

CHAPTER 7

Body waves and mantle tomography

1. The need for parameterization.

In various sections of these notes, we have solved a variety of inverse problems. In chapter 4, we solved problems of the form:

$$d_i \pm \sigma_i = \int_0^R G_i \delta m dr \quad (7.1)$$

where d_i is a datum such as the difference between an observed frequency of free oscillation of the earth and one calculated for a starting model, G_i is some continuous kernel which we can compute and δm is a continuous model perturbation. When we have relatively few data, it is possible to avoid parameterization of the model and make an expansion of the form:

$$\delta m = \sum_{i=1, N} a_i G_i(r) \quad (7.2)$$

where N is the number of data. Inserting this into equation 1 gives

$$\mathbf{d} = \mathbf{\Gamma} \cdot \mathbf{a} \quad \text{where} \quad \Gamma_{ij} = \int_0^R G_i G_j dr \quad (7.3)$$

and $\mathbf{\Gamma}$ is a matrix which is of dimension $N \times N$. Equation 3 can be solved in a variety of ways (e.g., we can impose smoothness constraints on the model perturbation or on the total model) and we can explore the trade off with fit to the data. As also demonstrated in chapter 4, we can look at the ability of our data to resolve features of our models.

Unfortunately, once N exceeds a few thousand, the computational burden of dealing with huge matrices becomes too great. The conventional way around this is to parameterize the model by expanding it in a set of basis functions where the number of parameters is chosen to be computationally manageable:

$$\delta m = \sum_{i=1, M} a_i f_i(r) \quad (7.4)$$

where M is the number of parameters. Of course, in 3D tomography, the basis functions f are functions of r , θ , and ϕ . Substituting 4 into equation 1 gives

$$\mathbf{d} = \mathbf{A} \cdot \mathbf{a} \quad \text{where} \quad A_{ij} = \int_0^R G_i f_j dr \quad (7.5)$$

and \mathbf{A} is a matrix which is of dimension $N \times M$.

The choice of basis functions in equation 4 can impact the kinds of models we can recover and can also impact the computational difficulty of solving equation 5. In global tomography, the choice of

basis functions has, for many years, involved the use of spherical harmonics for parameterizing lateral variations:

$$\delta m = \sum_{s,t} \delta m_s^t(r) Y_s^t(\theta, \phi) \quad (7.6)$$

where the radial expansion coefficients $\delta m_s^t(r)$ are further parameterized either in global functions (e.g. Legendre polynomials or Chebychef polynomials) or as local functions (e.g. layers or B-splines). The reason for the choice of spherical harmonics is that these are efficient for parameterizing the long-wavelength structure which dominates many of the seismic datasets and are the natural basis for interpreting mode structure coefficients (chapter 5). Later in this chapter, we shall describe an efficient algorithm for computing matrix elements for body wave travel times for the spherical harmonic basis (similar to the algorithm for surface wave phase measurements discussed in chapter 6). Unfortunately, a consequence of using global bases is that every datum effectively becomes sensitive to the entire model so that the matrix \mathbf{A} is very dense. This means that global models using spherical harmonics are typically limited to about 10,000 model parameters – if we divide the mantle up into roughly 20 layers (see below), each layer could have about 500 parameters. A spherical harmonic expansion up to degree l has $(l + 1)^2$ expansion coefficients so this means l is limited to about 21. If we recall that

$$ka = l + \frac{1}{2} = \frac{2\pi a}{\lambda} \quad (7.7)$$

where k is wavenumber, λ is wavelength, and $2\pi a$ is the circumference of the earth (about 40,000 km), we find that the minimum wavelength we can capture is about 2000 km. Unfortunately, dynamically interesting structures such as slabs typically have much smaller dimensions than this (and many of our data are, in principle, sensitive to small wavelength structure). This has motivated the use of local bases in global tomography (e.g. blocks of uniform lateral dimension, equal area blocks, non-uniform distribution of blocks mimicing data sampling, tesslations, etc).

Why are local bases so useful? Let us suppose we have several hundred thousand travel time measurements. To a fairly good approximation, ray theory can be used to interpret such data (see below) so each datum is sensitive to only a small fraction of the total number of parameters in the model (i.e. along a particular ray). For example, using blocks of lateral dimension 4 degrees at the equator (this corresponds to a surface wavelength of about 880km or an l of about 45 if we had done a spherical harmonic expansion) gives roughly 2500 blocks per layer for a total of 50,000 model parameters for a 20 layer model. However, each datum samples only about 1% or less of the blocks so each row of the matrix \mathbf{A} will have less then 500 non-zero entries. Sparse matrix techniques can then be efficiently used to solve equation 5.

2. Finding a model

Let us suppose we are solving the problem

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{d} \quad (7.8)$$

for the vector \mathbf{x} . Further, we shall assume that we have divided each row of this system of equations by the observation error on the datum so that the data vector \mathbf{d} has a covariance matrix which is just \mathbf{I} (i.e. we are assuming our data are statistically independent from each other. If our sytem of equations (8) were well-conditioned, we might just find the least-squares solution:

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{d} \quad (7.9)$$

which minimizes $(\mathbf{A} \cdot \hat{\mathbf{x}} - \mathbf{d})^2$. In reality, \mathbf{A} is usually not well-conditioned and $\mathbf{A}^T \mathbf{A}$ is even worse (the condition number is effectively squared) so the solution (9) is rarely chosen. One way around squaring the condition number is to use a singular value decomposition (SVD) on equation 8. The matrix \mathbf{A} is decomposed into singular values and matrices of left and right eigenvectors:

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T \quad (7.10)$$

where \mathbf{U} has dimension $N \times N$ and \mathbf{V} has dimension $M \times M$ and $\mathbf{\Lambda}$ is a $M \times N$ with non-zero diagonal elements. Note that $\mathbf{U}^T \mathbf{U} = \mathbf{I}$ and $\mathbf{V}^T \mathbf{V} = \mathbf{I}$. The least-squares solution in terms of the SVD is

$$\hat{\mathbf{x}} = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{d} = \mathbf{A}^+ \mathbf{d} \quad (7.11)$$

where \mathbf{A}^+ can be thought of as the (generalized) inverse of \mathbf{A} . If \mathbf{A} is not well-conditioned, it will have some small singular values which will generally lead to some poorly determined contributions to $\hat{\mathbf{x}}$. To see why this is so, consider the covariance matrix of the model. To get the model we are taking a linear combination of data: $\mathbf{A}^+ \mathbf{d}$. Now \mathbf{d} has covariance matrix \mathbf{I} so $\hat{\mathbf{x}}$ has covariance matrix:

$$\mathbf{A}^+ \mathbf{I} (\mathbf{A}^+)^T = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{U} \mathbf{\Lambda}^{-1} \mathbf{V}^T = \mathbf{V} \mathbf{\Lambda}^{-2} \mathbf{V}^T \quad (7.12)$$

The square roots of the diagonal elements of this matrix are the errors on our model parameters. Clearly, small singular values are going to make these errors large. One way to avoid this is to exclude small singular values from the sums implicit in equations 11 and 12 but this will mean that $\mathbf{A}^+ \mathbf{A}$ will no longer be \mathbf{I} . In fact, substituting 8 into 11 gives

$$\hat{\mathbf{x}} = \mathbf{A}^+ \mathbf{A} \mathbf{x} = \mathbf{R} \mathbf{x} \quad (7.13)$$

and the matrix $\mathbf{R} = \mathbf{A}^+ \mathbf{A}$ is sometimes called the "resolution matrix". In a perfectly resolved system, $\mathbf{R} = \mathbf{I}$ but, in general, each model element estimated will be a linear combination of all the model elements. For the truncated SVD approximation to the generalized inverse, $\mathbf{R} = \mathbf{V} \mathbf{V}^T$. We use the resolution matrix to estimate how much we are "blurring" the model.

The process of throwing away small singular values is an example of "regularization" of the inverse problem. It is not a commonly used method [though you may have used this when you did moment tensor retrieval in practical 3] because the model we end up with doesn't satisfy any particularly sensible optimization criterion. Usually we seek a model which has some property optimized and still adequately satisfies the data. For example, we might seek a model which has minimum first or second derivative. Let \mathbf{D} be some "roughening" operation on the model. Then we might want to minimize

$$f = (\mathbf{A} \mathbf{x} - \mathbf{d})^T (\mathbf{A} \mathbf{x} - \mathbf{d}) + \lambda (\mathbf{D} \mathbf{x})^T \mathbf{D} \mathbf{x} \quad (7.14)$$

where the parameter λ controls the degree of smoothing. Expanding out the brackets and taking the derivative with respect to \mathbf{x} and setting to zero gives

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{A}^T \mathbf{d} \quad (7.15)$$

Clearly, setting λ to zero gives us our least-squares result. Comparing equations 11 and 15 gives $\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{A}^T$ and we can use 12 and 13 to estimate the model covariance matrix and the resolution matrix. Increasing λ will result in models which have a smaller value of $\mathbf{x}^T \mathbf{D}^T \mathbf{D} \mathbf{x}$. One choice for \mathbf{D} is \mathbf{I} which results in a process called "ridge regression" and ends up minimizing the Euclidean length of the solution vector. This turns out to be a bad thing to do in tomography as it results in models which have wildly underestimated amplitudes. A good choice for \mathbf{D} is the first difference operator which in 1D looks like:

$$\begin{pmatrix} 1 & -1 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & \dots \\ 0 & 0 & 1 & -1 & \dots \end{pmatrix} \quad (7.16)$$

In tomography, we use this for smoothing in the radial direction and we use a form which minimizes the sum of the first differences between a block and its four nearest neighbors laterally for lateral smoothing. In practice, very different degrees of radial and lateral smoothing are required in the tomography problem because radial and lateral length scales are so different for mantle structure.

We have already complained about forming matrix products like $\mathbf{A}^T \mathbf{A}$ when the matrices are ill-conditioned and, in any case, making $\mathbf{A}^T \mathbf{A}$ can itself be time consuming (and may remove the sparsity). In practice, we construct the following equivalent system:

$$\begin{pmatrix} \mathbf{A} \\ \lambda^{\frac{1}{2}} \mathbf{D} \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{d} \\ \mathbf{0} \end{pmatrix} \quad (7.17)$$

and solve this rectangular system using SVD – or more likely a solver which takes advantage of the sparseness of the matrices \mathbf{A} and \mathbf{D} .

One final technical point about solving equation 17 is that we can help the conditioning of the system by solving a slightly different system:

$$\mathbf{C} \mathbf{y} = \begin{pmatrix} \mathbf{A} \\ \lambda^{\frac{1}{2}} \mathbf{D} \end{pmatrix} \mathbf{W} \mathbf{y} = \begin{pmatrix} \mathbf{d} \\ \mathbf{0} \end{pmatrix} \quad \text{where } \mathbf{y} = \mathbf{W}^{-1} \mathbf{x} \quad (7.18)$$

for \mathbf{y} then getting \mathbf{x} from $\mathbf{x} = \mathbf{W} \mathbf{y}$. \mathbf{W} can be chosen in a variety of ways – one is to make it a diagonal matrix such that the Euclidean lengths of the columns of \mathbf{C} are the same – this makes the range of singular values of \mathbf{C} much less extreme and also speeds up convergence of some of the iterative techniques we discuss in the next section. This process of weighting is sometimes called "preconditioning" of the system and whole books have been written on the topic.

We now consider some "iterative" techniques for solving large systems of (hopefully) sparse equations. Such techniques can operate on one row of the matrix at a time (and are sometimes called row-action methods)

3. True iterative techniques

For simplicity, we go back to equation 8: $\mathbf{A} \mathbf{x} = \mathbf{d}$ though we are more likely to be solving something like equation 18 in practice. Let \mathbf{x}^q be the q 'th iterate and define the residual vector

$$\mathbf{r}^q = \mathbf{d} - \mathbf{A} \cdot \mathbf{x}^q \quad (7.19)$$

Now we want to perturb \mathbf{x}^q to get a better answer. One way to do this is to work one equation at a time. Let $\Delta \mathbf{x}^q$ be the desired perturbation. We choose $\Delta \mathbf{x}^0$ to be the perturbation that makes the first element of \mathbf{r}^0 be zero, $\Delta \mathbf{x}^1$ is chosen to make the second element of \mathbf{r}^1 zero and so on – we then cycle through the equations until we get convergence. To get a unique perturbation, we choose the one that has $\|\Delta \mathbf{x}^q\|$ minimized. Thus we minimize

$$(A_{ij} \Delta x_j^q - r_i^q)^2$$

Then

$$\Delta x_j = \frac{A_{ij} r_i}{\sum_k A_{ik}^2} \quad (7.20)$$

This is the original procedure of Kaczmarz and is not terribly efficient. One popular modification to this is to compute the correction for each row (as above) and then average all the corrections to get a mean $\Delta \mathbf{x}$:

$$\Delta x_j = \frac{1}{M} \sum_{i=1}^M \frac{A_{ij} r_i}{\sum_k A_{ik}^2} \quad (7.21)$$

where M is the number of non-zero elements in A_{ij} . This process is called the Simultaneous Iterative Reconstruction Technique (SIRT) and is still commonly used. Some modifications are described in Hager and Clayton, 1989. A general family of SIRT methods is given by

$$\Delta x_j = \frac{\Omega}{\gamma_j} \sum_{i=1}^M \frac{A_{ij} r_i}{\rho_i}$$

where

$$\gamma_j = \sum_i |A_{ij}|^\alpha, \quad \rho_i = \sum_k |A_{ik}|^{2-\alpha}$$

with $0 < \Omega < 2$ and $0 < \alpha < 2$. Hager et al use ($\alpha = 1, \Omega = 1$). It turns out that SIRT as described above converges to a solution which is not the least squares solution of the original system of equations and some weighting must be applied to correct this (van der Sluis and van der Vorst, 1987). SIRT works well in practice but it is now more common to use a conjugate gradient method – one particular variant called LSQR has become popular in seismic tomography, probably because it was popularized in the mid 80's by Guust Nolet.

4. Gradient (Projection) techniques

Consider the function defined by

$$f(\mathbf{x}) = \frac{1}{2} (\mathbf{A} \cdot \mathbf{x} - \mathbf{d})^2 \quad (7.22)$$

In two dimensions ($\mathbf{x} = x_1, x_2$), f is a surface which has hills and valleys. Expanding out this function gives

$$\begin{aligned} f &= \frac{1}{2} (\mathbf{A} \cdot \mathbf{x} - \mathbf{d})^T (\mathbf{A} \cdot \mathbf{x} - \mathbf{d}) \\ &= \frac{1}{2} [\mathbf{d}^T \cdot \mathbf{d} + \mathbf{x}^T \cdot \mathbf{A}^T \cdot \mathbf{A} \cdot \mathbf{x} - 2\mathbf{x}^T \cdot \mathbf{A}^T \cdot \mathbf{d}] \end{aligned}$$

Now define the square symmetric matrix $\mathbf{B} = \mathbf{A}^T \cdot \mathbf{A}$ and the vector $\mathbf{b} = \mathbf{A}^T \cdot \mathbf{d}$ then

$$f = \frac{1}{2} [\mathbf{d}^T \cdot \mathbf{d} + \mathbf{x}^T \mathbf{B} \cdot \mathbf{x} - 2\mathbf{x}^T \cdot \mathbf{b}]$$

The first term on the right is just the length of the data vector so we define the misfit function $\phi(\mathbf{x})$ as the last two terms:

$$\phi(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{B} \cdot \mathbf{x} - \mathbf{x}^T \cdot \mathbf{b} \quad (7.23)$$

(This is the same function as f with all the same hills and valleys but with an offset removed.)

The gradient of ϕ with respect to \mathbf{x} is simply

$$\nabla\phi(\mathbf{x}) = \mathbf{B} \cdot \mathbf{x} - \mathbf{b} \quad (7.24)$$

At any point \mathbf{x}_k on the surface, the downhill slope is given by

$$-\nabla\phi(\mathbf{x}_k) = \mathbf{b} - \mathbf{B} \cdot \mathbf{x}_k = \mathbf{r}_k \quad (7.25)$$

and is actually zero at a solution which fits the data ($\mathbf{B} \cdot \mathbf{x} - \mathbf{b} = 0$)

Our procedure is to find \mathbf{x} by moving in a sequence of directions which take us down the misfit surface. Let

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{u}_k \quad (7.26)$$

where \mathbf{u}_k is a direction we choose to go in. We can find the value of λ_k (assuming \mathbf{u}_k is specified) that minimizes

$$\phi(\mathbf{x}_k + \lambda_k \mathbf{u}_k)$$

$$\phi = \frac{1}{2} (\mathbf{x}_k + \lambda_k \mathbf{u}_k)^T \cdot \mathbf{B} \cdot (\mathbf{x}_k + \lambda_k \mathbf{u}_k) - (\mathbf{x}_k + \lambda_k \mathbf{u}_k)^T \cdot \mathbf{b}$$

so

$$\frac{\partial\phi}{\partial\lambda_k} = \mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{x}_k + \lambda_k \mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{u}_k - \mathbf{u}_k^T \cdot \mathbf{b} = 0$$

so

$$\mathbf{u}_k^T \cdot (\mathbf{B} \cdot \mathbf{x}_k - \mathbf{b}) + \lambda_k \mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{u}_k = 0$$

$$\lambda_k = \frac{\mathbf{u}_k^T \cdot \mathbf{r}_k}{\mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{u}_k} \quad (7.27)$$

The next question is how to specify \mathbf{u}_k . If we choose $\mathbf{u}_k = \mathbf{r}_k$ we get the "steepest descent algorithm" (remember \mathbf{r} is the local downhill direction – see equation 25):

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{r}_k \quad \text{where} \quad \lambda_k = \frac{\mathbf{r}_k^T \cdot \mathbf{r}_k}{\mathbf{r}_k^T \cdot \mathbf{B} \cdot \mathbf{r}_k} \quad (7.28)$$

This isn't always a very good idea since it is possible to go from one side of the valley to another – rather than going down the middle. A better method is to choose directions so that they are "conjugate" (perpendicular in some sense) to all previous directions.

Reconsider equation 26:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{u}_k$$

Note that \mathbf{x}_{k+1} is actually a linear combination of all the directions taken to date: $\mathbf{u}_1 \dots \mathbf{u}_k$ – if there are N model parameters, then the final \mathbf{x} can be completely specified by an expansion in N (orthogonal) directions:

$$\mathbf{x} = \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \cdots + \lambda_N \mathbf{u}_N$$

If the directions were truly orthogonal to each other, we could just dot this equation with the transpose of the j 'th \mathbf{u} and that would pick out the j 'th term. It turns out that this isn't computationally helpful – but it is helpful to make the directions "B-orthogonal" which means that

$$\mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{u}_j = 0$$

Applying this to the above equation gives

$$\mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{x} = \mathbf{u}_k^T \cdot \mathbf{b} = \lambda_k \mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{u}_k$$

A conjugate-gradient algorithm can now be developed. We start with $\mathbf{x}_1 = 0$ and compute $\mathbf{r}_1 = \mathbf{b}$. For the first direction, we choose steepest descent so $\mathbf{u}_1 = \mathbf{r}_1$ and we get λ_1 from equation 27. We are now at point \mathbf{x}_2 and can compute \mathbf{r}_2 . In steepest descents, \mathbf{r}_2 would be our next direction but this is not "B-orthogonal" to the previous direction. To achieve this, we let the new direction be

$$\mathbf{u}_{k+1} = \mathbf{r}_{k+1} + \gamma_k \mathbf{u}_k \tag{7.29}$$

Dotting through by $(\mathbf{B} \cdot \mathbf{u}_k)^T$ gives

$$\gamma_k = -\frac{\mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{r}_{k+1}}{\mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{u}_k}$$

This form for γ_k is not computationally optimal as we shall see. To get our final algorithm, we first note that the \mathbf{r} 's can be computed recursively. Multiply equation 26 by \mathbf{B} and subtract \mathbf{b} from both sides:

$$\mathbf{B} \cdot \mathbf{x}_{k+1} - \mathbf{b} = \mathbf{B} \cdot \mathbf{x}_k - \mathbf{b} + \lambda_k \mathbf{B} \cdot \mathbf{u}_k$$

so

$$\mathbf{r}_{k+1} = \mathbf{r}_k - \lambda_k \mathbf{B} \cdot \mathbf{u}_k \tag{7.30}$$

We can further manipulate the above formulae to get some identities which allow us to compute λ_k and γ_k more efficiently. First, note that we recover equation 27 from equation 30 if we require $\mathbf{u}_k^T \cdot \mathbf{r}_{k+1} = 0$. Forcing this to be true and dotting \mathbf{r}_{k+1}^T into equation 29 gives the result that $\mathbf{r}_k^T \cdot \mathbf{u}_k = \mathbf{r}_k^T \cdot \mathbf{r}_k$. Furthermore, if we dot \mathbf{r}_{k+1}^T into 30 and use equation 27 for λ_k and the above formula for γ_k , we get

$$\mathbf{r}_{k+1}^T \cdot \mathbf{r}_{k+1} = \mathbf{r}_{k+1}^T \cdot \mathbf{r}_k - \lambda_k \mathbf{r}_{k+1}^T \cdot \mathbf{B} \cdot \mathbf{u}_k = \mathbf{r}_{k+1}^T \cdot \mathbf{r}_k + \gamma_k \mathbf{u}_k^T \cdot \mathbf{r}_k \tag{7.31}$$

Similarly, dotting $(\mathbf{B} \cdot \mathbf{u}_{k+1})^T$ into 29 shows that $\mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{u}_k = \mathbf{r}_k^T \cdot \mathbf{B} \cdot \mathbf{u}_k$. Dotting \mathbf{r}_k^T into 30 and using this result allow us to show that $\mathbf{r}_{k+1}^T \cdot \mathbf{r}_k = 0$. These identities allow us to compute γ_k and λ_k as

$$\gamma_k = \frac{\mathbf{r}_{k+1}^T \cdot \mathbf{r}_{k+1}}{\mathbf{r}_k^T \cdot \mathbf{r}_k} \quad \lambda_k = \frac{\mathbf{r}_k^T \cdot \mathbf{r}_k}{\mathbf{u}_k^T \cdot \mathbf{B} \cdot \mathbf{u}_k} \tag{7.32}$$

The algorithm can now be written (taking $\mathbf{x}_1 = 0$)

```

k = 0
r1 = b
u1 = r1
x1 = 0
begin loop
k = k + 1
w = B · uk
λ = rkT · rk / ukT · w
xk+1 = xk + λuk
rk+1 = rk - λw
γ = rk+1T · rk+1 / rkT · rk
uk+1 = rk+1 + γuk
end loop

```

Note that there is only one matrix-vector multiply per iteration. M iterations of this process would give the exact solution (in the absence of roundoff) but it is anticipated that much fewer than M iterations will be required to get an acceptable solution.

The algorithm described above is the standard CG algorithm – Golub and Van Loan (Chapter 10) 1996 give an extensive discussion of the theory. This is not in the best form for numerical application since it uses the "normal" equations $\mathbf{B} \cdot \mathbf{x} - \mathbf{b}$ which, as we have already noted, can square the condition number and introduce instability. We would like to go back to the rectangular system in equation 8. Remember, even just forming \mathbf{B} can turn a sparse \mathbf{A} matrix into a dense \mathbf{B} matrix though the sparseness can be retained by computing $\mathbf{B} \cdot \mathbf{u}$ as $\mathbf{A}^T \cdot (\mathbf{A} \cdot \mathbf{u})$. An equivalent sparse square system can be written down:

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{r} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix}$$

and used to develop algorithms which do not implicitly use the normal equations and which are stable when systems are not well-conditioned (e.g. LSQR). We leave this as an exercise to the reader.

One final point: knowing when to stop iterative techniques can be a bit of an art form. Typically, much of the misfit to the data is taken up in the first few iterations but convergence to a stable model can take much longer. In particular, where we include a smoother (as in equation 18), it seems that the effect of the smoother becomes more apparent at later iterations even though the fit to the data does not change much. Several stopping criteria for LSQR have been suggested (see original papers by Paige and Saunders) but it pays to be conservative and to iterate longer than you think you need to!

5. Resolution and error analysis

In section 2, we discussed resolution and error and gave results in terms of the generalized inverse of \mathbf{A} (equations 12 and 13). How do we go about computing resolution and error when \mathbf{A}^+ is not available (as when using an iterative technique). Some have suggested using a rough estimate of \mathbf{A}^+ (e.g. Nolet et al, 1999, GJI, v138, p36) using a one-step back projection which gives

$$\mathbf{A}^+ \simeq \mathbf{A}^T \mathbf{\Omega} \tag{7.33}$$

where $\mathbf{\Omega}$ is a diagonal matrix and

$$\Omega_{kk} = \frac{(\mathbf{A}\mathbf{A}^T)_{kk}}{\sum_{i=1}^N (\mathbf{A}\mathbf{A}^T)_{ik}^2} \quad (7.34)$$

It is not clear to us how well this performs in practice but we are often only interested in the overall nature of the resolution matrix and not precise values for its elements. Perhaps this is adequate for this.

One way of estimating the resolution matrix is to do an inversion where we set the m 'th element of the model vector \mathbf{x} to one and all the others to zero – call this vector \mathbf{x}_m . Now, compute $\mathbf{d}_m = \mathbf{A}\mathbf{x}_m$ and solve $\mathbf{A}\mathbf{x} = \mathbf{d}_m$ using exactly the same iterative algorithm as you used to get your true model. This process computes a single row (and column) of the resolution matrix corresponding to the m 'th model element. The complete resolution matrix can be computed by performing M such inversions – one for each model parameter. Clearly this is infeasible if we are talking about 50,000 model parameters but we can focus on key areas of the model where we are particularly interested in the resolvability of a particular structure.

A modification of the above process (which is sometimes called a "spike test") is to solve for some pattern to test resolution over a broad region. A common choice is to use a checkerboard pattern in one of the layers of the model. A synthetic data set is computed for this checkerboard model and then inverted using exactly the same iterative algorithm used to get the real model. The recovered checkerboard can indicate areas of problematic recovery in the layer being tested and can show leakage into adjacent layers above and below.

The estimation of the covariance matrix of the model can also be problematic but usually we are satisfied with the diagonal elements (the square roots of which are the standard deviations of the model parameters). It turns out that the best way to estimate these is to add a noise vector to the data vector $\mathbf{d} = \mathbf{d} + \mathbf{e}$ where the elements of \mathbf{e} are randomly chosen from a normal distribution with a unit standard deviation (remember, we divided all data by their errors initially). We then solve for a model using this perturbed data vector in our iterative procedure. We repeat this process many times (100 say) and then look at the standard deviations of the elements of the 100 models we have generated. Tests show that this process produces an excellent estimate of the diagonal elements of the model covariance matrix.

6. Measuring long period travel times

Historically, seismograms were recorded either at "long" periods or "short" periods. The reason for this is that a major source of motion of the ground is the "microseisms" which are due to nonlinear interactions of ocean waves causing pressure variations on the ocean floor. Microseisms have a main peak at 14 second period and a secondary peak at 7 seconds. It is actually the secondary peak that is mainly seen on seismometers. In the past, seismic recording systems did not have the dynamic range to record both the microseisms and the small seismic signals which ride on them. Thus instruments were designed to see periods shorter than 7 seconds (usually peak response at about 1 second) or periods longer than about 15 seconds. Modern seismic recording systems have enough dynamic range to be able to record the microseisms (so-called broad-band recording) but, for most earthquakes, we must still filter out the microseisms so we can see the small seismic signals.

On the short period side, body waves of dominant period 1 second are seen and the first arriving P wave can be accurately picked. Scattering by short-wavelength heterogeneity causes large "codas" which can obscure secondary arrivals. Many observations have been made and are collected by the International Seismological Commission (ISC) which has used them to make a more comprehensive tabulation of earthquake locations. Such data are also used in tomography. P wave tomography using the ISC data has been quite successful but the Swaves are more problematic. This is because S waves typically have

a lower frequency content due to attenuation and are more poorly recorded by short period instruments. Furthermore, ISC picks are usually made from vertical component instruments so interference of S by the SKS phase at distances beyond 80 degrees is a problem. This makes it very difficult to image S velocity in the lowermost mantle from ISC S picks.

Long-period data offer some advantages over the ISC data – in particular, codas from scattering are nearly nonexistent so later phases can be accurately picked. One picking algorithm is discussed in detail in the paper by Bolton and Masters (2001) and we shall give some demonstrations during the lecture. Differential times can also be picked but require corrections for relative attenuation and, sometimes, corrections for waveform distortion due to propagation effects. Again, we shall give some examples in the lecture.

7. Computing matrix elements for travel times using ray theory

Consider a ray through the Earth as shown in Figure 1.

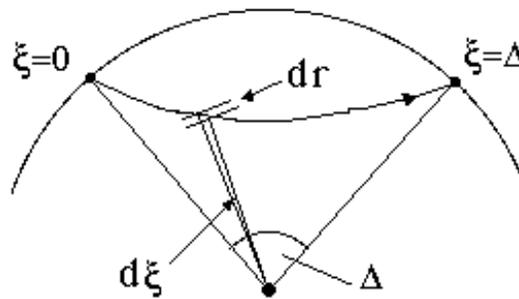


Fig 1

Now focus on the small segment of the ray which subtends the angle $d\xi$ at the center of the Earth (Fig. 2).

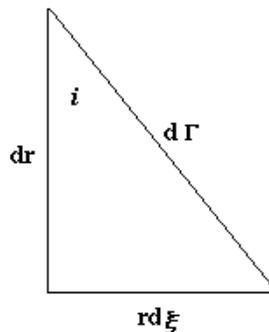


Fig 2

i is the angle the ray makes with the vertical and we know that the ray parameter is related to i by

$$p = \frac{r}{v} \sin i \tag{7.35}$$

where v is the velocity for the ray segment at radius r . The travel time of the ray is

$$T = \int_{\Gamma} \frac{1}{v} d\Gamma \quad (7.36)$$

where the integral is taken along the ray path. Fermat's principle states that a ray path between two points is a path of stationary time. Thus the travel time will not change to first order if the ray path is slightly perturbed. If we make a small perturbation in velocity structure there will be a change in travel time due to the change in velocity structure *and* due to the change in the ray path but the latter term is of second order and so can be neglected. We can therefore differentiate equation 36 giving:

$$\delta T = - \int_{\Gamma} \frac{\delta v}{v^2} d\Gamma \quad (7.37)$$

From figure 2, we have

$$\sin i = \frac{r d\xi}{d\Gamma} \quad \text{so} \quad d\Gamma = \frac{r^2}{pv} d\xi \quad (7.38)$$

and we can rewrite equation 37 as an integral over distance:

$$\delta T = \int_0^{\Delta} G(\xi) \delta v(\xi) d\xi \quad (7.39)$$

where

$$G(\xi) = - \frac{r^2}{pv^3} \quad (7.40)$$

This kernel is evaluated by keeping track of the depth of the ray for every position of arc length ξ . To do this we need an equation relating ξ to r . Reconsider Figure 2 and note that (using the equation for p)

$$r \frac{d\xi}{dr} = \tan i = \frac{\frac{vp}{r}}{\left(1 - \left(\frac{vp}{r}\right)^2\right)^{\frac{1}{2}}} \quad (7.41)$$

or

$$\frac{d\xi}{dr} = \frac{p}{r} \left(\frac{r^2}{v^2} - p^2 \right)^{-\frac{1}{2}} \quad (7.42)$$

On an aspherical Earth where we have used a local block parameterization, we step finely along in distance starting from a specific source position to a specific receiver position. At each point, we compute the radius we are at using equation 42 and then evaluate the kernel using equations 39 and 40. We also keep track of which block we are in at each step along the ray then integrate the contributions to each block at the end.

If we have parameterized the model laterally in spherical harmonics, the following algorithm gives an efficient way of computing the matrix elements. Suppose we represent the velocity perturbation by:

$$\delta v(r, \theta, \phi) = \sum_{s,t} \delta v_s^t(r) Y_s^t(\theta, \phi) \quad (43)$$

[$s = 0$ gives the perturbation to the spherically averaged Earth]. We now have:

$$\delta T = \sum_{s,t} \int_{\mathbf{x}_s}^{\mathbf{x}_r} G(\xi) \delta v_s^t(r) Y_s^t(\theta, \phi) d\xi \quad (44)$$

where r, θ, ϕ are points along the ray path and \mathbf{x}_s is the source position and \mathbf{x}_r is the receiver position.

Equation 41 is easy to evaluate if we transform the coordinate system (*i.e.*, move the North pole) so that the ray is in the equatorial plane. Spherical harmonics in one coordinate system transform to spherical harmonics in another according to a formula of the form:

$$Y_s^t(\theta', \phi') = \sum_{m=-s}^s B_{mt} Y_s^m(\theta, \phi) \quad (45)$$

Note that the harmonic degree, s , is not included in the sum so that, for example, a degree 2 harmonic in the rotated coordinate system is a linear combination of the 5 degree 2 harmonics in the the original coordinate system. The computation of the transformation matrix, B_{mt} , is tedious but straightforward (Edmonds, 1960, ch.4) and an efficient algorithm for its computation is given in Masters and Richards-Dinger (1999).

Using equations 40 and 42, we can easily rotate coordinate systems:

$$\delta v(r, \theta', \phi') = \sum_{s,t} \delta v_s^{t'}(r) Y_s^t(\theta', \phi') = \sum_{s,t} \delta v_s^{t'}(r) \sum_{m=-s}^s B_{mt} Y_s^m(\theta, \phi)$$

If we let

$$\delta v_s^m = \sum_t \delta v_s^{t'} B_{mt}$$

we get

$$\delta v(r, \theta', \phi') = \sum_{s,m} \delta v_s^m(r) Y_s^m(\theta, \phi)$$

thus changing the coordinate system gives us the same form as equation 40 for δv but with new expansion coefficients $\delta v_s^{t'}$. If we rotate the ray path to the equatorial plane we have

$$Y_s^t(\theta, \phi) \implies Y_s^t\left(\frac{\pi}{2}, \xi\right) = X_s^t\left(\frac{\pi}{2}\right) e^{it\xi}$$

where X_s^t is now a constant along the ray path. In the rotated system, equation 41 is

$$\delta T = \sum_{s,t} X_s^t\left(\frac{\pi}{2}\right) \int_0^{\Delta} G(\xi) \delta v_s^{t'}(r) e^{it\xi} d\xi \quad (46)$$

$$\text{where } \delta \mathbf{v}_s' = \mathbf{B}^{-1} \delta \mathbf{v}_s$$

The integral can be evaluated numerically providing we keep track of $r(\xi)$ (by integrating $dr/d\xi$). In fact it is most efficient to calculate tables of the form

$$I_s^t(\Delta) = \int_0^{\Delta} G(\xi) \delta v_s^t(r) e^{it\xi} d\xi$$

If we define $\mathbf{B}' = \mathbf{B}^{-1}$ then our equation becomes

$$\delta T = \sum_{s,t} X_s^t \left(\frac{\pi}{2} \right) \sum_m B'_{t,m} I_s^m(\Delta)$$

which can be rapidly evaluated.

When we come to doing the inverse problem, we would expand the coefficients for velocity, $\delta v_s^t(r)$ in a set of basis functions (e.g. B-splines or shells) so we can write

$$\delta v_s^t(r) = \sum_k a_{ks}^t g_k(r)$$

where the g_k 's are some known functions and the a 's are some constants. Then

$$I_s^t(\Delta) = \sum_k a_{ks}^t \int_0^\Delta G(\xi) g_k(r) e^{it\xi} d\xi = \sum_k a_{ks}^t J_{ks}^t(\Delta) \quad \text{say}$$

and our equation for a travel time residual becomes:

$$\delta T = \sum_k a_{ks}^t \sum_{s,t} X_s^t \left(\frac{\pi}{2} \right) \sum_m B'_{t,m} J_{ks}^m(\Delta)$$

which is in a form suitable for inversion.

So far, we have been considering the effect of a ‘‘volume perturbation’’ in velocity. There may also be perturbations in the levels of discontinuities which, if the velocity is different on either side, produce travel time anomalies. The formula for δT for a transmitted ray if a boundary is moved by δr is

$$\delta T = -\frac{\delta r}{r} \left[\left(\frac{r^2}{v^2} - p^2 \right)^{\frac{1}{2}} \right]_{-}^{+} \quad (47)$$

where $[f]_{\pm}^{\pm}$ indicates the value of f below subtracted from the value of f above the discontinuity. For a reflected ray, we get

$$\delta T = -\frac{2\delta r}{r} \left(\frac{r^2}{v^2} - p^2 \right)^{\frac{1}{2}} \quad (48)$$

for a topside reflection (so v is the velocity just above the discontinuity) and

$$\delta T = \frac{2\delta r}{r} \left(\frac{r^2}{v^2} - p^2 \right)^{\frac{1}{2}} \quad (49)$$

for a bottomside reflection (so v is the velocity just below the discontinuity). Note that the perturbation in the radius of a discontinuity can also be expanded in spherical harmonics:

$$\delta r = \sum_{s,t} \delta r_s^t Y_s^t(\theta, \phi) \quad (50)$$

and, in general, there are different expansion coefficients for each discontinuity in the model.

8. Importance of earthquake location in tomography

It turns out that our (in)ability to locate earthquakes accurately means that we have a source of noise in our tomographic problem which can rival the signal from 3D structure (at least for P-wave tomography). We can estimate the uncertainty due to event mislocation by considering the following equation

$$\delta t = \frac{\partial t}{\partial x} \delta x + \frac{\partial t}{\partial y} \delta y + \frac{\partial t}{\partial z} \delta z + \delta t_0,$$

where δx , δy , δz are errors in event location, δt_0 is the error in origin time, and δt is the resulting error in travel time. Analyses of mislocations of events located by independent means leads to estimates of the length of a typical mislocation vector, ϵ_X of $\simeq 14$ – 18 km. To convert this number to a typical change in epicentral distance, we assume that the stations are uniformly distributed around the event so that stations in a direction perpendicular to the mislocation vector see no change in epicentral distance while stations in the direction of mislocation will see the full value. Assuming a cosinusoidal dependence as a function of azimuth suggests that, on average, the error in epicentral distance is $\simeq \epsilon_X / \sqrt{2}$. Assuming that δx and δy do not co-vary (as suggested by an analysis of the differences of the NEIC and ISC locations), the error in the travel time due to the error in each of x and y is

$$\frac{p\epsilon_X}{\sqrt{2}},$$

where p is the ray parameter. It is well known that errors in depth and origin time do covary with $\delta t_0 \simeq \delta z / 9$ (depth in kilometers). Since $\partial t / \partial z$ is negative, the errors in origin time and depth tend to cancel in their contribution to the total error and the errors in x and y dominate the error budget. We now assume a typical depth uncertainty of about 10 km and find that σ_X is .6–1.2 seconds for P waves at epicentral distances of about 70° for mislocation vectors of length 10–20 km. The corresponding estimate of σ_X for S waves is 1.6–2.5 seconds. As we shall see below, these numbers rival the signals from 3D structure.

These results mean that we cannot ignore earthquake mislocation in our tomography and we must either relocate events or make our data insensitive to event location. Consider the travel time residuals for one event:

$$\delta \mathbf{t} = \mathbf{A} \delta \mathbf{h} + \mathbf{B} \delta \mathbf{v} \quad (7.51)$$

We could iteratively solve this equation first by relocating the events then solving separately for velocity structure then relocating again but now including the new velocity structure. Convergence is usually attained after a few iterations. Alternatively, we can seek linear combinations of the data for each event which, to first order, are insensitive to the event location. This reduces to finding \mathbf{P} such that

$$\mathbf{P} \delta \mathbf{t} = \mathbf{P} \mathbf{A} \delta \mathbf{h} + \mathbf{P} \mathbf{B} \delta \mathbf{v} = \mathbf{P} \mathbf{B} \delta \mathbf{v} \quad (7.52)$$

i.e., we want $\mathbf{P} \mathbf{A} = 0$. Note that if \mathbf{A} has the SVD $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$ then $\mathbf{P} = \mathbf{G} (\mathbf{I} - \mathbf{U} \mathbf{U}^T)$ where \mathbf{G} is any matrix. We choose \mathbf{G} so that the new data $\delta \mathbf{t}' = \mathbf{P} \delta \mathbf{t}$ are statistically independent. If $\delta \mathbf{t}$ has a covariance matrix \mathbf{I} then $\delta \mathbf{t}'$ has covariance matrix $\mathbf{G} (\mathbf{I} - \mathbf{U} \mathbf{U}^T) \mathbf{G}^T$ (since $(\mathbf{I} - \mathbf{U} \mathbf{U}^T) = (\mathbf{I} - \mathbf{U} \mathbf{U}^T)^T$ and $(\mathbf{I} - \mathbf{U} \mathbf{U}^T) (\mathbf{I} - \mathbf{U} \mathbf{U}^T) = (\mathbf{I} - \mathbf{U} \mathbf{U}^T)$). Thus, if $(\mathbf{I} - \mathbf{U} \mathbf{U}^T)$ has the eigenvalue decomposition $\mathbf{R} \mathbf{\Omega} \mathbf{R}^T$ then choosing $\mathbf{G} = \mathbf{\Omega}^{-\frac{1}{2}} \mathbf{R}^T$ leads to the desired covariance matrix which is \mathbf{I} . It is interesting that the eigenvalues of $(\mathbf{I} - \mathbf{U} \mathbf{U}^T)$ are one or zero and we lose four eigenvalues during the projection process – we have effectively used up four data to remove sensitivity to location.

The alternative process is to relocate initially, solving

$$\delta \mathbf{t} = \mathbf{A} \delta \mathbf{h} \quad (7.53)$$

which, if we have used a SVD would lead to a mislocation vector

$$\delta \hat{\mathbf{h}} = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{U}^T \delta \mathbf{t} \quad (7.54)$$

and equation 48 would become

$$\delta \mathbf{t} - \mathbf{A} \delta \hat{\mathbf{h}} \equiv (\mathbf{I} - \mathbf{U} \mathbf{U}^T) \delta \mathbf{t} = \mathbf{B} \delta \mathbf{v} \quad (7.55)$$

Note this is similar to the projection method where $\mathbf{P} = \mathbf{G}(\mathbf{I} - \mathbf{U} \mathbf{U}^T)$ except that we have not taken account of the fact that we have "used some data up" in doing the relocation and we have not consistently operated on the \mathbf{B} part of equation 48 as we did with the projection method. Ignoring these niceties does still leave us with a \mathbf{B} matrix which is sparse whereas, in the projection method, each new travel time is a linear combination of all the travel times for that event so that $\mathbf{P} \mathbf{B}$ is no longer as sparse as we would like.

9. Interpretation of long-period body wave travel times

We are almost ready to invert our travel times – ideally, we would like to incorporate both short- and long-period data in our inversions but, first, we must think about the relation between the two.

In our previous tomographic studies, we have noted a significant offset between the travel times of long-period data and those of traditional 1Hz waves. Our P data are offset from the predictions of PREM (using NEIC locations) by about 4 seconds. This is reduced to less than 2 seconds if we use consistently determined locations and 1D model (e.g. the EHB locations and model AK135: Engdahl et al, 1998, Kennett et al, 1995). This remaining offset is a clear function of moment of the event. Fig 3 shows the mean travel time residual per event binned by moment of the event (red line). There is a clear linear relationship between the offset and log moment, and the offset is almost zero at a moment of 10^{24} dyne cm – this implies agreement with the ISC data from which AK135 was built. If we correct for lateral heterogeneity using a 3D model, we find that AK135 is too fast by about a second – this is because of the preponderance of fast paths in the ISC data set – and we have an offset of 1 second at a moment of 10^{24} dyne cm (black line in fig. 3).

The linear relationship is not that expected from the usual source scalings (e.g. Kanamori and Anderson, 1975) where source time is expected to go like $(\text{moment})^{1/3}$. This relationship is clearly seen in the centroid times of the CMT solutions. If we plot the centroid time (corrected for a 2.5 second shift to account for the shift induced by using NEIC locations with the PREM model) versus log moment, we find the expected $(\text{moment})^{1/3}$ variation with a 1 second shift at a moment of 10^{24} dyne cm (Ekstrom, personal communication). This relationship is plotted on Fig. 3 as the green lines. Clearly, our measured offset agrees with the predicted shift from the centroid times at relatively low moments but diverges for the bigger events. This behavior is to be expected when looking at band limited data and is accurately predicted using measurements from synthetic seismograms including source time functions based on the CMT centroid times.

If we empirically correct for the source offset and for 3D structure, we again find that AK135 gives quite a good fit to the long-period teleseismic P data except for systematic deviations at both long and short distances (Fig 4). Note that deep events behave a little differently and a different source time/moment relationship is needed that predicts shorter offsets than for shallow events implying that deep events are systematically "faster" than shallow events.

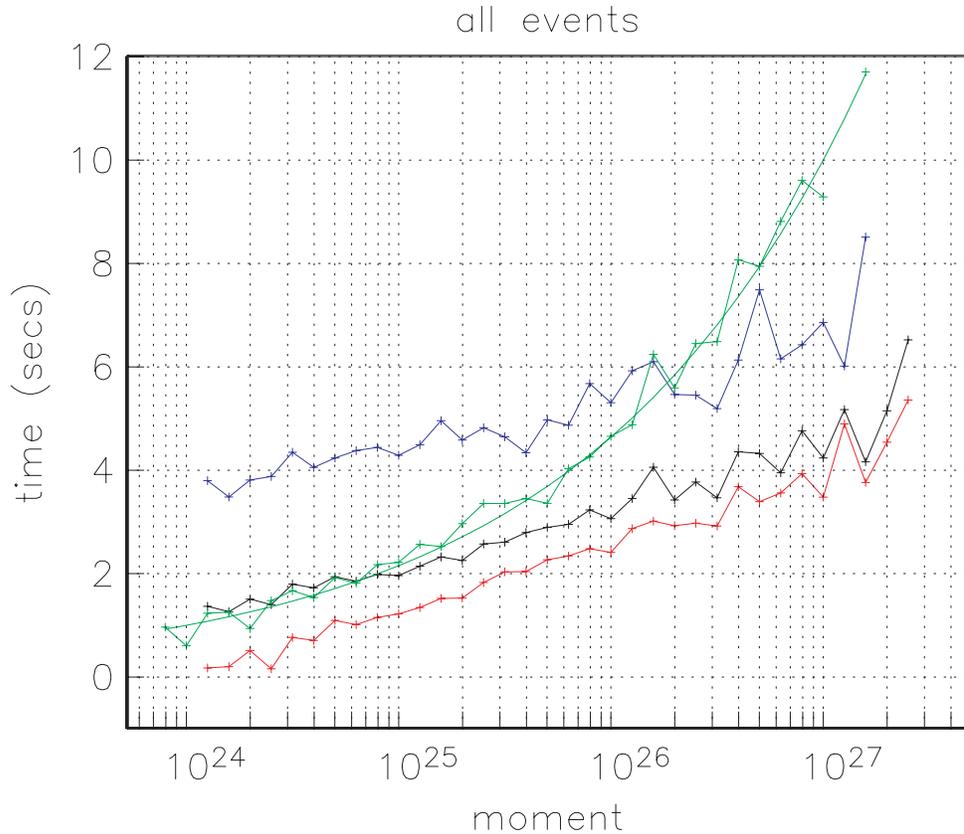


Figure 3. Mean residual offset (relative to AK135) per event plotted as a function of event moment. The red line is for raw P data using the EHB locations. The black line is for P data but corrected for 3D structure. The green line is the empirical $(\text{moment})^{1/3}$ expected for the source process time. The blue line is the same as the black line but now for S.

Also shown on Fig. 3 are the average S residuals per event plotted as a function of binned moment (blue line). As perhaps should be expected, the slope of the relationship is the same as that for P but there is an offset. Clearly, differential S-P times are insensitive to source process time (to first order) and can be used to isolate other effects. For example, the difference between S and P shown in Fig. 3 may be due to physical dispersion or it may be due to imperfections in AK135.

The reader may be wondering what all these source effects have on the recovery of 3D structure. All successful tomographic inversions include some treatment for the relocation of earthquakes (as discussed above) and relocation takes care of much of the effect we have just described. Even a directivity effect can be largely modelled by a relocation of the event. Of course, the locations derived from short and long period data will be different and so separate locations must be determined.

10. Importance of the crust

see appendix

11. Finite frequency kernels and Born approximation

maybe next time....

EHB locations with 3D and source corrections

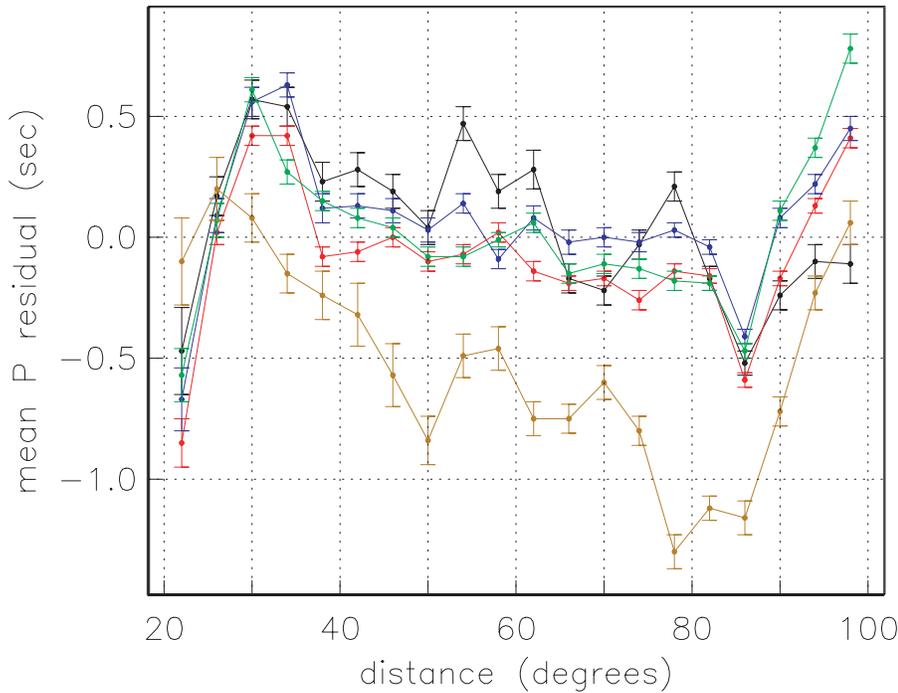


Figure 4. Average residuals relative to AK135 after correction for 3D structure and an empirical source size correction (taken from the black line in Fig. 2). The data have been grouped by event depth. Black (0–15km); red (15–35km); blue (35–55km); green (55–300km); gold (300–800km).

12. Anisotropy

maybe next time...

The following appendix is the text of a paper published in 2000 which describes some inversions of all the datasets considered in this class: mode structure coefficients, surface wave phase velocity, and P and S absolute and differential travel times. Combining all these data sets together allows us to look at joint models of compressional and shear velocity (or, more conveniently, bulk sound speed and shear velocity) which allows us to investigate the physical origin of some of the seismic velocity anomalies in the mantle.

There are also some comparisons to models in the literature. In the intervening six years, there is a growing consensus about the general size and shape of seismic velocity anomalies in the mantle and the earlier models based on ISC data (particularly ISC S data) have fallen somewhat into disfavor. The transition zone of the Earth remains one of the hardest areas to image since the upper mantle is controlled by surface waves and the lower mantle by body waves – this means that the TZ shows the least correlation between the various models in the literature

1. Appendix: The relative behavior of V_s , V_c , and V_p in the mantle

Abstract

Despite immense progress in imaging seismic velocity anomalies in the mantle over the past 15 years, we still know relatively little about their physical cause. One way to shed some light on this problem is to investigate the *relative* amplitudes of compressional and shear velocity anomalies in the mantle. Unfortunately, the amplitudes of velocity anomalies can be quite sensitive to the data sets and imaging techniques employed. It is therefore usually meaningless to take two models from the literature and do a simple comparison. In this paper, we consider joint modeling of P and S data sets and compare with some recent results from the literature. Some robust patterns are beginning to emerge which allow us to identify regions of the lower mantle which are anomalous. Such regions seem to be associated with large-scale upwelling in the mantle and may indicate chemical interaction with the core.

2. INTRODUCTION

During the past 15 years, our view of the seismic structure of the interior of the Earth has improved immensely. There is now general consensus on the gross 3D structure of the mantle: the upper mantle and lowermost mantle are characterized by large amplitude, long-wavelength structure while the mid-mantle contains low-amplitude slab-like fast features which surround broad regions of slow velocities.

Of course, we wish to go beyond the stage of merely mapping the mantle and proceed to the next step of interpreting the maps in terms of the thermal and chemical structure. One way of doing this is to investigate the *relative* behavior of the compressional and shear velocities in the mantle since we have a reasonable understanding of how this should vary for purely (sub-solidus) thermal effects from laboratory experiments.

To do this, we need to have good estimates of the *size* of velocity anomalies in the mantle as well as their shape. The size of anomalies can often be quite poorly determined and can depend on the data sets used as well as details of the inversion procedure. Usually, data sets have very uneven sampling characteristics, and to take advantage of the densest sampling, tomographic models are often grossly overparameterised. This can lead to instability in the inversion and to strong dependence on regularizing criteria. A common consequence is that amplitudes are underestimated. Often, the relative sampling of S and P data sets can be quite different so the degree of smoothing for each velocity type is different resulting in spurious values for their relative behavior.

We believe that one promising approach is to use many different data types such as mode structure coefficients, surface wave dispersion, and absolute and differential travel times, jointly in the inversions. To give a feel for what can be reliably determined from such data, we spend a significant fraction of this paper discussing different data types and the patterns of heterogeneity that they reveal. We also compare long- and short-period body wave data sets to convince ourselves that these data do not "see" fundamentally different Earths.

We follow this with a discussion of high-resolution modeling where we find that such inversions are relatively stable with no strong dependence on inversion technique provided the systems to be solved are reasonably well-conditioned. Some experiments in joint modeling follow where we concentrate on joint inversions of shear velocity and bulk sound speed. When we compare our results with those from the literature, we find good agreement with an inversion which uses multiple data types in a similar fashion to our experiments but poor agreement with inversions based primarily on ISC travel time data. We speculate on reasons for the disagreements and give arguments in favor of those models with larger velocity anomalies in the mid mantle and transition zone. We conclude with some preliminary results about the relative behavior of shear and compressional velocity in the mantle and speculate about their cause.

3. THERMODYNAMIC BACKGROUND

We shall be considering three seismic velocities: the compressional velocity, v_p , the shear velocity, v_s , and the bulk sound speed, v_c . These velocities are related to the bulk modulus, K_s , the rigidity, μ , and the density, ρ , by

$$K_s = \rho v_c^2 = \rho(v_p^2 - \frac{4}{3}v_s^2); \quad \mu = \rho v_s^2 \quad (1)$$

It therefore follows that

$$\frac{\delta v_p}{v_p} = \gamma \frac{\delta v_s}{v_s} + (1 - \gamma) \frac{\delta v_c}{v_c}; \quad \gamma = \frac{4}{3} \frac{v_s^2}{v_p^2} \quad (2)$$

where $\gamma = 0.40 \pm 0.02$ for plausible models of the mantle. Clearly, the variation in compressional velocity is almost as sensitive to variation in shear velocity (rigidity) as to variation in bulk sound speed (bulk modulus). Several studies have used the relative variation in shear and compressional velocity as a diagnostic of the physical cause of an anomaly but it may be better to use the relative variation of shear velocity to bulk sound speed. We therefore define:

$$R = \frac{\partial \ln v_s}{\partial \ln v_p}; \quad \zeta = \frac{\partial \ln v_s}{\partial \ln v_c} \quad (3)$$

These are easily shown to be related by

$$R = \frac{\zeta}{\zeta\gamma + 1 - \gamma}; \quad \zeta = \frac{(1 - \gamma)R}{1 - \gamma R} \quad (4)$$

An advantage of using the logarithmic ratio, R , is that its value, as determined by laboratory experiments, varies little from material to material. *Anderson and Isaak* [1995] summarize the high temperature experimental results for a wide variety of materials and find R between 0.9 and 1.6 (Table 1). This result has been the basis for using $R \approx 1.3$ in seismic studies where it is necessary to scale between shear and compressional velocities [*Masters et al.*, 1982; *Forte & Peltier*, 1987; *Woodward & Masters*, 1992; *Masters et al.*, 1996]. ζ has a mean value of about 1.5 but it varies from material to material more than R does. For this reason, we will focus on R in the following.

Table 1. High T experimental data, Θ is the Debye temperature

Substance	R	ζ	T/Θ
NaCl	1.2	1.3	2.0
KCl	1.2	1.3	2.0
MgO	1.2	1.6	1.5
CaO	1.2	1.5	1.5
MnO	1.6	2.4	1.0
Al ₂ O ₃	1.4	2.2	1.5
Mn ₂ SiO ₄	1.2	1.3	1.3
Co ₂ SiO ₄	1.0	0.9	1.3
Fe ₂ SiO ₄	0.9	0.9	1.2
Mg ₂ SiO ₄	1.2	1.4	2.0
Olivine	1.2	1.4	2.0
Pyrope	1.1	1.0	1.3
Grossular	1.3	1.8	1.5
Mg ₂ Al ₂ O ₄	0.9	0.8	1.1

Karato [1993] has pointed out that the laboratory value for R may not be appropriate for application to the Earth since anelastic as well as anharmonic effects are important. In real materials, relaxation processes occur which affect the seismic velocity and make it dependent on frequency. Since relaxation processes are usually thermally activated, a change in temperature changes the characteristic frequency of the process and so changes the elastic velocities. For a frequency-independent Q , the seismic velocity depends on frequency and temperature (T) as [*Kanamori and Anderson, 1977*]

$$v(\omega, T) = v_0(T) \left[1 + \frac{1}{\pi Q} \ln(\omega\tau) \right] \quad (5)$$

where v_0 is a reference velocity corresponding to the unrelaxed state and the relaxation time, τ , is a function of temperature of the form

$$\tau(T) = \tau_0 \exp(H^*/R_g T) \quad (6)$$

where H^* is the activation enthalpy and R_g is the gas constant. Differentiating equation (5) with respect to temperature gives

$$\frac{\partial \ln v}{\partial T} = \frac{\partial \ln v_0}{\partial T} - \frac{1}{\pi Q} \frac{H^*}{R_g T^2} \quad (7)$$

where attenuation has been assumed to be weak ($Q \gg 1$). The first term in equation (7) corresponds to the anharmonic effect while the second term represents anelastic effects. The importance of the latter will be greatest in zones of high attenuation where Q is low. The size of the effect is also strongly dependent on the activation enthalpy, H^* , which is poorly known in the lower mantle. Using data for olivine, *Karato* [1993] estimates that R (which is 1.2 in the infinite frequency limit) will increase to about 1.6 for a Q_μ of 100 and to 1.8 for a Q_μ of 50. He also estimates that this effect will be important throughout the mantle giving R values of the order of 1.7 (though the estimate is very uncertain).

Several of the seismic results described below suggest an R which is higher than that found in laboratory experiments and this fact has stimulated some theoretical work on the properties of lower mantle phases. *Agnon & Bukowinski* [1990] and *Isaak et al.* [1992] show that a value of $R \simeq 2 - 2.5$

for a possible lower mantle constituent (MgO) is consistent with a theoretically predicted, pressure-induced decrease in the Anderson-Grüneisen parameter $\delta_S = -(1/\alpha K_s) (\partial K_s / \partial T)_P$. (Here α is the coefficient of volume thermal expansivity, K_s is the adiabatic bulk modulus, and P is pressure.) A more recent analysis by *Karato et al.* [1999] again indicates higher values for R for purely thermal effects with a value of 1.8 at the top of the lower mantle rising to perhaps as much as 2.2 at the base of the mantle. Such values are still too low to cause bulk sound speed to be negatively correlated with shear velocity (equation 4) but this is a feature of several lower mantle models (see below).

In the next section, we consider some of the data sets that go into the construction of tomographic models. In particular, it is of interest to make rough estimates of R from the data using

$$R \approx \frac{v_s \delta t_s}{v_p \delta t_p} \quad (8)$$

where δt_s and δt_p are perturbations to the spherically averaged body wave travel times. This relationship between R and the travel time perturbations is derived from Fermat's principle and assumes that the perturbations in both the P and S seismic velocities which cause the travel time anomalies are co-located [*Souriau & Woodhouse, 1985*].

4. DATA TYPES

7.5 Body Wave Travel Times. Historically, travel times have been measured from short-period vertical-component instruments on which P arrivals typically have a dominant period of 1 s and S arrivals have a dominant period of about 4 s. Many such instruments distributed around the world still record such data. The arrival times are measured locally and then compiled and distributed by the International Seismological Centre (ISC). These data are published in the Bulletins of the International Seismological Centre and subsets of this data set (typically the first-arriving P wave times which now number several million) have been used to refine existing 1-dimensional models [*Dziewonski & Anderson, 1981; Kennett & Engdahl, 1991; Morelli and Dziewonski, 1993; Kennett et al., 1995*] and in tomographic studies of 3-dimensional structure [*Clayton & Comer, 1983; Dziewonski, 1984; Creager & Jordan, 1986; Morelli et al., 1986; Morelli & Dziewonski, 1987; Shearer et al., 1988; Inoue et al., 1990; Pulliam, 1991; Pulliam et al., 1993; Vasco et al., 1993, 1994, 1995; Vasco and Johnson, 1998; Su et al., 1994; Su and Dziewonski, 1997; Van der Hilst et al., 1997; Zhou, 1996; Kennett et al., 1998; Bijwaard et al., 1998*].

The strength of the ISC data set is its size but the fact that it is composed of readings from many different instruments (of varying quality) made by many different operators means that it is a very noisy and inhomogeneous data set. *Gudmundsson et al.* [1990] estimate the signal to noise (S/N) ratio of the ISC P -data to be less than 1.0 at local and regional distances and approximately 2 for teleseismic P . A major reprocessing of the ISC data set has been undertaken by *Engdahl et al.* [1998] (hereafter EHB) with particular attention paid to the problem of misidentification of depth phases for direct phases. It is this reprocessed data set that has been used in the most recent tomographic modeling [*Van der Hilst et al., 1997; Bijwaard et al., 1998; Kennett et al., 1998*].

ISC S wave times are typically very noisy because their low-frequency character is not well-recorded by short-period instruments and phase misidentification (particularly with SKS at distances beyond 80 degrees) is common since picking is typically done only from vertical component recordings. This makes it extremely difficult to use the ISC data set to do a comparative study of the lateral variations in P and S velocity though there have been a few recent efforts [*Robertson & Woodhouse, 1995, 1996; Kennett et al., 1998; Vasco & Johnson, 1998*].

Since the mid 1970's, digital recordings of relatively long-period body waves (about 20 s dominant period) have become available and have been analyzed in many different ways. The waveforms can

be directly modeled for aspherical structure using a variety of theoretical approximations [Woodhouse & Dziewonski, 1984; Dziewonski & Woodhouse, 1987; Tanimoto, 1990; Li and Romanowicz, 1995, 1996; Su & Dziewonski, 1991]. Alternatively, absolute and differential travel times can be estimated directly from the data and have been used in more traditional tomographic studies [Woodward & Masters, 1991a,b,1992; Masters *et al.*, 1996].

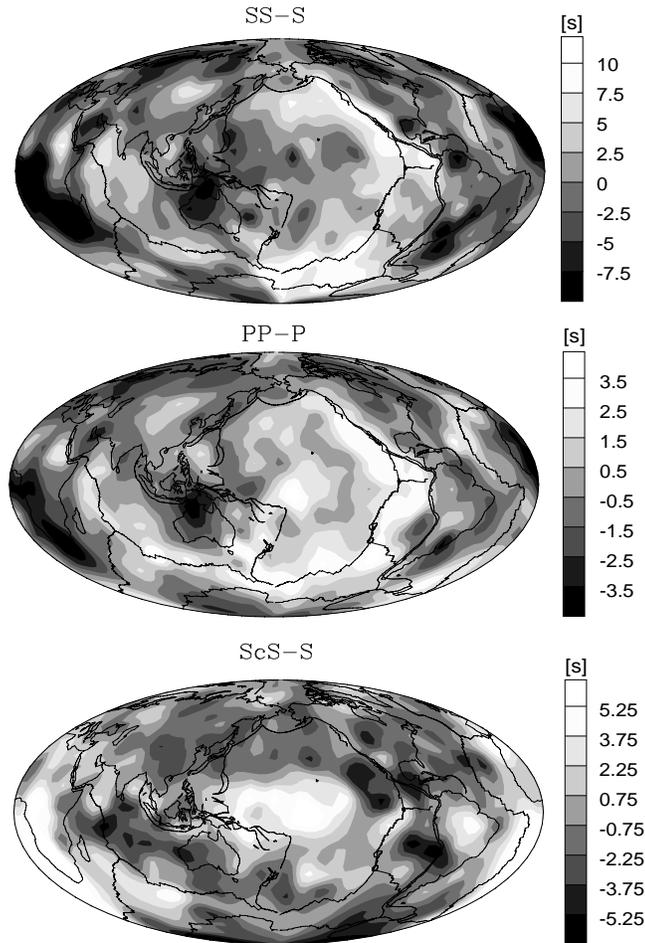


Figure 1 Travel time residual patterns for $SS - S$, $PP - P$, and $ScS - S$ differential times. These maps are made by lightly smoothing the pattern of residuals obtained by plotting the residual at the bouncepoint of SS , PP , and ScS respectively. The $SS - S$ and $PP - P$ maps are highly correlated (though there are significant differences) with an $SS - S$ residual being typically 3 times a $PP - P$ residual. Many of the features seen in the $SS - S$ and $PP - P$ maps are clearly related to near-surface tectonics while the $ScS - S$ map is quite different.

The fact that the long-period data have longer wavelengths than the waveforms used for the ISC data set has both advantages and drawbacks. An advantage is that it is possible to construct data sets with little spatial aliasing though, of course, resolving short-wavelength structure is much more difficult. Another

advantage is that the long-period data show few of the effects of scattering so evident in short-period data. This makes it very easy to pick secondary phases since codas are practically nonexistent. Another advantage is that both P and S long-period waveforms have similar frequency content as this is now controlled by the instrument response rather than the attenuation characteristics of the Earth. This makes the effects of physical dispersion much easier to assess. Currently we have measured approximately 41,000 S and 38,000 P long-period travel times at teleseismic distance ranges [Bolton & Masters, 1999], 18,000 $SS - S$ differential times, 12,000 $PP - P$ times, and 8000 $ScS - S$ times. Fig. 1 illustrates the residual patterns for $SS - S$, $PP - P$, and $ScS - S$. These maps are made by lightly smoothing the pattern of residuals obtained by plotting the residual at the bouncepoint of SS , PP , and ScS respectively. The $SS - S$ and $PP - P$ maps are highly correlated (though there are significant differences) with an $SS - S$ residual being typically 3 times a $PP - P$ residual [Woodward and Masters, 1991a]. We take this as representative of upper mantle structure (leading to an R value of about 1.7). Many of the features seen in the $SS - S$ and $PP - P$ maps are clearly related to near-surface tectonics while the $ScS - S$ map is quite different. Here, we see slow areas under Africa and the central Pacific with a ring of fast residuals around the Pacific. Not surprisingly, these features are reproduced in the lowermost mantle of almost all recent tomographic models.

Crustal structure, particularly structure directly under the receiver can contribute quite strongly to the measured absolute times. One way to account for this effect is to apply “station corrections” (i.e., mean station residuals usually corrected for uneven azimuthal distribution) to the data. Such corrections have been calculated for the ISC data set by a number of authors [e.g., Dziewonski and Anderson, 1983; Toy, 1989]. Of course, such corrections remove more than just the signal of the crust, they also effectively remove the upper mantle and possibly some of the lower mantle as well. Since we are interested in structure throughout the mantle, we elect to not use station corrections but correct for the crust using the global crustal model, CRUST 5.1, of Mooney *et al.*, [1998]. It is, however, interesting to note that most studies of station corrections find a typical 3:1 ratio between the S and P station means in agreement with the $SS - S/PP - P$ value discussed above.

Fig. 2 shows histograms of the S and P residuals binned by ray turning depth. This figure shows the remarkable increase in variance for S residuals as one samples deeper in the mantle which is apparently absent in the P data. It is noteworthy that the ISC defines a teleseismic S as anything within 7.5 seconds of the Jeffreys-Bullen (JB) predicted times. It is quite clear from Fig. 2 that long-period S residuals span a much larger range than this, and this is also true of short period S [Robertson and Woodhouse, 1995]. Consequently, any inversion which restricts attention to the ISC-defined S arrivals will underestimate the size of S anomalies in the mantle.

The increase in variance in S residuals with ray turning depth is mainly due to the appearance of very slow S residuals for rays which bottom under the central Pacific and under Africa. We illustrate this in Fig. 3 where we bin rays which have similar turning points and plot the resulting smoothed residuals at the turning point. The slow areas in the S map are much more pronounced than in the P map – this suggests that we may have to appeal to an effect such as partial melting to explain the relative behavior of P and S in these regions. If we plot co-located P and S mean residuals against each other for a variety of turning depth ranges, we obtain Fig. 4. Slopes are calculated by solving the “Least-Squares Cubic” [York, 1966] for the best fitting line. For the rays turning between 670-1100km we find a slope of about 3.3 ($R \sim 1.8$), in close agreement with that found for the long-period $SS - S/PP - P$ ratio. Between 1100-2000km the slope increases to about 4 ($R \sim 2.2$). We see a steady increase in the slope starting at about 2000km depth. For the deepest turning rays (2600-2890km), we see a further increase to a slope of about 7 ($R \sim 3.8$). Clearly, such high values of R are very different from what we expect for normal thermal effects.

7.6 Comparing Long and Short Period Data . In the following sections, we will find that there are

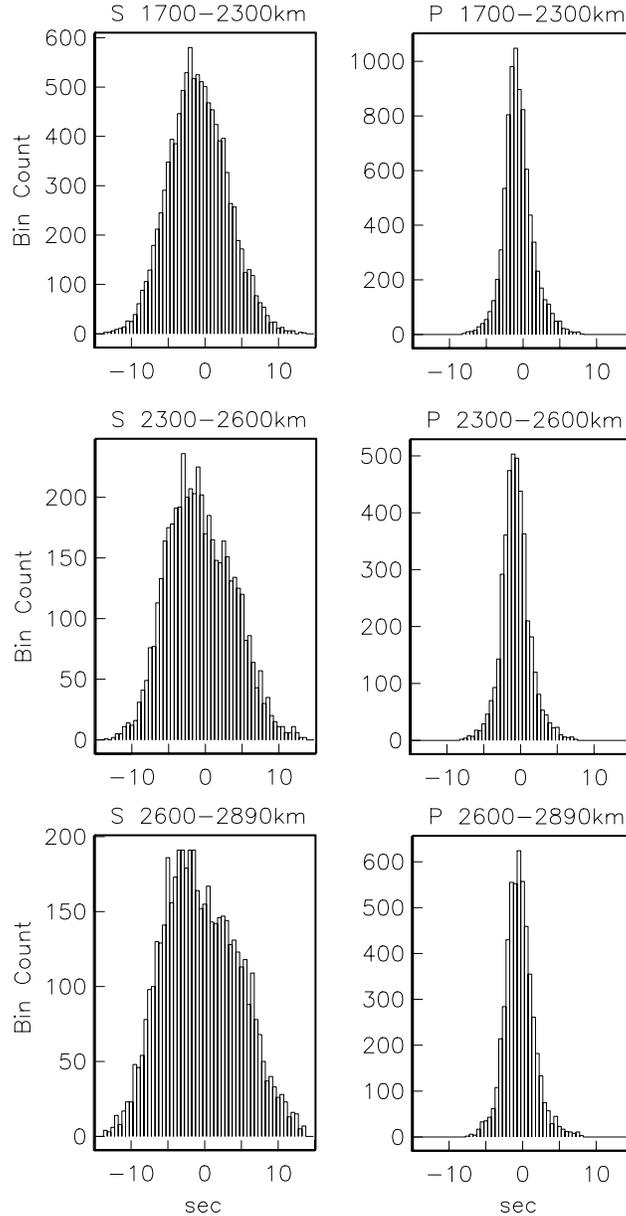


Figure 2 Histograms of the detrended P and S residuals as a function of ray turning depth. Note the significantly larger variance for S compared to P residuals. The variance of S residuals increases with depth where a good fraction of the data have a signal of more than 7.5 sec, the cut-off for S picks in the ISC catalog.

quite large differences in the amplitudes of S and P anomalies in the mantle determined by different researchers. It is of interest to see if this is due to fundamental differences in the data sets employed, or if it is a product of data selection and/or modeling technique. To clarify the comparison, we restrict attention to stations which report short period data to the ISC and also record broad-band digital data from which our long-period picks have been made. We form “summary rays” by grouping residuals from clusters of earthquakes and then plot the median residual of the summary ray at its turning point. All

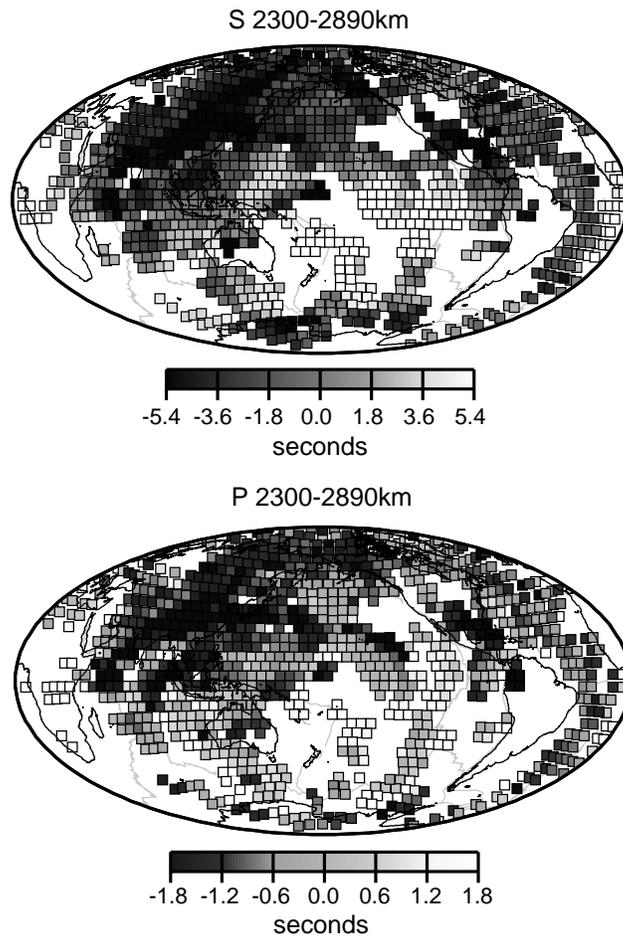


Figure 3 5° spherical cap averages of our detrended S and P residuals plotted at their ray turning points for rays which bottom in the lower 600 km of the mantle. The regions of slow S residuals under the Pacific and Africa are clearly visible though these are more muted in the P map.

residuals are computed using the EHB locations. Fig. 5 shows a comparison of the resulting smoothed residual patterns for station MAT/MAJO. This station is unusual in that it has a relatively large number of S picks (~ 4000) as well as P picks. The patterns in both S and P are very similar though the amplitude of the patterns in the short period data are a little smaller than in the long period data. It is interesting to explore the internal consistency of the data by looking at the variability within each summary ray. For summary rays with more than a few tens of measurements, we find that the ISC P data have a standard deviation of about 0.4–0.5 seconds while the long-period data are about twice this. The situation is reversed for S where the standard deviation of a group of rays constituting a summary ray is about 1.3 seconds for the long-period data and about twice this for the ISC data. Also, as noted above, there are few ISC S data beyond 80 degrees and these tend to be extremely noisy. MAT has a respectable number of S picks but other stations, e.g. NWA/NWAO, have only a few hundred S picks that are confined to a small azimuth range (Fig. 6). This suggests that studies which use ISC data and conservatively confine attention to only those paths that have common S and P will be restricted to using a relatively small portion of the data.

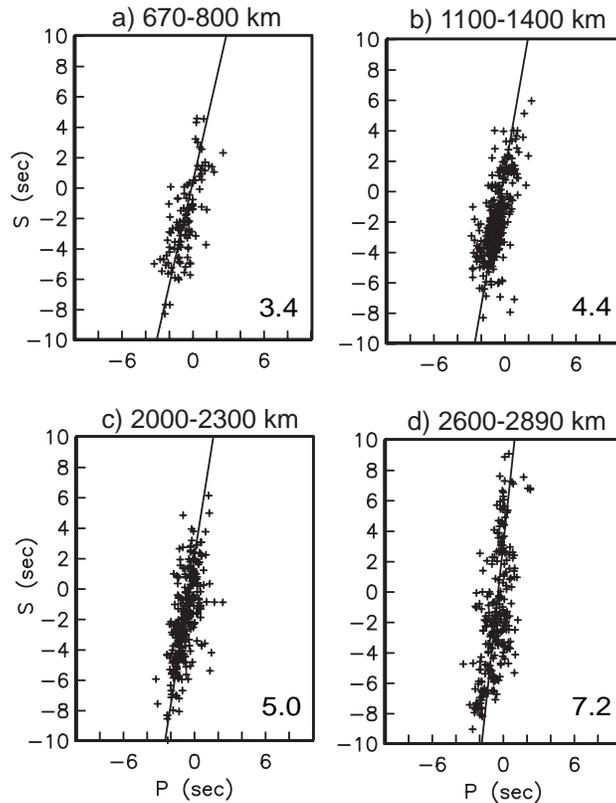


Figure 4 Scatter plots of S and P cap averages of residuals binned by turning depth (there are a minimum of 5 picks per cap). The ratio of the residuals (i.e., the slope) increases steadily with ray turning depth with a sharp increase for the deepest turning rays. The slopes for a-d are given in the lower right corner.

We conclude from this brief comparison that residual patterns in long and short period data are indeed similar and our analysis of errors suggests that the optimal approach may be to use ISC P data combined with long-period S data (as was done by *Su and Dziewonski*, [1997]).

7.7 Surface Waves . It is extremely difficult to constrain structure in the upper mantle and transition zone using body waves alone unless triplication phases of multiple surface bounce arrivals are included [Grand, 1994]. Such an analysis is delicate and time-consuming and a simpler way to constrain near-surface structure is to include surface waves.

Global surface wave phase velocity maps have been constructed by numerous workers by measuring surface wave dispersion [e.g. *Montagner & Tanimoto* 1990; *Laske and Masters*, 1996; *Ekström et al.*, 1997; *Trampert and Woodhouse*, 1995, 1996; *Zhang and Lay*, 1996] though some of these studies have also included polarization data to improve sensitivity to short wavelength structure. There is now good agreement between phase velocity maps for periods between 50 and 150 seconds up to relatively short wavelengths. An example is shown in Fig. 7.

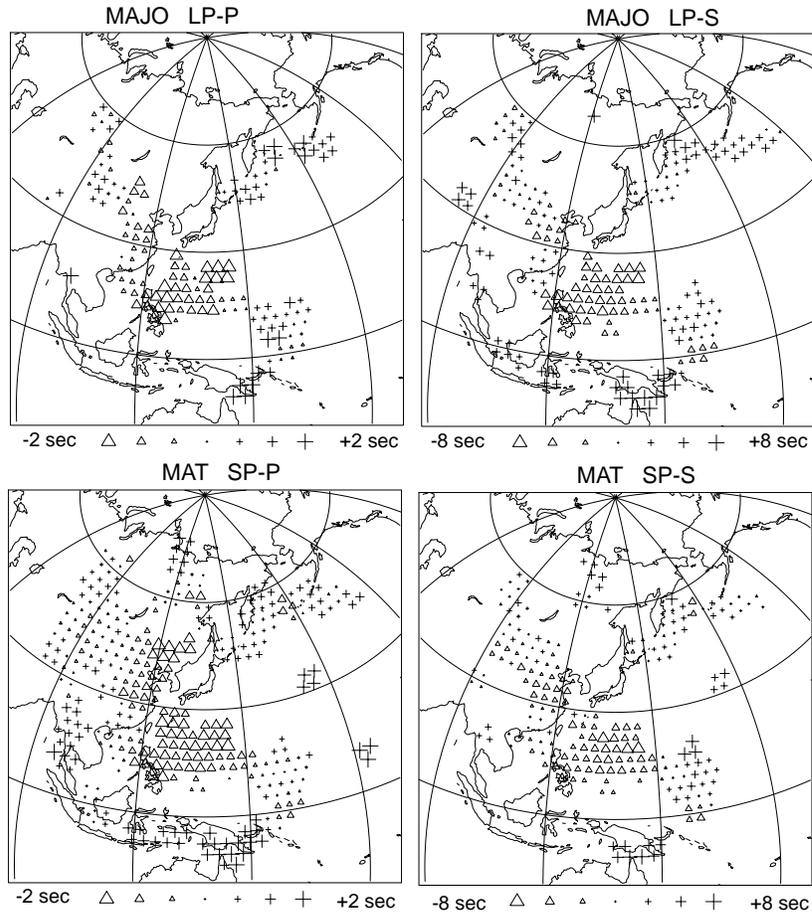


Figure 5 The medians of summary rays (see text) for the stations MAJO/MAT plotted at the surface projection of the summary ray turning points. MAJO is a broad band station, from which we have made long-period picks, co-located with MAT which is a short period station which reports to the ISC. There is extremely good correspondence between the long and short period picks though the coverage of the short period P is much better than for the long period data. There is also extremely good correlation between the patterns in P and S .

Global phase velocity maps are typically dominated by large-amplitude long-wavelength structure (Fig. 8). Though the spectral amplitudes of such maps are somewhat uncertain, it is probably fair to say that the maps of *Trampert and Woodhouse* [1996] are slightly too rough and that the amplitudes of *Zhang and Lay* [1996] are too low, hence placing reasonable error bounds on spectral amplitudes. The correlation of such maps at spherical harmonic degrees $\ell = 10$ and lower is typically well above the 99% confidence limit though it appears that the inclusion of higher orbit and great-circle dispersion data is needed to obtain results compatible with the mode structure coefficients (see below and Fig. 9). At harmonic degrees higher than $\ell = 16$, the correlation between maps of different workers is significantly reduced with the best correlation being between the maps of *Ekström et al.* [1997] and *Trampert and Woodhouse* [1995]. Both of these groups use a large amount of mainly minor arc phase data to obtain sensitivity to short wavelength structure. *Laske and Masters* [1996] have much less data though they use polarization data that give independent sensitivity to short wavelengths. It should be pointed out that

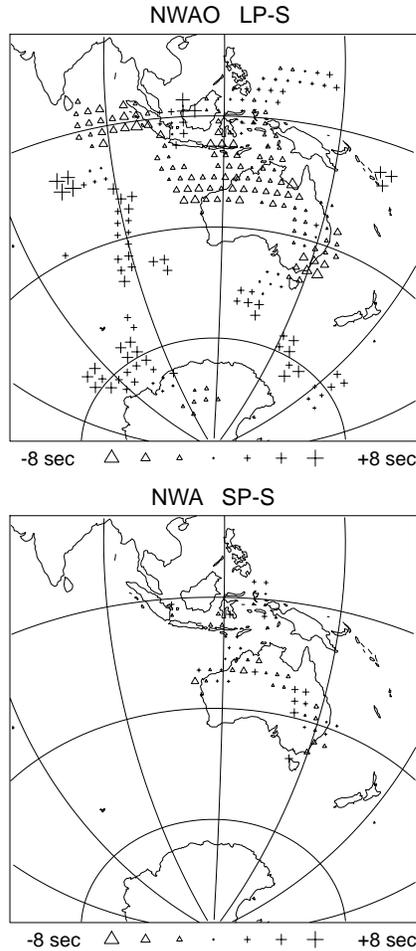


Figure 6 The medians of S summary rays for the stations NWAO/NWA plotted at the surface projection of the summary ray turning points. NWAO is a broad band station, from which we have made long-period picks, co-located with NWA which is a short period station which reports to the ISC. Note the small number of S picks for NWA and the small range of azimuth they cover.

the reduction in correlation is also accompanied by a reduction in amplitude so many of the differences at short wavelength are barely significant.

Surface waves have some sensitivity to density as well as to elastic velocities. The proper procedure in a tomographic study would be to include density as an additional free parameter in the modeling. The limited amount of data allows numerous trade-offs between (anisotropic) elastic velocities and density (see next section for a specific inversion) so most of our modeling assumes that density can be scaled to shear velocity using the relationship $dlv_s/dln\rho = 2.5$. The exact value has no impact on the inversion of Love waves (since these are virtually insensitive to density) but has a minor effect on the inversion of Rayleigh waves for structure in the top 250 km of the mantle.

7.8 Mode Structure Coefficients . Free oscillation structure coefficients [see e.g. *Ritzwoller et al.*, 1986, 1988; *Giardini et al.*, 1987,1988] provide constraints on the long-wavelength structure of the Earth with principal sensitivity to structure of even harmonic degree. This kind of data has been used previously to look at the relative variation of v_s to v_p [*Li et al.*, 1991] though the data sets available now are much more precise. In principle, these data are also sensitive to 3D density variations as well as to 3D elasticity though we shall find that the current data set is still not precise enough to make any

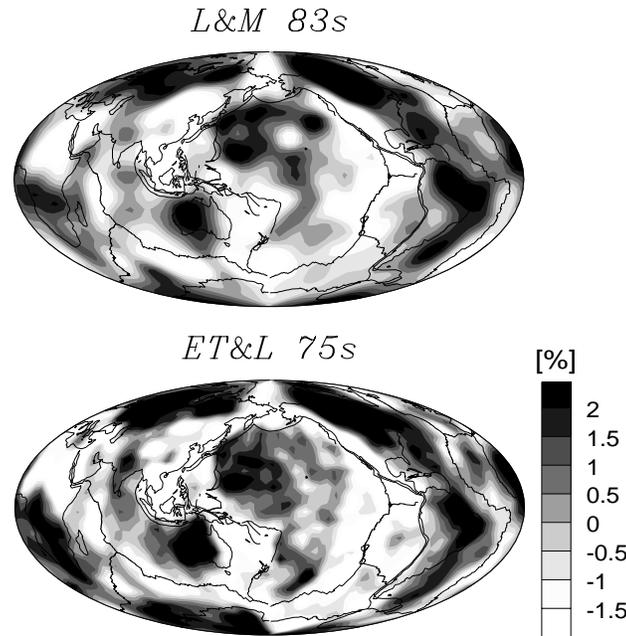


Figure 7 Phase velocity maps for Rayleigh waves at a frequency of about 12 mHz. We show maps of phase velocity perturbation with respect to the spherical average for two groups: L&M, *Laske and Masters* [1996]; and ET&L, *Ekström et al.* [1997]. At large scales, the maps are remarkably similar and show the typical features: seismically fast continental shields and old oceans and slow mid-ocean ridges. The perturbations are given in percent.

definitive conclusions about 3D density.

The analysis of free oscillations has seen something of a renaissance in the past few years because of the recent occurrence of several "great" earthquakes, including the great Bolivian earthquake of 1994 which was at a depth of 630km. We have combined data from several recent analyses [*He and Tromp*, 1996; *Resovsky and Ritzwoller*, 1998; *Um*, personal communication, 1996; *Masters et al.*, 1999] along with older peak-shift data [e.g. *Smith and Masters*, 1989] to give a new data set of structure coefficients of degrees 2,4 and 6 for 50 fundamental spheroidal mode, 26 spheroidal overtones, 31 fundamental toroidal modes, and 12 toroidal overtones. The structure coefficients of degree 2 are by far the best determined and also show the largest signal. Fig. 9 shows these coefficients for the fundamental spheroidal modes. It should be noted that the structure coefficients for surface-wave equivalent modes and coefficients in the expansion of phase velocity maps are not independent data. Indeed, for large harmonic degree, we find excellent agreement between structure coefficients estimated from travelling- and standing-wave analyses which gives us confidence in both data sets.

Given the interest in 3D density, it is useful to perform an inversion using the mode data alone. This inversion is parameterized by 18 shells in the mantle with shear velocity and bulk sound velocity expanded in spherical harmonics of degree 2,4, and 6. We scale perturbations in density to be proportional to perturbations in shear velocity. The resulting model is shown in Fig. 10 where it is clear that the bulk sound speed becomes increasingly negatively correlated with the shear velocity as one goes deeper in the mantle. This is a feature of all our joint inversions (see below). The model shown in Fig. 10 gives an excellent fit to the mode data and has an overall χ^2/N of 0.9 with no systematic misfit of any mode branch. Clearly, our current mode data set can essentially be fit with the assumption that

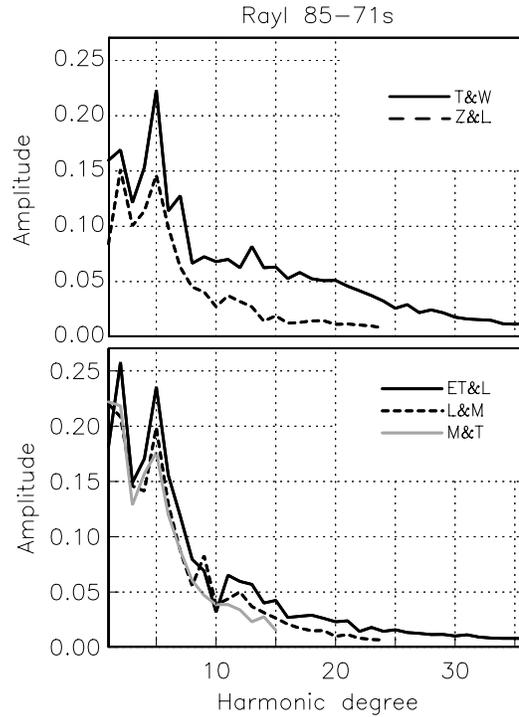


Figure 8 Amplitude spectra for Rayleigh wave phase velocity maps at around 12 mHz. The maps are: ET&L: *Ekström et al.* [1997] (75s); L&M: *Laske and Masters* [1996] (83s); T&W: *Trampert and Woodhouse* [1996] (80s); M&T: *Montagner and Tanimoto* [1990] (71s); Z&L: *Zhang and Lay* [1996] (85s).

density is proportional to shear velocity. This result contradicts that of *Ishii and Tromp* [1999] who find that density is negatively correlated with shear velocity in the lower mantle. To see if such a model is possible, we inverted the mode data again but now with the assumption that density is proportional to bulk sound speed (with a scaling corresponding to an Anderson-Grüneisen parameter, δ_S , of 2.5). Because the recovered bulk sound speed is still negatively correlated with shear velocity in the lower mantle, this model implicitly has a density that is also negatively correlated with shear velocity – now in agreement with the results of *Ishii and Tromp* [1999]. Furthermore, this model fits the mode data almost as well as the model with density proportional to shear velocity.

We take these results to imply that the current mode data set cannot distinguish between models in which density is positively or negatively correlated to shear velocity in the lower mantle (see also *Resovsky and Ritzwoller* [1999]). A further improvement in the precision of the structure coefficients will be required before 3D density can be resolved. The mode data do however robustly establish that bulk sound speed is negatively correlated with shear velocity in the lowermost mantle – at least at long wavelengths.

9. A HIGH-RESOLUTION INVERSION FOR SHEAR VELOCITY

We begin by inverting the long-period shear wave data sets, surface waves and structure coefficients described above for a so-called “high-resolution” shear-wave model of the mantle. This model is parameterized by 18 layers of roughly 100 km thickness in the upper mantle and transition zone and 200 km thickness in the lower mantle. Each layer is divided into equal-area blocks of dimension 4 degrees at the equator. Such a model has about 46,000 model parameters. This is many less than the number of

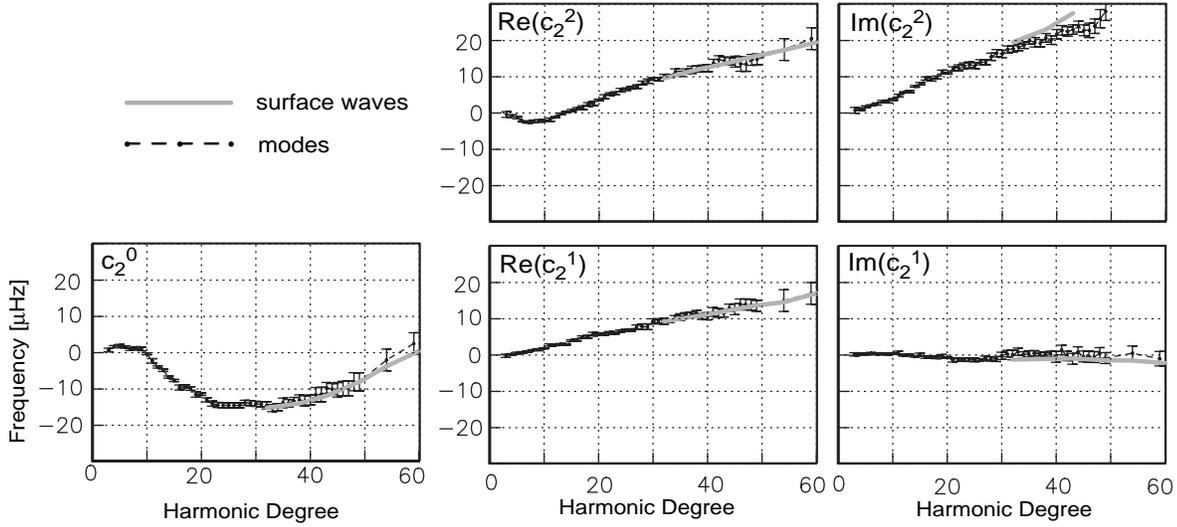


Figure 9 Degree 2 structure coefficients for fundamental spheroidal modes. The solid lines are estimates made from travelling wave studies at high harmonic degree. The agreement with the mode estimates is excellent.

data we are employing but there are still regions of the model that are not well sampled.

Fig. 11 shows how much each block is sampled by the body waves in the inversion. Bearing in mind that the surface waves constrain well the structure in the upper mantle, we find that the only poorly sampled regions are parts of the southern hemisphere in the lowermost mantle. The recent expansion of the global seismic network may allow us to fill these holes (our data sets include events from 1976 to 1994) but their present existence should be borne in mind when interpreting the results.

Earthquake mislocation is a major source of error in our datasets and it is important to treat this properly in the inversion. This is particularly true of teleseismic P data where mislocation effects are similar in size to the signal from 3D structure (a typical mislocation error is about 0.9 sec but the signal from 3D structure is only about 1 sec) but is less important for S waves. We handle mislocation by projecting the data for each event to make linear combinations of the original data which are, to first order, insensitive to the location of the event [e.g. *Masters et al.*, 1996]. Consider a vector of travel-time residuals for an event, $\delta\mathbf{t}$:

$$\delta\mathbf{t} = \mathbf{A} \cdot \delta\mathbf{h} + \mathbf{B} \cdot \delta\mathbf{v} \quad (9)$$

where $\delta\mathbf{h}$ is a four vector describing a perturbation in location and origin time of the event, and $\delta\mathbf{v}$ is a vector of velocity perturbations describing the 3D velocity structure. We operate on the equation above with a projection matrix \mathbf{P} chosen so that $\mathbf{P} \cdot \mathbf{A}$ is zero and the covariance matrix of the projected data $\mathbf{P} \cdot \delta\mathbf{t}$ is diagonal (so that the projected data are independent). This results in the loss of four data per event. We now have to solve the system

$$\mathbf{P} \cdot \delta\mathbf{t} = \mathbf{P} \cdot \mathbf{B} \cdot \delta\mathbf{v} \quad (10)$$

This process is somewhat extreme as it also projects out all structure which can be modeled by an

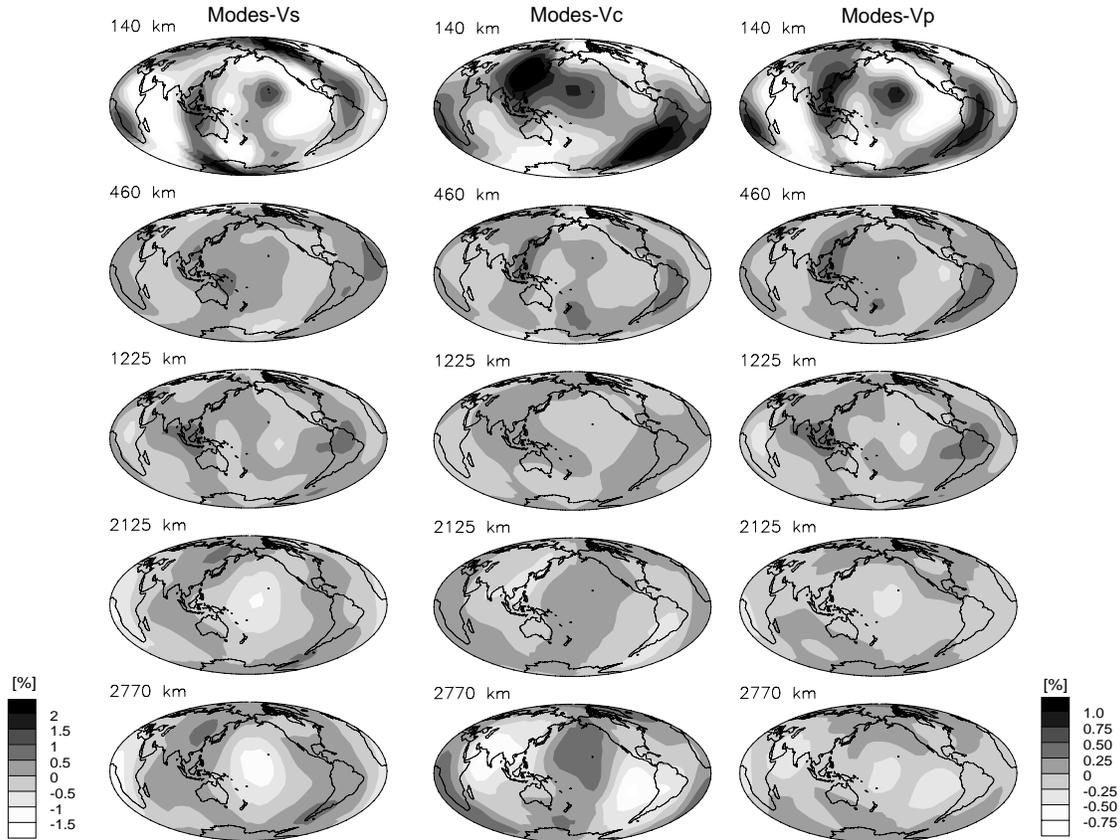


Figure 10 Cross-sections of the Earth at various depths showing perturbations of shear velocity, v_s , bulk sound speed, v_c , and compressional velocity, v_p (in percent). This model was constructed using the mode-structure coefficients alone. Note the strong anticorrelation of the bulk sound speed and shear velocity near the base of the mantle. The scale for v_c is the same as that of v_p .

equivalent shift in the source. We have found, however, that the results of S inversions vary little if we do the projection or not so the effects of mislocations are not dominant. (The same is not true for P inversions where the projection step can lead to quite different inversion results.)

Our inversions employ the LSQR algorithm as described by *Nolet* [1987] and we include first-difference smoothing both radially and laterally. There has been some controversy about the use of iterative inversion algorithms in seismic tomographic problems. Both SIRT [*Dines and Lytle*, 1979] and LSQR have been used in the geophysical literature though our experience has been that LSQR performs very well. As in any inversion, iterative techniques perform better when the system is well-constrained. Some recent inversions have attempted to recover over 250,000 block parameters with as little as 16,000 data. Such systems tend to converge poorly and, naturally, the results are a strong function of the smoothing employed. With the data sets at hand and our choice of parameterization, we have many more data than unknowns and convergence of the inversion is quite rapid and fairly insensitive to the degree of smoothing. In tests using synthetic data, we have found that the solution grows in amplitude and gradually asymptotes to the correct value as we iterate further. However, even after a few hundred iterations, it is possible for amplitudes near the bottom of the mantle to be slightly underestimated. This should be borne in mind when interpreting the results.

We show the results of two modeling experiments to highlight the importance of including surface-

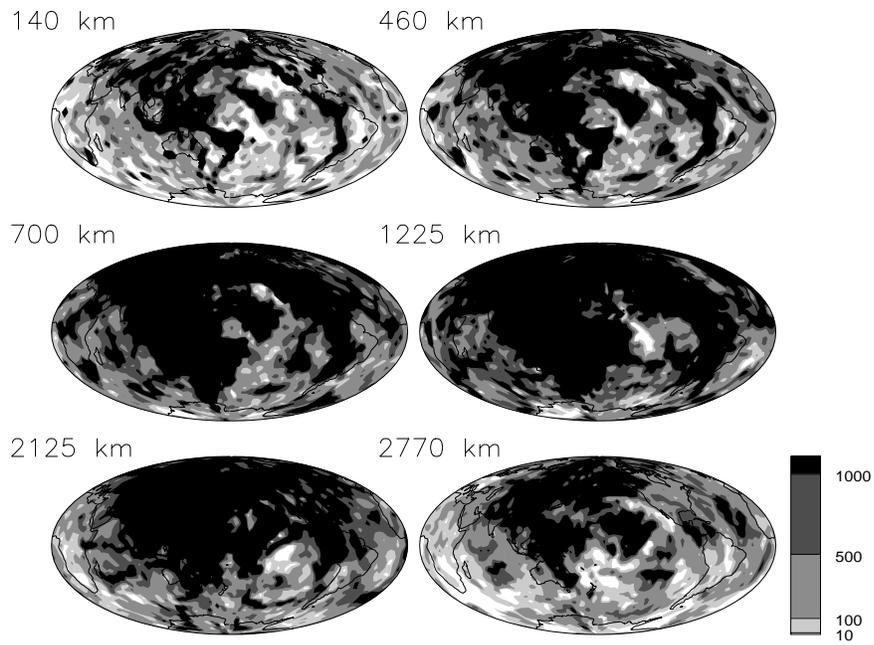


Figure 11 A graphical representation of the matrix connecting shear velocity perturbations to absolute and differential shear travel time data sets. Darker grays represent better sampling. Note that some parts of the lowermost mantle in the southern hemisphere remain poorly sampled.

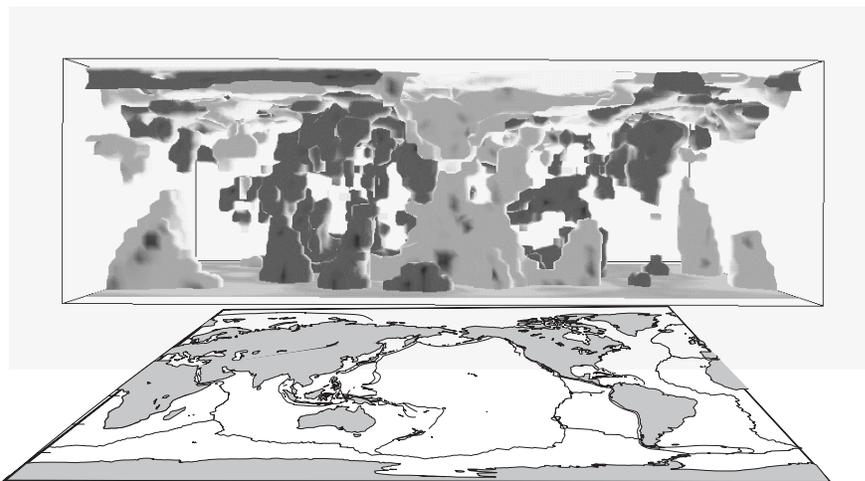


Figure 12 Isovelocity surfaces for values of -1.0% (light) and $+0.6\%$ (dark) reveal the general nature of the high resolution model SB4L18 (Plate 2): large-scale low velocity regions surrounded by fast slab-like regions which extend across the whole mantle. (Graphics courtesy of P. Tackley).

wave data. Plate 1 shows the result of fitting the long-period body-waves only and is reminiscent of

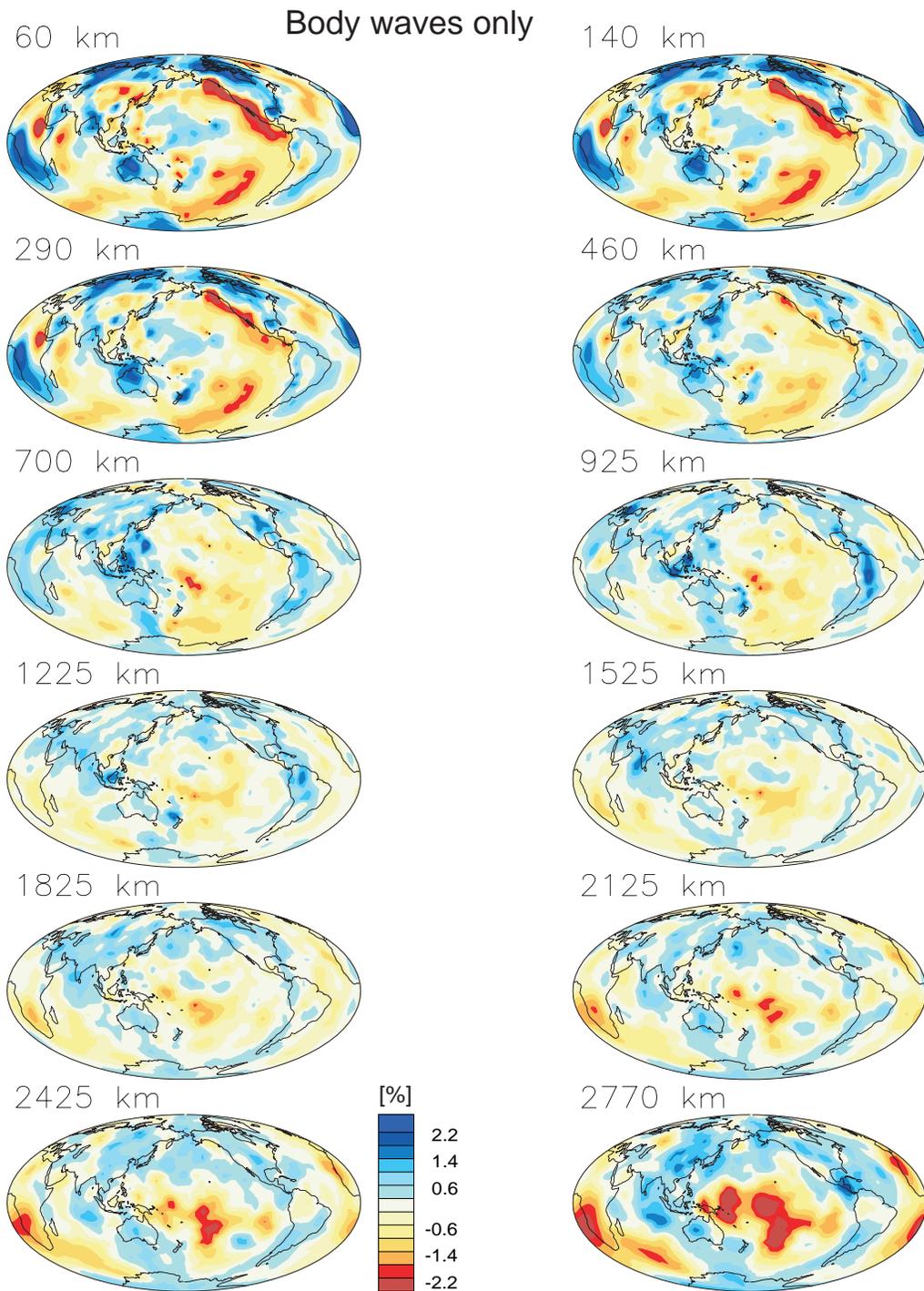


Plate 1 Cross-sections of the Earth at various depths showing perturbations of shear velocity (in percent). This high-resolution model was constructed using body wave data alone.

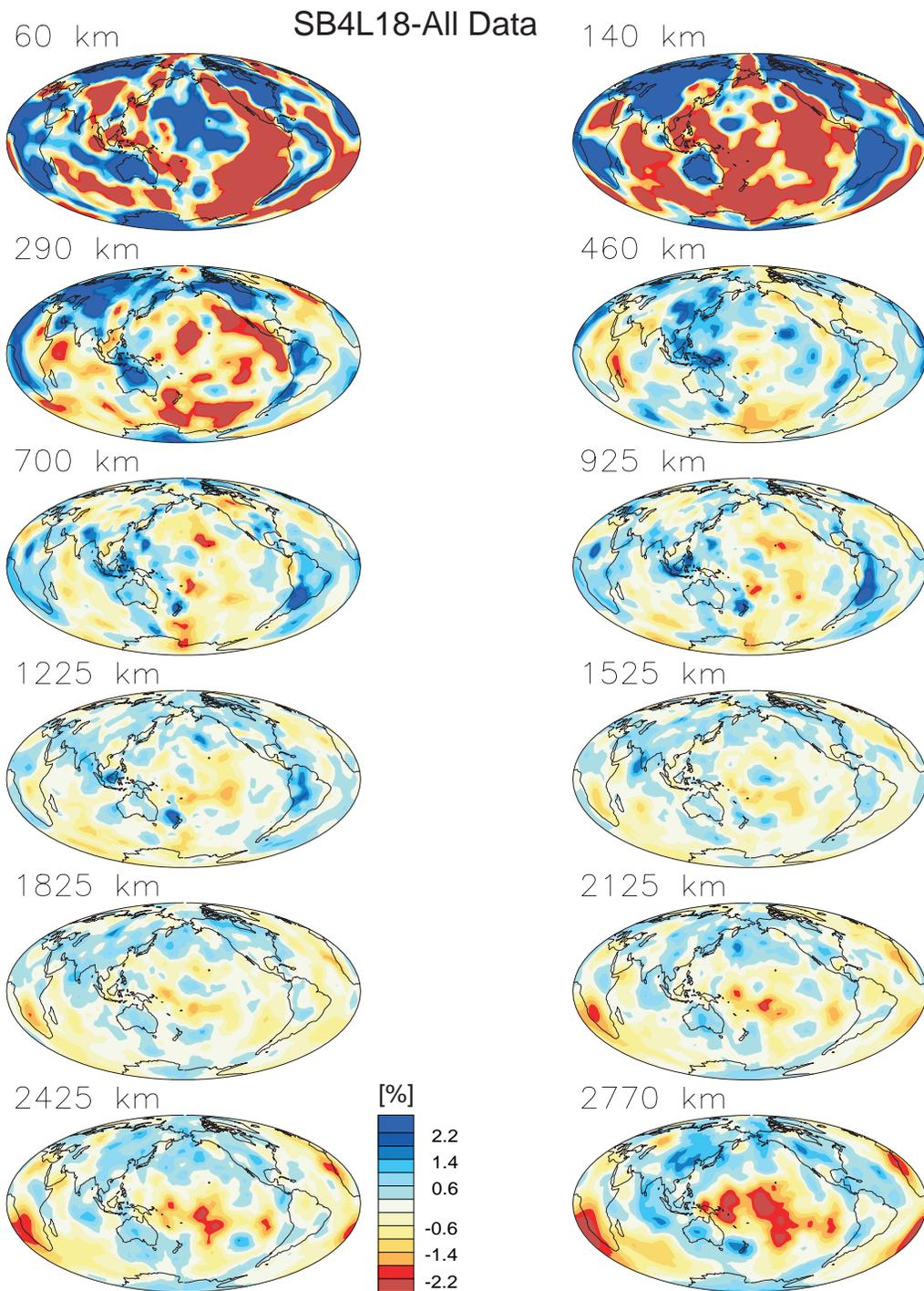


Plate 2 Cross-sections of the Earth at various depths showing perturbations of shear velocity (in percent). This high-resolution model (SB4L18) was constructed using surface-wave and free-oscillation data as well as body wave data. Compare with Plate 1. Note that the upper mantle is completely different but the anomalies in the lower mantle appear to be remarkably similar.

many models found by inverting ISC data. Plate 2 shows the effect of including the surface waves which demand high amplitude structure in the upper mantle. It is remarkable that the presence of this structure does not really change the picture in the lower mantle where slab-like features are readily apparent. In fact, both models fit the body-wave data about equally well though only the latter model provides an acceptable fit to the surface-wave and free-oscillation data. We can conclude two things from this experiment – the first is that surface waves should be included to get reasonable results in the upper mantle, the second is that the pattern of slab-like features imaged in the lower mantle is remarkably robust.

Fig. 14 compares the amplitudes in our model with those of the recent model by *Grand* [1994] and also shows the correlation between the two models (Fig. 16). The correlation is remarkably good in both the upper and lowermost mantle but is poorest in the transition zone. Furthermore, the amplitudes in the Grand model in the middle mantle and transition zone are extremely small.

Clearly, it is important to know which (if either) model is correct if we are to start interpreting the physical cause of the anomalies we image. We currently favor the larger amplitudes of our model for the following reasons:

1) The surface wave data are underpredicted by the Grand model, particularly at low frequencies where sampling of the transition zone is strongest (see Fig. 22).

2) Our differential travel time data are poorly fit by the Grand model (Table 2) including the $ScS - S$ which are dominantly sensitive to the lower mantle.

3) Inversions of structure coefficients result in models with the larger amplitudes [see also *Resovsky and Ritzwoller*, 1998].

4) Inversions of synthetic data constructed from the Grand model using our ray geometry reproduce its low amplitudes. This suggests that our data coverage and quality are sufficient to distinguish between the models.

Clearly, we need to clarify why seismologists can generate models of such different amplitudes (even though the shapes tend to be well-correlated). The first order of business is to ensure that the data sets that go into the different inversions are compatible and this is the subject of ongoing research.

Fig. 12 shows an image of our "high-resolution" model (courtesy of Paul Tackley) which elucidates its general features – notably the large-scale low-velocity regions in the lower mantle surrounded by slab-like structures. It is also interesting to note that the "slabs" in this image do not appear to fade out a few hundred kilometers above the CMB which is a feature of some tomographic models and has been the subject of some speculation in the recent literature [*van der Hilst and Karason*, 1999].

10. JOINT INVERSION OF THE P AND S DATA SETS

7.11 Constrained Inversions . There have been many recent attempts to jointly invert for shear velocity and bulk sound speed, or shear velocity and compressional velocity. These vary in sophistication and complexity with the simplest inversions assuming that v_p and v_s anomalies have the same geographic pattern but that their relative amplitudes are allowed to vary as a function of depth. Both ISC [*Robertson & Woodhouse*, 1996] and long-period data (including surface waves) [*Bolton*, 1996] have been treated in this fashion. Contamination of the S residuals by SKS is severe for the ISC data beyond 80° so the *Robertson & Woodhouse* [1996] analysis ends at a depth of 2000km. They also confine attention to long-wavelength models of structure (spherical harmonic degree 6). Their results indicate a linear dependence of R on depth between 600km and 2000km with R ranging between from 1.7 to 2.5 (Fig. 13). The Bolton study was very similar in that long-wavelength structure was sought (up to harmonic degree 8) but now the whole mantle could be analyzed because of the use of hand-picked S arrivals from transverse components. His results are also shown in Fig. 13 along with the estimates from the

crude analysis of the raw data we obtained from Fig. 4. It is remarkable how consistent a picture these different studies give and all agree that R increases monotonically in the lower mantle to a value that can not be explained by sub-solidus thermal effects.

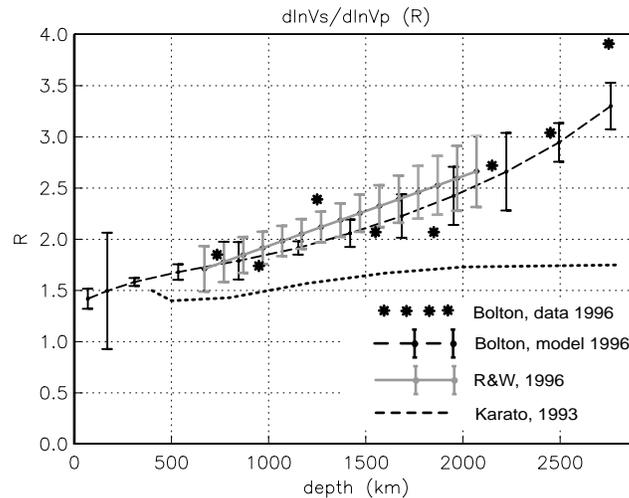


Figure 13 $R = dlnv_s/dlnv_p$ for two joint models which assume that the geographical shape of the compressional velocity anomaly is perfectly proportional to the shape of the shear velocity anomaly. Also shown are the values obtained directly from the data binned by ray turning depth (see Fig. 4). The resulting R values rapidly increase toward the CMB and are far too large to be explained by ordinary thermal effects [Karato, 1993].

However, we are hesitant to accept these large global values of R at face value. It seems clear that the increase in R is associated with the regions in the lower mantle where there are very large (negative) v_s anomalies accompanied by much smaller v_p anomalies. These regions actually seem to be quite localized (Fig. 3) though they are usually some of the best sampled (e.g. the central Pacific) which might lead to bias in the inversion. We therefore now turn attention to inversions where the geographic shapes of the velocity anomalies are not required to be the same.

7.12 Unconstrained Inversions . For the purposes of this paper, we have performed a joint bulk sound speed - shear velocity inversion of the data sets described above. In the interests of computational expediency, we have parameterized both velocities by a mantle divided into 18 layers each of which is divided into equal area blocks of dimension 10 degrees at the equator (this is equivalent to a spherical harmonic expansion of about harmonic degree 20). Again, we employ the LSQR algorithm and we include first-difference smoothing both radially and laterally.

The reason for the relatively coarse parameterization is that we must now perform the earthquake-location projection step to obtain reliable P models. When we have many data (both P and S) per event, the projected data are each linear combinations of *all* the original data and can sample a significant fraction of the model. This makes the matrices significantly less sparse and the inversion becomes much more of a computational burden.

Our favorite joint model is shown in Plate 3 (SB10L18) and the fit of this model to our data is shown in Table 2. The amplitude of this model is compared with others in Fig. 14. This model demonstrates a robust feature of all joint models we have obtained: the correlation between the bulk sound speed and the shear velocity becomes negative in the lowermost mantle (Fig. 15). This feature was first noted by Su & Dziewonski [1997] though their negative correlation is stronger and extends to shallower depths than in our model. Another feature of the joint models is that the compressional velocity is strongly

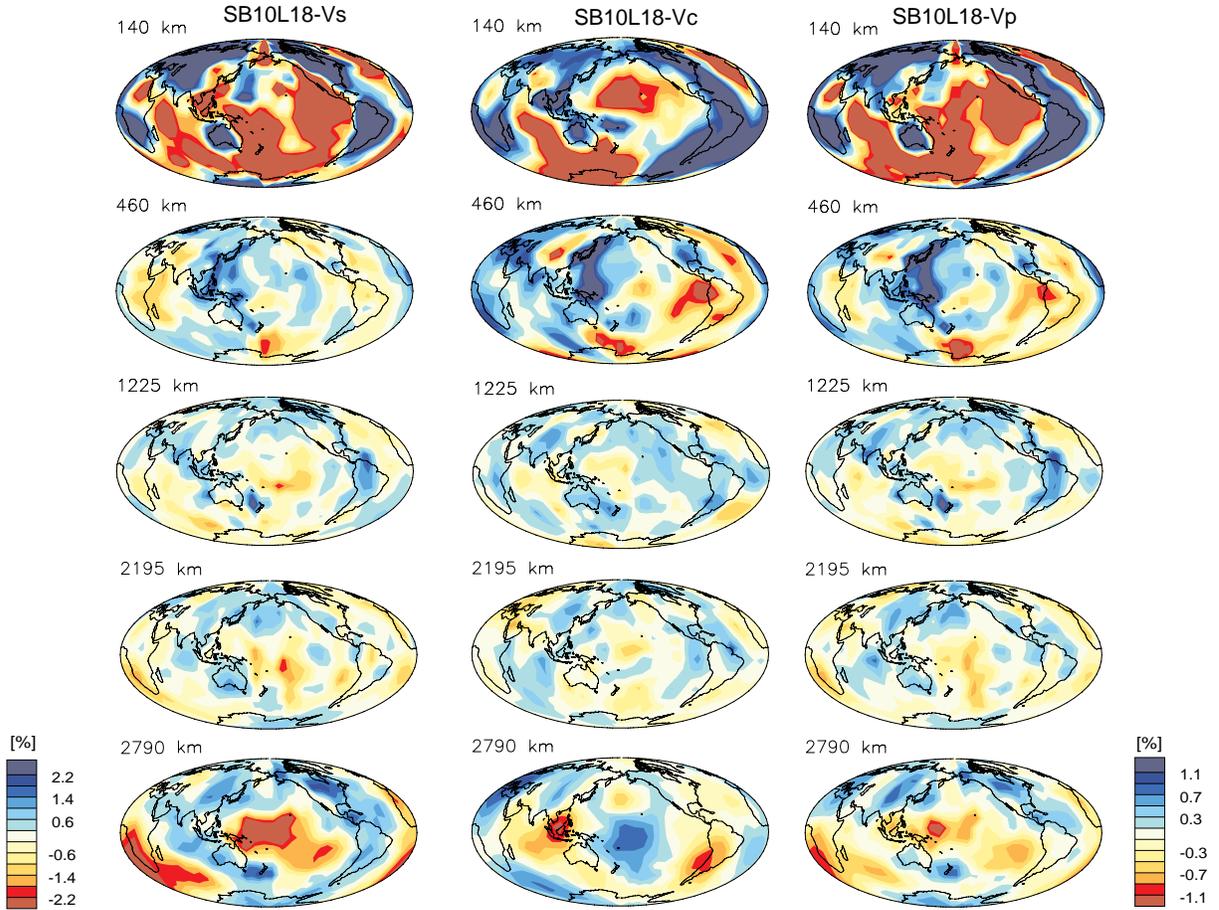


Plate 3 Cross-sections of the Earth at various depths showing perturbations of shear velocity, bulk sound speed and compressional velocity (in percent). This jointly-inverted model (SB10L18) was constructed using surface-wave and free-oscillation data as well as body wave data. Note the anticorrelation of bulk sound speed with shear velocity near the base of the mantle. Compressional velocity is strongly correlated with shear velocity throughout the mantle. The color scale for v_c is the same as that of v_p .

correlated with the shear velocity throughout the mantle (Fig. 15). We also show the correlation between the various shear velocity models in Fig. 16.

Table 2. Fit of models to S datasets, VR is the variance reduction in percent.

Model	$SS - S$		$ScS - S$		unproj. S	
	χ^2/N	VR	χ^2/N	VR	χ^2/N	VR
SB4L18	1.1	73%	2.0	66%	4.2	41%
SB10L18	1.3	69%	2.0	66%	4.2	41%
Grand	2.5	41%	3.9	39%	4.2	41%
KWH	3.8	14%	5.3	20%	7.1	8%
V&J	4.0	10%	6.5	-1%	5.7	23%
S&D	2.5	42%	2.9	50%	4.9	34%

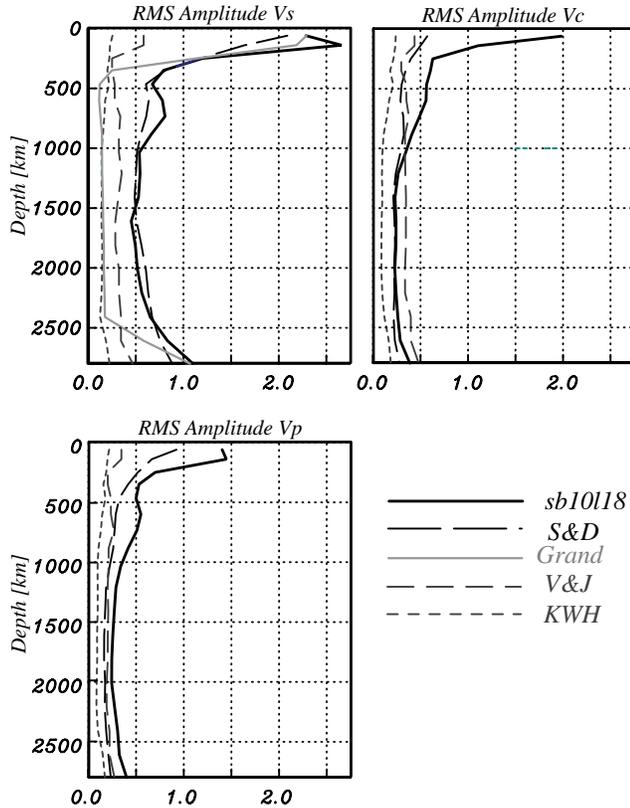


Figure 14 RMS amplitudes of shear and compressional velocities and bulk sound speed for several recent models. Note how the Grand and KWH models have extremely small shear velocity amplitudes in the mid mantle which leads to disagreement with several data sets. The models are: SB10L18: this study; S&D: *Su and Dziewonski* [1997]; V&J: *Vasco and Johnson* [1998]; KWH: *Kennett et al.* [1998]; Grand: *Grand* [1994]. The amplitude of SB4L18 is almost identical to that of SB10L18.

We have also computed a spherically-averaged R value as a function of depth for this and other models. We have noticed that different authors compute R in different ways. Our choice is to take all blocks in each shell which have non-zero values of both compressional and shear velocity perturbations and take their ratio. A histogram of these pixel-by-pixel estimates is made (Fig. 17) and the median value chosen as the best value for each shell. The values that we now obtain are smaller than found in the previous section but still show the increasing trend with depth (Fig. 18, curve SB10L18). This may confirm our suspicion that some relatively local regions with extreme values of R biased the inversions of the previous section. Our spherically averaged values of R are now much more consistent with the predictions of normal thermal effects with the possible exception of the lowermost mantle.

7.13 Comparison With Other Models . We now compare four recent joint models of velocities in the mantle. Three of these are products of joint bulk sound speed/ shear velocity inversions: this inversion, SB10L18 (Plate 3); *Su and Dziewonski* 1997, (S&D, Fig. 19), and *Kennett et al.* 1998, (KWH, Fig. 20)]. One model resulted from a joint compressional/shear velocity inversion [*Vasco and Johnson*, 1998 (V&J, Fig. 21)].

All the joint v_c/v_s inversions give models for which v_p is highly correlated with v_s (Fig. 15). This seems to be a natural outcome of this parameterization while joint v_p/v_s inversions lead to much less correlation. For three models v_c is negatively correlated with v_s in at least part of the lower mantle. The exception is model KWH where v_c is positively correlated with v_s throughout the whole mantle.

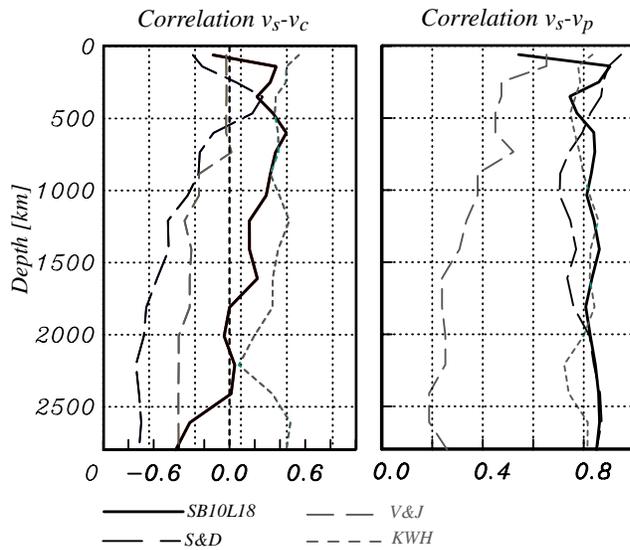


Figure 15 The correlation of v_s-v_c and of v_s-v_p as a function of depth in the mantle for a variety of models. Note that all models exhibit negative correlation for v_s-v_c near the base of the mantle except for model KWH.

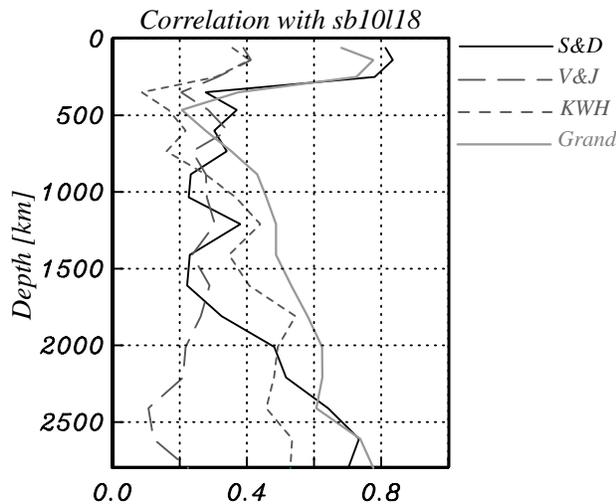


Figure 16 The correlation for v_s of SB10L18 with other models, as a function of depth. Note the excellent correlation of SB10L18 with S&D and Grand in the upper mantle. The correlation is significantly decreased in the lower mantle but rises towards the bottom. Also note that SB10L18 and the V&J model correlate poorly at all depths.

There is a wide variety in the behavior of R with depth. Both S&D and SB10L18 show a general increase of R with depth (Fig. 18) with the values of S&D being very similar to those determined from the unconstrained inversions. It is interesting to note that S&D use a degree 12 spherical harmonic expansion which is significantly more constrained than the parameterization of SB10L18 and which might partly contribute to the difference. S&D and SB10L18 also have very similar amplitudes for the anomalies in the mantle (Fig. 14) which might be partially explained by the fact that both of these models are built using surface wave and long-period body waves data sets though S&D also include ISC P data. The two models based on ISC data, KWH and V&J, are very different from the previous two

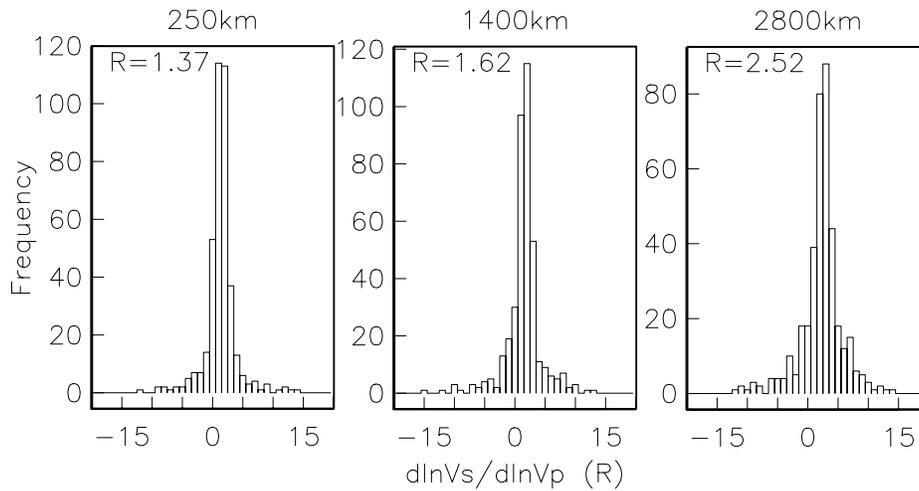


Figure 17 Histograms of R computed pixel-by-pixel at various depths in the mantle. This example is for model SB10L18 and we show R near the top, center, and base of the mantle. The mean value is quite robust and is used as an estimate of the spherically averaged value of R .

models and quite different from each other. In particular, the amplitudes of the shear anomalies in the KWH model are extremely small and the V&J shear model does not correlate with other models in the lower mantle.

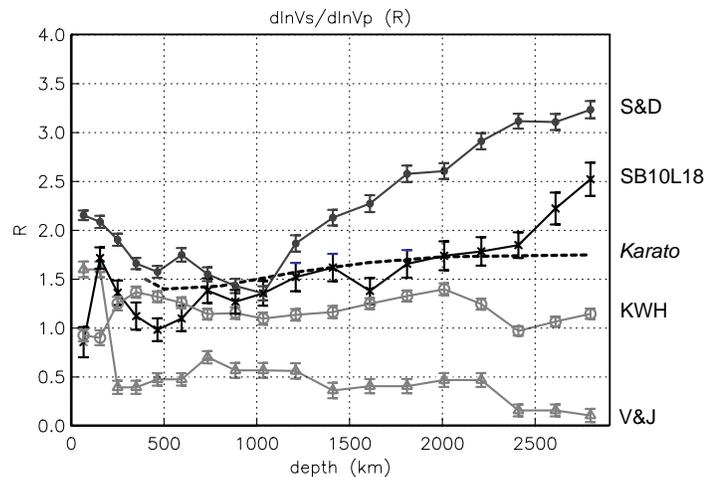


Figure 18 $R = dlnv_s/dlnv_p$ for the four unconstrained joint models. Model S&D gives answers very similar to the constrained inversions. Model SB10L18 shows a more modest increase of R across the mantle. The remaining two models show much smaller values of R .

We would obviously like to be able to distinguish which, if any, of these models has the correct levels of heterogeneity in the mantle. This is quite difficult to do since the ISC-based models have incomplete upper mantles (due to poor data coverage). We therefore cannot quantify a comparison of model predictions with many of the global data sets such as mode structure coefficients. Here, we rather rely on a visual comparison of predicted and measured data. Fig. 22 shows Rayleigh wave phase velocity

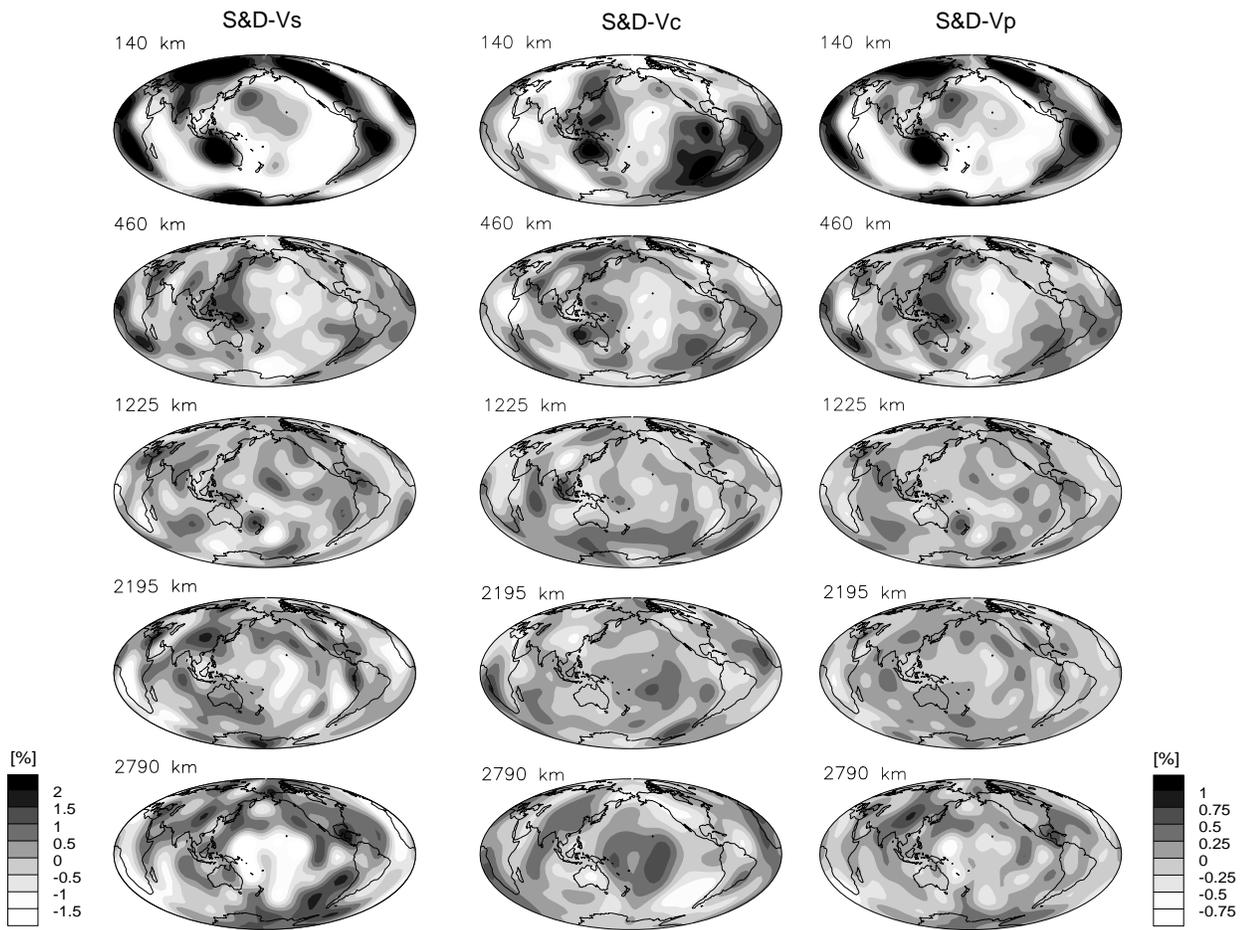


Figure 19 Cross-sections of the S&D model. For details see caption to Plate 3. Note the anticorrelation of bulk sound speed with shear velocity near the base of the mantle.

maps for 6 mHz as predicted by the models discussed here and the Grand model discussed in section 4. For the Grand model, the standard scalings were adopted to create the map ($dlnv_s/dlnv_p = 1.7$, $dlnv_s/dln\rho = 2.5$). We see that his model (as well as the two joint $v_s - v_c$ models that included surface waves in the modeling process) can reproduce the basic features of such maps. Obvious differences are: the S&D model produces smoothed phase velocity maps due to the $\ell = 12$ truncation of the spherical harmonic expansion and SB10L18 produces “blocky” maps due to the relatively large block size. It is interesting to note that the amplitudes in the maps for both Rayleigh and Love waves predicted by Grand’s model tend to be too small for very long periods (around 6mHz and below). This is probably caused by the rapid decrease of the rms amplitude of Grand’s model in the transition zone (Fig. 14). The two models that did not include surface waves in the modeling process (KWH and V&J) predict phase velocity maps whose amplitudes are about an order of magnitude too small – even allowing for their restricted sampling. While the KWH model predicts the location of most long-wavelength features correctly, the V&J model predicts several features not seen in the observed maps (e.g. fast velocities around New Zealand).

A data set that can be used to check the overall level of heterogeneity in the lower mantle is the set of $S_cS - S$ differential times since these are fairly insensitive to structure in the upper mantle.

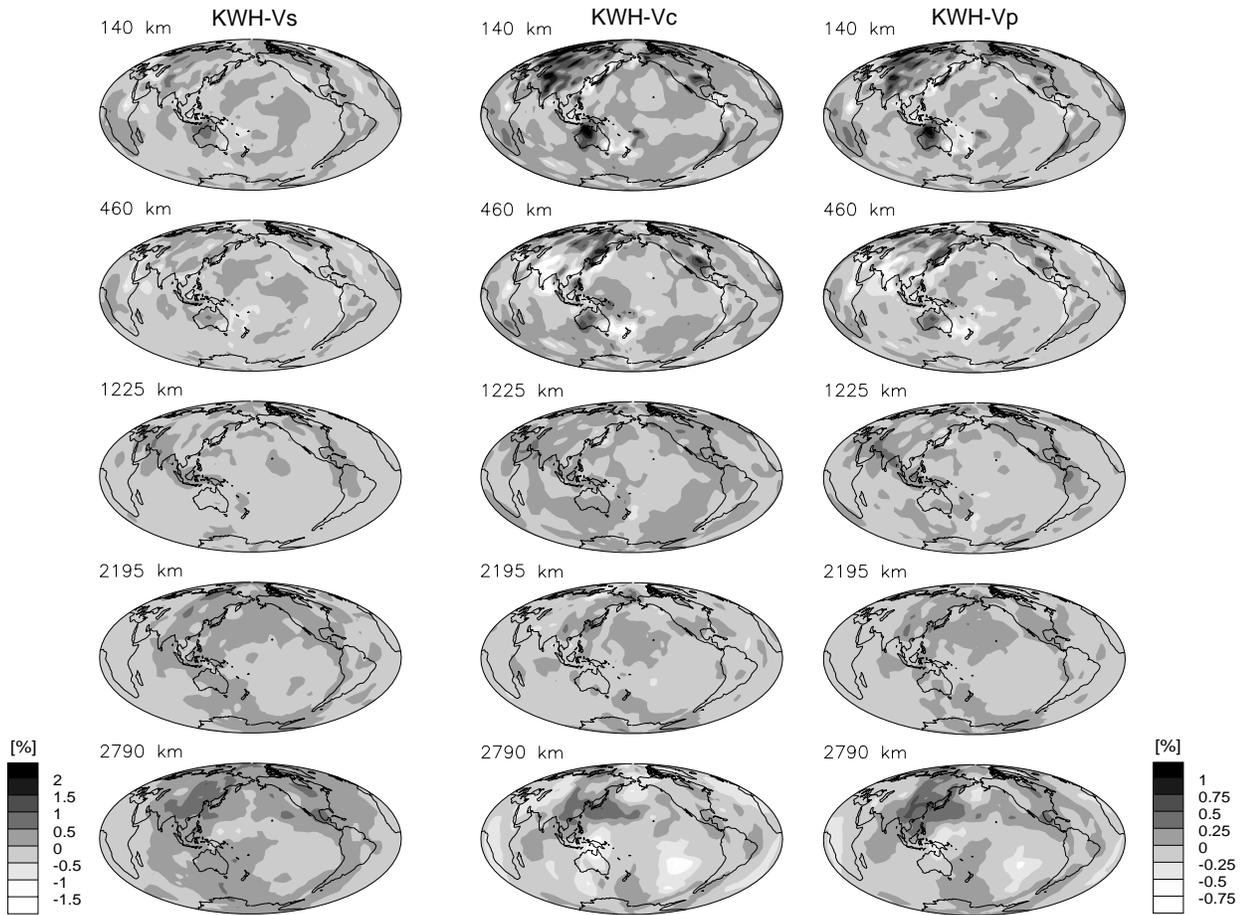


Figure 20 Cross-sections of the KWH model. For details see caption to Plate 3.

Not surprisingly, our model SB10L18 fits best with a 66% variance reduction, model S&D achieves a 50% variance reduction, model KWH a 20% variance reduction and model V&J a slightly negative variance reduction (Table 2). The poor performance of the last two models may reflect the difficulty of constraining S velocity in the lowermost mantle with ISC data alone (as discussed above).

14. DISCUSSION AND CONCLUSIONS

At this point, readers may be excused if they feel bewildered by the range of seismic models that we have presented. We shall therefore try to formulate some general observations though some of our conclusions (e.g. about the behavior of R) require discounting the amplitudes of the models built primarily from ISC data.

1) Our limited comparison indicates that ISC and long-period travel time residual patterns for various stations are compatible with each other. The signal level in the ISC P data is better than for the long-period P data but the signal level in the ISC S data is much worse than in the long-period S data. The range of S residuals found in both data sets is similar (up to 15 second residuals for teleseismic S) and greatly exceeds the range used by the ISC (± 7.5 sec) to identify teleseismic S . The fact that the KWH model restricted attention to S residuals smaller than 7.5 sec may have contributed to its very small

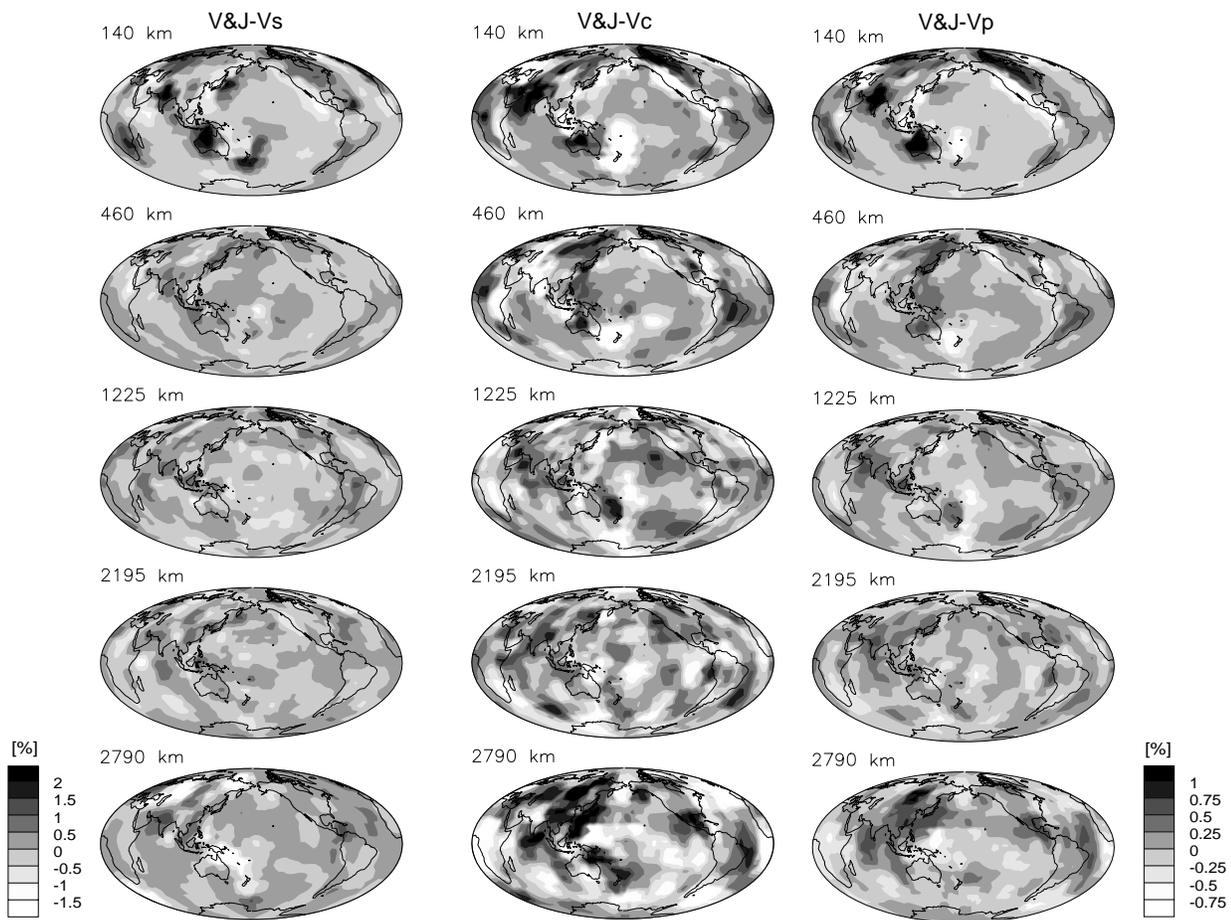


Figure 21 Cross-sections of the V&J model. For details see caption to Plate 3.

amplitudes.

2) The effects of earthquake mislocation can contribute errors to arrival times which are comparable in size with the signal from 3D structure for *P* waves. Hence, accounting for relocation in the inversion is critical.

3) We have generally found that the results of the inversions do not depend upon the inversion technique (e.g. iterative or direct) provided that the inverse problem is not grossly ill-conditioned. Gross over-parameterization can lead to poor convergence characteristics of some iterative solvers and, naturally, a strong dependence on regularization criteria.

4) Inversions of the current structure coefficient data set do not convincingly recover 3D density structure – in particular, we can not yet say whether density is positively or negatively correlated to shear velocity in the lowermost mantle.

5) The upper mantles of models based on travel times alone tend to have seismic anomalies that are greatly underestimated in magnitude (though inclusion of multiple bounce triplication phases [Grand, 1994] appears to be an effective strategy to combat this). Velocity anomaly patterns in the lower mantle are surprisingly robust to the inclusion of surface wave data which can radically change the models of the upper mantle.

6) Our high resolution models have amplitudes of velocity anomalies in the mid mantle consistent with

Rayleigh wave Phase Velocity Maps at 6 mHz

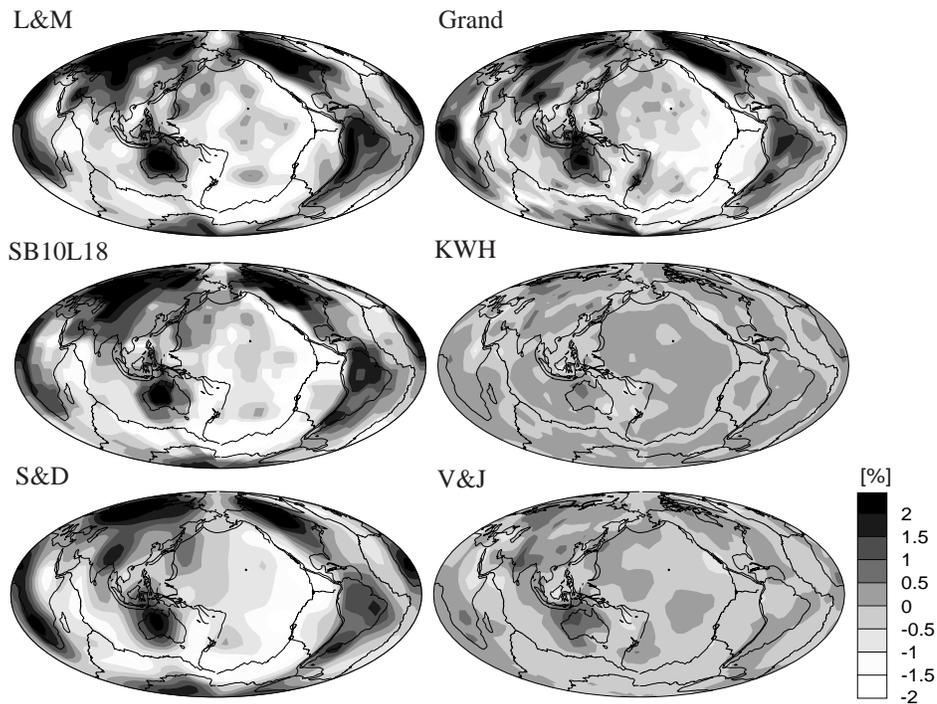


Figure 22 The phase velocity maps for Rayleigh waves at 6 mHz as predicted by the models discussed in this study. The L&M map is the measured map of *Laske and Masters* [1996]. Note the good agreement between predicted and observed maps for the models SB10L18 and S&D. Agreement is also quite good for the model of Grand though the amplitude is somewhat underestimated. The phase velocity variations are with respect to the spherical average and are given in percent.

previous low-resolution models. These amplitudes are compatible with both mode-structure coefficients and differential travel time data.

7) Our models show slab-like structures that extend throughout the mantle with little “fade-out” a few hundred kilometers above the CMB (see also Fig. 14)

8) Joint modeling of either ISC P and S data or long-period P and S data assuming identical (large-scale) patterns of velocity anomalies results in estimates of $R(z)$ which are quite high and increase towards the CMB. We speculate that extremely high values of R at the base of the mantle are a result of some relatively localized regions where large S velocity anomalies are accompanied by relatively minor P velocity anomalies.

9) More general unconstrained inversions lead to models in which perturbations in bulk sound speed negatively correlate with perturbations in shear velocity in at least part of the lower mantle (SB10L18 and S&D) though even here there is some discrepancy over the depth range over which the negative correlation exists. The S&D inversion shows a negative correlation over almost all the lower mantle, the modes-only inversion (Fig. 10) shows a negative correlation for about the bottom 800km of the lower mantle, while SB10L18 has the negative correlation over only the bottom 400 km of the lower mantle. There is some suggestion that the negative correlation is strongest for the longest wavelengths which

partially accounts for the discrepancy. At the present time, a mineral physicist looking for the "best" value of $R(z)$ to model should probably choose something between the curves for S&D and SB10L18 in Fig. 18.

10) In much of our analysis, we can identify a central Pacific region in the deep mantle as being particularly anomalous. The fact that this region is far from places of historic subduction is worthy of note – indeed the anomalous regions appears to be associated with a slow (warm) and presumably upwelling region in the lower mantle. Such regions also seem to be associated with ultra-low-velocity zones [ULVZ – see e.g. *Lay et al.*, 1998] which may indicate chemical interaction of the core and mantle. Perhaps entrainment of such chemical heterogeneity into the lowermost mantle (and/or partial melting associated with this chemical contamination) can give rise to the anomalous values of R that we observe.

Acknowledgments This research was supported by grants from the National Science Foundation. We are grateful to the network operators of the IRIS/IDA, IRIS/USGS, and GEOSCOPE networks, and the IRIS DMC for collecting and making easily available the data which made this study possible. We would also like to thank Paul Tackley for providing Fig. 12.

References

- Agnon, A., and M. Bukowinski, δ_s at high pressure and $d\ln v_s/d\ln v_p$ in the lower mantle, *Geophys. Res. Lett.*, *17*, 1149–1152, 1990.
- Anderson, O.L., and D.G. Isaak, Elastic constants of mantle minerals at high temperature, in *Mineral Physics and Crystallography, AGU Ref. Shelf #2*, ed. T.J. Ahrens, pp.64–97, 1995.
- Bijwaard, H., W. Spakman, and E.R. Engdahl, Closing the gap between regional and global travel time tomography, *J. Geophys. Res.*, *103*, 30,055–30,078, 1998.
- Bolton, H., Long period travel times and the structure of the mantle, *Ph.D. Thesis, U.C.S.D., La Jolla, California*, , 1996.
- Bolton, H., and G. Masters, Traveltimes of P and S from the Global Digital Seismic Networks; Implications for the Relative Variation of P and S Velocity in the Mantle, *J. Geophys. Res.*, submitted, 1999.
- Clayton, R.W., and R.P. Comer, A tomographic analysis of mantle heterogeneities from body wave travel time data, *EOS Trans. AGU*, *64*, 776, 1983.
- Creager, K.C., and T.H. Jordan, Large scale structure of the outermost core from P'_{DF} and P'_{AB} travel-times, *EOS Trans. AGU*, *67*, 311, 1986.
- Dines, K., and J. Lytle, Computerized geophysical tomography, *Proc. IEEE*, *67*, 1065–1073, 1979.
- 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 D.L. Anderson, Travel times and station corrections for P -waves at teleseismic distances, *J. Geophys. Res.*, *88*, 3295–3314, 1983.
- Dziewonski, A.M., and J.H. Woodhouse, Global images of the Earth's interior, *Science*, *236*, 37–48, 1987.
- Ekström, G., J. Tromp, and E.W. Larson, Measurements and models of global models of surface wave propagation, *J. Geophys. Res.*, *102*, 8137–8157, 1997.
- Engdahl, E.R., R.D. van der Hilst, and R.P. Buland, Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Bull. Seis. Soc. Am.*, *88*, 722–743, 1998.

- Forte, A.M., and W.R. Peltier, Mantle convection and core-mantle boundary topography, *J. Geophys. Res.*, *92*, 3645–3679, 1987.
- Giardini, D., X.-D. Li, and J.H. Woodhouse, Three-dimensional structure of the earth from splitting in free oscillation spectra, *Nature*, *325*, 405–411, 1987.
- Grand, S.P., Mantle shear structure beneath the Americas and surrounding oceans, *Journ. Geophys. Res.*, *99*, 11,591–11,621, 1994.
- Gudmundsson, O., J.H. Davis, and R.W. Clayton, Stochastic analysis of global traveltimes data: mantle heterogeneity and random errors in the ISC data, *Geophys. J. Int.*, *102*, 25–43, 1990.
- He, X., and J. Tromp, Normal mode constraints on the structure of the Earth, *J. Geophys. Res.*, *101*, 20,053–20,082, 1996.
- Inoue, H., Y. Fukao, K. Tanabe, and Y. Ogata, Whole mantle *P*-wave travel time tomography, *Phys. Earth Planet. Inter.*, *59*, 294–328, 1990.
- Isaak, D.G., O.L. Anderson, and R.E. Cohen, The relationship between shear and compressional velocities at high pressures: reconciliation of seismic tomography and mineral physics, *Geophys. Res. Lett.*, *19*, 741–744, 1992.
- Ishii, M., and J. Tromp, Normal mode and free-air gravity constraints on lateral variations in velocity and density of Earth's mantle, *Science*, *285*, 1231–1236, 1999.
- Kanamori, H., and D.L. Anderson, Importance of physical dispersion in surface wave and free oscillations problems: review, *Rev. Geophys. Space Phys.*, *15*, 105–112, 1977.
- Karato, S., Importance of anelasticity in the interpretation of seismic tomography, *Geophys. Res. Lett.*, *20*, 1623–1626, 1993.
- Karato, S., B. Karki, and R. Wentzcovitch, Origin of seismic velocity heterogeneity in the mantle, *EOS Trans AGU*, *80*, F939, 1999.
- Kennett, B.L.N., and E.R. Engdahl, Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.*, *105*, 429–465, 1991.
- Kennett, B.L.N., E.R. Engdahl, and R. Buland, Constraints on seismic velocities in the Earth from travel times, *Geophys. J. Int.*, *122*, 108–124, 1995.
- Kennett, B.L.N., S. Widiyantoro, and R.D. van der Hilst, Joint seismic tomography for bulk sound and shear wave speed in the Earth's mantle, *J. Geophys. Res.*, *103*, 12,469–12,493, 1998.
- Laske, G., and G. Masters, Constraints on global phase velocity maps by long-period polarization data, *J. Geophys. Res.*, *101*, 16,059–16,075, 1996.
- Lay, T., Q. Williams, and E.J. Garnero, The core–mantle boundary layer and deep Earth dynamics, *Nature*, *392*, 461–468, 1998.
- Li, X.-D., D. Giardini, and J.H. Woodhouse, The relative amplitudes of mantle heterogeneity in *P*-velocity, *S*-velocity and density from free-oscillation data, *Geophys. J. Int.*, *105*, 649–657, 1991.
- Li, X.-D., and B. Romanowicz, Comparison of global inversions with and without considering cross-branch modal coupling, *Geophys. J. Int.*, *121*, 695–709, 1995.
- Li, X.-D., and B. Romanowicz, Global mantle shear-velocity model developed using nonlinear asymptotic coupling theory, *J. Geophys. Res.*, *101*, 22,245–22,272, 1996.
- Masters, G., S. Johnson, G. Laske, and H. Bolton, A shear velocity model of the mantle, *Phil. Trans. R. Soc. Lond.*, *354A*, 1385–1411, 1996.
- Masters, G., T.H. Jordan, P.G. Silver, and F. Gilbert, Aspherical earth structure from fundamental spheroidal-mode data, *Nature*, *298*, 609–613, 1982.
- Masters, G., G. Laske, and F. Gilbert, Autoregressive estimation of the splitting matrix of free-oscillation multiplets, *Geophys. J. Int.*, in press, 1999.
- 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, 1990.

- Mooney, W.D., G. Laske, and G. Masters, CRUST 5.1: A global crustal model at 5° X 5°, *J. Geophys. Res.*, 103, 727–747, 1998.
- Morelli, A., A.M. Dziewonski, and J.H. Woodhouse, Anisotropy of the inner core inferred from *PKIKP* travel times, *Geophys. Res. Lett.*, 13, 1545–1548, 1986.
- Morelli, A., and A.M. Dziewonski, Topography of the core-mantle boundary and lateral homogeneity of the liquid core, *Nature*, 325, 678–683, 1987.
- Morelli, A., and A.M. Dziewonski, Body wave traveltimes and a spherically symmetric *P*– and *S*-wave velocity model, *Geophys. J. Int.*, 112, 178–194, 1993.
- Nolet, G., Seismic wave propagation and seismic tomography, in *In: Nolet, G. (Editor), Seismic Tomography, D. Reidel Publishing Company, Dordrecht*, pp.1–23, 1987.
- Pulliam, R.J., *Imaging Earth's interior: Tomographic inversions for mantle P-wave velocity structure*, Ph.D Thesis, U.C. Berkeley, 1991.
- Pulliam, R.J., D.W. Vasco, and L.R. Johnson, Tomographic inversions for mantle *P* wave velocity structure based on the minimization of l^2 and l^1 norms of International Seismological Centre travel time residuals, *J. Geophys. Res.*, 98, 699–734, 1993.
- Resovsky, J.S., and M.H. Ritzwoller, New and refined constraints on three-dimensional Earth structure from normal modes below 3mHz, *J. Geophys. Res.*, 103, 783–810, 1998.
- Resovsky, J.S., and M.H. Ritzwoller, Regularization uncertainty in density models estimated from normal mode data, *Geophys. Res. Lett.*, 26, 2319–2322, 1999.
- 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.
- Ritzwoller, M., G. Masters, and F. Gilbert, Constraining aspherical structure with low frequency interaction coefficients: Application to uncoupled multiplets, *J. Geophys. Res.*, 93, 6369–6396, 1988.
- Robertson, G.S., and J.H. Woodhouse, Evidence for proportionality of *P* and *S* heterogeneity in the lower mantle, *Geophys. J. Int.*, 123, 85–116, 1995.
- Robertson, G.S., and J.H. Woodhouse, Ratio of relative *S* to *P* velocity heterogeneity in the lower mantle, *J. Geophys. Res.*, 101, 20,041–20,052, 1996.
- Shearer, P.M., K.M. Toy, and J.A. Orcutt, Axi-symmetric earth models and inner core anisotropy, *Nature*, 333, 228–232, 1988.
- Smith, M.F., and G. Masters, Aspherical structure constraints from free oscillation frequency and attenuation measurements, *J. Geophys. Res.*, 94, 1953–1976, 1989.
- Souriau, A., and J.H. Woodhouse, A worldwide comparison of predicted *S*-wave delays from a three-dimensional upper mantle model with *P*-wave station corrections, *Phys. Earth Planet. Inter.*, 39, 75–88, 1985.
- Su, W.-J., and A.M. Dziewonski, Predominance of long-wavelength heterogeneity in the mantle, *Nature*, 352, 121–126, 1991.
- 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.
- Su, W.-J., and A.M. Dziewonski, Simultaneous inversion for 3D variations in shear and bulk velocity in the mantle, *Phys. Earth Planet. Inter.*, 100, 135–156, 1997.
- Tanimoto, T., Long-wavelength *S*-velocity structure throughout the mantle, *Geophys. J. Int.*, 100, 327–336, 1990.
- Toy, K.M., *Tomographic analysis of ISC travel time data for Earth structure*, Ph.D Thesis, Univ. Cal. San Diego, 1989.
- 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.
- Trampert, J., and J.H. Woodhouse, High-resolution global phase velocity distribution, *Geophys. Res. Lett.*, 23, 21–24, 1996.

- van der Hilst, R.D., S. Widiyantoro, and E.R. Engdahl, Evidence for deep mantle circulation from global tomography, *Nature*, 386, 578–584, 1997.
- van der Hilst, R.D., and K. Karason, Compositional Heterogeneity in the Bottom 100 Kilometers of Earth's Mantle: Toward a Hybrid Convection Model, *Science*, 283, 1885–1889, 1999.
- Vasco, D.W., L.R. Johnson, and J. Pulliam, Lateral variations in mantle velocity structure and discontinuities determined from P , PP , S , SS , and $SS - S_dS$ travel times residuals, *J. Geophys. Res.*, 100, 24,037–24,059, 1995.
- Vasco, D.W., and L.R. Johnson, Whole Earth structure estimated from seismic arrival times, *J. Geophys. Res.*, 103, 2633–2671, 1998.
- Vasco, D.W., R.J. Pulliam, and L.R. Johnson, Formal inversion of ISC arrival times for mantle P -velocity structure, *Geophys. J. Int.*, 113, 586–606, 1993.
- Vasco, D.W., R.J. Pulliam, L.R. Johnson, and P.S. Earle, Robust inversion of IASP91 travel time residuals for mantle P and S velocity structure, earthquake mislocations, and station corrections, *J. Geophys. Res.*, 99, 13,727–13,755, 1994.
- Woodhouse, J.H., and A.M. Dziewonski, Mapping of the upper mantle: three-dimensional modeling of earth structure by inversion of seismic waveforms, *J. Geophys. Res.*, 89, 5953–5986, 1984.
- Woodward, R.L., and G. Masters, Global upper mantle structure from long-period differential travel times, *J. Geophys. Res.*, 96, 6351–6377, 1991a.
- Woodward, R.L., and G. Masters, Lower mantle structure from $ScS - S$ differential travel times, *Nature*, 352, 231–233, 1991b.
- Woodward, R.L., and G. Masters, Upper mantle structure from long-period differential travel times and free-oscillation data, *Geophys. J. Int.*, 109, 275–293, 1992.
- York, D., Least squares fitting of a straight line, *Canadian J. Phys.*, 44, 1079–1086, 1966.
- Zhang, Y., and T. Lay, Global surface wave phase velocity variations, *J. Geophys. Res.*, 101, 8415–8436, 1996.
- Zhou, H.-W., A high resolution P wave model for the top 1200 km of the mantle, *J. Geophys. Res.*, 101, 27,791–27,810, 1996.