Upper and Lower Mantle Shear Velocity Modelling by Monte Carlo Inversion

M. H. Worthington,* J. R. Cleary and R. S. Anderssen

(Received 1973 June 4)†

Summary

Non-uniqueness bounds for shear velocity in the upper and lower mantle have been derived by Monte Carlo inversion, with constraints imposed by the data of Nuttli, Robinson & Kovach, and Hales & Roberts. The depth to the lower major discontinuity in the upper mantle is estimated to be 690 ± 25 km. Depths calculated for the corresponding discontinuity for compressional velocity are shallower than this, mainly as a result of differences in the interpretation of later arrivals in the travel-time data. Models of Hales & Roberts and Robinson & Kovach fall within our bounds for the lower mantle but the Fairborn mean model is excluded for depths greater than 2200 km.

Introduction

The overall purpose of the work described here is the determination of bounds for shear velocity in the mantle. Such bounds may be used, in conjunction with free oscillation data, to obtain gross Earth density models. No reliable upper mantle shear velocity data are available for oceanic regions, and the problem of deriving reliable gross Earth density models cannot be completely resolved until these are obtained (Dziewonski 1970; Worthington, Cleary & Anderssen 1972). On the other hand, upper mantle shear velocity profiles for the continental U.S. have been derived from S wave travel times by Ibrahim & Nuttli (1967) and Nuttli (1969) and from slowness (dT/dΔ) data by Robinson & Kovach (1972). The data from these authors are used as constraints in the inversion procedure now to be described.

Methodology

Johnson & Gilbert (1972) have approached the problem of non-uniqueness in the inversion of seismic body wave data using the methods of Backus & Gilbert (1967, 1968, 1970). Müller & Alsop (1972) and Wiggins, McMecan & Toksöz (1973) have pointed out the limitations of such methods, which are based upon linearization of the equations relating travel times and travel-time gradients to velocity within the Earth. They recommend that when travel-time data are inverted, methods should be used that take full account of the non-linearity in the inversion equations.

* Present address: Department of Geology, University of Oxford
† Received in original form 1973 March 2
FIG. 1. The model published in Nuttli (1969), and REDDOG-2 (continental-Tectonic) of Robinson & Kovach (1972), are shown. \(dT/d\Delta\) data is from Robinson & Kovach (1972). Travel-time data is from Nuttli (1969). Travel-time branches are labelled ABCDE as in Nuttli (1969). Also shown is the confidence limit for travel times and the permissible range of distance and time for the position of the A cusp.
We have used the Monte Carlo Inversion (MCI) technique to obtain a set of shear velocity models for the upper and lower mantle, based upon travel-time and $dT/d\Delta$ data. We do not invoke any linear approximation when calculating travel-times and travel-time gradients for the random models. However, as we have previously pointed out (Anderssen, Worthington & Cleary 1972), the success of the method depends largely upon the manner in which random models are generated between chosen a priori bounds for shear velocity. It is necessary in the interests of computational efficiency to restrict the production of physically unrealistic models without introducing a bias in the search over parameter space. The two-stage algorithm that we use to construct the random models is basically that described in Anderssen et al. (1972), and the justifications for its use in that paper are equally applicable to this study.

Travel-time data

$S$ wave travel-time data for distances within the range $0^\circ$ to $100^\circ$ have been tabulated by Ibrahim & Nuttli (1967), Nuttli (1969), Doyle & Hales (1967), Hales & Roberts (1970a), Fairborn (1969) and Randall (1970). Only the data of Ibrahim & Nuttli and Nuttli adequately cover the upper mantle and transition zone. We choose to constrain our upper mantle and transition zone models with the Nuttli (1969) data since these are based on times from nuclear explosions and consequently no errors of origin time and location are involved. The existence of later arrival branches extending to a distance of $40^\circ$, which we will refer to as the CD and DE branches (see Fig. 1), is a powerful constraint in our inversion procedure, and we note that Ibrahim & Nuttli (1967) and Nuttli (1969) clearly observed these branches.

For distances greater than $40^\circ$, we use the data of Hales & Roberts (1970a). These authors used the times of Nuttli (1969) to define the baseline for their data. Consequently, times calculated from their equation (2), which is a quadratic fitted to their observations, are also appropriate for our study. The bulletin data used by Randall (1970) give results similar to those of Hales & Roberts (apart from a baseline shift) but have the disadvantage that they extend only to $80^\circ$.

$S$ travel-time residuals abstracted from Tonto Forest Seismic Observatory (TFSO) seismic bulletins by Robinson & Kovach (1972) show an abrupt jump of about 7 s near $40^\circ$. In order to account for this, they have postulated significant lateral inhomogeneity between oceanic and continental regions down to depths of at least 1000 km. On the other hand, no such break has been observed in any of the $S$ travel-time studies listed above. In particular, no break occurs in the LASA $S$ travel-time data of Fairborn although there is a change from continental to oceanic paths at about $50^\circ$. We suggest that the TFSO data may be due to large negative $S$ source anomalies at distances beyond about $40^\circ$ for events occurring in the subduction zones of South America and the western Pacific margin (cf. the large negative $P$ anomalies found for these regions by Cleary & Hales 1966). We therefore take no further note of this result of Robinson & Kovach, although it obviously warrants further investigation.

d $T$/d$\Delta$ Data

d $T$/d$\Delta$ data for shear waves are provided by Fairborn (1969) and Robinson & Kovach (1972). Fig. 2 shows the mean straight line fitted to the Robinson & Kovach data for distances greater than $34^\circ$, together with the $2\sigma$ acceptability bounds for this line (defined by (7) of the Appendix) obtained by least squares. (We shall refer to these acceptability bounds as $2\sigma$ bounds). This line is almost identical to the mean line of Hales & Roberts (1970a) which was fitted to the first derivative of their travel-time data. The Fairborn data are low compared with these other two sets of data for
FIG. 2. $dT/d\Delta$ data of Robinson & Kovach (1972) are compared with the data of Fairborn (1969) and the equations (2) of Hales & Roberts (1970a). The solid line is the least squares mean of the Robinson & Kovach data for distances from 34° to 95°. The broken lines are the 2$\sigma$ bounds (i.e., 95 per cent confidence interval estimates) obtained by the method of least squares.

distances greater than 60 degrees, although most points fall within the 2$\sigma$ bounds. Hales & Roberts have noted a reduced slope in their $dT/d\Delta$ data at distances of 25° to 35°, and this is also observable in the Robinson & Kovach data and to a lesser extent in the Fairborn data.

The data of Hales & Roberts and Robinson & Kovach together represent a wide lateral region of the lower mantle. Because of the excellent agreement between the means of these data for distances greater than 34°, we conclude that there is a systematic error in the data of Fairborn (1969), and we base our lower mantle $dT/d\Delta$ constraints on the mean of the Robinson & Kovach data.

$dT/d\Delta$ data for the upper mantle are provided only by Robinson & Kovach (1972). In order to constrain our upper mantle models with these data in conjunction with the travel-time data of Nuttli, we assume that there is no significant lateral inhomogeneity in structure beneath the Basin and Range province and the Central U.S. below a depth of approximately 300 km. This assumption is supported by the findings of Wiggins & Helmberger (1973), Green & Hales (1968), and Massé, Landisman & Jenkins (1972). Fig. 1 shows a comparison of the models and data of Robinson & Kovach (1972) and Nuttli (1969). We have derived a suite of models that satisfy both sets of data within specified confidence limits. Because the $dT/d\Delta$ data are sparse, we allow for the possibility that some arrivals may have gone undetected, in a manner described in the next section.

Experimental procedure

Having determined the travel-time and $dT/d\Delta$ values for a random model lying between initial $a$ priori shear velocity bounds using the theory of Bullen (1961), we
Monte Carlo inversion

Table 1

<table>
<thead>
<tr>
<th>Depth (km)</th>
<th>Lower bounds km s⁻¹</th>
<th>Upper bounds km s⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>4.60</td>
<td>4.75</td>
</tr>
<tr>
<td>350</td>
<td>4.67</td>
<td>4.88</td>
</tr>
<tr>
<td>400</td>
<td>4.88</td>
<td>5.13</td>
</tr>
<tr>
<td>450</td>
<td>5.10</td>
<td>5.35</td>
</tr>
<tr>
<td>500</td>
<td>5.23</td>
<td>5.42</td>
</tr>
<tr>
<td>550</td>
<td>5.26</td>
<td>5.46</td>
</tr>
<tr>
<td>600</td>
<td>5.30</td>
<td>5.50</td>
</tr>
<tr>
<td>650</td>
<td>5.30</td>
<td>5.50</td>
</tr>
<tr>
<td>700</td>
<td>5.32</td>
<td>5.50</td>
</tr>
<tr>
<td>750</td>
<td>6.00</td>
<td>6.25</td>
</tr>
<tr>
<td>800</td>
<td>6.07</td>
<td>6.30</td>
</tr>
<tr>
<td>900</td>
<td>6.17</td>
<td>6.35</td>
</tr>
<tr>
<td>1000</td>
<td>6.30</td>
<td>6.49</td>
</tr>
<tr>
<td>1100</td>
<td>6.35</td>
<td>6.54</td>
</tr>
<tr>
<td>1200</td>
<td>6.42</td>
<td>6.61</td>
</tr>
<tr>
<td>1300</td>
<td>6.45</td>
<td>6.68</td>
</tr>
<tr>
<td>1400</td>
<td>6.54</td>
<td>6.72</td>
</tr>
<tr>
<td>1500</td>
<td>6.65</td>
<td>6.82</td>
</tr>
<tr>
<td>1600</td>
<td>6.78</td>
<td>6.93</td>
</tr>
<tr>
<td>1800</td>
<td>6.87</td>
<td>7.00</td>
</tr>
<tr>
<td>2000</td>
<td>6.99</td>
<td>7.10</td>
</tr>
<tr>
<td>2200</td>
<td>7.05</td>
<td>7.16</td>
</tr>
<tr>
<td>2400</td>
<td>7.08</td>
<td>7.24</td>
</tr>
</tbody>
</table>

require acceptability regions about the given travel time and $dT/d\Delta$ data before we can test whether this model is acceptable. The choice of acceptability regions is particularly difficult for the $dT/d\Delta$ data for distances less than 34°, where the data are few and triplication of the slowness curve is expected. Since a long $dT/d\Delta$ branch with very small slope represents arrivals of very low amplitude, the presence or absence of such branches may be difficult to determine. We avoid this problem by specifying upper and lower limits for $dT/d\Delta$ at the distances of each observed data point. Any acceptable $dT/d\Delta$ curve is then required only to pass between these bounds at the specified distances at least once in either the prograde or retrograde direction. At distances greater than 34°, control points are taken at 2-deg intervals along the mean line shown in Fig. 2, and the upper and lower bounds for $dT/d\Delta$ are the 2σ limits calculated for this mean line using equation (7) of the Appendix. For distances less than 34° the curves are controlled by the Robinson & Kovach data points. Because of the sparsity of the data in this range no function on which to base least squares modelling is known. Therefore, we estimate the bounds using equation (6) of Appendix A. A full rationale for the use of these bounds is given in the Appendix.

The random model-generating algorithm described in Anderssen et al. (1972) was slightly modified to suit the present problem. A 35-km crust with a velocity of 3.5 km s⁻¹ was assumed. A low velocity zone (LVZ) was required beneath the crust. The 'lid' could extend from the 'moho' to a depth of 175 km and have a velocity of between 4.45 and 4.65 km s⁻¹. The shape of the LVZ was controlled only by the requirement that the first rays from below the LVZ should have a cusp that falls within the time and distance bounds shown in Fig. 1. With the exception of the velocity within the LVZ, which was not permitted to be less than 4.0 km s⁻¹, the values of our final solutions were constrained by our initial a priori bounds.

We applied the constraint that only one region of negative gradient could occur below the LVZ and above 800 km. However, we greatly limited the possible extent of negative gradients by requiring that data points should not fall within a shadow
FIG. 3. (a) Our upper mantle models that satisfy the initial conditions described in the text. (b) A set of models that have been selected for their fit to all the later arrival travel-time data.
Monte Carlo inversion

zone. Anderson & Julian (1969) proposed that arrivals on the CD and DE branches (see Fig. 1) observed by Ibrahim & Nuttli (1967) could result from diffracted energy. We regard this interpretation as improbable, because amplitudes for arrivals on these branches observed by Nuttli (1969) do not show any systematic decrease with distance and are consistently greater than the amplitudes of the corresponding first arrivals. Two of our models shown in Fig. 3(a) have narrow low velocity zones between 650 and 700 km. These give rise to a shadow zone between 40° and 45° on the CD branch and there is no reason to eliminate them from our solutions.

Other conditions imposed as criteria for an acceptable solution were:

1. $dT/dA$ constraints as described above were required to be satisfied at all points.

2. $2\sigma$ limits (derived using (7) of the Appendix) were obtained for the Nuttli first arrival travel-time data by fitting three straight lines over the distance ranges 0° to 20°, 20° to 26°, and 26° to 40°. At distances less than 40° theoretical times were not permitted to deviate from these lines by more than $2\sigma (= 6 \text{ s})$.

3. A second arrival (CD) branch was required that fell within the $2\sigma$ limits at 39° as shown in Fig. 1. However, we discarded models where this branch extended to distances greater than 50°.

4. Models with an AB branch extending to distances greater than 35° were discarded.

The grid spacing in the upper mantle was 25 km. The spacing in the lower mantle is shown in Fig. 6.

Results and discussion

Fig. 3(a) shows the set of 22 derived models obtained after approximately 1,000,000 trials, which satisfy the conditions described in the previous section. The solutions are unconstrained by travel times of later arrivals other than those on the CD and DE branches, and this is reflected in the wide spread of models at depths less than 400 km. No model has a negative gradient immediately above the velocity discontinuity near 400 km because the $dT/dA$ points between 21 and 23 s deg$^{-1}$ will not be satisfied by a diffracted branch.

The later arrival travel-time data of Nuttli (1969) to which the AB and BC branches of his model are fitted take no part in constraining the models in Fig. 3(a). Consequently, the travel-time curves for some of these models bear no relation at all to these data. It is difficult to quantify what might be considered a satisfactory fit to these data. However, their distribution does at least define an approximate trend that a triplication, that results from a discontinuity near 400 km, should follow. The nine models shown in Fig. 3(b) are chosen from the models of Fig. 3(a) because they are in reasonable agreement with this trend, whereas none of the discarded models conformed to the later arrival data.

Bounds defined by the models of Fig. 3(b) are shown in Fig. 4 along with the Nuttli (1969) model, REDDOG2 (Tectonic-Continental) of Robinson & Kovach (1972) and US26 (Anderson & Julian 1969). As a result of the smoothing effect of the $dT/dA$ constraint on the Nuttli profile, the bounds provided by this study show the velocity discontinuity in the depth range 300–500 km to change gradually over approximately 150 km. This appears to be reasonable because Massé et al. (1972) found two discontinuities in velocity between depths of 300 and 450 km beneath the Basin and Range province and the Central U.S., and Simpson, Mereu & King (1973) strongly favour a broad and probably complex region of increasing velocity in this depth
Fig. 4. The extremal bounds for $P$ velocity of McMechan & Wiggins (1972) are shown above our bounds for $S$ velocity (solid lines). Also shown are the models HWNE of Helmberger and Wiggins (1971), the model of Johnson (1967), US26 of Anderson & Julian (1969), the model published in Nuttli (1969), and REDDOG-2 (Continental-Tectonic) of Robinson & Kovach (1972).

range, rather than one obvious discontinuity. This is in agreement with geochemical evidence (Ringwood 1970) that over a depth range of approximately 380–420 km magnesium-rich olivine would first form a solid solution of olivine and spinel components before transforming to the beta spinel phase and that these two phase transformations would be preceded at a lower pressure by the pyroxene-garnet transformation.

The shear velocity bounds in Fig. 4 show the depth to the deeper discontinuity in the transition zone to be $690 \pm 25$ km. On the other hand, an equivalent estimate of this depth for compressional velocity obtained by McMechan & Wiggins (1972) is $650 \pm 30$ km. Although there is some overlap of these bounds, the discrepancy between the means of these two estimates is due mainly to the presence of a CD branch extending to a distance of $40^\circ$ in the $S$ data, and the absence of such a branch in the $P$ data. We believe that our Monte Carlo procedure has exhausted the possibility that the discontinuity for $S$ might be brought to a shallower depth by varying the structure above it.

It is unlikely that lateral inhomogeneity can be an explanation for the incompatibility of the $P$ and $S$ wave discontinuities. All the models and bounds shown in Fig. 4 are based on data representing the mantle below the Basin and Range province or the Central U.S. and references have already been cited concerning the apparent similarity of these regions. An obvious possibility is that either the inversion in this study or the McMechan and Wiggins $P$ extremal inversion has been over-constrained by the CD and DE branch travel-time points. Unless there is good
reason to reject these data, the lower discontinuity for $S$ cannot be made any shallower. On the other hand, Simpson et al. (1973) have presented a $P$ model with a depth of 680–695 km for the lower discontinuity, based on data from the Warramunga array in Australia, which agree well with the Tonto Forest data of Johnson (1969) in the region of his CD and DE branches. The $dT/d\Delta$ bounds of McMechan & Wiggins (1972, following Johnson's interpretation) constrain the Johnson slowness observations of about 10 s deg$^{-1}$ from 24° to 28° to lie on the retrograde branch DE. This interpretation is different to that given by Helmberger & Wiggins (1971), who chose to confine these Johnson data to the CD branch, and is in conflict with Simpson et al.'s interpretation of their own data. Thus, the McMechan & Wiggins bounds for $P$ may be overconstrained in this region.

![Travel-time and $dT/d\Delta$ curves of our lower mantle models. Also shown are the confidence limits of time and $dT/d\Delta$.](https://academic.oup.com/gji/article-abstract/36/1/91/708202)
It should be noted that there is an order of magnitude difference in wavelength between the compressional and shear wave data referenced above. The possibility that simple ray theory is not wholly applicable for waves of approximately 100-km wavelength in the upper mantle warrants further investigation.

For the lower mantle, our bounds on first arrival travel times are defined by:

$$\pm 2\sigma_s = \pm 2(\gamma^2 + \sigma_0^2)^{\frac{1}{2}}$$

where $\sigma_0$ is the root mean square deviation of the Hales & Roberts data and $\gamma$ is the uncertainty of the baseline for this data. We require the mean line to the Nuttli data to coincide with the quadratic of equation (2) of Hales & Roberts (1970a). There is an uncertainty of approximately 1 s in the positioning of the quadratic function relative to the Nuttli mean line at distances between 30° and 40°. This results in a value of 1.45 s for $\sigma_s$. The travel-time and $dT/d\Delta$ curves of our acceptable models are shown in Fig. 5. The existence of triplications is limited only by the choice of grid spacing and the requirement that velocity should increase monotonically with depth.

Hales & Roberts (1970b) have pointed out the extent to which the Fairborn models fail to satisfy either their or his own travel-time data. The bounds that we have obtained, shown in Fig. 6, serve to demonstrate the degree of non-uniqueness associated with the Hales & Roberts (SLUTD-I) model (subject to the conditions stated above) and confirm that the Fairborn average model is not compatible with the Hales & Roberts data for distances greater than 60°. However, there is no conflict between the data and models of Robinson & Kovach and Hales & Roberts.

We intend in a future paper to return to the problem of obtaining non-uniqueness bounds for density within the Earth, and will use the constraints derived here in our inversion procedure for the determination of those bounds.

Acknowledgments

We wish to thank David Simpson for the use of his travel-time programme and plotting routine and for many helpful discussions.
Monte Carlo inversion

M. H. Worthington and J. R. Cleary:
Department of Geophysics and Geochemistry
Australian National University
Canberra

R. S. Anderssen:
Computer Centre
Australian National University
Canberra

References


### Appendix

**Acceptability regions**

Having determined, for a random model lying between given *a priori* bounds, its corresponding travel-time and $dT/d\Delta$ values, we require acceptability regions for the given travel-time and $dT/d\Delta$ data before we can test the acceptability of this random model. We base our method of solution on the following statistical problem which summarizes the situation:

Consider the data $y_i = \{y_i; i = 1, 2, ..., M+N\}$ defined on the grid

$$G = \{x_i; i = 1, 2, ..., M+N\}$$

where

(i) the observational errors $\{e_i\}$ in $\{y_i\}$ are independent and randomly distributed with mean zero and constant variance $\sigma^2$ (homoscedastic); or more briefly, the observational errors $\{e_i\}$ are $N(0; \sigma^2)$;

(ii) the data $\{y_i\}$ can be partitioned into the data sets $\{g_i\}$ and $\{f_i\}$ such that

$$\{g_i\} = \{g_i = y_i; i = 1, 2, ..., M\},$$

$$\{f_i\} = \{f_i = y_{M+i}; i = 1, 2, ..., N\};$$

(iii) a linear model $f(x)$ for the $f_i$ is known, viz.

$$f(x) = \sum_{k=1}^{K} a_k \phi_k(x)$$

with the $\phi_k(x)(k = 1, 2, ..., K)$ linearly independent; and

(iv) a model for the $\{g_i\}$ is unknown.

Letting $\alpha = [\alpha_1, \alpha_2, ..., \alpha_K]^T$ and $f = [f, f, ..., f_N]^T$, we can estimate the $\alpha_k$ in (1) using least squares to obtain the estimates

$$\alpha = (X^TX)^{-1}X^T f, \quad X_{k,f} = \phi_k(x_f),$$

where $X = [X_{1,f}, X_{2,f}, ..., X_{K,f}]^T$. 

---

Downloaded from https://academic.oup.com/gji/article-abstract/36/1/708/708202 by guest on 13 April 2019
and thus, the least squares solution

\[ \hat{f}(x) = \sum_{k=1}^{K} \delta_k \alpha_k(x); \]  

(3)

and thereby, can construct as the acceptability limits for the data \( \{f_i\} \) the corresponding \((1 - \beta)\) 100 per cent confidence intervals

\[ \hat{f}_i \pm t_{N-K}(\beta) \{\delta(X^T X)^{-1} X^T\}, \]  

(4)

where \([A]_{i,i}\) denotes the \(i\)-th diagonal element of the square matrix \(A\),

\[ \hat{f} = \frac{1}{N} f^T (I - X(X^T X)^{-1} X^T) \hat{f}, \]  

(5)

and \(t_{N-K}(\beta)\) is the two-tailed \(\beta\)-point of the \(t\)-distribution with \(N-K\) degrees of freedom. In order to obtain accurate results, the evaluation of these estimates must be based on either orthogonal factorization or modified Gram–Schmidt techniques (see, for example, Anderssen 1969; Wilkinson & Reinsch 1971).

This approach is not possible for the data \(\{g_i\}\), since a model for it is not known. However, the value of \(\hat{\sigma}^2\) determined for the data \(\{f_i\}\), viz. (5), can be regarded as an estimate for \(\sigma^2\). Since we assume that the errors \(\{e_i\}\) have mean zero, it follows that the data \(\{y_i\}\) is distributed normally with mean (expectation) \(E(y_i)\), and variance \(\sigma^2\). We can therefore regard the \(\{g_i\}\) as single estimates of the mean \(E(g_i)\), and thereby, can construct the acceptability limits for each \(g_i\) of the data \(\{g_i\}\) as the \((1 - \beta)\) 100 per cent confidence intervals for the distribution \(N(E(g_i); \sigma^2)\) viz.

\[ g_i \pm p(\beta) \hat{\sigma} \]  

(6)

with \(p(\beta)\) defined for \(N(E(g_i); \sigma^2)\) by the condition:

\[
\text{Probability } [g_i - p(\beta) \hat{\sigma} \leq E(g_i) \leq g_i + p(\beta) \hat{\sigma}] = 1 - \beta.
\]

For example, when \(\beta = 0.05\), \(p(\beta) = 1.96 \sim 2.0\) (see Kendal & Stuart (1967), Chapter 20). It should be noted at this stage that (6) is valid only when the sample size \(N\) is so large that the estimate \(\hat{\sigma}^2\) may for practical purposes be taken as coincident with \(\sigma^2\).

We can also use these acceptability limits for the \(\{f_i\}\) except that in this case we have a better estimate at \(x_i\) of \(E(x_i)\) than \(f_i; \) viz. \(f_i\). Thus the corresponding acceptability limits for the data \(\{f_i\}\) become

\[ f_i \pm p(\beta) \hat{\sigma} \]  

(7)

References

