Spectra of mantle shear wave velocity structure

J. Huw Davies, Olafur Gudmundsson, and R. W. Clayton

gyrometric Laboratory, 2521 California Institute of Technology, Pasadena, CA 91125, USA

Accepte 1991 September 26. Received 1991 May 21

SUMMARY

We applied the stochastic method of Gudmundsson, Davies & Clayton (1990) (which was applied to ISC P-wave data) to teleseismic ISC S-wave data to obtain an independent estimate of mantle structure. We inverted the variance of S-wave traveltime residuals of bundles of rays to obtain a description of the spectrum of lateral heterogeneity as a function of depth through the mantle. The technique yields robust estimates of the traveltime scattering power (the product of a characteristic scalelength of heterogeneity and the mean square of slowness perturbations). We can estimate the characteristic scalelength (half-width), from the auto covariance which can be reconstructed from the spectra. Hence by division we can estimate the root mean square slowness. By extrapolating the variance of bundles of rays to bundles of zero cross-sectional area we can also estimate the scale-invariant signal (which is a plausible estimate of the noise in the data), which is removed from the data.

We find that most of the structure generating shear wave traveltime residuals is located in the uppermost mantle. About half of the structure is short scale (harmonic degree l<50). The large-scale structure (l>50) has a half-width of about 500 km in the upper half of the mantle. This S-wave half-width is consistent with the P-wave half-widths determined by Gudmundsson et al. (1990). The S-wave half-width in the lower half of the mantle is poorly constrained. It varies from 500 to 3000 km, which spans the better constrained value of 1200 km found by Gudmundsson et al. (1990) for P-waves. The incoherent scatter suggests that the signal-to-noise ratio of the S-wave data set is around 1.5.

Assuming that the compressional and shear wave velocity variations are correlated then the signal weighted value of the ratio d ln(Vs)/d ln(Vp) is ≈ 2, as also found in normal mode studies. This is much larger than the d ln(Vs)/d ln(Vp) = 0.8–1.4 suggested by laboratory experiments undertaken at atmospheric pressure. There is no evidence of periodicity in the traveltime autocovariance; this suggests little or no periodicity in the underlying convection. The short half-width through most of the mantle suggests high Rayleigh number convection, with its attendant small-scale structures. The power decreases by an order of magnitude or more in going from the upper mantle to the lower mantle, the same as found by Gudmundsson et al. (1990) for P-waves. This large difference suggests either a change in convective regime and/or a difference in the temperature sensitivity of elastic constants in both layers. That the increased short-scale amplitude at the top of the mantle may be due to the presence of seismic discontinuities, which is consistent with the observed 1/3 power law over a range of scales extending from 500 km to 5000 km.

Key words: body waves, inversion, stochastic, traveltime variance.

*Presently at: Institute of Theoretical Geophysics, Department of Earth Sciences, University of Cambridge, Downing St, Cambridge CB2 1QZ, UK.
†Presently at: Research School of Earth Sciences, Australian National University, GPO Box 4, Canberra, ACT 2601, Australia.
1 INTRODUCTION

Since the inception of plate tectonics, we have been unable to elucidate the details of how it is powered by mantle convection. With the advent of deterministic (tomographic) models of lateral seismic variations (Dziewonski 1986; Clayton & Comer 1983; Hager & Clayton 1989) we have the first chance of imaging directly the driving forces of mantle convection and hence constraining plate motion and establishing how plate tectonics is powered. The body wave tomography models quoted above have in principle a higher spatial resolution than surface wave or free coda studies because of the shorter wavelengths of body waves. However, current lower mantle body wave studies suffer from a lack of data, and hence there are many questions related to the reliability of these tomographic models, including the following. How much random noise is there in the data and how well is it eliminated from the model? How much aliasing is there, due to the finite parameterization (Dziewonski 1984, t = 6; Clayton & Comer 1983), cell size = 250 km, t = 360? How well do the station corrections account for the crustal and upper mantle structure? How are the results affected by the uneven sampling?

Convection models which have referred to as DGC, presented a theory to invert the scatter in the traveltimes of seismic waves to obtain the spectra of lateral seismic velocity variations (Davies 1988). Following Davies, Gudmundsson & Clayton (1988), we also estimate the error in the ISC data set. This estimated error is subtracted from the data before computing the model, while the model parameterization is extended to the smallest scale avoiding aliasing. The model methods the whole mantle and attempts to account for the upper mantle. Hence the stochastic that traveltime models can be used to answer some of the questions introduced above regarding the results of deterministic inversion.

Convection models have advanced, but they are still generic in that they do not claim to know the initial or boundary conditions, while it would be able to simulate the behaviour of the actual Earth. Hence deterministic seismic models are of limited direct use (with the possible exception of prediction simulations), while stochastic seismic models provide directly spectral information which can be used as constraints. The ISC data set contained very large numbers of data (P = 4 million, S = 0.5 million) of a very heterogeneous nature (arrivals picked by different observers, on different seismometers, at unique sites); hence the data are well suited for a statistical treatment. Since we are also combining greater numbers of data to evaluate fewer parameters, we should more efficiently test and model of any random error left in the data. All of the body wave lower mantle models discussed above have derived the mantle P-wave velocity structure. In this paper we use S-waves to obtain an independent estimate of mantle structure; i.e. its shear velocity structure.

Several estimates of the ratio of ln (Vp/Vs)/ln (Vp) in the literature have been suggested to be about 2 (Dziewonski & Woodhouse 1988). This value is much higher than that obtained by theoretical structure and the grid; Anderson (1987) has suggested that such values in the lower mantle can be explained as the result of the effect of

2 EVALUATION OF TRAVELTIME VARIANCE

As discussed above, the heart of the stochastic method involves evaluating the scatter in the traveltime residuals of a bundle of rays. We define the bundle of rays, by means of an unique source and receiver cell on the surface of the Earth. We follow the procedure outlined in GDC, where we divided the earth into a grid of equal area cells. Grids were generated at scales varying from one box for the world to Earth to boxes which are 2° on a side. Rays are binned together if they fall in the same epicentral distance (E) range (the distance between source and receiver) and the same source depth (Z) bin. A summary ray is a collection (a bundle) of rays that share the same epicentral distance and source depth bins, and also the same region and source grid cells. The bins are 4° wide in epicentral distance, extending from 30° to 75°. The bins in source depth range are 0.1° to 0.6°, 0.60--1.00, 1.00--2.00, 2.00--3.00, and 3.00--700 km. The scale of a grid is defined as the angular radius of a circle on the Earth's surface, which has the same surface area as the Earth's surface. This may be denoted by g and reported in degrees. We use data collected by the ISC from 1964 to 1986. We select S-wave arrivals from events that have at least 50 phases reported because these would be better located. All events located at 0 or 33 km in depth are ignored, since the ISC frequently assigns poorly located events to these depths. We use shear wave arrivals out to only 75° to avoid contamination by SKS-waves, whose traveltime curve crosses S at about 80°. The ray parameter for this method is the angle that the ray makes with the observer, and the ray parameter for this method is the angle that the ray makes with the observer.

Figure 1. Plot of Jeffreys-Bullen shear wave traveltime residuals as a function of epicentral distance. Notice that their distribution is uniform over teleseismic distances (90° to 100°) with the exception of the vertical streak due to the misidentification of SKS in S from 0° to 90°. arrivals from events that have at least 50 phases reported because these would be better located. All events located at 0 or 33 km in depth are ignored, since the ISC frequently assigns poorly located events to these depths. We use shear wave arrivals out to only 75° to avoid contamination by SKS-waves, whose traveltime curve crosses S at about 80°. The ray parameter for this method is the angle that the ray makes with the observer, and the ray parameter for this method is the angle that the ray makes with the observer.

The mean number of rays summed for each grid is given in equation (1). Note that n indicates a function of b, i.e., the total number of summary rays depends on the grid orientation. The above equation may seem complicated, but it is in fact very simple, since it is only an average variance for all the summary rays with at least four component rays, found for all cells and orientations of the grid.

The uncertainty of the estimate is derived from the scatter between the estimates of the six rotated grids. At small scales with a grid size of 25° a reasonable estimate could be obtained if there are more than 25 summary rays, i.e., n > 25.

Each epicentral distance and source bin is characterized by a reference ray, which is located at the mean epicentral depth and travels the mean epicentral distance for that data bin. The reference rays are traced through the spatially symmetric Jeffreys-Bullen model, which forms a reasonable model for locating ray paths at teleseismic epicentral distances. From Fig. 1 we see that the residuals are symmetric about the mean ray path for teleseismic rays (5° - 30°) demonstrating that there is little bias in the reference model. In this study we use 6 reference rays, i.e., 6 epicentral distance bins, each with 6 epicentral depth bins.

The raw data is represented in Fig. 2 and 3. Half the data is shown in Fig. 2 and the other half is shown in Fig. 3. The scale axis is logarithmic, allowing the data at all scales to be reasonably viewed. Interestingly the data approximately describes a unimodal distribution and an extrapolating variance to small scales. In Fig. 3 we present data for two selected depth bins close to the origin with a linear scale, the same data, but with a different scale is shown on the reasonable straight line. We also observe a general trend of decreasing variance at decreasing scale as would be expected, since the rays are travelling over similar structures. Note the knee in the data that occurs at a scale of 5°-10°. For uniform sampling this would imply that the signal-averaged scalelength is of the same order. Since this sampling is clustered the effective grid size is less than the actual grid size, hence the actual scale length is found to be shorter. We also observe that this implies an increase with increasing source depth. This is an expected since path length decreases as source depth is increased. The variance decreases weakly with distance. This implies that the strength of heterogeneity must decrease with depth, since if the Earth had constant heterogeneity throughout then the variance would increase with depth [see equation (11) below]. Finally, we observe that the variance approaches a finite value at small scales. We interpret this as an estimate of the natural variance. Since the variance at all scales is observed, the data curves follow our intuition in their
INTRODUCTION

Since the inception of plate tectonics, we have been unable to elucidate the details of how it is powered by mantle convection. With the advent of deterministic (tomographic) models of lateral seismic variations (Dziewonski 1986; Clayton & Comer 1983; Hager & Clayton 1989) we have the first chance of imaging directly the driving forces of mantle convection and, hence, constraining models of convection and establishing how plate tectonics is powered. The body wave tomography models quoted above have in principle a higher spatial resolution than surface wave or free oscillation studies because of the shorter wavelengths of body waves. However, current lower mantle body wave studies suffer from the same problem. There are many questions related to the reliability of these tomographic models, including the following. How much random noise is there in the data and how well is it eliminated from the model? How much aliasing is there, due to the finite parametrization (Dziewonski 1984, t = 6; Clayton & Comer 1983), cell size =250 km, fs = 360? How well do the station corrections account for the crustal and upper mantle structure? How are the effects of the uneven sampling? Geophysical models that differ slightly referred to as GDC, presented a theory to invert the scatter in the traveltimes of seismic waves to obtain the spectra of lateral seismic velocity variations. Following Davies, Gudmundsson & Clayton (1988), we also estimate the error in the ISC data set. This estimated error is subtracted from the data before computing the model, while the model parametrization is extended to the smallest scale avoiding aliasing. The model makes the whole mantle and attempts to account for anisotropic mantle. Hence the stochastic travel time models can be used to answer some of the questions introduced above regarding the results of deterministic inversion.

Convection models have advanced, but they are still generic in that they do not claim to know the initial or boundary conditions, but rather to be able to simulate the behaviour of the actual Earth. Hence deterministic seismic models are of limited direct use (with the possible exception of generation simulations), while stochastic seismic models provide directly spectral information which can be used as constraints. The ISC data set contains very large numbers of data (P = 4 million, S = 0.5 million) of a very heterogeneous nature (arrivals picked by different observers, on different seismometers, at unique sites); hence the data are well suited for a statistical treatment. Since we are also combining greater numbers of data to evaluate fewer parameters, we should more effectively test any model of random error left in the data. All of the body wave lower mantle models discussed above have derived the mantle P-wave velocity structure. In this paper we use S-waves to obtain an independent estimate of mantle structure; i.e. its shear velocity structure.

Seismological studies of the ratio of ln(Vs)/ln(Vp) in the literature have been suggested to be about 2 (Dziewonski & Woodhouse 1988). This value is much higher than that obtained from laboratory and pressure tests. Anderson (1967) has suggested that such values in the lower mantle can be explained as the result of the effect of compression on the elastic properties. He has since been supported by the molecular dynamics calculations of Aagnar & Budowski (1991) which have improved calculations of different period data (Dziewonski & Woodhouse 1987), or data with an unknown noise level (Davies & Gudmundsson 1988) or compare P- and S-wave sets of the same period range and estimate the noise.

The hearth of this stochastic method is evaluating the scatter in the traveltime residuals of rays that travel similar paths. Consider two earthquakes occurring at the same location. They should have the same traveltime to all stations. Since there are two Rayleigh waves, motion will be a reflection of errors in picking the arrival time, errors in locating the source, errors in the clock at the stations etc. Consider now an Earth with laterally uniform velocities at all depths except for one layer, where the velocity is lower by a constant amount in one region. Consider two rays that are sufficiently close together, that they both travel equal distances through the anomalous region. They will record the same residual. Next let the rays be spaced strongly, such that only one ray goes through the anomalous region. In this case the two residuals will be different. By discovering at what separation the difference appears we can estimate the size of the anomalous region. From the magnitude of the difference between the residuals we can also evaluate the magnitude of the region of ray perturbation, assuming the ray length through the region. We can investigate the depth variations of velocity variations in the lower mantle by comparing the scatter in the traveltime residuals of rays that travel equal distances, since rays that travel further penetrate deeper into the Earth. Equally bundles with deep source depths sample only half the upper mantle compared to bundles with shallow source depths, hence this allows us to investigate the depth variations of velocity variations in the upper mantle.

EVALUATION OF TRAVELTIME VARIANCE

As discussed above, the heart of the stochastic method involves evaluating the scatter in the traveltime residuals of a bundle of rays. We define the bundle of rays, by means of an unique source and receiver cell on the surface of the Earth. We follow the procedure outlined in GDC, where we divided the earth into a grid of equal area cells. Grids were generated at scales varying from one box for the whole Earth to boxes on which "r" is a side. Rays are binned together if they fall in the same epicentral distance (A) range (the distance between source and receiver) and between the same source depth (Z) bin. A summary ray is a collection (a bundle) of rays that share the same epicentral distance and source depth bins, and also the same receiver cell. The bins are 4° wide in epicentral distance, extending from 31° to 75°. The bins in source depth are 0, 50–60, 100–200, 200–300, and 400–700 km. Since the grid is defined as the angular radius of a circle on the Earth's surface, which has the same number of cells and size, this is a more practical cell system, denoted by Z6 and represented in degrees. We use data collected by the ISC from 1964 to 1986. We select S-wave arrivals from events that have at least 50 phases reported because these would be better located. All events located at 0 or 33 km in depth are ignored, since the ISC frequently assigns poorly located events to these depths. We use shear wave arrivals out only to 75° to avoid contamination by 58°-wave, whose traveltime curve crosses 5 at about 80°. The length of the traveltime curve is 1 as the ray that is directed down and slightly to the right at this distance. Having no data beyond 75° means that we do not sample the lower 450 km of the mantle. Hence, we have no information on the shear wave velocity structure at the core-mantle boundary (CMB). We selected only shear wave arrivals whose 58°-wave traveltime residuals (dt) lie between bearing of +90° (mean = 1.0s) to remove obviously incorrect residuals (e.g., 1 min errors, phase mispick etc.).

Due to problems of tipplings, we only use teleseismic phases. Hence, all the rays have similar paths in the upper mantle. Consequently we would have poor radial resolution in the upper mantle. However, enough rays occur at all depths in the upper mantle and we gain some radial resolution by comparing the scatter in bundles of rays at different source depths. We compare rays with a shallow source depth has a longer path length on the source-side leg than rays with a deep source depth. Hence, it samples shallow heterogeneities that are not sampled by the deep source depth bundle.

To measure the degree of correlation between the traveltimes of the components of a summary ray, we evaluate the variance of the residuals as follows:

\[ \sigma^2 = \sigma^2_S \sigma^2_P \sigma^2_{S-P} \]

where \( \sigma^2_S \) is the traveltime residual of the ith component ray, \( \sigma^2_P \) is the mean residual for the kth summary ray, \( \sigma^2_{S-P} \) is the variance of the kth summary ray, and \( n_k \) is the number of component rays in the summary ray.

We only use summary rays which have at least four component rays (n ≥ 4). Then we average the variance of all the summary rays at a given grid size, epicentral distance and source depth bin. To improve the variance estimate we repeated the process on the same grid, but with all the cells displaced one-sixth of their width in longitude. This is done six times and the mean of the six is used as the estimate of the variance.

Here p is the summation index over the n_k summary rays that are from the same A, and Z bin, and Z is the summation index over the n_k = 6 grid rotations. The variance of the individual summary rays \( \sigma^2_{S-P} \) is evaluated in equation (1). Note that \( n_k \) is a function of q, i.e., the total number of summary rays depends on the grid orientation. The above equation may seem complicated, but it is in fact very simple, since it is only an average variance for all the summary rays with at least four component rays, found for all cells and orientations of the grid.

The uncertainty of the estimate is derived from the scatter between the estimates of the six rotated grids. At small scales with grid size 60 km, the uncertainty may not be significant if there are more than 25 summary rays, i.e., n_k > 25.

Each epicentral distance and source bin is characterized by a reference ray, which is located at the mean epicentral depth and travels the mean epicentral distance for that data bin. The reference rays are traces through the spherically symmetric Jeffreys-Bullen model. Our goal is to find a reasonable model for locating ray paths at teleseismic epicentral distances. From Fig. 1, we see that the residuals are symmetric with respect to the location for teleseismic rays (A = 30°) demonstrating that there is little bias in the reference model. In this study we use 6 reference rays, 4 epicentral distance bins, each with six epicentral depth bins.

The raw data \( \sigma^2_{S-P} \) are presented in Figs 2 and 3. Half the data are not shown, but all are presented in Fig. 2. The scale axis is logarithmic, allowing the data at all scales to be reasonably viewed. Interestingly the data approximately describe a power-law behaviour on a log-log plot, extrapolating to variance at small scales. In Fig. 3 we present data for two selected depth bins close to the origin with a linear scale. The data is again consistent with the power-law behaviour of a reasonable straight line. We also observe a general trend of decreasing variance at decreasing scale as would be expected, since the rays are travelling over ever smaller paths. Note the knee in the data that occurs at a scale of 5°-10°. For uniform sampling this would imply that the signal-averaged scalelength is of the same order. Since the sampling is clustered the effective grid size is less than the actual grid size, hence the actual scalelength is found to be shorter. We also observe that the variance decreases with increasing source depth. This is expected since path length decreases as source depth is increased. The variance varies weakly with depth. This implies that the strength of heterogeneity must decrease with depth, since if the Earth had constant heterogeneity throughout then the variance would increase with increasing depth [see equation (11) below]. Finally, we observe that the variance approaches a finite small value at small scales. We interpreted this as a result of the fact that is calculated at each grid point. The data curves follow our intuition in their
behaviour and are reasonably coherent from curve to curve. This gives us hope that the data can be reasonably modelled in terms of Earth structure.

We estimate the intercept by extrapolating all the data for a given depth bin of a scale smaller than $S^5$, assuming that the decay is linear and that the slope is fixed as $b = a + c$ (where $a$ is the intercept, $c$ is the slope, and $d$ is the depth). The linear decay is supported by the fact that the data at the smallest scales are well described by a straight line, as seen in Fig. 3. Assuming that most of the small-scale signal is located at shallow depths close to the receiver, we weight the slope by $1/cos(i)$, which is proportional to the path length near the surface. The different source depth bins require different intercept estimates. This suggests that the intercept signal is not solely due to different estimates of the smallest scale signal for different depth bins, but reflects a real decrease in noise with increasing source depth.

In Fig. 4 we illustrate the intercept estimates for all the reference rays. We find that the estimates decrease with source depth and are relatively constant as a function of epicentral distance. The average variance for the intercepts for sources shallower than 100 km is $\approx S^5$ as compared to a maximum total signal at the largest scales of $\approx S^7$. The uncertainty in estimating the intercept is high. Hence, the residual data arising from subtracting the intercepts from the original data have a proportionately higher uncertainty. The extrapolation of the intercept could be done logarithmically rather than linearly. The straight line that the data describe at larger scales in Fig. 2 supports this choice. The logarithmic intercepts are shown in Fig. 4(b). We have used both sets of intercepts estimates in our inversion of the residual variance for the statistics of Earth structure. We observed that the differences in the results are restricted to small-scale structure. The residual variance data that result from the subtraction of linear intercept estimates (Fig. 4a) are shown in Fig. 5. It is the residual variance data, following the subtraction of linear intercept estimates, that are used to constrain the spectrum of mantle heterogeneity via a linearized inversion procedure. Note that the intercept estimates include errors arising from incorrect hypocentral parameters, hence their contamination of estimates of structure is limited. Before developing the inversion procedure though we must derive a forward model to relate the variance to the lateral variations of velocity structure: this is outlined next.

3. LINEAR MODEL RELATING TRAVELTIME VARIANCE TO SPECTRA OF MANTLE STRUCTURE

Using ray theory we develop a linear theory to relate the variance of the traveltime residuals of the different bins of rays at different grid sizes to the spectra of the seismic heterogeneity as a function of depth through the mantle. We do this in two stages: first we relate the variance of the sum of the rays to the autocovariance of traveltime residuals, and second we relate the autocovariance of the traveltime residuals to the autocovariance of the slowest perturbations in the medium. Since the details of the derivation of the forward model have been given previously in GIDC, we shall only outline the derivation.

The relationship between the variance of the sum of rays $\sigma^2(\theta, \lambda, Z)$ and the autocovariance of the traveltime residuals $\langle T(\theta) T(\phi) \rangle$, where $\phi$ is the separation of the endpoints, is geometrical, and is represented by a weighting function $w(\theta, \phi)$. To account for non-uniform sampling we also introduce the autocovariance of the sampling function $\langle B(\theta, \phi) \rangle$. By assuming that the sampling is independent of the structure, we can separate $w$ and $B$, and hence we get

$$\sigma(\theta) = \int w(\theta, \phi) |T(\theta) - T(\phi)| \sigma(B(\theta, \phi)) d\phi$$.

By using ray theory, and assuming that our reference model is linearly close to the real Earth, we can apply Fermat's principle and linearly relate the traveltime residual linearity to the slowest perturbations along the ray path

$$\Delta t(\theta) = \int \mu_t(x) ds$$


Figure 2. Variance of a selection of reference rays plotted versus grid size. Note the logarithmic scale for the grid size axis. The variance decreases regularly as scale of grid decreases, and the data from the deepest depth bins (open symbols) are lower than the shallower data; e.g. the filled squares, circles and triangles.

Figure 3. Variance of a selection of data points plotted versus the scale of grid size, for small grid sizes. Note scale axis is linear.

Figure 4. Intercept estimates versus epicentral distance. (a) Linear extrapolation. (b) Logarithmic extrapolation.

Figure 5. Data after removal of the linearly extrapolated intercepts of Fig. 4a.)
behaviour and are reasonably coherent from curve to curve. This gives us hope that the data can be reasonably modelled in terms of Earth structure.

We estimate the intercept by extrapolating all the data for a given depth bin of a scale smaller than 5°, assuming that the decay is linear and that the slope is fixed as $b = \frac{1}{\cos(i)}$, where $i$ is the angle of incidence at the receiver. The linear decay is supported by the fact that the data at the smallest scales are well described by a straight line, as seen in Fig. 3. Assuming that most of the small-scale signal is located at shallow depths close to the receiver, we weight the slope by $\frac{1}{\cos(i)}$, which is proportional to the path length near the surface. The different source depth bins require different intercept estimates. This suggests that the incoherent signal is not solely due to different estimates of the smallest scale signal for different depth bins, but reflects a real decrease in noise with increasing source depth.

In Fig. 4 we illustrate the intercept estimates for all the reference rays. We find that the estimates decrease with source depth and are relatively constant as a function of epicentral distance. The average variance for the intercepts for sources shallower than 100 km is $\sim 0.05$ s$^2$ as compared to a maximum total signal at the largest scales of $\sim 0.08$ s$^2$. The uncertainty in estimating the intercept is high. Hence, the residual data arising from subtracting the intercepts from the original data have a proportionately higher uncertainty. The extrapolation of the intercept could be done logarithmically rather than linearly. The straight line that the data describe at larger scales in Fig. 2 supports this choice. The logarithmic intercepts are shown in Fig. 4(b). We have used both sets of intercepts estimates in our inversion of the residual variance for the statistics of Earth structure. We observed that the differences in the results are restricted to small-scale structure. The residual variance data that result from the subtraction of linear intercept estimates (Fig. 4a) are shown in Fig. 5. It is the residual variance data, following the subtraction of linear intercept estimates, that are used to constrain the spectrum of mantle heterogeneity via a linearized inversion procedure. Note that the intercept estimates include errors arising from incorrect hypocentral parameters, hence their contamination of estimates of structure is limited. Before developing the inversion procedure though we must derive a forward model to relate the variance to the lateral variations of velocity structure; this is outlined next.

3. LINEAR MODEL RELATING TRAVELTIME VARIANCE TO SPECTRA OF MANTLE STRUCTURE

Using ray theory we develop a linear theory to relate the variance of the traveltime residuals of the different bins of rays at different grid sizes to the spectra of the seismic heterogeneity as a function of depth through the mantle. We do this in two stages; first we relate the variance of the summation rays to the autocovariance of traveltime residuals, and second we relate the autocovariance of the traveltime residuals to the autocovariance of the slowest perturbations in the medium. Since the details of the derivation of the forward model have been given previously in GICD, we shall only outline the derivation.

The relationship between the variance of the summation rays $\sigma^2(\theta, \lambda, Z)$ and the autocovariance of the traveltime residuals $\rho^2(\alpha, \beta)$ is geometrical, and is represented by a weighting function $w(\theta, \lambda, Z)$. To account for non-uniform sampling we also introduce the autocovariance of the sampling function $B(\theta, \lambda)$. By assuming that the sampling is independent of the structure, we can separate $w$ and $B$, and hence we get

$$\sigma^2(\alpha) = \int \rho^2(\alpha, \beta) B(\theta, \lambda) w(\theta, \lambda, Z) d\beta$$

(3)

By using ray theory, and assuming that our reference model is linearly close to the real Earth, we can apply Fermat’s principle and linearly relate the traveltime residual linearity to the slowest perturbations along the ray path

$$\Delta t(\eta) = \int \Delta t(x) ds$$

(4)

The use of ray theory enables us to relate the variance of the traveltime residuals to the spectra of the slowest perturbations in the medium. This relationship is used to constrain the spectrum of mantle heterogeneity via a linearized inversion procedure. The intercept estimates include errors arising from incorrect hypocentral parameters, hence their contamination of estimates of structure is limited. Before developing the inversion procedure though we must derive a forward model to relate the variance to the lateral variations of velocity structure; this is outlined next.
where $dR$ is the traveltime residual, $U$ is the slowness ($1/\text{velocity}$) and $S$ is the path length. Hence the autocorrelation function of the traveltime residuals, $T(a)$, can be related to the autocovariance of the medium, $R(x)$ (which is assumed to be a function of separation only and not of direction, i.e., we assume isotropy) as follows:

$$T(a) = E[R(b)R(a+b)],$$

where $R = [r_1, r_2]$ is the separation of the endpoints, and $E$ is the expectation operator. By substitution of Fermat’s equation we get

$$T(a) = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} E[U(x)U(y)] |S_x - S_y| \, dx \, dy,$$

(5)

where $D = \text{grad} \, U + \text{grad} \, U^T$.

To evaluate one of the integrals we approximate the ray geometry by two parallel rays and assume that the radius of curvature of the rays is greater than the scalelength of the anisotropies; then it can be shown that

$$T(a) = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{1}{2} \left( \frac{\partial^2 R}{\partial x^2} \right) |S_x - S_y| \, dx \, dy,$$

(6)

where $C$ is a constant of order 2; its exact value depends upon the medium autocorrelation function but is exactly $(\pi/2)\sqrt{2}$ if it is Gaussian, while $\chi_{1/2}$ is the half-width of $R(p)$, i.e., $(R_1 - 0.5R_0)$. For the spherical Earth we have parameterized the slowness perturbations as follows:

$$\delta S(x, y, z) = \sum_{i, j=1}^{\infty} \rho_{ij} \delta Y_{ij}(x, y, z),$$

(7)

where $Y_{ij}(x, y, z)$ are the fully normalized spherical harmonics, $\rho_{ij}$ is the colatitude and $\phi_{ij}$ is the longitude, $\rho_{ij}$ are the coefficients of the appropriate spherical harmonic and $m$ is harmonic order. Hence the medium autocorrelation function can be written as

$$R(x, y) = \frac{1}{2\pi} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} |S_x - S_y| \, dx \, dy,$$

(8)

where

$$\rho_{ij} = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \rho_{ij} \delta Y_{ij}(x, y, z),$$

(9)

where $\lambda$ denotes complex conjugate, $\lambda$ is the angular separation of the two points and $\rho_{ij}$ is the Legendre function of degree $l$. Hence by combining the equations above we find the equation relating residual summary ray variance, $\sigma^2$, (with the estimated intercepts removed) at scale $\theta$, to the power spectrum, $Q(z)$, of the structure:

$$\sigma^2(\theta) = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{1}{2\pi} \left( \frac{\partial^2 R}{\partial x^2} \right) |S_x - S_y| \, dx \, dy.$$

(10)

The above equation can be cast in matrix form as

$$D = \text{GRF}^T,$$

(11)

where $D = \text{GRF}^T$.

Figure 6. Trade-off surface for inversion. The contoured surface illustrates how a sum of the model error and the resolution lengths of both the depth and density matrices varies with the depth (alpha $\alpha$) and spectral (beta $\beta$) damping parameters. The cross with the horizontal and vertical bars shows the position of the favoured solution (Fig. 9) while the diagonal cross shows the damping parameters for the alternative solution (Fig. 10); both are obviously close to the minima. Since the normalizations and weighting of the different components are slightly arbitrary it is obviously treacherous to search for an exact minima. Solutions from this region of parameter space were similar. An exception is shown in Fig. 10.

$$\mathbf{F} = \mathbf{R}^T \mathbf{S} \mathbf{R},$$

(12)

where $\mathbf{F}$ are the matrices of the left and right eigenvectors corresponding to non-zero singular values and $\mathbf{R}$ are the diagonal matrices of the singular values. Equations (15) and (16) can be rewritten as

$$\mathbf{X} = \mathbf{V} \mathbf{A} \mathbf{F} + \mathbf{U} \mathbf{B} \mathbf{F}.$$

(13)

where $\mathbf{U}$, $\mathbf{V}$, $\mathbf{B}$, $\mathbf{A}$ and $\mathbf{F}$ are $\mathbf{D}$, $\mathbf{S}$ and $\mathbf{R}$.

Notice that damped least-squares modelling leads to global damping, i.e., the same value of the damping parameters is used in deriving the whole model. We have poor depth resolution in the upper mantle (since no rays bottom out), while we have relatively large error variance in the lower mantle (since the model is small there). Thus, we would prefer to tune the trade-off differently in these two parts of the model. This problem can be partially alleviated by a suitable choice of parameterization. Given the large uncertainties in the finely parameterized model STP1 of GDC we use a coarse depth parameterization in our model, similar to that of model STP2 of GDC. The depth boundaries in the model, giving the radial parameterization are at 61, 300, 540, 670, 970, 1470 and 2490 km depth, i.e. seven depth layers.

The spectral parameterization was chosen to be parabolic, i.e., we write the depth parameters to give one spectral model parameter, with the number of harmonics summed into a single model parameter increasing approximately parabolically with the harmonic degree of the mid-point. We invert for 50 such parameters, the first 11 representing single harmonics, with the following parameters representing degree 2, with the number increasing parabolically until the last parameter represents 40 harmonic degrees, from 460 to 500. Notice that the smallest model parameter we use has a scale of 7. Hence, our data do not constrain any power present below this scale.

We define a simple measure of resolution length in both depth and spectra for display purposes as follows:

$$l_i = 1 - \frac{1}{\lambda} \left( 1 - R_{\lambda}^2 \right) \sum_{j} R_{ij}^2,$$

(14)

where $l_i$ is the simple measure of the resolution length of the $i$th element, $\lambda$ is the size of the element, i.e., when considering spatial resolution this is the thickness of the corresponding layer in km, while when considering spectral resolution it is the number of harmonic degrees combined in the spectral parameter, and $R_{ij}$ is the sum of all the elements of the $i$th column of the appropriate resolution matrix as defined below.

The spectral and (radial) resolution matrices are

$$\mathbf{S} = \mathbf{Q}^T (\mathbf{Q}^T \mathbf{D} \mathbf{Q} + \mathbf{Q} \mathbf{D}^T \mathbf{Q})^{-1}$$

(15)

and

$$\mathbf{R} = \mathbf{Q}^T (\mathbf{Q}^T \mathbf{D} \mathbf{Q} + \mathbf{Q} \mathbf{D}^T \mathbf{Q})^{-1} \mathbf{Q} \mathbf{D}^T \mathbf{Q},$$

(16)

The spectral resolution (see Fig. 7a) behaves nearly linearly with harmonic degree. This is due to the uneven sampling of the scale depth. The depth resolution (see Fig. 7b) is nearly identical to the radial parameterization in the upper mantle. The poorest resolution occurs just above the 670 km discontinuity and in the shallowest layer. The resolution is relatively poor in the upper mantle, but
where \( \sigma \) is the traveltime residual, \( U \) is the slowness (1/velocity) and \( S \) is the path length. Hence the autocorrelation function of the traveltime residuals, \( T(\sigma) \), can be related to the autocovariance of the medium, \( R(\sigma) \), which is assumed to be a function of separation only and not of direction, i.e. we assume isotropy as follows:

\[
T(\sigma) = \int \rho(\sigma) d\rho(\sigma)
\]

where \( p = \|p\| \) is the separation of the endpoints, and \( E \) is the expectation operator. By substitution of Fermat’s equation we get

\[
T(\sigma) = \frac{1}{\sqrt{2\pi}} \int \exp\left(-\frac{x^2}{2}\right) dx = \frac{1}{\sqrt{2\pi}}
\]

which states that the data, \( D \), can be related to the model, \( X \), by a set of kernel matrices. One is \( G_0 \), which relates to the variance of the ith reference ray to the value of the model at the ith depth. The second if \( G_2 \), which relates the variance at scale \( s \) to the model’s ith harmonic degree.

An intuitive explanation for why the variance of traveltimes is proportional to the product of the variance of slowness variations and a scale factor of the slowness variations integrated along the ray length can be developed by analogy to the random walk of a drunkard. Consider a 1-D medium consisting of equal length segments in which the slowness perturbations are of constant magnitude, but random in sign. A ray path through such a medium can be thought of as a random walk. We can think of a ray stepping from one anomaly to the next with the cumulative traveltime residual increasing or decreasing by a fixed amount. This is equivalent to the 1-D random walk, where the drunk can go to the right (positive residual) or to the left (negative residual). The most probable result is a zero residual (no digression), but the variance of the residuals is proportional to the number of steps times the square of the residual of a single step, as in a random walk. The residual of a single step is proportional to the product of the slowness perturbation and its width (scalelength). The number of steps is proportional to the ray length divided by the perturbation width. Hence, the observed traveltime variance is proportional to the ray length, the scalelength and the square of the slowness perturbation.

4 INVERSION

In the previous section we have developed a linear formulation for the forward problem describing the effect of lateral heterogeneity, \( X(n) \), the product of the half-width of the correlation function of slowness variations and the variance of the scale of the slowness perturbations, hereafter referred to as the scattering power) on the variance of the summation rays, \( D_i \). Hence we can invert equation (11) using any of a number of linear inversion techniques. We chose to use a damped least-squares method, which gives

\[
X = \mathbf{G}^T \mathbf{F}^T \mathbf{F} \mathbf{G} + \alpha \mathbf{I}^{-1} \mathbf{F}^T \mathbf{F} \mathbf{X}
\]

which can also be written as

\[
X = \mathbf{G}^T \mathbf{G} + \alpha \mathbf{I}^{-1} \mathbf{F}^T \mathbf{F} \mathbf{X}
\]

This is a 2-D inverse problem with two rather than one kernel matrix and a model matrix rather than a model vector. The sampling of the data is the same as before, but the two singular values require the singular vectors as well.

where \( \mathbf{G} \) and \( \mathbf{F} \) are the matrices of the left and right eigenvectors corresponding to non-zero singular values, and \( \mathbf{D} \) are the diagonal matrices of the singular values. Using equations (15) and (16), (14) can be rewritten as

\[
X = \mathbf{G}^T \mathbf{A} + \mathbf{E}^T \mathbf{F} \mathbf{X}
\]

where we have used that \( \mathbf{E}^T \mathbf{E} = \mathbf{I} \) or \( \mathbf{E} = 1 \), and a similar transformation to the one that is used to relate equation (13) to equation (14). Notice that the only matrices which need to be inverted are diagonal, and hence, can be inverted trivially. This reduces the cost of computations that require repeated inversion, such as a search for suitable damping parameters.

Lagrangian minimization techniques minimize the prediction error, i.e., the sum of the squares of the differences between the data and the data as predicted by the model. Damped least-squares inversion minimizes a linear sum of the prediction error and the L2 norm of the model (in this case we can think of our model matrix disassembled into a single long vector, in order to define an L2 norm). The damping parameters determine the relative importance of data prediction and model minimization. We selected the damping parameters such that model error was minimized, while depth and spectral resolution were maximized. We investigated a number of different linear combinations of the above model measures and found that the trade-off surface changed little. An example of a trade-off surface is shown in Fig. 6. Solutions from work with a constant (defined by the smallest contour) were investigated and found to be largely insensitive to the exact choice of damping. The small cross with vertical and horizontal bars shows the damping parameters chosen for the presented solution. The cross with diagonal bars shows the damping parameters used in an alternate solution to illustrate which features of the solution are robust. This choice of damping effectively limits the solution to nine spectral degrees of freedom and 5-6 radial degrees of freedom.

Notice that damped least-squares modelling leads to global damping, i.e., the same value of the damping parameters is used in deriving the whole model. We have poor depth resolution in the upper mantle (since no rays bottom there), while we have relatively large error variance in the lower mantle (since the model is small there). Thus, we would prefer to tune the trade-off differently in these two parts of the model. This problem can be partly alleviated by a suitable choice of parameterization. Given the large uncertainties in the finely parametrized model STP1 of GDC we use a coarse depth parametrization in our model, similar to that of model STP2 of GDC. The depth boundaries in the model, giving the radial parametrization are at 67, 300, 540, 670, 970, and 2450 km depth, i.e. seven depth layers.

The spectral parametrization was chosen to be parabolic, i.e., we optimize the spectral degrees to give one spectral model parameter, with the number of harmonics summed into a single model parameter increasing approximately parabolically with the harmonic degree at the midpoint. We invert for 50 such parameters, the first 11 representing single harmonics, with the following parameters representing harmonic degree, with the number increasing parabolically until the last parameter represents 40 harmonic degrees, from 450 to 500. Notice that the smallest set of layers we use has a scale of 7. Hence, our data do not constrain any power present below this scale.

We define a simple measure of resolution length in both depth and spectra for display purposes as follows:

\[
L_i = A_i + \frac{1}{\sum_{j=1}^{N-1} A_j}
\]

where \( L_i \) is the simple measure of the resolution length of the \( i \)th element, \( A_i \) is the size of the element, i.e. when considering spatial resolution this is the thickness of the corresponding layer in km, while when considering spectral resolution it is the number of harmonic degrees combined in the spectral parameter, and \( N \) is the sum of all the elements of the \( i \)th column of the appropriate resolution matrix as defined below.

The spectral and (radial) resolution matrices are

\[
\mathbf{S} = \mathbf{G}^T \mathbf{G} + \alpha \mathbf{I}^{-1} \mathbf{F}^T \mathbf{F} \mathbf{S}
\]

and

\[
\mathbf{R} = \mathbf{G}^T \mathbf{G} + \alpha \mathbf{I}^{-1} \mathbf{F}^T \mathbf{F} \mathbf{R}
\]

The spectral resolution (see Fig. 7a) behaves nearly linearly with harmonic degree. This is due to the uneven sampling of the scale depth. The depth resolution (see Fig. 7b) is nearly identical to the radial parametrization in the lower mantle. The poorest resolution occurs just above the 670 km discontinuity and in the shallowest layer. The resolution is relatively poor in the upper mantle, but
because of the coarse parameterization of the lower mantle, each model parameter is nearly perfectly resolved.

In evaluating the F matrix, the clustering of stations and events was taken into account, i.e. $R(\theta, \lambda)$ was evaluated. This was done by evaluating the frequency of event pairs and station pairs of a given separation, which is proportional to $R(\theta, \lambda)w(\theta, \lambda)$. Hence, we can derive $B$ by dividing the histograms by the known kernel functions, $w(\theta, \lambda)$. This was repeated over different scales. Much like GDC we found that the functions, $B$, are reasonably described by power laws of separation (see Fig. 8) with the exponent decreasing as scale increases (at vanishing scales the exponent must be infinite). The variation in the exponent with scale was found to be well fit by a linear relation. Since this is computationally very expensive the distribution was evaluated for the data of a single reference ray (51°–55° and 1–32 km), and was assumed to hold for all other reference rays. The distribution for the S-waves was found to be a weaker function of scale than for the P-waves (GDC). This is not unexpected since the P-wave data set is approximately 5 to 6 times larger, while the pattern of seismicity and station distribution is little changed between the two data sets.

Using kernels based upon a uniform sampling distribution would lead to overestimates of correlation length, and hence, underestimates of slowness perturbations. The S-waves show a marked deviation from a power-law relation at the largest scales, see the curve for scale 90° in Fig. 8. The curve shows a large deficiency of pairs at large separations. Therefore, we have few data at large scale, which are related to large separations. Hence, sensitivity to the largest scale heterogeneities is limited. This follows from the fact that no major zones of seismicity are antipodal; equally there are no antipodal continents.

5 RESULTS

The results of our inversion of residual variance for the statistics of Earth structure are presented in Figs 9(a), (b), (c), (d), (e) and (f). In Fig. 9(a) we present the total scattering power (the product of half-width and the slowness variance). It is highest in the upper mantle and decreases with increasing depth. There are two exceptions to this behaviour. The surface value is low, and there is an
Figure 7. Radial (a) and spectral (b) resolution matrices, on the left, with the simple measure of resolutions as given by equations (19) and (20) on the right. Note both radial figures (a) have the same vertical depth axis, while both of the spectral (b) figures have the same vertical harmonic degree axis.

because of the coarse parametrization of the lower mantle, each model parameter is nearly perfectly resolved.

In evaluating the F matrix, the clustering of stations and events was taken into account, i.e. $R(\Theta, \lambda)$ was evaluated. This was done by evaluating the frequency of event pairs and station pairs of a given separation, which is proportional to $R(\Theta, \lambda) w(\Theta, \lambda)$. Hence, we can derive $B$ by dividing the histograms by the known kernel functions, $w(\Theta, \lambda)$. This was repeated over different scales. Much like GDC we found that the functions, $B$, are reasonably described by power laws of separation (see Fig. 8) with the exponent decreasing as scale increases (at vanishing scales the exponent must be infinite). The variation in the exponent with scale was found to be well fit by a linear relation. Since this is computationally very expensive the distribution was evaluated for the data of a single reference ray (51°–55° and 1–32 km), and was assumed to hold for all other reference rays. The distribution for the S-waves was found to be a weakly function of scale than for the P-waves (GDC). This is not unexpected since the P-wave data set is approximately 5 to 6 times larger, while the pattern of seismicity and station distribution is little changed between the two data sets.

Using kernels based upon a uniform sampling distribution would lead to overestimates of correlation length, and because of the coarse parametrization of the lower mantle, each model parameter is nearly perfectly resolved.

In evaluating the F matrix, the clustering of stations and events was taken into account, i.e. $R(\Theta, \lambda)$ was evaluated. This was done by evaluating the frequency of event pairs and station pairs of a given separation, which is proportional to $R(\Theta, \lambda) w(\Theta, \lambda)$. Hence, we can derive $B$ by dividing the histograms by the known kernel functions, $w(\Theta, \lambda)$. This was repeated over different scales. Much like GDC we found that the functions, $B$, are reasonably described by power laws of separation (see Fig. 8) with the exponent decreasing as scale increases (at vanishing scales the exponent must be infinite). The variation in the exponent with scale was found to be well fit by a linear relation. Since this is computationally very expensive the distribution was evaluated for the data of a single reference ray (51°–55° and 1–32 km), and was assumed to hold for all other reference rays. The distribution for the S-waves was found to be a weakly function of scale than for the P-waves (GDC). This is not unexpected since the P-wave data set is approximately 5 to 6 times larger, while the pattern of seismicity and station distribution is little changed between the two data sets.

Using kernels based upon a uniform sampling distribution would lead to overestimates of correlation length, and because of the coarse parametrization of the lower mantle, each model parameter is nearly perfectly resolved.

In evaluating the F matrix, the clustering of stations and events was taken into account, i.e. $R(\Theta, \lambda)$ was evaluated. This was done by evaluating the frequency of event pairs and station pairs of a given separation, which is proportional to $R(\Theta, \lambda) w(\Theta, \lambda)$. Hence, we can derive $B$ by dividing the histograms by the known kernel functions, $w(\Theta, \lambda)$. This was repeated over different scales. Much like GDC we found that the functions, $B$, are reasonably described by power laws of separation (see Fig. 8) with the exponent decreasing as scale increases (at vanishing scales the exponent must be infinite). The variation in the exponent with scale was found to be well fit by a linear relation. Since this is computationally very expensive the distribution was evaluated for the data of a single reference ray (51°–55° and 1–32 km), and was assumed to hold for all other reference rays. The distribution for the S-waves was found to be a weakly function of scale than for the P-waves (GDC). This is not unexpected since the P-wave data set is approximately 5 to 6 times larger, while the pattern of seismicity and station distribution is little changed between the two data sets.

Using kernels based upon a uniform sampling distribution would lead to overestimates of correlation length, and

![Figure 8](image)

**Figure 8.** Plot of the normalized log pair frequency as a function of separation of the stations and events for different scales. This is evaluated for the data in the epicentral distance bin from 31° to 33°, and the shallowest epicentral depth bin. The scales are labeled degrees above the curves while the negative numbers below represent the least-squares estimates of the linear gradients of the curves.

**5 RESULTS**

The results of our inversion of residual variance for the statistics of Earth structure are presented in Figs 9(a), (b), (c), (d), (e) and (f). In Fig. 9(a) we present the total scattering power (the product of half-width and the slowness variance). It is highest in the upper mantle and decreases with increasing depth. There are two exceptions to this behaviour. The surface value is low, and there is an

![Figure 9](image)

**Figure 9.** Results. (a) Total scattering power (the product of half-width and the square of the mean slowness), the width of the box represents one standard deviation error. (b) Normalized spectra: the dashed line represents negative power and the number in each box represents the mean radius of that model bin. (c) Short-scale scattering power ($r < 50$); notice power concentrated in uppermost mantle. (d) Large-scale scattering power ($r < 50$); Power is concentrated in the upper mantle and decays with depth in the lower mantle. Notice more power below than above 620 km discontinuity. (e) Slowness perturbations away from reference model. Notice pattern similar to large-scale power. (f) Half-width estimates of large-scale correlation function, around 300–600 km for uppermost mantle. Poorly constrained in lower mantle.
increase in power across the 670 km discontinuity. Both exceptions occur in regions of poor depth resolution.

In Fig. 9(e) we present the lowest 25 harmonic degrees of the spectra for all seven depth bins. They are normalized to unit height. Dashed curves signify negative power, an unphysical result. The negative values are small. Their magnitude is probably a fair indication of the real uncertainty in our solution (including random and systematic data errors and forward modelling errors).

Since the spectral resolution deteriorates with increasing harmonic degree, we decided to split the spectrum into two parts; a long-wavelength part and a short-wavelength part. We note that we can make a reasonable estimate of the half-width of the long-wavelength part, whereas the estimate derived from the whole spectrum would probably not be as meaningful. The cut-off between the two parts was arbitrarily chosen at harmonic degree 50.

In Fig. 9(e) we present the short-scale power, the part of the total power that is due to structure of harmonic degree >50. Nearly all this power is concentrated in the second depth bin (60–300 km depth). There is very little small-scale power in the lower mantle. Some of the values (i.e., the third depth bin) are negative. In Fig. 9(d) we present the integral long-scale power, due to structure of harmonic degree <50. It is largest in the third layer (i.e., from a depth of 300–540 km), and decays through the lower mantle. The layer just below the 670 km discontinuity has more power than the layer above. The negative small-scale power and the huge large-scale power in the third layer suggests that the spectrum for this layer is extremely focused in the low harmonic degrees. A comparison with the other layers of both the P-wave study (GDSC) and the present S-wave study suggests that this is aberrant.

In Fig. 9(f) we present the half-width, which is estimated by evaluating the scale at which the large-scale correlation function has fallen to half its peak value. By ‘large-scale correlation function’, we mean a correlation function that is constructed from only the large-scale portion of the power spectrum. By dividing this estimate of the half-width into the large-scale power (Fig. 9(c)) we obtain the estimate of root mean square RMS slowness, shown in Fig. 9(e).

The main difference between solutions with alternate values of the damping parameter is that the long-wavelength half-width in the lowermost mantle is now 400 km as opposed to 300 km; otherwise the results are very similar. Hence, the large half-width in the lower mantle in Fig. 10 is not a robust feature. It is probable that the long-wavelength half-width in the lowermost mantle is close to the better constrained estimate of 1200 km obtained by GDSC from P-wave data.

In Fig. 11 we compare the data predictions of the model and the original residual variance data. The predicted data lie within the error bars of the data, but in Fig. 11(b) they appear to be offset by a constant, i.e., a different intercept estimate could lead to a better data fit.

The variance reduction by the model in Fig. 9 is 94 per cent. The chi-squared value ($\chi^2$) is approximately equal to the number of degrees of freedom (945) divided by the number of data (66 × 15 = 990), effective number of model parameters 45 (five degrees of freedom in depth and nine spectral degrees of freedom). Hence, the model fit is significant and highly unlikely to be the result of a random distribution of data. Note that there is a potential confusion in comparing the variance reduction of this model with that of deterministic models since our initial data are already variances of travel times.

A significant part of the original signal is taken up in the intercept estimates and does not enter into the structural model. The intercepts and model are not evaluated simultaneously but sequentially. This leads to a simpler and more stable inversion scheme. Different schemes of estimating the intercepts can lead to large differences in the estimates. To see how these differences affect the structural model, we inverted a data set where the intercepts were evaluated using logarithmic rather than linear extrapolation. These intersect estimates are lower and the variance left to be inverted is higher (average variance is 6.5σ as against 4.4σ). The variance reduction (95 per cent) and $\chi^2$ value are similar to the model derived from data using linear intercept estimates. The primary effect on the model is an increase in short-scale power in the depth range from 50 to 300 km (see Fig. 12). The short-scale power in the lower mantle is also slightly increased. The large-scale power is virtually unchanged (see Fig. 13), except for the half-width in the poorly resolved shallowest depth bin. These differences demonstrate the trade-off between intercept estimates and small-scale power due to the lack of data at small scales.

The model error variance is low throughout the model. We note, however, that it is possible that the data contain errors which are spatially coherent, and thus depend on scale length. These errors would not contribute to the intercept estimates, which are measures of the incoherent noise, and are indistinguishable from the residual variance, which we interpret as structural. The only way to confirm the absence of appreciable coherent noise (systematic error) is to obtain an equivalent model using an independent data set. Our results will be shown to be consistent with previous work, suggesting that systematic errors are not a severe problem; this is confirmed by the limited amount of unphysical negative power (Fig. 9b). Negative power arises from trying to fit data where the variance decreases as scale- or path length increases.

Because of the large damping in the upper mantle due to the poor resolution we attempted to model the data using only shallow structure, to see whether we were under-estimating upper mantle structure and whether the data required structure in the lower mantle. Good data fits were obtained provided that at least two layers were included (91 per cent variance reduction; chi squared slightly greater than the number of degrees of freedom). The small-scale structure was unchanged, while the large-scale structure increased by up to 50 per cent and RMS slowness by up to 20 per cent. Thus, little is sacrificed in data fit as we simplify the model greatly and restrict it to the upper mantle. This
increase in power across the 670 km discontinuity. Both exceptions occur in regions of poor depth resolution.

In Fig. 9(b) we present the lowest 25 harmonic degrees of the spectra for all seven depth bins. They are normalized to unit height. Dashed curves signify negative power, an unphysical result. The negative values are small. Their magnitude is probably a fair indication of the real uncertainty in our solution (including random and systematic data errors and forward modelling errors).

Since the spectral resolution deteriorates with increasing harmonic degree, we decided to split the spectrum into two parts: a long-wavelength part and a short-wavelength part. We can then make a reasonable estimate of the half-width of the long-wavelength part, whereas the estimate derived from the whole spectrum would probably not be as meaningful. The cut-off between the two parts was arbitrarily chosen at harmonic degree 50.

In Fig. 9(c) we present the short-scale power, the part of the total power that is due to structure of harmonic degree >50. Nearly all this power is concentrated in the second depth bin (60–300 km depth). There is very little small-scale power in the lower mantle. Some of the values (i.e., the third depth bin) are negative. In Fig. 9(d) we present the integral long-scale power, due to structure of harmonic degree <50. It is largest in the third layer (i.e., from a depth of 300–540 km), and decays through the lower mantle. The layer just below the 670 km discontinuity has more power than the layer above. The negative small-scale power and the huge large-scale power in the third layer suggests that the spectrum for this layer is extremely focused in the low harmonic degrees. A comparison with the other layers of both the P-wave study (GDSC) and the present S-wave study suggests that this is abberant.

In Fig. 9(e) we present the half-width, which is estimated by evaluating the scale at which the large-scale correlation function has fallen to half its peak value. By 'large-scale correlation function,' we mean a correlation function that is constructed from only the large-scale portion of the power spectrum. By dividing this estimate of the half-width into the large-scale power (Fig. 9d) we obtain the estimates of root mean square RMS slowing, shown in Fig. 9(e).

The main difference between solutions with alternate values of the damping parameter is that the long-wavelength half-width in the lowermost mantle is now 400 km, as opposed to 300 km, otherwise the results are very similar. Hence, the large half-width in the lower mantle in Fig. 10 is not a robust feature. It is probable that the long-wavelength half-width in the lowermost mantle is close to the better constrained estimate of 1200 km obtained by GDSC from P-wave data.

In Fig. 11 we compare the data predictions of the model and the original residual variance data. The predicted data lie within the error bars of the data, but in Fig. 11(b) they appear to be offset by a constant, i.e., a different intercept estimate could lead to a better data fit.

The variance reduction by the model in Fig. 9 is 94 per cent. The chi-squared value (χ²) is approximately equal to the number of degrees of freedom (945) [number of data (66 × 15 = 990)], effective number of model parameters 45 (five degrees of freedom in depth and nine spectral degrees of freedom). Hence, the model fit is significant and highly unlikely to be the result of a random distribution of data. Note that there is a potential confusion in comparing the variance reduction of this model with that of deterministic models since our initial data are already variances of traveltime.

A significant part of the original signal is taken up in the intercept estimates and does not enter into the structural model. The intercepts and model are not evaluated simultaneously but sequentially. This leads to a simpler and more stable inversion scheme. Different schemes of estimating the intercepts can lead to large differences in the estimates. To see how these differences affect the structural model, we inverted a data set where the intercepts were evaluated using logarithmic rather than linear extrapolation. These intercept estimates are lower and the variance left to be inverted is higher (average variance is 6.5σ² as against 4.4σ²). The variance reduction (95 per cent) and χ² value are similar to the model derived from data using linear intercept estimates. The primary effect on the model is an increase in short-scale power in the depth range from 50 to 300 km (see Fig. 12). The short-scale power in the lower mantle is also slightly increased. The large-scale power is virtually unchanged (see Fig. 13), except for the half-width in the poorly resolved shallowest depth bin. These differences demonstrate the trade-off between intercept estimates and small-scale power due to the lack of data at small scales.

The model error variance is low throughout the model. We note, however, that it is possible that the data contain errors which are spatially coherent, and thus depend on scale-length. These errors would not contribute to the intercept estimates, which are measures of the incoherent noise, and are indistinguishable from the residual variance, which we interpret as structural. The only way to confirm the absence of appreciable coherent noise (systematic error) is to obtain an equivalent model using an independent data set. Our results will be shown to be consistent with previous work, suggesting that systematic errors are not a severe problem; this is confirmed by the limited amount of unphysical negative power (Fig. 9b). Negative power arises from trying to fit data where the variance decreases as scale- or path length increases.

Because of the large damping in the upper mantle due to the poor resolution we attempted to model the data using only shallow structure, to see whether we were under-estimating upper mantle structure and whether the data required structure in the lower mantle. Good data fits were obtained provided that at least two layers were included (91 per cent variance reduction, chi-squared slightly greater than the number of degrees of freedom). The small-scale structure was unchanged, while the large-scale structure increased by up to 50 per cent and RMS slowdowns by up to 20 per cent. Thus, little is sacrificed in data fit as we simplify the model greatly and restrict it to the upper mantle. This
structure. However, the sparseness of data at the small separations is probably a more severe limitation to our resolution at small scales.

It is evident that rays in a summary ray can be modelled as parallel is questionable at large scale. Only an expensive synthetic study of a wide range of models using the theory to verify its validity is the problem, it should be restricted to the lowest harmonics. We also have incomplete sampling, which could manifest itself in the correlation of structure and sampling. In particular, most of the deep seismic sources are in a few subduction zones, which are known to have a unique seismic signature, i.e. a minimum in power at large scale. The results may be biased towards these parts of the Earth.

Caution is also suggested by certain features in the data that are unexplainable by the theory of GDC. At the large scale, the variance frequently decreases. To explain this, the theory requires unphysical negative model power at large scales. We suggest that this may be explained, as mentioned above, by deviations in the ray geometry from parallelism as assumed by the theory, at scales larger than the maximum expected from the model power explanation. In the large-scale variance is the biasing, e.g. more towards oceanic structure as scales increase. A third feature, in our opinion unlikely explanation, is that the dip at large scale is a reflection of a very large-scale (harmonic degree 2 or 3) periodic convection pattern. We studied the effect of this feature using a large number of the large-scale models that the variance become flat. The resulting model is very similar to the preferred model. There are some differences, particularly in the upper mantle model. It is possible that GDC is a more satisfactory interpretation of the results. In Fig. 12, especially near the 570 km discontinuity. The increase in large-scale power is expected. It gives a quantitative indication. These differences correspond to RMS slowdowns of 1.5 per cent in the oceanic, and 2.5 per cent in the continental areas.

The low level of power in the shallow layer in our model may be artificial. It may be caused by artificial large in a tomographic model data depth. The shallow results in low residual variance, or it could be due to biased sampling. Shallow earthquakes are distributed globally, with a deep power from 60 to 600 km depth. The differences between different models are not significant because of the more oceanic and continental regions.

Collison zones may be more heterogeneous than other provinces of the Earth. This trend has been observed in the results of interferometric and seismological studies. This shows that the difficulties facing traveltime inversions for lateral heterogeneity of the lower mantle, the core–mantle boundary and the core.

Before discussing implications arising from these results, we should like to remind the reader that the simple model presented here gives an excellent fit to the data, as do the P-wave results of GDC. Furthermore, the statistical models derived from both S and P data are similar in pattern. We should also like to remind the reader that the data have been averaged by structure-dependent sampling. The post-resolution of shallow heterogeneity and limited spectral resolution are also (potentially) greater problems in smaller-scale depth sampling. The accuracy of the scattering power results are probably of order 20–30 per cent (as estimated from differences to 20 per cent power at different intermediate independent negative power etc). This is to be contrasted with high precision (from error bars estimated from estimates of error data), and we would have to consider diffraction to obtain the whole suggestions of overestimating the degree of heterogeneity in the lower mantle by power leakage as a result of poor vertical resolution in the upper mantle. Note that the RMS slowness in the lower mantle is already very low, within two standard deviations of zero, and that this test did favour, if only slightly structure in the lower mantle from the improved model. This does illustrate the difficulties facing traveltime inversions for lateral heterogeneity of the lower mantle, the core–mantle boundary and the core.

In the above discussion, we have considered the impact of the use of the full high-frequency limit while we use finite-frequency data. Hence, we have intrinsic averaging and a minimum resolution, which is generally deteriorates with depth. We would have to consider diffraction to obtain the whole structure. However, the sparseness of data at the small separations is probably a more severe limitation to our resolution at small scales.
structure. However, the sparseness of data at the small separations is probably a more severe limitation to our resolution at small scales.

One might argue that rays in a summary ray can be modelled as parallel is questionable at large scale. Only an expensive synthetic study of a wide range of models using the actual ray parameters and the value of the seismic problem, it should be restricted to the lowest harmonics. We also have incomplete sampling, which could manifest itself in the correlation of structure and sampling. In particular, most of the deep seismic sources are in a few sub-sections zones, which are known to have a unique seismic signature. If a localised pattern of results may be biased towards these parts of the Earth.

Caution is also suggested by certain features in the data that are unexplainable by the theory of GDC. At the largest scale the variance frequently decreases. To explain this, the theory requires unphysical negative model power at large scales. We suspect that this may be explained, as mentioned above, by deviations in the ray geometry from parallel as assumed by the theory, at scales larger than the maximum explored. This unphysical explanation is that the sampling is biased, e.g., more towards oceanic structure as scales increase. A third feature, which in our opinion unlikely explanation, is that the dip at large scale is a reflection of a very large-scale (harmonic degree 2 or 3) periodic convection pattern. We studied the effect of this feature on the large and more moderate parts of the variance that the variance become flat. The resulting model is very similar to the preferred model. There are some differences, particularly in the large and moderate parts. As pointed out by GDC a more subtle effect of the large-scale structure arises from the fact that the component rays of a summary ray originate at different source depths below a given source depth bin. There is heterogeneity at the scalelength of the path difference, then this will introduce scale-independent scatter. Also, it is reasonable to expect more backscatter from shallow events than from deep events, since their arrivals are frequently complex and emergent, while deep events generally have simpler impulsive arrivals.

If we assume that the depth variation of the intercepts is primarily due to small-scale structure and finite depth binning, then it is reasonable to take the depth intercept estimates as a measure of the non-structure noise in the data. For teleseismic P-waves this is 0.25 Hz (GDC) and for 4-5 Hz. Estimates of the structural variance are around 1 Hz for P-waves (GDC) and 1 Hz for S-waves. Hence, we estimate the signal-to-noise ratio of teleseismic NC S-wave data as V/S<1.4 as opposed to the V/S<2.5=2 estimate of GDC for P-waves. If we use logarithmic intercept estimates (GDC) we use the signal-to-noise ratio changes to V/S<2.3 for the S-wave data.

2.5 Discussion of seismic velocities
Two striking features of the shear-wave velocity heterogeneity are its concentration in the upper mantle compared to the lower mantle, and its similarity to the results of GDC for P-waves. The results suggest that the large-scale slowow fluctuations in the upper mantle are more than an order of magnitude larger than the large-scale fluctuations in the mid-lower mantle. The halfwidth of the P- and S-wave studies are similar throughout the uppermost mantle, and in the lowermost mantle where the S-wave halfwidth is not robust we suggest that the P-wave halfwidth of around 1200 km might be a reasonable estimate; i.e., about 2-3 times larger than the 400-600 km of the upper half of the mantle.

Tanimoto (1990) derived a long-wavelength, whole-mantle shear velocity model using surface waves. He finds a minimum power around a depth of 200 km, with variations of the order of 0.7 per cent peak to peak, i.e., RMS amplitude variations of 0.2 per cent. In his study, the shear wave velocity is expanded in spherical harmonics to degree 6 (scalelength of order 5000 km). His results suggest RMS variations of order 0.4 per cent (assuming a halfwidth of 1200 km from GDC), these are larger as expected given that our model extends to higher harmonic degree.

The distribution of seismicity over the Earth prohibits a global, upper mantle, shear velocity model based on teleseismic first arrival body waves. However, a number of regional shear velocity models have been derived using high-resolution waveform modelling. For instance, Grand & Helmerger (1984) derived a model for the Canadian shield (SNA) and another model (TNA) appropriate for younger terranes in the North American basin. These two models are appropriate for younger terranes in the North American basin. These two models are different from the Canadian shield model due to the different geological setting.

The resolution of shear body wave studies were regional. Global comparisons can be made with the global, upper mantle, shear velocity models derived from surface waves. One of the main assumptions of the shear body wave data is that the models represent the average properties of the earth's interior. If we assume that the shear velocity models are valid, then the models represent the average properties of the earth's interior. If we assume that the shear velocity models are valid, then the models represent the average properties of the earth's interior. If we assume that the shear velocity models are valid, then the models represent the average properties of the earth's interior.
Free oscillations have also been used to estimate heterogeneity. The study of spherical, fluid-filled models is most probably used for aspheric models. However, the problem of small-scale structures in the pattern of the radial variation of the strength of heterogeneity between the two studies. Both show strong heterogeneity, and both show the approach the lowermost mantle. Both studies have a decrease in the scattering power just above the 670 km discontinuity and an increase just below it. The region above the 670 km discontinuity is poorly resolved (see Fig. 7), and hence best in evidence. For this reason we make this feature in both studies that makes it worthy of mention.

Because of the difference in resolution (spectral depth) of the P- and S-wave data sets, the inverted models for the S- wave data are increasingly different. A quantitative comparison of the local values of the ratio of the P- and S-wave may be inappropriate. The heavy damping of the shallowest layers in the S- wave study, where the P- wave study has relatively lower damping, power, and which have pressure resolved in the lower mantle. They do, however, increase the advantage of averaging global volumes, and hence, provide useful information about the lower mantle heterogeneity, e.g., Wolf and Barreto (1989), which predict that double layers of S- wave studies have a different allocation of power between long- and short-wavelength components compared to the P- wave study. In conclusion, S- wave studies of the lower mantle have not yet been investigated and it seems that the global average of this ratio integrated over a range in layer 1.5.

Other studies have also found values of the P- and S-wave ratio in the mantle. These studies are consistent with mantle velocity heterogeneity to that farther for S- wave velocities. This may be due to the use of different models of heterogeneity. Davies and Clayton (1986) and Davies and Clayton (1986) also find that the mantle velocity heterogeneity is not comparable for S- wave data. This contrasts favourably with the heterogeneity found in the mantle.

Both the Wolf and Wolf and Barreto (1989) data show that the level of small-scale structure in the lower mantle is significantly different from that in the upper. This is consistent with the findings of Wolf and Wolf and Barreto (1989) that the ratio of the P- and S-wave velocities is correlated. If the ratio of the P- and S-wave velocities is correlated, then it is possible that the magnitude of wave propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) and Wolf and Wolf and Barreto (1989) note that the ratio of the P- and S-wave velocities is correlated. We cannot provide a correlation for the ratio of the P- and S-wave velocities. However, Wolf and Wolf and Barreto (1989) have not shown any observational support for the ratio of the P- and S-wave velocities. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed. Nolte (1987) and Wolf and Wolf and Barreto (1989) note that the ratio of the P- and S-wave velocities is correlated. We cannot provide a correlation for the ratio of the P- and S-wave velocities. However, Wolf and Wolf and Barreto (1989) have not shown any observational support for the ratio of the P- and S-wave velocities. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed.

Nolte (1987) and Wolf and Wolf and Barreto (1989) note that the ratio of the P- and S-wave velocities is correlated. We cannot provide a correlation for the ratio of the P- and S-wave velocities. However, Wolf and Wolf and Barreto (1989) have not shown any observational support for the ratio of the P- and S-wave velocities. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed. Nolte (1987) and Wolf and Wolf and Barreto (1989) note that the ratio of the P- and S-wave velocities is correlated. We cannot provide a correlation for the ratio of the P- and S-wave velocities. However, Wolf and Wolf and Barreto (1989) have not shown any observational support for the ratio of the P- and S-wave velocities. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed.

3.2 Comparison with GDC

A comparison of our S-wave results to the P-wave results of GDC (1984) shows a high level of correlation. Most striking is the consistency in the pattern of the radial variation of the strength of heterogeneity between the two studies. Both show strong heterogeneity, and both show the approach the lowermost mantle. Both studies have a decrease in the scattering power just above the 670 km discontinuity and an increase just below it. The region above the 670 km discontinuity is poorly resolved (see Fig. 7), and hence best in evidence. For this reason we make this feature in both studies that makes it worthy of mention.

Because of the difference in resolution (spectral depth) of the P- and S-wave data sets, the inverted models for the S-wave data are increasingly different. A quantitative comparison of the local values of the ratio of the P- and S-wave may be inappropriate. The heavy damping of the shallowest layers in the S-wave study, where the P-wave study has relatively lower damping, power, and which have pressure resolved in the lower mantle. They do, however, increase the advantage of averaging global volumes, and hence, provide useful information about the lower mantle heterogeneity, e.g., Wolf and Barreto (1989), which predict that double layers of S-wave studies have a different allocation of power between long- and short-wavelength components compared to the P-wave study. In conclusion, S-wave studies of the lower mantle have not yet been investigated and it seems that the global average of this ratio integrated over a range in layer 1.5.

Other studies have also found values of the P- and S-wave ratio in the mantle. These studies are consistent with mantle velocity heterogeneity to that farther for S-wave velocities. This may be due to the use of different models of heterogeneity. Davies and Clayton (1986) and Davies and Clayton (1986) also find that the mantle velocity heterogeneity is not comparable for S-wave data. This contrasts favourably with the heterogeneity found in the mantle.

Both the Wolf and Wolf and Barreto (1989) data show that the level of small-scale structure in the lower mantle is significantly different from that in the upper. This is consistent with the findings of Wolf and Wolf and Barreto (1989) that the ratio of the P- and S-wave velocities is correlated. If the ratio of the P- and S-wave velocities is correlated, then it is possible that the magnitude of wave propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) and Wolf and Wolf and Barreto (1989) note that the ratio of the P- and S-wave velocities is correlated. We cannot provide a correlation for the ratio of the P- and S-wave velocities. However, Wolf and Wolf and Barreto (1989) have not shown any observational support for the ratio of the P- and S-wave velocities. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed. Nolte (1987) and Wolf and Wolf and Barreto (1989) note that the ratio of the P- and S-wave velocities is correlated. We cannot provide a correlation for the ratio of the P- and S-wave velocities. However, Wolf and Wolf and Barreto (1989) have not shown any observational support for the ratio of the P- and S-wave velocities. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed. Nolte (1987) and Wolf and Wolf and Barreto (1989) note that the ratio of the P- and S-wave velocities is correlated. We cannot provide a correlation for the ratio of the P- and S-wave velocities. However, Wolf and Wolf and Barreto (1989) have not shown any observational support for the ratio of the P- and S-wave velocities. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed. Nolte (1987) and Wolf and Wolf and Barreto (1989) note that the ratio of the P- and S-wave velocities is correlated. We cannot provide a correlation for the ratio of the P- and S-wave velocities. However, Wolf and Wolf and Barreto (1989) have not shown any observational support for the ratio of the P- and S-wave velocities. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed.

Nolte (1987) and Wolf and Wolf and Barreto (1989) note that the ratio of the P- and S-wave velocities is correlated. We cannot provide a correlation for the ratio of the P- and S-wave velocities. However, Wolf and Wolf and Barreto (1989) have not shown any observational support for the ratio of the P- and S-wave velocities. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed. Nolte (1987) and Wolf and Wolf and Barreto (1989) note that the ratio of the P- and S-wave velocities is correlated. We cannot provide a correlation for the ratio of the P- and S-wave velocities. However, Wolf and Wolf and Barreto (1989) have not shown any observational support for the ratio of the P- and S-wave velocities. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed. Nolte (1987) and Wolf and Wolf and Barreto (1989) note that the ratio of the P- and S-wave velocities is correlated. We cannot provide a correlation for the ratio of the P- and S-wave velocities. However, Wolf and Wolf and Barreto (1989) have not shown any observational support for the ratio of the P- and S-wave velocities. The ratio of the P- and S-wave velocities is not dependent on the mantle or upper mantle structure or velocity heterogeneity as we previously showed.
Free oscillations have now been used to estimate heterogeneity. The study of sphenoidal, sedimentary wall rock, free oscillations and free oscillations by Masters et al. (1982) suggests a large quadrupolar pattern in the transition region. Splitting of free oscillation modes has also been used to detect for asperpherical structure (e.g., Riftweller, Masters & Gilbert 1986). They found perturbations in density in the lower mantle, with one mode being the combined effect of the Earth's density variations and shear velocity variations they suggest variations of up to 0.9 per cent in S-wave velocities. This is for a lower mantle velocity of 8 km/s. When they have a poor depth resolution in the lower mantle. They do have the advantage of averaging global volumes, and hence, provide a useful complement to higher resolution studies, e.g., body wave studies, where potential problems with biased sampling are more prominent. Giardini, Li & Woodhouse (1988), similarly looked at the splitting of long-period normal modes. They show that the splitting functions are consistent with the mantle models of heterogeneity in the mantle (Dziewonski 1984; Woodhouse & Dziewonski 1984) that provided that the splitting function is a combination of lower mantle P and S velocity models (Davies & Clayton 1988, comparison of P and S velocity models (Davies & Clayton 1988) and comparison of P and S velocity models (Davies & Clayton 1988)). We can get a model of the lower mantle P-wave model to an S-wave model to give the best match for free oscillation splitting functions (Giardini et al. 1988). Li, Giardini & Woodhouse (1993) constrain the b in d ln(V)/d ln(V) to be greater than 0.18 for free oscillation data at the 95 per cent confidence level. Our upper mantle is rather than the residual velocities of order 10 km/s should be modelled for S-waves and of order 15 km/s for P-waves. Hence, if the shear and compressional wave velocities are correlated, they are consistent with wave propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects. Nolte (1987) proposed that for the Earth's core, the correlation between shear and compressional wave velocity is usually because of swell propagation, i.e., due to intrinsic averaging and wavefront healing because of low-frequency effects.
of shear and compressional velocities at intermediate scales.

6.6 Comparisons to convection models

What seismic variations might one expect from a convection model? Jarvis (1985) and Jarvis & Peltier (1986, 1989) considered this for the case of a steady, incompressible, constant viscosity, 2-D, unit aspect-ratio convection. They found the boundary layers to have red spectra and that the spectra became progressively whiter towards the middle of cells. Also, the variations had the largest magnitude at the boundaries. At high Rayleigh numbers they found that the convection becomes more vigorous and the boundary layers thinnier. Hence, the spectra become whiter, and stronger.

Towards the top of the mantle we have an increase in heterogeneity and a slight decrease in scale-length. The increased heterogeneity is expected for boundary layers, but we would expect an increase rather than a decrease in scale-length. Our inability to discern changes in characteristic length scales could be due to poor spectral resolution, poor vertical resolution of the data, or this boundary layer is also a site of compositional variations. GDC present similar evidence for the core–mantle boundary. This evidence though, was questioned by Glazmaier (1989), in a synthetic study that showed that due to poor coverage a similar feature can be reproduced in the inversion if the data are contaminated with realistic noise.

Glazmaier (1988) presented one of the more complex simulations of mantle convection to date and evaluated spectra of the thermal variations. He considered compressible 3-D convection, with large Rayleigh numbers. He found large variations in behaviour as he increased the Rayleigh number at $Ra = 10^5$. At $Ra = 10^4$ he found a network of narrow, cold downwellings and broad regions of upwellings, but at $Ra = 10^5$ he found instead hot, plumes in a broad region of warm downwellings. His model produced maximum temperature deviations of about 200 K at the largest Rayleigh number. His computations were truncated at harmonic degree 65. At $Ra = 10^5$ he found that the thermal variance decreased by nearly an order of magnitude from its peak at degrees 3–10 down to degree 50. It is a linear function of degree, and for degree 50 is approximately a linear function of the harmonic degree, beyond $j = 10$. In Fig. 15 we present the histogram of the normalized spectrum versus harmonic degree and find a similar linear behaviour. This spectrum has been defined as power per harmonic degree as in Glazmaier (1988) rather than as conventional power per order [i.e. power per harmonic degree $(2j + 1)$], and it is a depth-weighted average through the mantle. Two curves are presented, for the different damping parameters. The equivalent curves for the model resulting from the data with logarithmic intercept estimates removed, and the lifted asymptote data values are nearly identical to the one with a similar damping parameter. The difference between the two curves suggests that we cannot rigorously estimate the slope, but both show a decrease in thermal variance with depth. If the similarity in form is real, it can potentially provide a constraint on mantle convection and viscosity. A comparison of these curves to Glazmaier's work would suggest a Rayleigh number of either $10^5$ or $10^6$.

Comparing the slope of a limited portion of the seismic spectra to the thermal variance of convection calculations is an indirect method to estimate the Rayleigh number of the convective regime of the Earth. It can be reasonable, provided that the thermal variance of convection is a strong function of the Rayleigh number, but a weak function of other potential complications (depth and temperature-dependent viscosity, Glazmaier's simulations are for constant viscosity, plates, etc.). We would also have to investigate that the large-scale seismic spectra reflected only temperature variations. If the comparison is reasonable and correct, then the above values would suggest that convection was layered, since a Rayleigh number of order $10^6$ would be suggested by whole mantle convection.

7 CONCLUSION

Gudmundsson et al. (1990) developed a method to image the spectrum of the Earth's heterogeneity as a function of depth from traveltimes and data and applied it to ISC P-wave data. We have applied the same method to the ISC S-wave data. The method yields estimates of incoherent noise in the data, which we find to be about 3% for shallow events and 2% for deep events (measured in terms of one standard deviation). The spatially coherent signal in the data is of the order of 3x. Thus, we estimate a signal-to-noise ratio of slightly larger than unity for the teleseismic, ISC, S-wave data set.

We observe that the Earth's seismic heterogeneity is concentrated in the upper 400 km of the mantle. This holds for the S-wave results presented in this paper and the P-wave results presented by GDC. This could be the result of a decrease in the absolute magnitude of the temperature derivatives of seismic velocity at high pressure or a dramatic deviation away from uniform convection, due to significant radial variations in material properties, e.g., viscosity, coefficient of thermal expansion or thermal diffusivity.

The pattern of the depth variation of the strength of heterogeneity is similar for shear and compressional (GDC) velocities. Assuming that the two velocities are spatially correlated, a signal averaged value of d ln Vp/d ln Vs [of at least 2 is required in this paper. The data have estimates of noise removed and are from the same period range. This value is appreciably larger than that expected from low-pressure laboratory measurements (0.8–1.4). It is unclear from this study whether this ratio holds for the lower mantle as well as the upper mantle. There is no strong evidence of periodicity in the traveltimes autocovariance. Hence, the underlying flow is unlikely to be periodic and the convection is unlikely to be steady. The short length scale at the surface suggests that the upper surface of the Earth is a compositional as well as a thermal boundary layer. The derived spectra have a similar dependence with harmonic degree as the spectra resulting from compressive convection in spherical geometry at Rayleigh numbers of $10^5$ and $10^6$ (Glazmaier 1988). As far as the simulation is a reasonable model of mantle convection, and the comparison is legitimate (i.e. seismic heterogeneity is the result of temperature between $10^5$ and $10^6$), then this result favours layered convection.

8 ACKNOWLEDGMENTS

This work was supported by NSF awards EAR83-17623 and EAR83-53371. HD wishes to acknowledge the support of an NERC post-doctoral fellowship, during the final stages of writing. Malcolm Sambridge and Bruce Buffett read an earlier draft of this paper, which led to significant improvements.

REFERENCES


Duffy, T. S. & Atrens, T. J., 1990. Sound velocities at high
of shear and compressional velocities at intermediate scale lengths.

6.6 Comparisons to convection models

What seismic variations might one expect from a convection model? Jarvis (1985) and Jarvis & Peltier (1986, 1989) considered this for the case of a steady, incompressible, constant viscosity, 2-D, unit aspect ratio convection. They found the boundary layers to have red spectra and that the spectra became progressively whiter towards the middle of cells. Also, the variations had the largest magnitude at the boundaries. At high Rayleigh numbers they found that the convection becomes more vigorous and the boundary layers thinner. Hence, the spectra become whiter, and stronger.

Towards the top of the mantle we have an increase in heterogeneity and a slight decrease in scale length. The increased heterogeneity is expected for boundary layers, but we would expect an increase rather than a decrease in scale length. Our inability to discern changes in characteristic length scales could be due to poor spectral resolution, poor vertical resolution or that the boundary layer is also a site of compositional variations. GDC present similar evidence for the core-mantle boundary. This evidence though, was questioned by Gudmundsson (1989), in a synthetic study that showed that due to poor coverage a similar feature can be reproduced in the inversion if the data are contaminated with realistic noise.

Glatzmaier (1988) presented one of the more complex simulations of mantle convection to date and evaluated spectra of the thermal variation. He considered compressible 3-D convection, with large Rayleigh numbers. He found large variations in behaviour as he increased the Rayleigh number. At $Ra = 10^7$ he found a network of narrow, cold downwellings and broad regions of upwellings, but at $Ra = 10^8$ he found instead hot plumes in a broad region of downwelling. His model produced maximum temperature deviations of about 200 K at the largest Rayleigh number. His computations were truncated at harmonic degree 45. At $Ra = 10^9$ he found that the thermal variance decreased by nearly an order of magnitude from its peak at degrees 3-10 out to degree 50. While at $Ra = 10^8$ he found that the harmonic thermal variance had only decreased by a factor of 2. This is qualitatively similar to the results of the much simpler 2-D convection models of Jarvis & Peltier (1986).

The lack of short-scale power in the lower mantle can be explained by the convective features being similar or smaller than the averaging ray width (100 km), and being insufficiently sampled, or that they do not exist and only large-scale features exist. The first two explanations agree with the very narrow features observed in high Rayleigh number, convective system with temperature-dependent viscosity. The third explanation is the opposite and implies very weak convection, i.e., a low Rayleigh number, possibly due to increasing viscosity with depth (Sammis et al. 1977) or decreasing coefficient of expansion (Anderson et al. 1987, Chelikani & Bohor 1989) with depth.

In Fig. 14 we present the logarithm of the spectra versus the logarithm of the harmonic degree. These curves are not linear, and therefore do not suggest a power law. If we forced the spectra to be fit by a power law with the exponent a function of scale, it would become more negative as the harmonic degree increased. In contrast, many surface fields are well described by power laws, e.g., heat flow varies like $l^{-0.6}$ (l is harmonic degree), which is much flatter than gravity, which varies like $l^{-2.5}$, or the toroidal or poloidal velocity of the plates, which vary as $l^{-2.5}$, or surface topography, which varies like $l^{-1.2}$ (Kaula 1980).

Glatzmaier (1988) found in his simulations that the logarithm of the thermal variance out to degree 50 is approximately a linear function of the harmonic degree, beyond $l > 10$. In Fig. 15 we present the logarithm of the normalized spectrum versus harmonic degree and find a similar linear behaviour. This spectrum has been defined as power per harmonic degree as in Glatzmaier (1988) rather than as conventional power per order [i.e., power per harmonic degree $(2l+1)$], and is a depth-weighted average through the mantle. Two curves are presented, for the different damping parameters. The equivalent curves for the model resulting from the data with logarithmic intercept estimates removed, and the lifted asymptote data values are nearly identical to the one with a similar damping parameter. The difference between the two curves suggests that we cannot rigorously estimate the slope, but both curves suggest that the thermal variance increases with harmonic degrees 10 and 50. If the similarity in form is real, it can potentially provide a constraint on mantle convection and viscosity. A comparison of these curves to Glatzmaier’s work would suggest a Rayleigh number of either $10^8$ or $10^9$.

Comparing the slope of a limited portion of the seismic spectra to the thermal variance of convection calculations is an indirect method to estimate the Rayleigh number and hence, the convective regime of the Earth. It was, however, reasonable, provided that the thermal variance of convection is a strong function of the Rayleigh number, but a weak function of other potential complications (depth and temperature-dependent viscosity, Glatzmaier’s simulations are for constant viscosity, plates, etc.). We would also have to note that the large-scale seismic spectra reflected only temperature variations. If the comparison is reasonable and correct, then the above values would suggest that convection was layered, since a Rayleigh number of order $10^8$ would be suggested by whole mantle convection.

7 CONCLUSION

Gudmundsson et al. (1990) developed a method to image the spectrum of the Earth’s heterogeneity as a function of depth from traveltime data and applied it to ISC P-wave data. We have applied the same method to the ISC S-wave data. The method yields estimates of inherent noise in the data, which we find to be about 3s for shallow events and 5s for deep events (measured in terms of one standard deviation). The spatially coherent signal in the data is of the order of 3s. Thus, we estimate a signal-to-noise ratio of slightly larger than unity for the teleseismic, ISC S-wave data set.

We discover that the Earth’s seismic heterogeneity is concentrated in the upper 400 km of the mantle. This holds for the S-waves results presented in this paper and the P-wave results presented by GDC. This could be the result of a decrease in the absolute magnitude of the temperature derivatives of seismic velocity at high pressure or a dramatic deviation away from uniform convection, due to significant radial variations in material properties, e.g., viscosity, coefficient of thermal expansion or thermal diffusivity.

The pattern of the depth variation of the strength of heterogeneity is similar for shear and compressional (GDC) velocities. Assuming that the two velocities are spatially correlated, a signal averaged value of $log(V_c/V_s)$ or $log(V_p/V_s)$ of at least 2 is required. The data have estimates of noise removed and are from the same period range. This value is appreciably larger than that expected from low-pressure laboratory measurements (0.5-1.4). It is unclear from this study whether this ratio holds for the lower mantle as well as the upper mantle. There is no strong evidence of periodicity in the traveltime autocovariance. Hence, the underlying flow is unlikely to be periodic and the convection is unlikely to be steady. The short length scale at the surface suggests that the upper surface of the Earth is a compositional as well as a thermal boundary layer. The derived spectra have a similar dependence with harmonic degree as the spectra resulting from convection in spherical geometry at Rayleigh numbers of $10^6$ and $10^7$ (Glatzmaier 1988). As far as the simulation is a reasonable model of mantle convection, and the comparison is legitimate (i.e., seismic heterogeneity is the result of temperature between $l = 10$ and 50), then this result favours layered convection.

8 ACKNOWLEDGMENTS

This work was supported by NSF awards EAR83-17623 and EAR85-31371. HD wishes to acknowledge the support of an NERC post-doctoral fellowship, during the final stages of writing. Malcolm Sambridge and Bruce Buffett read an earlier draft of this paper, which led to significant improvements. Comments from the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA.

REFERENCES


Duffy, T. S. & Atrens, T. J. 1990. Sound velocities at high
Early Devonian (pre-Acadian) magnetization directions in Lower Old Red Sandstone of south Wales (UK)

J. E. T. Channell, 1 C. McCabe 2 and N. H. Woodcock 3

1 Department of Geology, University of Florida, Gainesville, FL 32611, USA
2 Department of Geology and Geophysics, Louisiana State University, Baton Rouge, LA 70803, USA
3 Department of Earth Sciences, University of Cambridge, Cambridge, CB2 3EQ, UK

Accepted 1991 September 30. Received 1991 September 30; in original form 1991 May 8

SUMMARY

Two components of magnetization have been resolved from the Lower Old Red Sandstone of south Wales. All 39 sampled sites lie north of the Variscan front, and all but three are largely unaffected by Variscan folding. Thirteen sites yield a high unblocking temperature magnetization component which pre-dates Acadian (mid-Devonian) folding. The site mean direction for this component (Dec.: 232°, Inc.: 31.9°, α95: 8.5°) yields a pole position (Lat.: 7.7°S, Long.: 306.7°E) which we assign as an Early Devonian age. A lower unblocking temperature component is present at 38 of the sampled sites. This component post-dates both the Acadian and Variscan folding, has reversed polarity and a mean direction (Dec.: 193.8°, Inc.: 7.2°, α95: 2.9°) which corresponds to a pole position (Lat.: 40.4°S, Long.: 338.3°E) consistent with a Late Carboniferous or Early Permian age.

The Early Devonian pole position confirms the existence of a Siluro-Devonian inflexion in the apparent polar wander path (APWP) for southern Britain which coincides with an inflexion in the APWP for Scotland (Britain north of the Iapetus suture), indicating the closure of the Iapetus Ocean by Siluro-Devonian time. The Early Devonian palaeolatitude (175° S)° is consistent with an Early Devonian configuration in which Gondwana, Laurentia, Baltica and Avalonia were part of a single supercontinent.

Key words: Early Devonian, Old Red Sandstone, palaeomagnetism, Wales.

1 INTRODUCTION

The Old Red Sandstone of south Wales was the subject of a classic analysis of multicomponent magnetization in the early days of a palaeomagnetic revolution (Creech 1962; Chalmers & Creer 1964; Chamaulau 1964). The high unblocking temperature mean direction (Dec.: 66°, Inc.: -37°) given by Chalmers & Creer (1964) is quoted in all Palaeozoic apparent polar wander paths (APWP) for southern Britain south of the Iapetus suture (e.g., Briden & Duff 1981; Briden, Turnbull & Watts 1984; van der Voo 1986; Kröp, White & van der Voo 1990; Trench & Toorik 1991). This pole position is particularly important to the APWP (and hence to the closure history of the Iapetus Ocean) due to the paucity of Siluro-Devonian palaeomagnetic data from southern Britain and to the fact that the pole defines an inflexion in the path. In the Chamaulau & Creer (1964) study, the sites are located mainly (but not entirely) within the Variscan belt, and although the presence of a pre-folding magnetization component was recognized, the component directions are not well resolved and the mean direction is poorly defined. Subsequent palaeomagnetic studies of the Old Red Sandstone from south Wales (McClelland Brown 1983; Stearns & Van der Voo 1987), using modern methods for resolving magnetization components, have been performed (Stearns & Van der Voo 1987) and synfold (McClelland Brown 1983) magnetization components. The mean direction of the pre-folding component (Dec.: 194°, Inc.: -35°) gives a pole position (Lat.: 39°S, Long.: 334°E) which lies on the Late Carboniferous part of the APWP, and is therefore considered to be a local fault (Stearns & Van der Voo 1987). The synfaulting magnetization directions must also be Carboniferous in age, but have more westerly declinations of about 40°.