An Informative Solution to a Seismological Inverse Problem

(Earth's interior/structure/mass/moment of inertia)

FREEMAN GILBERT*, ADAM DZIEWONSKI†, AND JAMES BRUNE*

* Institute of Geophysics & Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, Calif. 92037, and † Department of Geological Sciences, Harvard University, Cambridge, Massachusetts 02138

Communicated by Walter Munk, February 22, 1973

ABSTRACT Preliminary results are presented that infer that 2 sec should be added to the tabular values for \( P \) phases and 4 sec to the tabular values for \( S \) phases of seismic travel times. From seismic evidence, the radius of the inner core of the Earth is 1229–1250 km; the radius of the outer core is 3482–3485 km. Data are presented relating resolving power with error of measurement for the Earth's mantle.

To infer the mechanical structure of the interior of the Earth from seismological data has been an important scientific activity for most of this century. Recently, a large amount of new data has become available and a greatly augmented dataset now exists. While the augmentation of the seismological dataset continues, we present in this report some preliminary results on a study of the existing dataset.

We have used 497 gross Earth data in this study. A gross Earth datum is a measured property of the whole Earth. Examples of gross Earth data are the Earth's mass, the frequency of one of the Earth's elastic-gravitational normal modes, the travel-time of a pulse from a particular source to a particular receiver, the rate of diurnal variation, the phase lag of one of the bodily tides, the surface impedance of the magnetic daily variation, etc. Our dataset is restricted to include only seismological data in addition to the Earth's mass and moment of inertia. We have 368 normal mode frequencies, simply called modes, and 127 travel-times, simply called rays. From the work of Dziewonski and Gilbert (ref. 1 and manuscript in preparation), we have 44 \( T_1 \) modes \((l \leq 46)\), 50 \( S_1 \) modes \((l \leq 50)\), 9 \( T_2 \) modes, and 111 \( S_2 \) modes \((n \leq 23, l \leq 20)\). From the work of Brune (2) and Brune and Gilbert (manuscript in preparation), we have 154 \( T_1 \) modes \((6 \leq n \leq 18, 8 \leq l \leq 275)\) that have been derived from the phase spectra of multiply reflected SH pulses. For ray data we have used 16 PKIKP, \( \theta = 50^\circ(2^\circ)180^\circ \) (3, 4) 22 SKS, \( \theta = 84^\circ(2^\circ)126^\circ \) and 22 SKKS, \( \theta = 84^\circ(2^\circ)126^\circ \) (5a) and 34 P, \( \theta = 27^\circ(2^\circ)93^\circ \) (6). All of the ray data have been converted to the form advocated by Johnson and Gilbert (7).

The novelty of our study arises from our use of the large number of spheroidal and toroidal overtones. Many of the spheroidal overtones are high-\( Q \) overtones, discovered by Dratler, Farrell, Block, and Gilbert (8). Also included are those used to infer that the Earth's inner core is solid (1a). The toroidal overtones contain information about the shear speed in the middle mantle, and the location of the core-mantle boundary.

Since the modal data are averaged values, from many stations and, in some cases, many sources, they are interpreted by invoking the diagonal sum rule (DSR) (9). According to this rule the mean value of all eigenfrequencies belonging to a multiplet is the multiplet's degenerate eigenfrequency belonging to the spherically averaged Earth. This means that, correct to first-order in a perturbation expansion, the modal data are not contaminated by lateral heterogeneities in the Earth. Although a similar sum rule, based on Fermat's principle, applies to the ray data (see, for example, ref. 10) it has not been used by seismologists. Moreover, the ray data are biased, much more so than the modal data, by the restriction that almost all stations are on continental platforms. The bias in the ray data is often called the "baseline" problem. The shape of a travel time graph may be well determined, but not its reference level or baseline. Fortunately, we can use modal data to establish baselines.

Our first task is to develop credible models of the Earth that agree with the data. While it is not clear whether we can succeed in this task, we are encouraged because we already have models that are fairly good fits to the data. We seek to refine the models to be better fits. Backus and Gilbert (11a) have discussed this inverse problem in some detail. A rather general presentation has been given by Sabatier (ref. 12, pp. 9-8, 9-9). We want to add a perturbation to a model to obtain an improved model that is predicted to fit the data within, say, one standard error. In addition, the perturbation must be small with respect to some norm. The choice of norm is, to a large extent, arbitrary, and a successful choice depends on luck, insight, and experience. It is our experience that to demand that the perturbation be smooth provides an acceptable norm. We minimize \( N_1 \), where

\[
N_1 = \int [d2\dot{m}/dx^2]dx
\]

where \( \dot{m} \) is the perturbation.

Because of the redundancy of the data, we use only the statistically significant subset found by the ranking and winnowing procedure of Gilbert (9b). The composition of the subset is determined by decomposing the inner product matrix of Fréchet kernels, and the inner product depends on the chosen norm. Thus, the number of data in the subset, as well as their standard errors, depends on the norm. In the sequel we use the norm described above and we exclude from the statistically significant subset any ranked datum whose relative variance is greater than unity. In this way we can solve what appear to be very ill-conditioned problems.

We remark in passing the similarity between our procedure, singular value decompositions (13) and the construction of a generalized inverse operator (14). We also remark in passing
that the decomposition of a real symmetric matrix of order 500 takes about 4 min on a modern computer (CDC7600), so that continued augmentation of the seismological dataset is welcome.

In our first inversion we used 198 gross Earth data; mass and moment of inertia (hereafter called \( m_0 \) and \( m_1 \), 44 \( S_T \), 50 \( S_T \), 9 \( T_T \), and 93 of the 111 \( S_d \). To each datum was assigned the standard error (SEM) of Dziewonski and Gilbert (1), except that no relative standard error was allowed to be less than \( 5 \times 10^{-4} \). With this restriction, and using the norm producing the smoothest perturbation, there are 38 data in the statistically significant subset (relative standard error less than unity). Each datum in the subset is said to be a significant Earth datum (SED). Thus, we say that in our first inversion we used 198 gross Earth data (GED) containing 38 significant Earth data (SED).

It should be realized that if we have two datasets containing respectively, \( G_1 \) and \( G_2 \) GED and \( S_1 \) and \( S_2 \) SED, then the combined dataset containing \( G_1 + G_2 \) GED will usually contain less than \( S_1 + S_2 \) SED.

To give a coarse measure of the extent to which a model fits the data, we adopt a suggestion of Backus (15) and introduce the concept of the credibility of the model. Let \( \delta \gamma_i \) be the difference between the computed value and the mean of the measured values of the \( i \)th datum, and let \( \sigma_i \) be the standard error of the \( i \)th datum. Define chi-squared

\[
\gamma_i = \frac{\delta \gamma_i}{\sigma_i}, \quad \psi = N^{-1} \sum_{i=1}^{N} \gamma_i^2 = N^{-1} \chi^2
\]

\[ C = \exp(-\frac{1}{2} \psi) \]

We call \( C \) the credibility of the model. In effect, \( C \) is a measure of how well the model fits the data. It has been found convenient to use an extended definition of credibility. Let \( n(0) \) be a real number. Define

\[
l_i = \frac{\delta \gamma_i}{\sigma_i} - n, \quad u_i = \frac{\delta \gamma_i}{\sigma_i} + n
\]

\[
\gamma_i^{(n)} = 0 \text{ if } l_i \times u_i < 0; \quad \min (|l_i|, |u_i|)
\]

otherwise

\[
\psi_1 = N^{-1} \sum_{i=1}^{N} \gamma_i^{(n)}^2, \quad C_n = \exp(-\frac{1}{2} \psi_1). \quad [1]
\]

If the \( i \)th datum is fit within \( \psi \leq \psi_0 \), then \( C_n = 1 \). We call \( C_n \) the credibility of the model at the \( n \) - \( \sigma \) level.

In our first inversion we used model UTD124A' (1b) as the starting model. For the 198-member dataset \( C_n = 0.44, 0.71, 0.87 \) for \( n = 1, 2, 3 \), respectively. In addition, 13 of the 93 \( S_0 \) overtones have computed eigenfrequencies that agree with the observed values within \( 5 \times 10^{-4} \) (relative error). For these 13 modes we say \( |\delta \gamma| < 5 \times 10^{-4} \). After three interactive improvements to the model we have \( C_n = 0.81, 0.94, 0.99 \) for \( n = 1, 2, 3 \), respectively. Also, of the 93 \( S_0 \) overtones, 45 have \( |\delta \gamma| < 5 \times 10^{-4} \).

At each step of the iterative improvement we construct the smoothest perturbation to the model such that \( C_1 = 1 \) for the 38 SED in the statistically significant subset of the data. Because the inverse problem is nonlinear, we do not expect \( C_1 = 1 \) for the entire dataset. Furthermore, contaminations in the data, not removed by averaging over the source-receiver net, will cause \( C_1 > 1 \), as will misidentifications in the spectra. In fact, the closer \( C_1 \) is to unity the more we can believe in the compatibility of the dataset and its freedom from contamination. It is a matter of opinion whether \( C_1 = 0.81 \) is good enough. Certainly \( C_1 = 0.99 \) is rarely unacceptable.

We consider our first inversion to be successful, and turn our attention to the resolving power of the dataset (11b, 11d).

To construct averaging kernels, we use the Dirichlet criterion (11b, 9b). It has been chosen for three reasons. First, it allows us to consider several parameter distributions simultaneously. An integral such as

\[
q = \int_0^1 \alpha(x)A(x)dx + \int_0^1 \beta(x)B(x)dx
\]

can be rewritten

\[
q = \int_0^1 \alpha(x)A(x)dx + \int_1^2 \beta(x-1)B(x-1)dx
\]

where

\[
\nu(x) = \alpha(x) \quad 0 < x < 1, \quad \beta(x-1) \quad 1 < x < 2
\]

\[ H(x) = \alpha(x) \quad 0 < x < 1, \quad B(x-1) \quad 1 < x < 2. \]

The ordering of \( \nu \) and \( H \) is not important when we use the Dirichlet criterion. Second, the quantity of available data is large enough so that the problem of sidebands of the averaging kernel is not significant. Third, computing time is vastly reduced (9b) by use of the Dirichlet criterion.

For the linear problem of resolution we have the integral relations

\[
\gamma_i = \int_0^1 m(x)G_i(x)dt
\]

between the data \( \gamma_i \), the model \( m(x) \), and the data kernel \( G_i(x) \). For the nonlinear problem \( m \) is replaced by \( \delta m \), the difference between two models, \( \gamma_i \), is replaced by \( \delta \gamma_i \), and \( G_i(x) \) is the Fréchet kernel for the \( i \)th datum. We have arranged that

\[
\int_0^1 G_i(x)G_j(x)dx = \delta_{ij}.
\]

Table 1. Spread \( S \) as a function of radius \( r \) for \( \epsilon = 5 \times 10^{-3} \)

<table>
<thead>
<tr>
<th>( r )</th>
<th>( S_\rho )</th>
<th>( S_\alpha )</th>
<th>( S_\beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>680</td>
<td>70</td>
<td>1660</td>
</tr>
<tr>
<td>800</td>
<td>620</td>
<td>70</td>
<td>710</td>
</tr>
<tr>
<td>1200</td>
<td>410</td>
<td>380</td>
<td>810</td>
</tr>
<tr>
<td>1600</td>
<td>550</td>
<td>530</td>
<td>-</td>
</tr>
<tr>
<td>2000</td>
<td>510</td>
<td>400</td>
<td>-</td>
</tr>
<tr>
<td>2400</td>
<td>450</td>
<td>400</td>
<td>-</td>
</tr>
<tr>
<td>2800</td>
<td>380</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3200</td>
<td>390</td>
<td>150</td>
<td>-</td>
</tr>
<tr>
<td>3600</td>
<td>320</td>
<td>230</td>
<td>260</td>
</tr>
<tr>
<td>4000</td>
<td>320</td>
<td>280</td>
<td>260</td>
</tr>
<tr>
<td>4400</td>
<td>290</td>
<td>190</td>
<td>240</td>
</tr>
<tr>
<td>4800</td>
<td>300</td>
<td>90</td>
<td>200</td>
</tr>
<tr>
<td>5200</td>
<td>280</td>
<td>190</td>
<td>120</td>
</tr>
<tr>
<td>5600</td>
<td>240</td>
<td>80</td>
<td>150</td>
</tr>
<tr>
<td>6000</td>
<td>170</td>
<td>330</td>
<td>220</td>
</tr>
</tbody>
</table>

The subscripts \( \rho, \alpha, \) and \( \beta \) refer to spreads for density, \( P \)-wave speed, and \( S \)-wave speed, respectively.
Then the best “delta-like” Dirichlet kernel is
\[ D(x,x_0) = \sum_i G_i(x) G_i(x_0). \]
We want to make an averaging kernel
\[ A(x,x_0) = \sum_i G_i(x) A_i(x_0) \]
that minimizes \( \Delta \)
\[ \Delta = \int_0^1 [D(x,x_0) - A(x,x_0)]^2 dx \]
and \( \epsilon^2 \)
\[ \epsilon^2 = \sum_i G_i(x_0) \sigma_i^2 = \sum_i A_i(x_0) \nu_i. \]
We minimize a linear combination of \( \Delta \) and \( \epsilon^2 \). Let
\[ \int_0^1 D(x,x_0) dx = \sum_i G_i(x_0) G_i(x_0) = D(x_0,x_0) = d_D^{-1} \]
\[ \int_0^1 D(x,x_0) A(x,x_0) dx = \sum_i G_i(x_0) A_i(x_0) = d_A^{-1} \]
\[ \int_0^1 A(x,x_0) dx = \sum_i A_i(x_0) A_i(x_0) = A_A^{-1}. \]
If we think of \( D(x,x_0) \), which is symmetrical about \( x_0 \), as looking like a resonance curve, then \( d_D^{-1} \) is the height of \( D \) and \( d_A \) is the width of \( D \) at the “half-power” level. Alternatively, if \( D(x,x_0) \) has a “boxcar” shape, then, clearly, \( d_D^{-1} \) is the height and \( d_A \) is the width of \( D \). We have not proved that \( D \) is unimodular, but, as the \( G_i(x) \) become a “more complete” set of functions, \( D \) will approach unimodularity. We shall adopt \( d_A \) as a reasonable estimate of the width of \( D \). Similarly, \( A_A^{-1} \) is taken as the width of \( A \).

We choose the \( A_i(x_0) \) to minimize \( J \)
\[ J = \sum_i [G_i(x_0) - A_i(x_0)]^2 + \lambda \left( \sum_i A_i(x_0)^2 \nu_i - \epsilon^2 \right). \]
The result is
\[ A_i(x_0) = \frac{G_i(x_0)}{1 + \lambda \nu_i}. \]

The Lagrange multiplier, \( \lambda \), lies in the range \((0, \infty)\). Backus (15) has shown that \( \lambda \) can be given a Bayesian interpretation, and Johnson and Gilbert (7a) have used that interpretation in their investigation of the inverse travel-time problem. In this preliminary report we shall be content with the interpretation based on the concept of resolving power, or averaging length.

From Eqs. [2] and [3], we see that \( d_A^{-1} \geq 0 \). We write \( \Delta \) in the form
\[ \Delta = \sum_i [G_i(x_0) - A_i(x_0)]^2 = d_D^{-1} - 2d_A^{-1} + A_A^{-1} \geq 0 \]
Let \( S^{-1} = 2d_A^{-1} - A_A^{-1} \), then \( \Delta = d_D^{-1} - S^{-1} \).

We adopt \( S \) as a convenient estimate of the width of \( A \) and call it the spread or averaging length of \( A \). From the foregoing equations, \( \Delta \) is an increasing function of \( \lambda \) and \( \epsilon^2 \) is a decreasing function of \( \lambda \). Since \( d_A \) is fixed, \( S \) increases as \( \Delta \) increases. It is, therefore, easy to see that \( \epsilon^2 \) is a monotonically decreasing function of \( S \). The minimum value of \( S \) is \( d_D^{-1} \). The corresponding maximum value of \( \epsilon^2 \) is
\[ \epsilon^2_{\text{max}} = \sum_i G_i(x_0) \nu_i. \]

Backus and Gilbert (11d) have termed the relationship between \( \epsilon^2 \) and \( S \) a tradeoff curve. Rather than present tradeoff curves, we choose to fix \( \epsilon^2 \leq \epsilon^2_{\text{max}} \) for a given \( x_0 \) and then to calculate \( S \). We have
\[ \epsilon^2 = \sum_i x_i G_i^2(x_0) \frac{1}{(1 + \lambda \nu_i)^2}. \]
For a given value of \( \epsilon^2 \) we find \( \lambda \) in Eq. [4] by Newton’s method. Then we have \( A_i(x_0) \) in Eq. [3] and can easily calculate \( \Delta \) and \( S \).

The credible model obtained by inverting 198 gross Earth data is termed model C198.

For model C198 with \( \epsilon = 5 \times 10^{-4} \) (relative error) values of \( S \) for the P-wave speed \( \alpha \) are 600 km < \( S \) < 1200 km. Both the density \( \rho \) and the S-wave speed \( \beta \) are unconstrained. The travel time of PKIKP(F), the time it takes a P-wave to cross a diameter, is
\[ T = 2 \int_0^\infty a^{-1}(r) dr, \]
where \( r_0 \) is the radius of the Earth. If the uncertainty in \( \alpha \) is \( \alpha \) then the uncertainty \( \delta T \) in \( T \) is \( \delta T \). Since \( T \) is about 1200 sec, \( \delta T = 0.6 \) sec for \( \epsilon = 5 \times 10^{-4} \). This means that modal data alone constrain the travel time of PKIKP(F) to \( \pm 0.6 \) sec. If we decrease \( \epsilon \) until the largest \( S \) satisfies \( S = r_0 \), we find \( \epsilon = 2.5 \times 10^{-4} \). Therefore, modal data determine the travel time of PKIKP(F) to \( \pm 0.3 \) sec. For model C198 that travel time is 1213.1 sec.

We must emphasize here that 0.3 sec is a conservative estimate of the uncertainty. We have not allowed the relative standard error of any datum to be less than \( 5 \times 10^{-4} \). Of the 93 S1 overtakes, 39 have experimental standard errors smaller than \( 5 \times 10^{-4} \) and 45 of them have \( \delta = 5 \times 10^{-4} \). Relaxing this constraint on the errors would reduce \( \epsilon^2 \) for a fixed \( S \). This would reduce the uncertainty with which the modal data constrain the travel time of PKIKP(F).

It is a conservative position to take to state that the travel time of PKIKP(F) is 1213.1 \( \pm 0.3 \) sec.

Turning now to the density distribution, \( \rho(r) \), we have the following values of radius \( r \) and spread \( S \) presented as \( r, S \) in km:
- \( \epsilon = 10^{-2} \)
  - 600, 473; 1200, 326; 1800, 367; 2400, 325; 3000, 316; 3600, 329; 4200, 276; 4800, 306; 5400, 254; 6000, 192.

In this calculation both \( \alpha \) and \( \beta \) were not constrained. Thus, the density in the mantle is known within 1% when averaged over about 300 km. Even in the core, including the inner core, the spreads do not greatly exceed 300 km.

For the S-wave speed \( \beta \), the spreads are near 450 km in the lower mantle and 300 km in the upper mantle for \( \epsilon = 10^{-4} \). This result means that model C198 should provide a good baseline for mantle S-waves with an uncertainty less than \( 10^{-4} \) or about 0.7 sec.

In the inner core with \( \epsilon = 10^{-4} \) the spreads for \( \beta \) are 600, 670; 1200, 616. The average value of \( \beta \) for model C198 is 3.6 km sec \(^{-1} \). This result raises a very interesting question.

Julian, Sheppard, and Davies (16) have reported observations of PKJKP, where \( J \) stands for propagation through the inner core as an S-wave. To explain their observations they require the average value of \( \beta \) in the inner core to be 3.0 km sec \(^{-1} \). It is not possible to fit their data and the modal data simultaneously if the pulses observed by them are identified as PKJKP. However, if the pulses observed by them are PKJKS or as SKJKP, then the two data sets are
more nearly compatible. If this interpretation is correct one
should observe the phase PKJKP roughly 120 to 180 sec
earlier than SKJKP or PKJKS. A search for such observations
is being made (Julian and Davies, personal communication).

In our second inversion we have used model C198 as the
starting model. We have used the experimentally determined
standard errors without restricting them to be less than
5 \times 10^{-4}. This choice adds 2 SED. We have used 16
(baseline corrected) PKIKP, adding 1 SED and 154 \alpha T_1
overtones (multiplied reflected SH-pulses), adding 2 SED. All together
there are 376 GED containing 44 SED. Alone the 154 \alpha T_1
overtones contain 5 SED. That they contribute only 2 SED to the
data set means that there is a large degree of overlap or
redundancy.

After one iterative improvement we have a model, A376,
with \sigma = 0.81, 0.94, 0.98 for n = 1, 2, 3, respectively. Both
models A376 and C198 are equally credible with respect to the
datasets used to derive them. For the 154 \alpha T_1 overtones
95 or 67% have |\delta \gamma| > \sigma, 149 or 96% have |\delta \gamma| \leq 2 \sigma,
and 153 or 99% have |\delta \gamma| \leq 3 \sigma. Furthermore, 76 of these overtones
have \delta \gamma \leq 0 and 78 have \delta \gamma \geq 0. We conclude that the 154 \alpha T_1
overtones are thoroughly compatible with the remaining
modal data. Together they provide an accurate baseline for
S waves.

To investigate the resolving power of the 376-member
data set we must quell the integrable singularities in the
Fréchet kernels of the ray data. Backus and Gilbert (11c),
Backus (15), and Johnson and Gilbert (7a) have dealt with
this problem and we omit the details. As a consequence of the
improved resolving power, we can now show that the travel time of PKIKP(F)
is 1213 \pm 0.2 sec. Studies of PKIKP by Bolt (4), using the Herrin (17) P tables as a baseline, give
the travel time as 1211 sec. We have shown, then, that + 2
sec is the baseline correction for PKIKP. From this we infer
that the baseline correction for P should be +2.0 sec and
for S should be +3.7 sec (most studies use the same stations
for P and S baselines).

The resolving power of the present dataset shows that the
baseline correction for P is +2.0 \pm 0.4 sec, for S is +3.7 \pm 0.9 sec, and for both SKS and SKKS is +3.4 \pm 0.9 sec.

These baseline corrections have been made to the ray data
which, along with 10 \delta S_1 overtones, are now added to the
data set to provide a total of 497 gross Earth data. Adding
111 ray data and 10 modal data to the dataset increases the
number of SED from 44 to 45. Another way to put this is to
segregate the dataset and find that m_0 + m_2 have 2 SED, all
368 modes have 41 SED, and all 127 rays have 19 SED. The
modes have 24 SED not contained in the rays + m_0 + m_2,
while the rays have 2 SED not contained in the modes + m_0
+ m_2. Reduction of \sigma for the rays would give them more SED
but would not necessarily reduce the redundancy in the data-
set.

From these results it is easy to be persuaded that the
Earth’s observed normal mode eigenfrequencies, particularly
the overtones, provide a wealth of information about the
mechanical structure of the interior. In fact, the information
so provided exceeds that derived from rays alone.

In our third inversion we have used the 497 gross Earth
data to derive model B497. The credibility \sigma = 0.83, 0.95,
0.99 for n = 1, 2, 3, respectively. Of the 111 \delta S_1 overtones 54
have |\delta \gamma| < 5 \times 10^{-4}.

To calculate the resolving power of the 497 GED for model
B497, each \sigma is restricted so that \sigma \geq |\delta \gamma|. In this way \sigma = 1,