation (A17) of *Woodhouse* [1980]) is omitted for simplicity and can be added back in as needed, kernels $R_s^{(1)}$ and $R_s^{(2)}$ are defined in *Woodhouse* [1980], $\delta A_{st}, \delta C_{st}, \delta L_{st}, \delta N_{st}, \delta F_{st}$ are the coefficients in the spherical harmonic expansion of heterogeneity in the material parameters of transverse isotropy, A, C, L, N, F are the associated kernels $\bar{A}_s, \bar{C}_s, \bar{L}_s, \bar{N}_s, \bar{F}_s$; and $\tilde{A}_s, \tilde{C}_s, \tilde{L}_s, \tilde{N}_s, \tilde{F}_s$ are given by

$$\bar{A}_s = ff' B^{(0)+}_{l'sl} \quad \text{(C4)}$$

$$\bar{C}_s = \dot{U}\dot{U}' B^{(0)+}_{l'sl} \quad \text{(C5)}$$

$$\bar{L}_s = (XX' + ZZ')B^{(1)+}_{l'sl} + (ZX' - XZ')iB^{(1)-}_{l'sl} \quad \text{(C6)}$$

$$\bar{N}_s = -ff' B^{(0)+}_{l'sl} + \frac{1}{r^2}(VV' + WW')B^{(2)+}_{l'sl} + \frac{1}{r^2}(WV' - VW')iB^{(2)-}_{l'sl} \quad \text{(C7)}$$

$$\bar{F}_s = [\dot{U}f' + \dot{U}'f]B^{(0)+}_{l'sl} \quad \text{(C8)}$$

$$\tilde{A}_s = ff' B^{(0)+}_{l'sl} \quad \text{(C9)}$$

$$\tilde{C}_s = -\dot{U}\dot{U}' B^{(0)+}_{l'sl} + \frac{1}{r}[V\dot{U}' B^{(1)+}_{ll's} + V'\dot{U} B^{(1)+}_{l'ls}] + \frac{1}{r}[\dot{U}W' - \dot{U}'W]iB^{(1)-}_{l'sl} \quad \text{(C10)}$$

$$\tilde{L}_s = [XX' + ZZ' - \dot{V}X' - \dot{V}'X - \dot{W}Z' - \dot{W}'Z]B^{(1)+}_{l'sl} + [ZX' - XZ' + \dot{V}Z' - \dot{V}'Z + X\dot{W}' - \dot{W}X']iB^{(1)-}_{l'sl} \quad \text{(C11)}$$

$$\tilde{N}_s = -ff' B^{(0)+}_{l'sl} + \frac{1}{r^2}(VV' + WW')B^{(2)+}_{l'sl} + \frac{1}{r^2}(WV' - VW')iB^{(2)-}_{l'sl} \quad \text{(C12)}$$

$$\tilde{F}_s = \frac{1}{r}(Vf' B^{(1)+}_{ll's} + V'fB^{(1)+}_{l'ls}) + \frac{1}{r}(fW' - Wf')iB^{(1)-}_{l'sl} \quad \text{(C13)}$$

with

$$f = \frac{1}{r}[2U - l(l+1)V] \quad \text{(C14)}$$

$$f' = \frac{1}{r}[2U' - l'(l'+1)V'] \quad \text{(C15)}$$

and

$$B^{(N)\pm}_{l'sl} = \frac{1}{2}[1 \pm (-1)^{l'+s+l}]\left[\frac{(l'+N)!(l+N)!}{(l'-N)!(l-N)!}\right]^{1/2} \times (-1)^N \begin{pmatrix} l' & s & l \\ -N & 0 & N \end{pmatrix}. \quad \text{(C16)}$$

For the asymptotic case $l >> s$ and $l' >> s$, which is assumed in NACT [*Li and Tanimoto*, 1993], we have

$$B^{(N)\pm}_{l'sl} = 0, \text{ if } N = 1, 2 \quad \text{(C17)}$$

and we may rewrite (C1) in the form of

$$Z^{kk'}_{mm'} = \int \delta\omega^2_{kk'}(\theta,\phi)Y_l^{m*}(\theta,\phi)Y_{l'}^{m'}(\theta,\phi)d\Omega. \quad \text{(C18)}$$

where

$$\delta\omega^2_{kk'} = \sum_{st}\left\{\int_0^a [\delta z_{st}]r^2 dr - \sum_d r_d^2 h_d^{st}[z_s]^+_-\right\}. \quad \text{(C19)}$$

The effect of the hydrostatic ellipticity is given by [*Dahlen*, 1968]

$$\delta A_{st} = \delta_{s2}\delta_{t0}\sqrt{\frac{4\pi}{5}}\cdot\frac{2}{3}r\epsilon\dot{A} \quad \text{(C20)}$$

$$\delta C_{st} = \delta_{s2}\delta_{t0}\sqrt{\frac{4\pi}{5}}\cdot\frac{2}{3}r\epsilon\dot{C} \quad \text{(C21)}$$

$$\delta L_{st} = \delta_{s2}\delta_{t0}\sqrt{\frac{4\pi}{5}}\cdot\frac{2}{3}r\epsilon\dot{L} \quad \text{(C22)}$$

$$\delta N_{st} = \delta_{s2}\delta_{t0}\sqrt{\frac{4\pi}{5}}\cdot\frac{2}{3}r\epsilon\dot{N} \quad \text{(C23)}$$

$$\delta F_{st} = \delta_{s2}\delta_{t0}\sqrt{\frac{4\pi}{5}} \cdot \frac{2}{3} r\epsilon \dot{F} \tag{C24}$$

$$\delta \rho_{st} = \delta_{s2}\delta_{t0}\sqrt{\frac{4\pi}{5}} \cdot \frac{2}{3} r\epsilon \dot{\rho_0} \tag{C25}$$

$$h_d^{st} = -\delta_{s2}\delta_{t0}\sqrt{\frac{4\pi}{5}} \cdot \frac{2}{3} r\epsilon \tag{C26}$$

$$\delta \phi_{st} = \delta_{s2}\delta_{t0}\sqrt{\frac{4\pi}{5}} \cdot \frac{2}{3}[r\epsilon g_0 - \frac{1}{2}\Omega^2 r^2]. \tag{C27}$$

where $\epsilon = \epsilon(r)$ is the hydrostatic ellipticity, the solution of Clairaut's equation [*Jeffreys*, 1970].

Appendix D: Equivalence of Legendre and Chebyshev Polynomials for Inverse Problems

For a given non negative integer N, we define $\mathcal{P}_N[-1, 1]$ to be the function space spanned by the polynomial set $\{r^0, r^1, \ldots, r^N; r \in [-1, 1]\}$. Both Chebyshev polynomial set $\{b_n(r); r \in [-1, 1], n = 0, 1, 2, \ldots, N\}$ and Legendre polynomial set $\{b'_n(r); r \in [-1, 1], n = 0, 1, 2, \ldots, N\}$ [e.g., *Abramowitz and Stegun*, 1965] are complete sets of $\mathcal{P}_N[-1, 1]$. Thus there exists an $N \times N$ invertible mapping matrix $\mathbf{M}$ such that

$$\mathbf{b}'(r) = \mathbf{M}\mathbf{b}(r) \tag{D1}$$

and

$$\mathbf{b}(r) = \mathbf{M}^{-1}\mathbf{b}'(r), \tag{D2}$$

where the nth ($0 \le n \le N$) element of function vector $\mathbf{b}(r)$ is $b_n(r)$ and the nth element of function vector $\mathbf{b}'(r)$ is $b'_n(r)$.

For any $p(r) \in \mathcal{P}_N[-1, 1]$, we may expand it in terms of both $\{b_n(r)\}$ and $\{b'_n(r)\}$:

$$\begin{aligned} p(r) &= \sum_{n=0}^{N} p_n b_n(r) = \mathbf{p}^T \mathbf{b}(r) \\ &= \sum_{n=0}^{N} p'_n b'_n(r) = \mathbf{p}'^T \mathbf{b}'(r), \end{aligned} \tag{D3}$$

where $\mathbf{p}$ and $\mathbf{p}'$ denote the coefficient vector (of dimension N) of Chebyshev expansion and Legendre expansion, respectively and the superscript T denotes matrix transpose. By virtue of (D1) and (D2), we have

$$\mathbf{p} = \mathbf{M}^T \mathbf{p}' \tag{D4}$$

and

$$\mathbf{p}' = (\mathbf{M}^{-1})^T \mathbf{p}. \tag{D5}$$

Let us expand $h(r) \in \mathcal{P}_N[-1, 1]$ in terms of $\{b_n(r)\}$ with the coefficients $\mathbf{h}$ and $g(r) \in \mathcal{P}_N[-1, 1]$ in terms of $\{b'_n(r)\}$ with the coefficients $\mathbf{g}'$. If $\mathbf{h}$ and $\mathbf{g}'$ satisfy

$$\mathbf{h} = \mathbf{M}^T \mathbf{g}' \tag{D6}$$

or

$$\mathbf{g}' = (\mathbf{M}^{-1})^T \mathbf{h}, \tag{D7}$$

it is easy to prove that $h(x) = g(x)$.

Now suppose that we have two solutions to the inverse problem, $y(r) \in \mathcal{P}_N[-1, 1]$ and $z(r) \in \mathcal{P}_N[-1, 1]$, obtained by iterative application of the recursion (4), where $y(r)$ is represented by its Chebyshev expansion coefficients $\mathbf{y}$ and $z(r)$ is represented by its Legendre expansion coefficients $\mathbf{z}'$. We want to demonstrate that $y(r) = z(r)$ if we have the same a priori constraint.

The a priori constraint is given by the specification of the a priori expected model and its probability distribution [*Tarantola and Valette*, 1982]. Suppose the a priori expected model $y_0(r) = z_0(r)$ has a coefficient vector $\mathbf{y}_0$ for Chebyshev expansion, then its coefficient vector for Legendre expansion is

$$\mathbf{z}'_0 = (\mathbf{M}^{-1})^T \mathbf{y}_0. \tag{D8}$$

The probability distribution of the a priori model is given by the covariance matrix, which is defined, using Chebyshev expansion, as

$$\mathbf{C}_y = < \delta\mathbf{y}_0 \delta\mathbf{y}_0^T >, \tag{D9}$$

where the angle brackets denote the expectation value of the enclosed quantity. When Legendre expansion is used, we have

$$\begin{aligned} \mathbf{C}'_z &= < \delta\mathbf{z}'_0 \delta\mathbf{z}'^T_0 > \\ &= (\mathbf{M}^{-1})^T < \delta\mathbf{y}_0 \delta\mathbf{y}_0^T > \mathbf{M}^{-1} \\ &= (\mathbf{M}^{-1})^T \mathbf{C}_y \mathbf{M}^{-1}. \end{aligned} \tag{D10}$$

We demonstrate below, by induction, that $z(r) = y(r)$ (i.e., $\mathbf{z}'_\infty = (\mathbf{M}^{-1})^T \mathbf{y}_\infty$, where the subscripts indicate the iteration number). Since we already have $\mathbf{z}'_0 = (\mathbf{M}^{-1})^T \mathbf{y}_0$, we need only prove that $\mathbf{z}'_{i+1} = (\mathbf{M}^{-1})^T \mathbf{y}_{i+1}$ for $\mathbf{z}'_i = (\mathbf{M}^{-1})^T \mathbf{y}_i$ given. From (4), we have

$$\begin{aligned} \mathbf{z}'_{i+1} &= \mathbf{z}'_i + (\mathbf{A}'^T_i \mathbf{C}_d^{-1} \mathbf{A}'_i + \mathbf{C}'^{-1}_z)^{-1} \\ &\times [\mathbf{A}'^T_i \mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{f}(\mathbf{z}'_i)) - \mathbf{C}'^{-1}_z(\mathbf{z}'_i - \mathbf{z}'_0)], \end{aligned} \tag{D11}$$

where $\mathbf{A}'_i$ is the matrix of partial derivatives

$$\mathbf{A}'_i = \left[\frac{\partial \mathbf{f}(\mathbf{z}')}{\partial \mathbf{z}'}\right]_{\mathbf{z}'=\mathbf{z}'_i} = \left[\frac{\partial \mathbf{f}(\mathbf{y})}{\partial \mathbf{y}}\right]_{\mathbf{y}=\mathbf{y}_i} \left[\frac{d\mathbf{y}_i}{d\mathbf{z}'_i}\right] = \mathbf{A}_i \mathbf{M}^T \tag{D12}$$

with

$$\mathbf{A}_i \equiv \left[\frac{\partial \mathbf{f}(\mathbf{y})}{\partial \mathbf{y}}\right]_{\mathbf{y}=\mathbf{y}_i}. \tag{D13}$$

Thus, after some algebra, using (D8) and (D10), (D11) becomes:

$$\begin{aligned} \mathbf{z}'_{i+1} &= (\mathbf{M}^{-1})^T \{\mathbf{y}_i + (\mathbf{A}_i^T \mathbf{C}_d^{-1} \mathbf{A}_i + \mathbf{C}_y^{-1})^{-1} \\ &\times [\mathbf{A}_i^T \mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{f}(\mathbf{y}_i)) - \mathbf{C}_y^{-1}(\mathbf{y}_i - \mathbf{y}_0)]\} \\ &= (\mathbf{M}^{-1})^T \mathbf{y}_{i+1}. \end{aligned} \tag{D14}$$

In summary, the solution to the inverse problem is independent of whether we choose Chebyshev polynomials and Legendre polynomials as basis functions, if we have the same a priori constraint on the solution. By "the same a priori constraint," we mean that it is the

same function vector in $\mathcal{P}_N[-1,1]$ with the same probability distribution. However, the explicit coefficient expression of the same function in terms of different basis functions is different and subject to transforms, such as (D8). The coefficient expression of the same probability distribution, the covariance matrix, is subject to transformations such as (D10).

Acknowledgments. We thank Bob Woodward for providing his travel time data set and the corresponding software and Adam Dziewonski for helpful suggestions on merging the waveform and travel time data sets. We are very grateful to Wei-jia Su and Bob Woodward for their help in our reading digital seismic data from Harvard Seismic Archive Facility. We thank the Harvard and Scripps groups for making their models available to us in digital form. Suggestions by L. Zhao and another anonymous reviewer have helped improve the manuscript. This research was partially supported by NSF grant EAR-9304492. This Seismographic Station contribution 96-06. Model SAW12D is available in digital form upon request.

References

Abramowitz, M., and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, Mineola, N.Y., 1965.

Aki, K. and P. G. Richards, *Quantitative Seismology: Theory and Methods*, W. H. Freeman, New York, 1980.

Bunge, H.P., M. A. Richards, and G. Glatzmeier, Effect of depth-dependent viscosity on the platform of 3-D spherical mantle convection, *Nature,379*,436-438, 1996.

Corrieu, V., Y. Ricard and C. Froidevaux, Converting mantle tomography into mass anomalies to predict the Earth's radial viscosity, *Phys. Earth Planet. Inter.*, *84*, 3-13, 1994.

Dahlen, F. A., The normal modes of a rotating, elliptical earth, *Geophys J. R. Astron. Soc.*, *16*, 329-367, 1968.

Durek, J. J., M. H. Ritzwoller and J. H. Woodhouse, Constraining upper mantle anelasticity using surface wave amplitude anomalies, *Geophys J. Int.*, *114*, 249-272, 1993.

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, 1984.

Dziewonski, A. M., and D. L. Anderson, Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, *25*, 297-356, 1981.

Dziewonski, A. M. and R. L. Woodward, Acoustic imaging at the planetary scale, in *Acoustical Imaging*, vol. 19, pp. 785-797, edited by H. Ermert and H. P. Harjes, Plenum, New York, 1992.

Dziewonski, A.M., B. H. Hager, and R. J. O'Connell, large-scale heterogeneities in the lower mantle, *J. Geophys. Res.*, *82*, 239-255, 1977.

Dziewonski, A. M., G. Chou, and J. H. Woodhouse, Determination of earthquake source parameters form waveform modelling, *J. Geophys. Res.*, *86*, 2825-2852, 1981.

Dziewonski, A. M., A.M. Forte, W.-J. Su, and R. L. Woodward, Seismic tomography and geodynamics, in *Relating Geophysical Structures and Process: Proceedings of the Jeffreys Symposium, Geophys. Monogr. Ser.*, vol. 76, edited by K. Aki and R. Dmowska, pp. 67-105, AGU, Washington, D.C., 1993.

Ekstrom, G., J. Tromp, and A. M. Dziewonski, Measurements and models of global surface wave propagation, *Seism. Res. Lett.*, *65*, 19, 1995.

Forte, A. M., Mantle convection and global geophysical observables, Ph.D. thesis, Univ. of Toronto, Toronto, Ont., 1989.

Forte, A., A. M. Dziewonski, and R. L. Woodward, Aspherical structure of the mantle, tectonic plate motions, nonhydrostatic geoid, and topography of the core-mantle boundary, in *Dynamics of the Earth's Deep Interior and Earth Rotation, Geophys. Monogr. Ser.*, vol. 72, edited by J.-L. Le Mouel, D. E. Smylie, and T. Herring, pp. 135-166, AGU, Washington, D.C., 1993.

Geller, R. J., and T. Hara, Two efficient algorithms for iterative linearized inversion of seismic waveform data, *Geophys. J. Int. 115*, 699-710, 1993.

Grand, S., Mantle shear structure beneath the Americas and surrounding oceans, *J. Geophys. Res.*, *99*, 11,591-11,621, 1994.

Hager, B. H., and M. Richards, Long-wavelength variations in the Earth's geoid: Physical models and dynamical implications, *Philos. Trans. R. Soc. London A*, *328*, 309-327, 1989.

Hager, B., R. W. Clayton, M. A. Richards, R. P. Comer and A. M. Dziewonski, Lower mantle heterogeneity, dynamic topography and the geoid, *Nature*, *313*, 541-545, 1985.

Inoue, H., Y. Fukao, K. Tanabe, and Y. Ogata, Whole mantle it P wave travel time tomography, *Phys. Earth Planet. Inter.*, *5*, 294-328, 1989.

Jackson, D. D., The use of 'a priori' data to resolve nonuniqueness in linear inversion, *Geophys. J. R. Astron. Soc.*, *57*, 137-157, 1979.

Jarvis, G. T., and W. R. Peltier, Lateral heterogeneity in the convecting mantle, *J. Geophys. Res.*, *91*, 435-451, 1986.

Jeffreys, H., *The Earth*, 5th ed., pp. 183-192, Cambridge Univ. Press, New York, 1970.

Johnson, S, P. Shearer, and G. Masters, A shear velocity model of the mantle, *Eos Trans. AGU, 75 (44)*, Fall Meet. Suppl. 475, 1994.

Jordan, T. H., The continental tectosphere, *Rev. Geophys.*, *13*, 1-12, 1975.

Jordan, T. H., Composition and development of the continental tectosphere, *Nature*, *274*, 544-548, 1978.

Karato, S. I., Importance of anelasticity in the interpretation of seismic tomography, *Geophys. Res. Lett.*, *20*, 1623-1626, 1993.

Li, X.-D., and B. Romanowicz, Comparison of global waveform inversions with and without considering cross-branch modal coupling, *Geophys. J. Int.*, *121*, 695-709, 1995.

Li, X.-D., and T. Tanimoto, Waveforms of long-period body waves in a slightly aspherical Earth model, *Geophys. J. Int.*, *112*, 92-102, 1993.

Liu, X.-F., W.-J. Su, and A. M. Dziewonski, Improved resolution of the lowermost mantle shear wave velocity structure obtained using *SKS-S* data, *Eos Trans. AGU 75* (16), Spring Meet. Suppl., 232, 1994.

Lognonne, P., and B. Romanowicz, Modelling of coupled normal modes of the Earth: the spectral method, *Geophys. J. Int.*, *102*, 365-395, 1990.

Love, A. E. H., *A Treatise on the Theory of Elasticity*, 4th ed., Cambridge Univ. Press, New York, 643pp., 1927.

Mitrovica, J. X., and W. R. Peltier, The radial profile of mantle viscosity (abstract), *Eos Trans. AGU*, *73* (14), Spring Meet. Suppl., 202, 1992.

Mitrovica, J. X., Haskell [1935] revisited: A revised constraint on mantle viscosity, *Eos Trans. AGU, 76(46)*, Fall Meet. Suppl., F578, 1995.

Mochizuki, K., Free oscillations and surface waves of an aspherical earth, *Geophys. Res. Lett.*, *13*, 1478-1481, 1986.

Montagner, J. P., and B. Romanowicz, Degrees 2-4-6 inferred from seismic tomography, *Geophys. Res. Lett.*, *20*, 631-634, 1993.

Montagner, J. P., and T. Tanimoto, Global anisotropy in the upper mantle inferred from the regionalization of phase velocities, *J. Geophys. Res.*, *95*, 4797-4819, 1991.

Nataf, H.C., I. Nakanishi, and D. L. Anderson, Measure-

ments of mantle wave velocities and inversion for lateral heterogeneities and anisotropy, *J. Geophys. Res.*, *91*, 7261-7307, 1986.

Park, J., Asymptotic coupled-mode expressions for multiplet amplitude anomalies and frequency shifts on an aspherical Earth, *Geophys. J. R. Astron. Soc.*, *90*, 129-169, 1987.

Phinney, R. A., and R. Burridge, Representation of the elastic-gravitational excitation of a spherical Earth model by generalized spherical harmonics, *Geophys. J. R. Astron. Soc.*, *34*, 451-487, 1973.

Ricard, Y., and C. Vigny, Mantle dynamics with induced plate tectonics, *J. Geophys. Res.*, *94*, 17,534-17,559, 1989.

Ricard, Y., M. Richards, C. Lithgow-Bertelloni, and Y. Le Stunff, A geodynamic model of mantle density heterogeneity, *J. Geophys. Res.*, *98*, 7690-7708, 1993.

Richards, M., and B. H. Hager, Geoid anomalies in a dynamic Earth, *J. Geophys. Res.*, *89*, 5987-6002, 1984.

Ritzwoller, M., G. Masters, and F. Gilbert, Observations of anomalous splitting and their interpretation in terms of aspherical structure, *J. Geophys. Res.*, *91*, 10,203-10,228, 1986.

Romanowicz, B., Multiplet-multiplet coupling due to lateral heterogeneity: asymptotic effects on the amplitude and frequency of the earth's normal modes, *Geophys. J. R. Astron. Soc.*, *90*, 75-100, 1987.

Romanowicz, B., The upper-mantle degree 2: Constraints and inferences from global mantle wave attenuation measurements, *J. Geophys. Res.*, *95*, 11051-11071, 1990.

Romanowicz, B., Seismic tomography of the earth's mantle, *Annu. Rev. Earth. Planet. Sci.*, *19*, 77-99, 1991.

Romanowicz, B., A global tomographic model of shear attenuation in the upper mantle, *J. Geophys. Res.*, *100*, 12,375-12,394, 1995.

Shearer, P., and G. Masters, Global mapping of topography on the 660 km discontinuity, *Nature*, *355*, 791-796, 1992.

Snieder, R., J. Beckers, and F. Neele, The effect of small-scale structure on normal mode frequencies and global inversions, *J. Geophys. Res.*, *96*, 501-515, 1991.

Stacey, F. D., *Physics of the Earth*, 2nd ed., John Wiley, New York, 1977.

Su, W.-J., and A. M. Dziewonski, Predominance of long-wavelength heterogeneity in the mantle, *Nature*, *352*, 121-126, 1991.

Su, W.-J., and A. M. Dziewonski, On the scale of mantle heterogeneity, *Phys. Earth Planet. Inter.*, *74*, 29-54, 1992.

Su, W.-j., R. L. Woodward, and A. M. Dziewonski, Degree-12 model of shear velocity heterogeneity in the mantle, *J. Geophys. Res.*, *99*, 6945-6980, 1994.

Tackley, P., D. J. Stevenson, G. A. Glatzmeier, and G. Schubert, Effects of an endothermic phase transition at 670 km depth in a spherical model of convection in the Earth's mantle, *Nature*, *361*, 699-704, 1993.

Tanimoto, T., Long wavelength S-wave velocity structure throughout the mantle, *Geophys. J. Int.*, *100*, 327-336, 1990.

Tarantola, A. and B. Valette, Generalized non-linear inverse problems solved using the least squares criterion, *Rev. Geophys.*, *20*, 219-232, 1982.

Trampert, J., and J. H. Woodhouse, Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds, *Geophys. J. Int.*, *122*, 675-690, 1995.

VanDerHilst, R., Effects of relative plate motion on the deep structure and penetration depth of slabs below the Izu Bonin and Mariana Island Arcs, *Earth Planet. Sci. Lett.*, *120*, 395-407, 1993.

VanDerHilst, R., Complex morphology of subducting lithosphere in the mantle beneath the Tonga trench, *Nature*, *374*, 154-157, 1995.

Woodhouse, J. H., The coupling and attenuation of nearly resonant multiplets in the Earth's free oscillation spectrum, *Geophys. J. R. Astron. Soc.*, *61*, 261-283, 1980.

Woodhouse, J. H., and F. A. Dahlen, The effect of a general aspherical perturbation on the free oscillations of the Earth, *Geophys. J. R. Astron. Soc.*, *53*, 335-354, 1978.

Woodhouse, J. H., and A. M. Dziewonski, Mapping the upper mantle: Three dimensional modelling of Earth structure by inversion of seismic waveforms, *J. Geophys. Res.*, *89*, 5953-5986, 1984.

Woodhouse, J. H., and A. M. Dziewonski, Models of the upper and lower mantle from waveforms of mantle waves and body waves, *Eos Trans. AGU*, *68*, 356-357, 1987.

Woodhouse, J. H. and D. Giardini, Inversion for the splitting function of isolated low order normal mode multiplets, *Eos Trans. AGU*, *66*, 300, 1985.

Woodhouse, J. H., and T. P. Girnius, Surface waves and free oscillations in a regionalized Earth model, *Geophys. J. R. Astron. Soc.*, *68*, 653-673, 1982.

Woodward, R. L., and T. G. Masters, Global upper mantle structure from long-period differential travel-times, *J. Geophys. Res.*, *96*, 6351-6377, 1991a.

Woodward, R. L., and T. G. Masters, Lower-mantle structure from ScS-S differential travel times, *Nature*, *352*, 231-233, 1991b.

Woodward, R. L., A. M. Forte, W.-J. Su, and A. M. Dziewonski, Constraints on the large-scale structure of the Earth's mantle, in *Evolution of the Earth and Planets, Geophys. Monogr. Ser.*, vol. 74, edited by E. Takahashi, R. Jeanloz, and D. Rubie, pp. 89-109, AGU, Washington, D.C., 1993.

Zhang, Y. S., and T. Tanimoto, Global Love wave phase velocity variation and its significance to plate tectonics, *Phys. Earth Planet. Inter.*, *66*, 160-202, 1991.

Zhang, Y. S., and T. Tanimoto, High-resolution global upper mantle structure and plate tectonics, *J. Geophys. Res.*, *98*, 9793-9823, 1993.

X.-D. Li, Molecular Dynamics, 928 East Arques Avenue, Sunnyvale, CA 94086.

B. Romanowicz, Seismographic Station, 475 McCone Hall, Berkeley, CA 94720. (e-mail: barbara@seismo.berkeley.edu)

(Received January 8, 1996; revised April 19, 1996;
accepted April 25, 1996.)