REVIEWS OF GEOPHYSICS AND SPACE PHYSICS, VOL. 10, NO. 1, PP. 251–285, FEBRUARY 1972

The General Linear Inverse Problem: Implication of Surface Waves and Free Oscillations for Earth Structure

RALPH A. WIGGINS[1]

Department of Earth and Planetary Sciences
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

The discrete general linear inverse problem reduces to a set of m equations in n unknowns. There is generally no unique solution, but we can find k linear combinations of parameters for which restraints are determined. The parameter combinations are given by the eigenvectors of the coefficient matrix. The number k is determined by the ratio of the standard deviations of the observations to the allowable standard deviations in the resulting solution. Various linear combinations of the eigenvectors can be used to determine parameter resolution and information distribution among the observations. Thus we can determine where information comes from among the observations and exactly how it constrains the set of possible models. The application of such analyses to surface-wave and free-oscillation observations indicates that (1) phase, group, and amplitude observations for any particular mode provide basically the same type of information about the model; (2) observations of overtones can enhance the resolution considerably; (3) the degree of resolution has generally been overestimated for many model determinations made from surface waves; and (4) computation of parameter and information resolution is such a simple extension of any inversion procedure based on perturbation parameters that such inversion studies are incomplete without considering resolution.

In this paper we shall consider the general inverse problem in the context of determining earth structure. Since any particular solution is highly nonunique, we must also attempt to determine the limitations or resolution of the solution and the relationship between each individual observation and the class of satisfactory solutions.

General treatments of the inverse problem have received considerable attention in the geophysical literature during the past few years. *Backus and Gilbert* [1967, 1968, 1970] have published three lengthy papers in which they present a detailed construction of a general inverse formalism in the context of determining the structure of the earth from free oscillation observations. *Parker* [1970] has given a very lucid summary of their work while determining the conductivity of the earth. *Backus* [1970*a*, *b*, *c*] has recast and expanded their inversion formalism in an abstract framework. *Gilbert* [1971] has recently pointed out some advantages of factoring the problem by eigenvalues and eigenvectors. *Der et al.* [1970] and *Franklin* [1970] have independently published treatments of the

[1] Now at the Department of Physics, University of Toronto, Toronto 5, Canada.

inverse problem in a stochastic framework. *Jordan and Franklin* [1971] have elaborated on Franklin's theory for applications to determination of earth structure. Finally, mention should be made of the use of Monte Carlo inversion techniques for solving similar problems. *Keilis-Borok and Yanovskaya* [1967], *Press* [1968, 1970*a*, *b*], *Wiggins* [1969], and *Anderssen* [1970] have published on various Monte Carlo problems.

The purpose of the present paper is to cast the general linear inverse problem into a (hopefully) simple matrix algebra and to try to show the relationships between the various works listed above. In addition, various examples taken from seismological problems will be presented to illustrate the type of questions that can be addressed by using the inverse theory.

A note about notation: Scalars will be referred to by upper or lower case italicized letters; vectors (always considered as column matrices) will be referred to by lower case boldface letters; and matrices will be referred to by upper case boldface letters. Throughout the text, I use the singular personal pronouns 'I' and 'my' to refer to my own opinions and preferences; I use the plural pronouns 'we,' 'us,' and 'our' or third-person constructions when making general statements and deductions.

We investigate the relationship between a set of m *observations* O_j where $j = 1 \cdots m$ and a set of n *parameters* P_i where $i = 1 \cdots n$ within the context of some model. The model that we assume must be capable of providing us with a functional (computational) relationship between the model parameters and the calculated values

$$C_j = \mathcal{F}_j(P_i) \qquad j = 1 \cdots m \tag{1}$$

One object of the investigation is to find a set of parameters that minimize the differences $O_j - C_j$. To effect this minimization, we linearize the problem by expanding the functionals in a Taylor's series and discarding all second- and higher-order terms

$$O_j - C_j = \sum_{i=1}^{n} \partial C_j/\partial P_i \, \Delta P'_i \qquad j = 1 \cdots m \tag{2}$$

The dual processes of parameterization of the model and linearization of the functionals have left us with a set of m simultaneous equations to solve for n unknown parameter corrections $\Delta P'_i$. The solution of the problem requires that we be able to compute the partial derivatives or variational parameters $\partial C_j/\partial P_i$ as well as the functional values C_j.

Since all the derivations are in terms of linear variational parameters, the results are strictly true only for small values of $\Delta P'_i$. If $\Delta P'_i$ must be large to satisfy the observations, then the results must be checked by expanding about the new parameters. The absolute size of 'small' that is allowable is highly dependent on the type of observations under consideration. Variational parameters for mass and moment of inertia are absolutely linear; for surface waves and free oscillations, they are slowly varying, whereas for body waves, they are very rapidly varying.

There are two interesting variations on equation 2 that have been examined

in the literature. *Kaula* (1966, chapter 5) examines the case in which there are corrections ΔO_j that must be estimated and added to the observations while estimating the parameters. These corrections arise from inadequacies in the assumed model. The second variation [*Franklin*, 1970; *Jordan and Franklin*, 1970; *Der et al.*, 1970] is concerned with the effects of additive random noise N_j on the observations. I shall consider this case later.

Given the above outline of the problem, the objectives can be stated more precisely:

1. *Reparameterization.* We shall find a new parameterization of the model such that each of the new parameters can be determined uniquely within certain variance limits. These new parameters will be constructed from linear combinations of the old parameters.
2. *Particular solution.* We seek a particular set of parameters $P_i + \Delta P'_i$ that will satisfy the observations within their variance.
3. *Resolving power.* By using the reparameterization determined in part 1 we wish to find the resolution of the observations in parameter space. Since we cannot determine the corrections for individual parameters uniquely, our approach to studying resolution is to try to find the smallest groups of parameters for which the average value can be determined.
4. *Information distribution.* Generally, each particular observation does not contribute independent information about the model. We shall determine the distribution of information among the observations. Such knowledge can then be used as a constraint on data acquisition and smoothing operations.

THEORETICAL DISCUSSION

In the introduction, I indicated how the general inverse problem can be reduced to the solution of a set of linear simultaneous equations. I shall begin the discussion of the properties of the solution by introducing a matrix notation:

$$\Delta\mathbf{c}' = \begin{bmatrix} O_1 - C_1 \\ O_2 - C_2 \\ \vdots \\ O_i - C_i \\ \vdots \\ O_m - C_m \end{bmatrix} \quad \Delta\mathbf{p}' = \begin{bmatrix} \Delta P'_1 \\ \Delta P'_2 \\ \vdots \\ \Delta P'_i \\ \vdots \\ \Delta P'_n \end{bmatrix} \quad \text{var}\{\Delta\mathbf{p}'\} = \begin{bmatrix} \text{var}\{\Delta P'_1\} \\ \text{var}\{\Delta P'_2\} \\ \vdots \\ \text{var}\{\Delta P'_i\} \\ \vdots \\ \text{var}\{\Delta P'_n\} \end{bmatrix} \tag{3}$$

$$\mathbf{A}' = \begin{bmatrix} \partial C_1/\partial P_1 & \partial C_1/\partial P_2 & \cdots & \partial C_1/\partial P_n \\ \partial C_2/\partial P_1 & & & \\ \vdots & & \partial C_i/\partial P_i & \vdots \\ \partial C_m/\partial P_1 & & \cdots & \partial C_m/\partial P_n \end{bmatrix}$$

Notice that each row of $\mathbf{A}'$ corresponds to one particular observation and that each column corresponds to a particular parameter.

Example: To help fix ideas, I shall illustrate the steps for constructing the solution to a general inverse problem by considering the specific problem of determining the shear-wave velocity structure of the earth from Rayleigh wave phase-velocity observations.

There are two principal types of surface waves generated by earthquakes and explosions: Love waves and Rayleigh waves. The depth of penetration of such waves depends on their period; longer-period waves penetrate deeper than shorter-period waves. In a radially heterogeneous medium, both types of waves are dispersed. The amount of dispersion for Rayleigh waves depends on the shear-wave velocity $\beta(z)$, the compressional-wave velocity $\alpha(z)$, and the density $\rho(z)$ over the range of depths z to which the wave penetrates. The dispersion for Love waves is controlled by only $\beta(z)$ and $\rho(z)$. This example assumes that $\alpha(z)$ and $\rho(z)$ are known and that we are seeking to determine $\beta(z)$.

When surface waves travel around the earth several times, various frequencies interfere constructively and destructively to produce discrete modes of free oscillation. Oscillations corresponding to Love-wave interference are called toroidal; oscillations corresponding to Rayleigh-wave interference are called spheroidal. (See *Bullen* [1965, chapters 5 and 8] or *Garland* [1971, chapters 4 and 8] for a more detailed introduction to surface waves and free oscillations.)

In this example we consider only the fundamental Rayleigh mode between periods of 100 sec and 350 sec. This period range corresponds to 78 spheroidal free-oscillation modes, which are labeled ${}_0S_{97}$ (100-sec period) to ${}_0S_{20}$ (350-sec period). Changes in the phase velocity of the Rayleigh wave can be related to changes in the period of the free oscillations. The variational parameters of *Wiggins* [1968] are directly applicable to this problem. Figure 1 is a plot of the partial derivatives

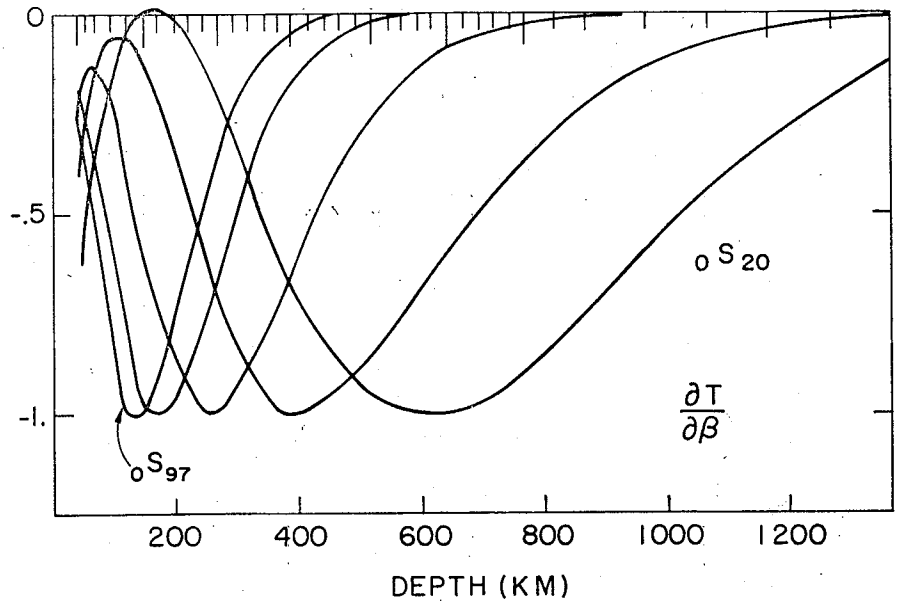

Fig. 1. Partial derivatives of the period T of selected fundamental spheroidal oscillation modes with respect to shear-wave velocity $\beta(z)$. The curves have been normalized so that the maximum displacement is —1.0.

(of T, the period of a particular free-oscillation mode, with respect to $\beta(z)$, the shear-wave velocity at a particular depth z) for selected modes as a function of depth. The derivatives are normalized so that the curve at any depth z gives the appropriate value for a layer of unit thickness. The actual layer thicknesses are indicated by the tick marks at the top of the plot. This figure is a direct graphical representation of a subset of the rows of the matrix $\mathbf{A}'$.

The simultaneous equations can be written in matrix form as

$$\underset{m\times n}{\mathbf{A}'}\ \underset{n\times 1}{\Delta\mathbf{p}'} = \underset{m\times 1}{\Delta\mathbf{c}'} \tag{4}$$

Among these m equations, there are some k independent equations where generally $k \ll n, m$. Such underdetermined systems have a generalized inverse [*Smith and Franklin*, 1969; *Penrose*, 1955; *Moore*, 1920] that minimizes both $|\boldsymbol{\varepsilon}'|^2 = |\mathbf{A}'\,\Delta\mathbf{p}' - \Delta\mathbf{c}'|^2$ and $|\Delta\mathbf{p}'|^2$ simultaneously ($|\boldsymbol{\varepsilon}'|^2$ = determinant of $\boldsymbol{\varepsilon}'^T\boldsymbol{\varepsilon}'$, $\boldsymbol{\varepsilon}'^T = \boldsymbol{\varepsilon}'$ transpose). That is, the generalized inverse selects the smallest change in $\Delta\mathbf{p}'$ that will satisfy the simultaneous equations. The mechanics of finding such inverses will be considered later.

If the generalized inverse is found directly without weighting the equations, the results will be highly dependent on the dimensionality of the components of the vectors $\Delta\mathbf{p}'$ and $\Delta\mathbf{c}'$. For example, if the thickness of a layer corresponding to $\Delta P'_i$ is very large, then the minimizations will favor large changes for that layer relative to the changes for thinner layers. We introduce an $n \times n$ weighting matrix $\mathbf{W}$. Formally, this matrix is the covariance matrix for the parameters. For now, we shall consider it to be a diagonal matrix whose elements W_{ii} are proportional to the dimensions of the parameters. If the parameters are mean values for homogeneous layers, the appropriate weighting terms are the values of the parameters divided by the layer thicknesses or volumes. The weighting is inversely proportional to the layer size because each variational parameter is proportional to the size. The choice of weights has the interesting property of making the lengths of the rows of $\mathbf{A}'\mathbf{W}^{1/2}$ invariant to the selection of layer thicknesses. In a later section I shall illustrate how this definition can be expanded to impose correlation properties on the solution. Instead of minimizing $|\Delta\mathbf{p}'|^2$, we now minimize $\Delta\mathbf{p}'^T\mathbf{W}^{-1}\,\Delta\mathbf{p}'$.

Similarly, we define an $m \times m$ covariance matrix $\mathbf{S}$ for the observations. If the observations are error free, then $\mathbf{S}$ is a diagonal matrix. Each element S_{ii} is proportional to the dimension of the observation O_i. If the observations are not error free, then $S_{ij} = \mathrm{Cov}\,\{O_iO_j\}/\sigma^2$, the covariance of the uncertainties in the observations divided by the problem variance. If the observational errors are uncorrelated, $\mathbf{S}$ is diagonal. We choose to minimize $\boldsymbol{\varepsilon}'^T\mathbf{S}^{-1}\boldsymbol{\varepsilon}'$ rather than $|\boldsymbol{\varepsilon}'|^2$ so that the corrections will be weighted proportionally to the certainty of the observations. *Kaula* [1966, chapter 5] has an excellent discussion of the rationale for using the minimization of $\boldsymbol{\varepsilon}'^T\mathbf{S}^{-1}\boldsymbol{\varepsilon}'$ when $\mathbf{S}$ is not diagonal. The standard deviation σ serves a dual role: it is an adjustment factor for the standard deviations of the observations that I shall refer to as the problem standard deviation, and it acts as a weighting factor between the dual minimizations.

If we rewrite the equation in transformed coordinates

$$\mathbf{A}\,\Delta\mathbf{p} = \Delta\mathbf{c} \tag{5}$$

where $\mathbf{A} = \mathbf{S}^{-1/2}\mathbf{A}'\mathbf{W}^{1/2}$, $\Delta\mathbf{p} = \mathbf{W}^{-1/2}\,\Delta\mathbf{p}'$, and $\Delta\mathbf{c} = \mathbf{S}^{-1/2}\,\Delta\mathbf{c}'$, then minimizing $|\boldsymbol{\varepsilon}|^2 = |\mathbf{A}\,\Delta\mathbf{p} - \Delta\mathbf{c}|^2$ and $|\Delta\mathbf{p}|^2$ is equivalent to minimizing $\boldsymbol{\varepsilon}'^T\mathbf{S}^{-1}\boldsymbol{\varepsilon}'$ and $\Delta\mathbf{p}'^T\mathbf{W}^{-1}\,\Delta\mathbf{p}'$ (since both $\mathbf{W}$ and $\mathbf{S}$ are symmetric and positive definite, there is no difficulty finding $\mathbf{W}^{\pm 1/2}$ or $\mathbf{S}^{\pm 1/2}$; generally $\mathbf{W}$ and $\mathbf{S}$ are diagonal, so that the radicals are trivial). In the sequel, we shall assume that the transformations defined by $\mathbf{W}$ and $\mathbf{S}$ have been made and shall only consider the unprimed equations. The variance of each of the transformed observations Δc_i is σ^2.

Let us now consider the various facets of the solution of the simultaneous equations.

1. *Reparameterization*

If there are k independent equations among the simultaneous equations $\mathbf{A}\,\Delta\mathbf{p} = \Delta\mathbf{c}$, then the matrix $\mathbf{A}$ has rank k and can be factored as

$$\underset{m\times n}{\mathbf{A}} = \underset{m\times k}{\mathbf{U}}\;\underset{k\times k}{\boldsymbol{\Lambda}}\;\underset{k\times n}{\mathbf{V}^T} \tag{6}$$

where $\mathbf{U}$ contains k eigenvectors $\mathbf{u}_i$ of length m associated with the columns (observations) of $\mathbf{A}$, $\boldsymbol{\Lambda}$ is a diagonal matrix of k eigenvalues λ_i ($\lambda_{i+1} \leq \lambda_i$ by convention), and $\mathbf{V}$ contains k eigenvectors $\mathbf{v}_i$ of length n associated with the rows (parameters) of $\mathbf{A}$. Lancsoz [1961] has a particularly lucid description of this eigenvalue problem. Since the eigenvectors are orthonormal,

$$\mathbf{AV} = \mathbf{U}\boldsymbol{\Lambda} \qquad \mathbf{A}^T\mathbf{AV} = \mathbf{V}\boldsymbol{\Lambda}^2$$

$$\mathbf{A}^T\mathbf{U} = \mathbf{V}\boldsymbol{\Lambda} \qquad \mathbf{AA}^T\mathbf{U} = \mathbf{U}\boldsymbol{\Lambda}^2$$

The nonnegative-definite symmetric matrix $\mathbf{A}^T\mathbf{A}$ is $n \times n$, and the matrix $\mathbf{AA}^T$ is $m \times m$. We can use either $\mathbf{A}^T\mathbf{A}$ or $\mathbf{AA}^T$ to find eigenvalues and eigenvectors of $\mathbf{A}$, depending on whether $n \leq m$ or $m < n$, respectively.

Now consider the geometrical significance of the eigenvectors $\mathbf{v}_i$. Madden (personal communication, 1969) has suggested the following interpretation. Suppose we try to find a unit length vector $\mathbf{v}$ that is most nearly parallel to the rows of $\mathbf{A}$ in the sense that $\mathbf{v}^T\mathbf{A}^T\mathbf{Av}$ is maximized. Expand $\mathbf{v}$ in the eigenvectors of $\mathbf{A}$

$$\mathbf{v} = \mathbf{Vb}$$

where

$$\mathbf{b}^T = [b_1 \cdots b_k]$$

$$\sum_i b_i^2 = 1$$

Then

$$\max \mathbf{v}^T\mathbf{A}^T\mathbf{Av} = \max \mathbf{b}^T\mathbf{V}^T\mathbf{V}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{Vb}$$

$$\max \mathbf{v}^T\mathbf{A}^T\mathbf{Av} = \max \mathbf{b}^T\boldsymbol{\Lambda}^2\mathbf{b}$$

$$\max \mathbf{v}^T\mathbf{A}^T\mathbf{Av} = \max (b_1^2\lambda_1^2 + \cdots + b_k^2\lambda_k^2)$$

Clearly, if the eigenvalues are arranged in order of decreasing size, the maximum value is obtained when $b_1 = 1$, $b_2 = \cdots = b_k = 0$. That is, the eigenvector $\mathbf{v}_1$

corresponding to the largest eigenvalue λ_1 is most nearly parallel to the rows of **A**. We can see by induction that the eigenvector $\mathbf{v}_2$ is the vector most nearly parallel to the remainder, etc. The corresponding case holds for the eigenvectors $\mathbf{u}_i$; $\mathbf{u}_1$ is the vector most nearly parallel to the columns of **A**, etc.

The factorization of **A** by eigenvalues and eigenvectors serves as an ideal method for representing and ranking the information contained therein. We can think of the eigenvectors $\mathbf{v}_i$ where $i = 1 \cdots k$ as a new parameterization of the model. These vectors represent a set of k specific linear combinations of the old parameters that are fixed by the observations. Likewise, the k eigenvectors $\mathbf{u}_i$ represent the set of k linear combinations of the observations that are relevant to the particular inverse problem.

To formalize the new parameterization, we define a new parameter correction vector

$$\mathbf{\Delta p}^* = \mathbf{V}^T \mathbf{\Delta p} \qquad \mathbf{\Delta p} = \mathbf{V}\, \mathbf{\Delta p}^* \tag{7}$$

Since there are only k parameters to determine in $\mathbf{\Delta p}^*$, they are determined uniquely by the simultaneous equations. The new parameter correction ΔP^*_i indicates the amount of the eigenvector (new parameter) $\mathbf{v}_i$ that must be added to the initial parameter vector **p** to solve equation 5.

In addition to the k fixed eigenvectors $\mathbf{v}_i$, there remain $n - k$ linear combinations of the parameters that are not determined by the set of equation 4 or 5. I shall refer to these vectors as *free vectors* $\mathbf{v}_{0i}$. The free vectors will also be constructed to be orthonormal.

Example: Figure 2 illustrates the eigenvalues λ_i and transformed eigenvectors $\mathbf{v}'_i = \mathbf{W}^{1/2}\mathbf{v}_i$ where $i = 1 \cdots 5$ for the **A** of the example problem. Notice how the vector $\mathbf{v}'_1$ at the top of the plot visually resembles the average of the vectors in Figure 1. The eigenvectors also show increasing complexity with decreasing magnitude. This seems to be a universal property for all inverse problems that the author has encountered.

2. *Particular Solution*

Given the eigenvalue-eigenvector factorization of **A**, a particular solution for equation 5 is easy to find in terms of the generalized inverse:

$$\begin{aligned} \mathbf{A}\, \mathbf{\Delta p} &= \mathbf{\Delta c} \\ \mathbf{U\Lambda V}^T \mathbf{\Delta p} &= \mathbf{\Delta c} \\ \mathbf{\Delta p} &= \mathbf{V\Lambda}^{-1}\mathbf{U}^T\, \mathbf{\Delta c} \\ \mathbf{\Delta p}^* &= \mathbf{\Lambda}^{-1}\mathbf{U}^T\, \mathbf{\Delta c} \\ \Delta P^*_i &= \lambda_i^{-1}\mathbf{u}_i^T\, \mathbf{\Delta c} \end{aligned} \tag{8}$$

Each of the new parameter corrections ΔP^*_i is determined by a particular linear combination of the observations.

The $k \times k$ variance matrix $\mathbf{X}^*$ of the new parameters $\mathbf{\Delta p}^*$ is

$$\mathbf{X}^* = \mathbf{\Lambda}^{-1}\mathbf{USU}^T\mathbf{\Lambda}^{-1}$$

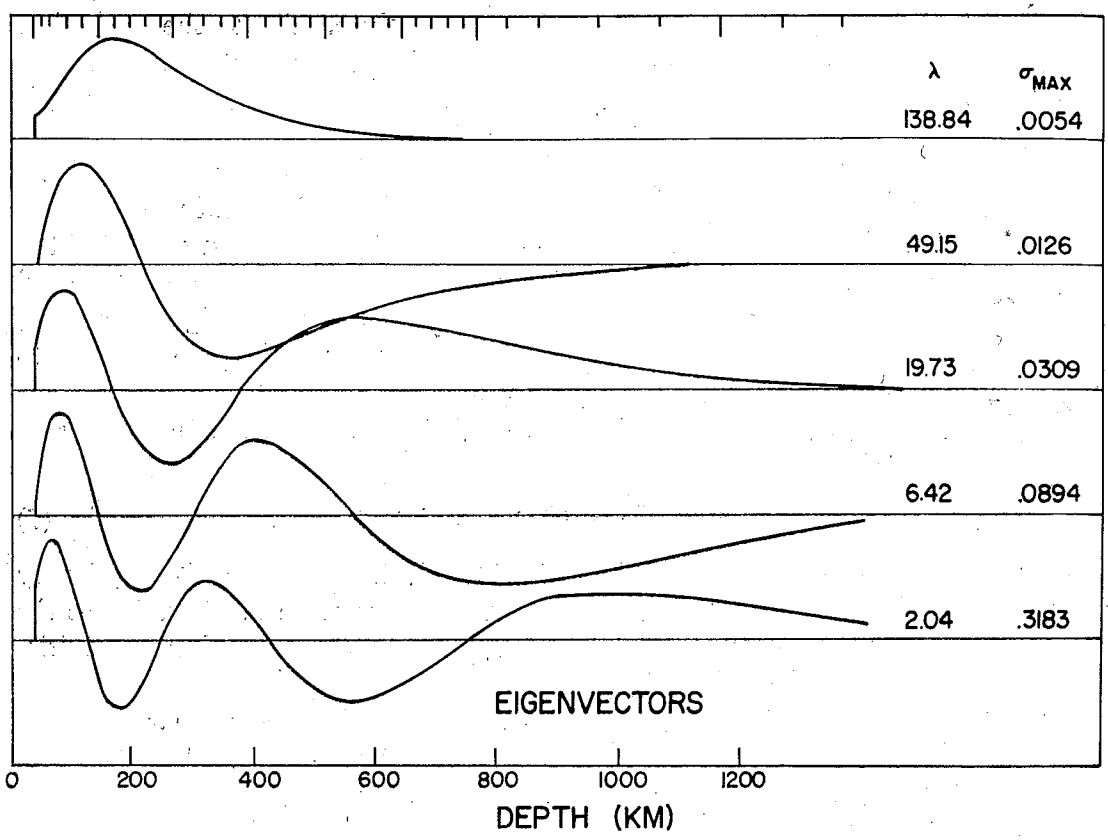

Fig. 2. Eigenvectors derived from a coefficient matrix based on the partial derivatives illustrated in Figure 1. The corresponding eigenvalues and solution standard deviations are given at the right.

where **S** is the variance matrix of the observations. (If $x = a_1y_1 + a_2y_2$, then Var $\{x\} = a_1^2$ Var $\{y_1\} + a_2^2$ Var $\{y_2\}$ [see *Johnston*, 1963, pp. 106–110; *Kaula*, 1966, pp. 100–103]. Since we have transformed the coordinates so that **S** is a unit matrix multiplied by σ^2, we have in the new coordinates

$$*\mathbf{X} = \sigma^2 \mathbf{\Lambda}^{-2}$$

$$\sigma^*_{pi}{}^2 = \text{Var}\,\{\Delta P^*_i\} = \sigma^2/\lambda_i^2 \tag{9}$$

and the variance of the corrections is

$$\sigma_{pi}{}^2 = \text{Var}\,\{\Delta P_i\} = \sigma^2 \sum_{j=1}^{k} V_{ij}{}^2/\lambda_j^2 \tag{10}$$

where V_{ij} are the elements of the eigenvector matrix **V**. Notice that this is the variance of the parameter corrections ΔP_i, not the variance of the parameters P_i. Since an arbitrary amount of any of the $n - k$ free vectors can be added to the parameter vector **p**, we can always find parameter adjustments that fall outside the range of the correction standard deviations.

There is some question about the interpretation of σ^2, the variance of the problem. One either assumes that $\sigma = \sigma_o$ (generally we choose $\sigma_o = 1$) is known from the estimation of the variances of the observations or that $\sigma = \sigma_\epsilon$ must be

estimated from the error of fit

$$\sigma_\epsilon^2 = |\boldsymbol{\varepsilon}|^2/m - k$$

$$(\boldsymbol{\varepsilon} = \mathbf{A}\,\Delta\mathbf{p} - \Delta\mathbf{c})$$

[*Johnston*, 1963, chapter 4; *Draper and Smith*, 1966, chapter 2]. If these two estimates of σ are not of the same order of magnitude, the problem is inconsistent. If $\sigma_\epsilon \ll \sigma_o$, then the variances estimated for the observations are too large compared to the internal consistency between observations. In this case, we would be justified in setting $\sigma = \sigma_\epsilon$. If $\sigma_o \ll \sigma_\epsilon$, either the variances of the observations were underestimated compared to the internal consistency of the observations or the model is inadequate to fit the observations. Whether or not the model is inadequate can generally be determined by examining the errors $\boldsymbol{\varepsilon}_i$ for trends.

In the discussion above we have assumed that $\mathbf{A}$ has k nonzero eigenvalues and that all the rest are identically zero. For many problems where $m \gg k$ the eigenvalues approach zero exponentially, and it is very difficult to distinguish between very small eigenvalues and zero eigenvalues. Thus we are generally faced with the task of selecting some cutoff value λ_{lim} and ignoring all eigenvalues less than that level. There are several approaches to this problem.

My own 'sharp cutoff' approach is to say that there is an upper limit on the standard deviations $\sigma_{p\text{-lim}}$ of the parameter corrections beyond which the corrections have no physical meaning. I search for the largest k for which each Var $\{\Delta P'_i\}$ is less than $\sigma^2_{p\text{-lim}}$. This then determines the number of degrees of freedom associated with the solution.

Example: The maximum of the untransformed parameter standard deviations

$$\sigma_{\max} = \max_i \{\Delta P'_i\}$$

are listed for each k value in Figure 2. These values are based on the assumption that the standard deviation for the period of each free oscillation mode is 0.5% of the period. The standard deviations of the ΔP^*_i are just the reciprocals of the eigenvalues. For the choice of $\mathbf{W}$ matrix used here, the $\sigma_{\max}$ are somewhat less than the σ^*_{pi}.

As will be shown in later parts of this section, the great advantage of the sharp cutoff strategy is that the computation of the relationships between (1) the standard deviations and resolution of parameter corrections and (2) the standard deviations and information distribution among the observations is direct and trivial. The disadvantages are that the solution is slightly biased in the presence of random noise in the observations and that the standard deviations of the parameter corrections may not be uniform for all the parameters. We shall see examples in the second part of this paper where the parameter correction standard deviations are nearly uniform and other examples where they are not. Most investigators have solved the nonuniform standard deviation problem by varying the effective number of degrees of freedom for each parameter estimated. An alternative method for equalizing the standard deviations is to alter the weighting matrix $\mathbf{W}$. Suppose that for some value of k the standard deviations of the parameter corrections is characterized by a smoothly varying envelope ω_i where $i = 1 \cdots n$. If we form a

diagonal matrix $\boldsymbol{\Omega} = \text{diag}\,[\omega_1 \cdots \omega_n]$ and a new weight matrix $\mathbf{W}^* = \boldsymbol{\Omega}^{-1/2}\mathbf{W}\boldsymbol{\Omega}^{-1/2}$, the envelope of the resulting variance for the problem solved with $\mathbf{W}^*$ rather than $\mathbf{W}$ will be uniform for k degrees of freedom.

Example: Consider a problem in which $n = 4$, $m = 1$, $k = 1$, $\mathbf{A}' = [2, 2, 1, 1]$, $\Delta\mathbf{c} = [1]$, and $\mathbf{S} = [1]$. If we solve the problem directly in the primed coordinates we have

$$\mathbf{u}'_1 = [1] \qquad \lambda'_1 = 10^{1/2} \qquad \mathbf{v}_1' = 1/(10)^{1/2}[2, 2, 1, 1]^T$$

$$\text{Var}\,\{\Delta P'^*_1\} = 1/10$$

$$\Delta\mathbf{p}' = 1/10[2, 2, 1, 1]^T$$

$$\text{Var}\,\{\Delta\mathbf{p}'\} = 1/100[4, 4, 1, 1]^T$$

Now if we solve the problem in the unprimed coordinates using a weighting matrix $\mathbf{W} = \text{diag}\,[\frac{1}{2}, \frac{1}{2}, 1, 1]$ that is proportional to the inverse of the standard deviations above, we have

$$\mathbf{A} = [\sqrt{2}, \sqrt{2}, 1, 1], \qquad \mathbf{u}_1 = [1] \qquad \lambda_1 = \sqrt{6} \qquad \mathbf{v}_1 = 1/\sqrt{6}\,[\sqrt{2}, \sqrt{2}, 1, 1]^T$$

$$\text{Var}\,\{\Delta P^*_1\} = 1/6$$

$$\Delta\mathbf{p} = 1/6[\sqrt{2}, \sqrt{2}, 1, 1]^T$$

$$\text{Var}\,\{\Delta\mathbf{p}\} = 1/36[2, 2, 1, 1]^T$$

$$\Delta\mathbf{p}' = 1/6[1, 1, 1, 1]^T$$

$$\text{Var}\,\{\Delta\mathbf{p}'\} = 1/36[1, 1, 1, 1]^T$$

Gilbert [1971] has taken a more indirect approach. He says that since we know the correction ΔP^*_i only within $\pm\sigma^*_{pi}$, we should take the smallest adjustment within that range. Thus if $\Delta P^*_i < -\sigma^*_{pi}$, he resets ΔP^*_i to $\Delta P^*_i + \sigma^*_{pi}$; if $\Delta P^*_i > \sigma^*_{pi}$, he resets ΔP^*_i to $\Delta P^*_i - \sigma^*_{pi}$; and if $-\sigma^*_{pi} \leq \Delta P^*_i \leq \sigma^*_{pi}$, he resets ΔP^*_i to zero. There are some difficulties with this approach. It tends to give models for which the calculated values are near the standard deviation limits (unless the beginning model already gives very small errors), thus ignoring any added knowledge gained from the internal consistency between the observations. A more subtle difficulty arises when Gilbert's scheme is used within the context of this paper. Since both ΔP^*_i and σ^*_{pi} are proportional to λ_i^{-1}, the inclusion of a particular ΔP^*_i is independent of the size of λ_i^{-1}. Thus we have no cutoff criterion at all. Gilbert successfully circumvents the possible bad effects of including very small eigenvalues by averaging the resulting parameter values with kernels determined from the eigenvectors with large eigenvalues.

Franklin [1970], *Jordan and Franklin* [1971], and *Der et al.* [1970] take another approach. They set up the equations in a stochastic framework to estimate the parameters in the presence of noise

$$\mathbf{A}'\Delta\mathbf{p}' + \mathbf{n}' = \Delta\mathbf{c}' \tag{11}$$

where $\mathbf{n}'$ is the $m \times 1$ vector of observational noise. *Franklin* [1970] obtains a best unbiased linear estimate of the parameter corrections in the form

$$\Delta \mathbf{p}' = \mathbf{W}\mathbf{A}'^{T}(\mathbf{A}'\mathbf{W}\mathbf{A}'^{T} + \mathbf{S})^{-1}\,\Delta \mathbf{c}' \tag{12}$$

If we transform the coordinates to eliminate $\mathbf{W}$ and $\mathbf{S}$, we have

$$\begin{aligned} \Delta \mathbf{p} &= \mathbf{A}^{T}(\mathbf{A}\mathbf{A}^{T} + \sigma^2\mathbf{I}_m)^{-1}\,\Delta \mathbf{c} \\ \Delta \mathbf{p} &= \mathbf{V}\mathbf{\Lambda}\mathbf{U}^{T}(\mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^{T} + \sigma^2\mathbf{I}_m)^{-1}\,\Delta \mathbf{c} \\ \Delta \mathbf{p}^* &= \mathbf{V}^{T}\,\Delta \mathbf{p} \end{aligned} \tag{13}$$

(The σ^2 is included because the variance of the transformed observations is σ^2.) Now if we assume that $\mathbf{A}\mathbf{A}^{T}$ has m nonzero eigenvalues (many may be very small), we can expand this solution in the form

$$\begin{aligned} \Delta \mathbf{p}^* &= \mathbf{\Lambda}\mathbf{U}^{T}(\mathbf{U}[\mathbf{\Lambda}^{-2} - \sigma^2\mathbf{\Lambda}^{-4} + \cdots]\mathbf{U}^{T})\,\Delta \mathbf{c} \\ \Delta \mathbf{p}^* &= \mathbf{\Lambda}(\mathbf{\Lambda}^2 + \sigma^2\mathbf{I}_m)^{-1}\mathbf{U}^{T}\,\Delta \mathbf{c} \\ \Delta P^*_i &= (\lambda_i/(\lambda_i^2 + \sigma^2))\mathbf{u}_i^{T}\,\Delta \mathbf{c} \\ \mathrm{Var}\,\{\Delta P^*_i\} &= [\sigma\lambda_i/(\lambda_i^2 + \sigma^2)]^2 \end{aligned} \tag{14}$$

When $\lambda_i^2 \gg \sigma^2$, the coefficient is approximately $1/\lambda_i$; when $\lambda_i^2 = \sigma^2$, the coefficient is $1/2\lambda_i$; and when $\lambda_i^2 \ll \sigma^2$, the coefficient is λ_i/σ^2. The variance of the problem σ^2 acts as a tapered cutoff that eliminates the effects of small eigenvalues. The number of degrees of freedom for this solution is $k = \sum \lambda_i^2/(\lambda_i^2 + \sigma^2)$.

The uncertainty of the observations may give unacceptably large standard deviations Var $\{\Delta P_i\}$ for the solution parameter corrections. For example, if the observations had no errors, all eigenvalues of any size would be included, and the solution would likely be unstable. *Der et al.* [1970] take the pragmatic approach of adjusting the size of σ^2 until acceptable standard deviations are determined for the parameters. *Jordan and Franklin* [1971] choose to alter the parameter covariance matrix $\mathbf{W}$. The idea is that one can force the adjustments of groups of parameters to be correlated in any solution obtained by entering off-diagonal terms in the $\mathbf{W}$ matrix. Let us assume that the dimensionality of all the parameters is the same, so that before introducing off-diagonal terms $\mathbf{W}$ is an identity matrix. Now suppose it is desired that the parameter adjustments $\Delta P'_i$ and $\Delta P'_j$ be identical. This condition can be imposed by setting $W_{ij} = W_{ji} = 1$. The effect of this change is to cause the large eigenvalues to grow larger and the small eigenvalues to grow smaller. If one is very clever in setting up the $\mathbf{W}$ matrix, he can force the eigenvalues to polarize into a group of large values and another group of small values. The problems with this approach will be further discussed in the next section.

In my opinion, the selection of a cutoff level according to the *Der et al.* [1970] strategy of varying σ^2 in a stochastic model setting is probably the optimum approach. The results obtained using a sharp cutoff do not differ greatly from the stochastic approach for most applications. I shall use this sharp cutoff strategy for the illustrations in this paper.

When solving nonlinear problems in a linear setting, one must frequently iterate to find a model by solving the equation $\mathbf{A}\,\Delta \mathbf{p} = \Delta \mathbf{c}$ several times in succession until a solution is determined. This multivariate Newton iteration process is notoriously unstable for many problems. R. Harter (work in preparation, 1971)

has pointed out that one effective method of stabilizing the search for the minimum is to limit the step size during each iteration by selecting a k value such that

$$\sum_{i=1}^{k} \Delta P^{*}{}_{i}^{2} \leq \Delta P^{2}{}_{\max}$$

where $\Delta P_{\max}$ is the maximum allowable step size.

3. *Parameter Resolution*

In part 1 of this section, we determined a new parameterization of the problem in terms of the eigenvectors $\mathbf{v}_i$. In part 2, we described two methods for determining the number of degrees of freedom k of the solution based on a maximum allowable standard deviation of the parameter corrections. If we think of a particular set of parameter corrections as a point in an n-dimensional space, the solution obtained in the last part fixes the position of the point in the k dimensions of the space spanned by the vectors $\mathbf{v}_i$. My purpose here is to characterize the nature of the solution by examining the nature of the vectors $\mathbf{v}_i$. Such a study is generally referred to as determining the resolution of the observations.

The object is to find linear combinations of $\mathbf{v}_i$ that will provide physically meaningful measures of the solution. If a parameter correction ΔP_j is perfectly resolvable, we can find a linear combination

$$\mathbf{v}^{*}{}_{j} = \sum_{i=1}^{k} b_{ij}\mathbf{v}_i \tag{15}$$

such that $V^{*}{}_{ji} = \delta_{ij}$ ($V^{*}{}_{ji}$ are the components of $\mathbf{v}^{*}{}_{i}$, and δ_{ij} is the Kronecker delta function). If the parameter correction ΔP_j is not perfectly resolvable, perhaps we can resolve a linear combination of physically adjacent parameters. If this is the case, then we shall say that the problem is characterized by compact resolution. In more complicated situations, the resolution may consist of linear combinations of parameters of different physical type (β and ρ, for example) or from widely separate parts of the model.

Example: Suppose a problem has 4 parameters ($n = 4$) and that there are 2 degrees of freedom in the solution ($k = 2$) characterized by the eigenvectors

$$\mathbf{v}_1 = (2/5)^{1/2}[1., 1., 0.5, 0.5]^T$$

$$\mathbf{v}_2 = (2/5)^{1/2}[-0.5, -0.5, 1., 1.]^T$$

We can find linear combinations such that

$$\mathbf{v}^{*}{}_{1} = \sqrt{2}/5\ (2\mathbf{v}_1 - \mathbf{v}_2) = 1/\sqrt{2}\ [1., 1., 0., 0.]^T$$

$$\mathbf{v}^{*}{}_{2} = \sqrt{2}/5\ (\mathbf{v}_1 + 2\mathbf{v}_2) = 1/\sqrt{2}\ [0., 0., 1., 1.]^T$$

Such a problem has compact resolution. Now suppose that the eigenvectors are

$$\mathbf{v}_1 = (2/5)^{1/2}[1., 0.5, 0.5, 1.]^T$$

$$\mathbf{v}_2 = (2/5)^{1/2}[-0.5, 1., 1., -0.5]^T$$

The linear combinations give

$$\mathbf{v}^*_1 = \sqrt{2}/5\,(2\mathbf{v}_1 - \mathbf{v}_2) = 1/\sqrt{2}\,[1., 0., 0., 1.]^T$$

$$\mathbf{v}^*_2 = \sqrt{2}/5\,(\mathbf{v}_1 + 2\mathbf{v}_2) = 1/\sqrt{2}\,[0., 1., 1., 0.]^T$$

This case does not have compact resolution.

There are several methods for determining the b_{ij} in equation 15. The simplest method consists of minimizing the difference between the desired vector $\mathbf{v}^*_j$ and a delta function

$$\min\left[\sum_{l=1}^{n}(V^*_{lj} - \delta_{lj})^2\right]$$

where

$$V^*_{lj} = \sum_{i=1}^{k} b_{ij} V_{li}$$

Since the $\mathbf{v}_i$ are orthonormal, the solution is $b_{ij} = V_{ij}$. More generally, we can define a resolution matrix

$$\mathbf{R} = \mathbf{V}\mathbf{V}^T$$

The jth column of this $n \times n$ matrix $\mathbf{r}_j$ is the least-squares solution for maximizing the jth parameter.

The formula given above for the resolution matrix is only valid for the case of a sharp cutoff, of course. For the stochastic solution, the resolution matrix is defined slightly differently

$$\mathbf{R} = \sum_j (\lambda_j^2/\lambda_j^2 + \sigma^2)\mathbf{v}_j\mathbf{v}_j^T$$

to accommodate the tapered cutoff. The rows of this matrix are equivalent to the resolution curves of *Der et al.* [1970]. For the rest of this part, we shall consider only the sharp cutoff case. There are equivalent tapered cutoff techniques for each of the cases described.

Example: The solid curves in Figure 3 show five representative columns from the resolution matrix for the example problem. The form of the curves indicates that for the most part this problem is characterized by compact resolution. The bottom curve, however, shows that there is some trade off between corrections very near the surface and corrections at great depth.

Since the matrix $\mathbf{R}$ was constructed from k orthonormal vectors, we can extract another set of k orthonormal vectors from $\mathbf{R}$ in a way that exhibits the resolution (the columns of $\mathbf{R}$ cannot be completely orthogonal as they stand unless $\mathbf{R}$ is the identity matrix). There are many ways to perform the extraction, depending on the particular need for the vectors. For this work, I have tried to select a set of k vectors that will enhance the delta-like behavior of the resolution matrix. My scheme consists of searching the matrix $\mathbf{R}$ for that vector with the greatest resolving power. This is a simple search, since the vector with the best resolving power is nearly always the vector with the maximum diagonal element. Suppose this vector is found in column j. The vector is normalized $\mathbf{v}^*_1 = \mathbf{r}_j/\sqrt{R_{jj}}$; then I

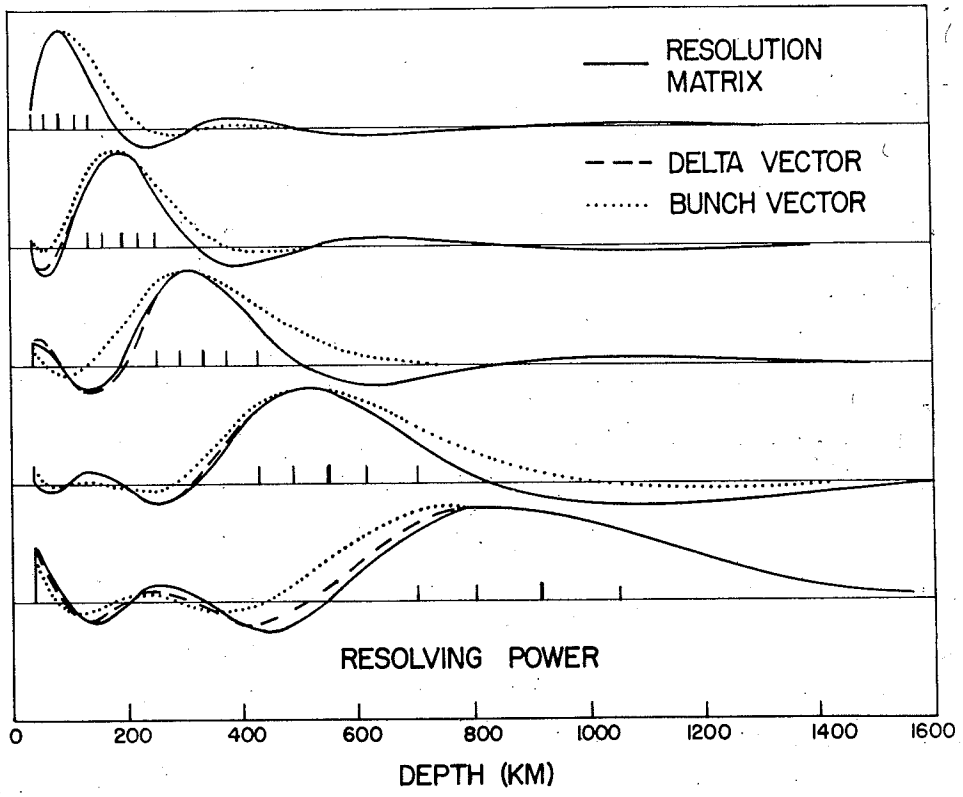

Fig. 3. Various measures of parameter resolution for the data set illustrated in Figure 1.

form a new matrix $\mathbf{R}_1 = \mathbf{R} - \mathbf{v}^*_1\mathbf{v}^{*T}_1$ and repeat the search. This process is repeated until I obtain k vectors. These vectors are referred to as 'delta vectors.'

Example: The long dashes in Figure 3 illustrate the delta vectors $\mathbf{v}^*_i$ extracted from the resolution matrix. In this example, the columns selected from the resolution matrix (the solid curves) were nearly orthogonal. Thus there is not much difference between the two sets of vectors. In some cases (see part 4 of this section, for example) the delta vectors extracted from the resolution matrix will have smaller side lobes than the resolution matrix columns. For this reason, the delta vectors generally give a better indication of the compactness of the resolving power than do the columns of the resolution matrix.

We can optimize the compactness of the representation by seeking linear combinations of the $\mathbf{v}_i$ that maximize the ratio of energy for a particular set of parameters relative to the total energy in the vector.

Let us define our desired vector $\mathbf{v}_b$ (called 'bunch vector') as a linear combination of the eigenvectors $\mathbf{v}_b = \mathbf{V}\,[c_1 \cdots c_k]^T$. We select a set of parameters I_1 and seek to maximize the total energy within this set. That is, we maximize the ratio

$$\lambda = \sum_{i \subset I_1} v_{bi}^2 \Big/ \sum_{i=1}^{k} v_{bi}^2$$

Taking the derivative with respect to c_i and equating to zero, we find $(\mathbf{U} - \lambda \mathbf{I}_I)$ $[c_1 \cdots c_i]^T = 0$ where

$$\mathcal{V}_{ij} = \sum_{i \in I_1} V_{ii} V_{ij}$$

are elements of the square $k \times k$ matrix $\mathcal{V}$. The solution is the eigenvector corresponding to the largest eigenvalue of the matrix $\mathcal{V}$. Notice that, if the parameter set I_1 includes only one parameter, this equation gives the same solution as we found for the resolution vectors earlier in this section.

Example: The dotted curves in Figure 3 illustrate some representative bunch vectors calculated for the example problem. The parameter sets were selected so that the central lobes approximately coincide with those of the delta vectors. These vectors are not orthogonal. They show somewhat better compactness than the delta vectors, but they add little new information about the resolving power of the initial data set.

If the delta or bunch vectors indicate that the system exhibits compact resolution, then it becomes relevant to estimate the number of adjacent parameters that must be grouped together to obtain a unique inverse. Of course, this number will be a function of the set of parameters chosen. This estimate can be obtained directly from the diagonal of the resolution matrix **R**. First note that the sum of the diagonal terms of **R** is Tr (**R**) = k, the number of eigenvectors used to form **R**. If we find two indices r and s such that

$$\sum_{i=r}^{s} R_{ii} = 1$$

the $r - s + 1$ parameter corrections $\Delta P_r \cdots \Delta P_s$ can be resolved uniquely if the system exhibits compact resolution. The depths that correspond to ΔP_r and ΔP_s will indicate the physical size of the resolved section.

Example: Consider again the four-parameter problem from earlier in this section. The first case, $\mathbf{v}_1 = (2/5)^{1/2}\,[1., 1., 0.5, 0.5]^T$ and $\mathbf{v}_2 = (2/5)^{1/2}\,[-0.5, -0.5, 1., 1.]^T$, gives a resolution matrix

$$\mathbf{R} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

and $\sqrt{2}\,\mathbf{v}^*_1 = \mathbf{r}_1 = \mathbf{r}_2$, $\sqrt{2}\,\mathbf{v}^*_2 = \mathbf{r}_3 = \mathbf{r}_4$. The second case, $\mathbf{v}_1 = (2/5)^{1/2}\,[1., 0.5, 0.5, 1.]^T$ and $\mathbf{v}_2 = (2/5)^{1/2}\,[-0.5, 1., 1., -0.5]^T$, gives the resolution matrix

$$\mathbf{R} = \begin{bmatrix} 0.5 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \end{bmatrix}$$

and $\sqrt{2}\,\mathbf{v}^*_1 = \mathbf{r}_1 = \mathbf{r}_4$, $\sqrt{2}\,\mathbf{v}^*_2 = \mathbf{r}_2 = \mathbf{r}_3$. The trace from either resolution matrix

indicates the same resolution distribution. Thus the characterization of the resolution by using the trace is only valid when the resolution is compact.

Example: The resolution ticks along the axes in Figure 3 correspond to sets of fractional limits r and s such that

$$\sum_{i=r}^{s} R_{ii} = 0.25$$

The resolution at the depth corresponding to the jth tick is given by the distance between tick $j - 2$ and tick $j + 2$. This example does not exhibit the ideal form of compact resolution. Nevertheless, the ticks give an excellent summary of the distribution of resolution.

Another method for characterizing the uniqueness and nonuniqueness is to examine the properties of the $n - k$ free vectors $\mathbf{v}_{0i}$ that are orthogonal to the k eigenvectors $\mathbf{v}_i$. Once we have constructed the resolution matrix $\mathbf{R}$ from the eigenvectors $\mathbf{v}_i$, we can find the resolution matrix $\mathbf{R}_0$ of the free vectors directly: $\mathbf{R}_0 = \mathbf{I}_n - \mathbf{R}$. From this resolving matrix, we can then extract $n - k$ free delta vectors as we just described above for $\mathbf{R}$. Such vectors can then be used in many different ways to adjust the model to fit criteria other than those used in the inversion without altering the fit to the observations. In most real problems the amount of each free vector that can be added is limited by physical considerations or by the nonlinear properties of the functional relationships between the parameters and the calculated values. Thus the free vectors are seldom completely free but have large standard deviations.

Backus and Gilbert [1968, 1970] and *Backus* [1970*b*] examine extensively several other techniques for extracting representations and measures of the resolving power of a set of observations. These approaches are difficult to describe in the notation used here but do not give results that differ significantly from those above.

Let us consider briefly the approach of *Jordan and Franklin* [1971] that was mentioned in part 2 of this section. They stabilize the solution by introducing off-diagonal cross-correlation terms in the covariance matrix $\mathbf{W}$. Apparently one must introduce correlation terms based on the achievable resolving power for a particular problem. Thus one is faced with the prospect of needing to know the solution in order to determine the solution. If guesses are made about the resolving power, either the solution will be unstable if the resolving power is overestimated or the solution will be overdetermined if the resolving power is underestimated. It seems preferable to establish a cutoff that limits the degrees of freedom of the solution.

An alternate application for the use of off-diagonal terms may be the establishment of correlations based on physical considerations. We expect density increases to be correlated with compressional velocity increases, for example. Although this is certainly a valid application, I prefer to first find a solution to the unconstrained general inverse problem and then modify that solution by adding free vectors $\mathbf{v}_{0i}$ where $i = 1 \cdots n - k$ to approximate the other physical constraints.

4. *Information Distribution*

Until now we have concentrated all our attention on the properties of the rows of the variational matrix $\mathbf{A}$. This has led to an understanding of the limits on and

nature of the solution in parameter space. Here we shift our attention to the columns of the matrix **A**. Since each index along a column corresponds to a particular observation, an analysis by columns will indicate the information distribution among the observations. Such an analysis indicates which data are providing the same information as other data.

The approach is exactly parallel to that used for the row analysis. Here we consider the eigenvectors $\mathbf{u}_i$ rather than $\mathbf{v}_i$. As before, we can compute an $m \times m$ resolution matrix $\mathbf{R} = \mathbf{U}\mathbf{U}^T$ and then apply all the techniques discussed in the last part to estimate the information distribution.

Example: Figure 4 illustrates resolution-matrix vectors, delta vectors, and resolution ticks for our sample problem. This figure gives a good indication of the improved compactness of the delta vectors over the resolution-matrix vectors. The meaning of the information distribution found can be stated as follows. If the uncertainty in determining the period of each free oscillation mode is 0.5% of the period and the maximum uncertainty that we can tolerate in the solution is 0.4 km/sec (~10%), the modes ${}_0S_{20}$ to ${}_0S_{24}$ provide one independent constraint on the solution, the modes ${}_0S_{25}$ to ${}_0S_{37}$ provide a second constraint, etc.

The use of resolution ticks provides us with a compact method for displaying the entire dependence between (1) uncertainty of the parameter corrections, (2)

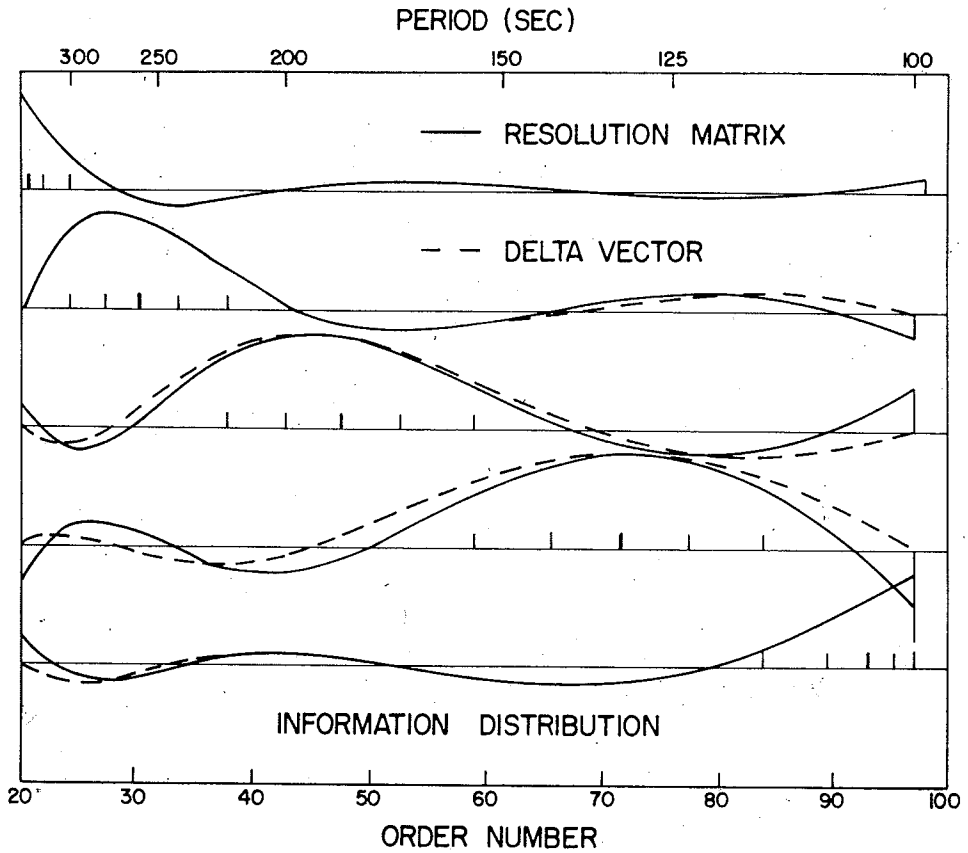

Fig. 4. Various measures of information distribution for the data set illustrated in Figure 1.

parameter resolution, and (3) information distribution. Figure 5 is such a display for the example problem.

The bottom box in Figure 5 shows information distribution plotted against the order number of the free oscillation modes under consideration. The first line has tick marks showing the distribution of the observations. The next five lines above show the resolution ticks for the cases $k = 2 \cdots 6$. The observations between ticks $j - 2$ and $j + 2$ (for any j) contribute $1/k$ of total information.

The middle box is a similar display showing parameter resolution. The ticks on the bottom line show the layer thicknesses that were used to parameterize the model. The upper five lines show the parameter resolution for the cases $k = 2 \cdots 6$. In some cases when the ticks are very close together, the lighter ticks are omitted.

For the moment, regard only the scalloped curves in the top box of the figure. This box shows the distribution of parameter uncertainty for various k values. Here we plot the standard deviations

$$\sigma'_{pkj} = \left(\sigma^2 W^2_{jj} \sum_{i=1}^{k} V_{ji}^2/\lambda_i^2\right)^{1/2}$$

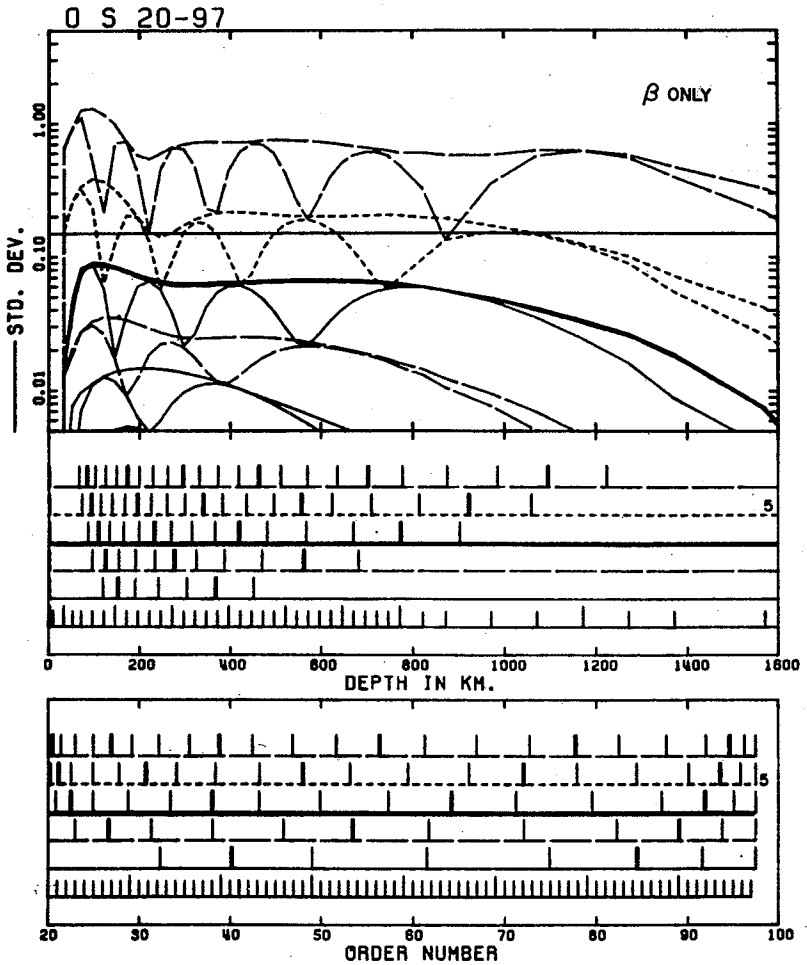

Fig. 5. Resolution diagram for long-period Rayleigh waves (periods 100 to 350 sec). See text for an explanation of the diagram.

for $k = 1 \cdots 6$. The dashes used to plot the curves correspond to the dashes used in the other boxes.

From the display of standard deviations, we see that if we want to know the corrections within 0.15 km/sec (the solid horizontal line in the top box), we can use only four eigenvectors. The corresponding lines in the two lower boxes give us the resolution in parameter space and observation space. If we could improve the accuracy of the observations, the effect on the plot would be simply to lower all the curves. Thus, lowering the observation standard deviations from 0.5% to 0.1% is equivalent to raising the cutoff level from 0.15 to 0.75 km/sec. The plot indicates that we would then be justified in using five degrees of freedom in the solution. The values of σ'_{pj7} are so large that this case will probably never be of practical importance.

The scalloped appearance of the standard deviation curves is caused by zero crossings in the eigenvectors. A close examination of the relationship between the two upper displays shows that at least one peak of the standard deviation curve is included within each resolution thickness. Since this peak value is the only value relevant to the interpretation of the curve, I have devised a smoothing scheme (based on finding the envelope of oscillatory functions) to improve the ease of reading the graphs. If another set of orthogonal vectors were determined from linear combinations of the k eigenvectors, the zero crossings and peaks would usually fall in different places, and the appearance of the scalloped curves would be altered. The envelope gives an estimate of the maximum values attained by all such linear recombinations of the eigenvectors. If we desired to equalize the standard deviations of the parameter corrections by modifying $\mathbf{W}$, an envelope such as is shown in Figure 5 should be used to effect the equalization.

This completes the formal discussion of the theory and computational procedures for determining resolving power. I have a few miscellaneous topics to consider before closing this section.

Until now the development has been discussed from the point of view of analyzing a set of preexisting observations. Frequently, after making such an analysis (and finding insufficient resolving power), we are faced with the question of what new observations to attempt in order to improve the resolving power. If we are able to compute the relevant partial derivatives, we can perform computational experiments to determine how small the variance of the new observations would need to be to add to the resolving power of the existing observations. If that variance is not too small, we would be justified in collecting the necessary data. Examples of such calculations will be given in the next section.

Another type of question that can be answered by computational experiment concerns the effect of the specific parameterization selected. There are two basic questions about parameterization. The first is concerned with whether the layer thicknesses selected to represent the model are sufficiently thin. This is essentially an aliasing problem. If, after computing the resolution matrix, we find that a single layer is nearly perfectly resolved, we have not selected the layering thin enough in the region near this layer to determine the exact shape of the resolution. If this occurs, the partial derivatives should be recomputed for thinner layers.

The second question about parameterization is whether we have selected all the relevant parameters for inversion. All the calculations shown in the examples have been concerned with determining the shear-wave structure $\beta(z)$ of the upper mantle from Rayleigh wave observations. Rayleigh wave behavior also depends on density $\rho(z)$ and compressional-wave velocity $\alpha(z)$. Once again we appeal to the pragmatic argument of the number of degrees of freedom determined. If we find more degrees of freedom in the solution when we include variational parameters for β and ρ than we find when including β alone, the observations contain independent information about β structure and ρ structure. If we do not find more degrees of freedom, then any change in the ρ structure will be completely linearly related to changes in the β structure. For this particular example, we do not observe an increase in the number of degrees of freedom for the Rayleigh wave observations and therefore need not include ρ in the inversion. Example calculation will be given in the next section.

Monte Carlo inversion. Monte Carlo inversion is a formalized trial-and-error procedure for finding models that are consistent with a set of observations. The procedure consists of generating random models within some predetermined limits and testing the calculated values of the functionals. All models for which the calculated values are within some multiple of the standard deviations of the observed values are saved. The usual result of such a procedure is a scatter diagram of the set of passing models. I shall suggest here how such results can be related to the type of analyses that have been presented in this paper for the linear inverse problem.

Let us consider an exercise. Suppose that we have a particular solution to a linear inverse problem $\mathbf{p}_0 = [P_1 \cdots P_n]^T$ together with the eigenvectors $\mathbf{v}_i$ and the variances $\sigma^*_i{}^2 = \text{Var}\,\{\Delta P^*_i\}$. Further, let us assign large variances $\sigma^*_{0i}{}^2$ to the free vectors $\mathbf{v}_{0i}$. We now generate a set of random models $\mathbf{p}_i$ where $i = 1 \cdots l$ by adding random amounts of each of the vectors $\mathbf{v}_i$ and $\mathbf{v}_{0i}$ to the particular solution $\mathbf{p}_0$. The random coefficient for adding each vector will be constrained to have zero mean and standard deviation σ^*_i or σ^*_{0i}, respectively.

The object of the exercise is to realize that we can recover the eigenvectors $\mathbf{v}_i$ and $\mathbf{v}_{0i}$ and their standard deviations from the models. We simply form the $1 \times n$ model matrix

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1{}^T - \mathbf{p}_0{}^T \\ \vdots \\ \mathbf{p}_l{}^T - \mathbf{p}_0{}^T \end{bmatrix}$$

If the standard deviations were distinct, the eigenvalues of this matrix will also be distinct and will be proportional to the standard deviations σ^*_i and σ^*_{0i} (arranged in descending numerical size), and the row eigenvectors will be equal to the corresponding vectors $\mathbf{v}_i$ and $\mathbf{v}_{0i}$. Notice that here the best-determined vectors have the largest standard deviations as opposed to the analysis of the $\mathbf{A}$ matrix.

If for each of the models $\mathbf{p}_i$ we compute a vector of calculated values $\mathbf{c}_i = [\mathcal{F}_1(\mathbf{p}_i) \cdots \mathcal{F}_m(\mathbf{p}_i)]^T$, we could also recover the $\mathbf{u}_i$ from the row eigenvectors of the matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}_1{}^T - \mathbf{c}_0{}^T \\ \vdots \\ \mathbf{c}_l{}^T - \mathbf{c}_0{}^T \end{bmatrix}$$

Hence we see that given an adequate sampling of random models, all of which are solutions of the inverse problem, we can estimate the same quantities that were previously determined by the general linear inverse: an average solution, the standard deviation of the parameter estimates, the resolving power in parameter space, and the information distribution among the observations. In addition, we *may* be able to establish absolute bounds on the range of all parameters.

DATA ANALYSES

In this section, we shall examine the properties of a large set of observations. These data include short-period Love waves, long-period Love and Rayleigh waves, free oscillations, total mass, and moment of inertia (see the first example of the last section for a brief introduction). The perturbation parameters for the

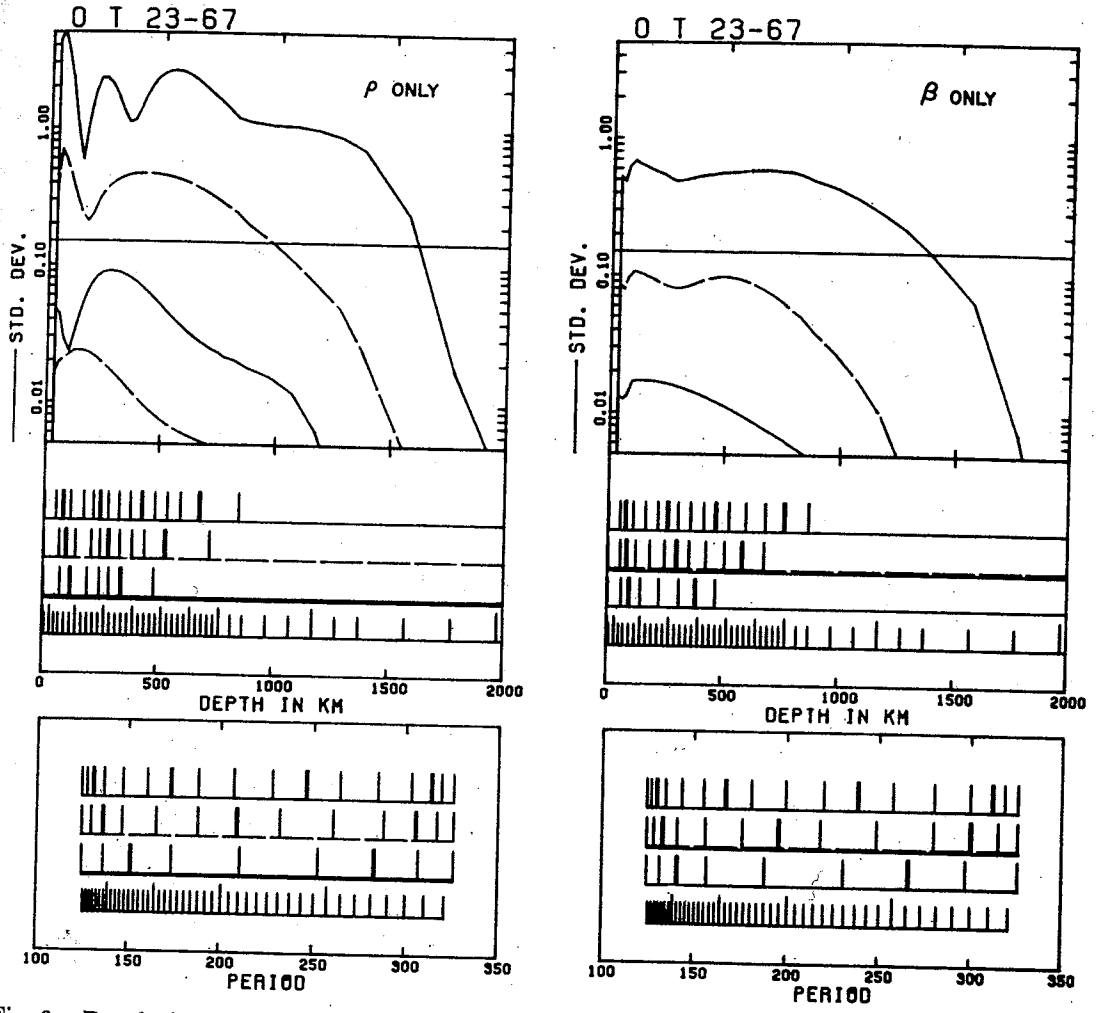

Fig. 6. Resolution diagrams for long-period Love waves (periods 125 to 350 sec) for inversions of β or ρ structure independently.

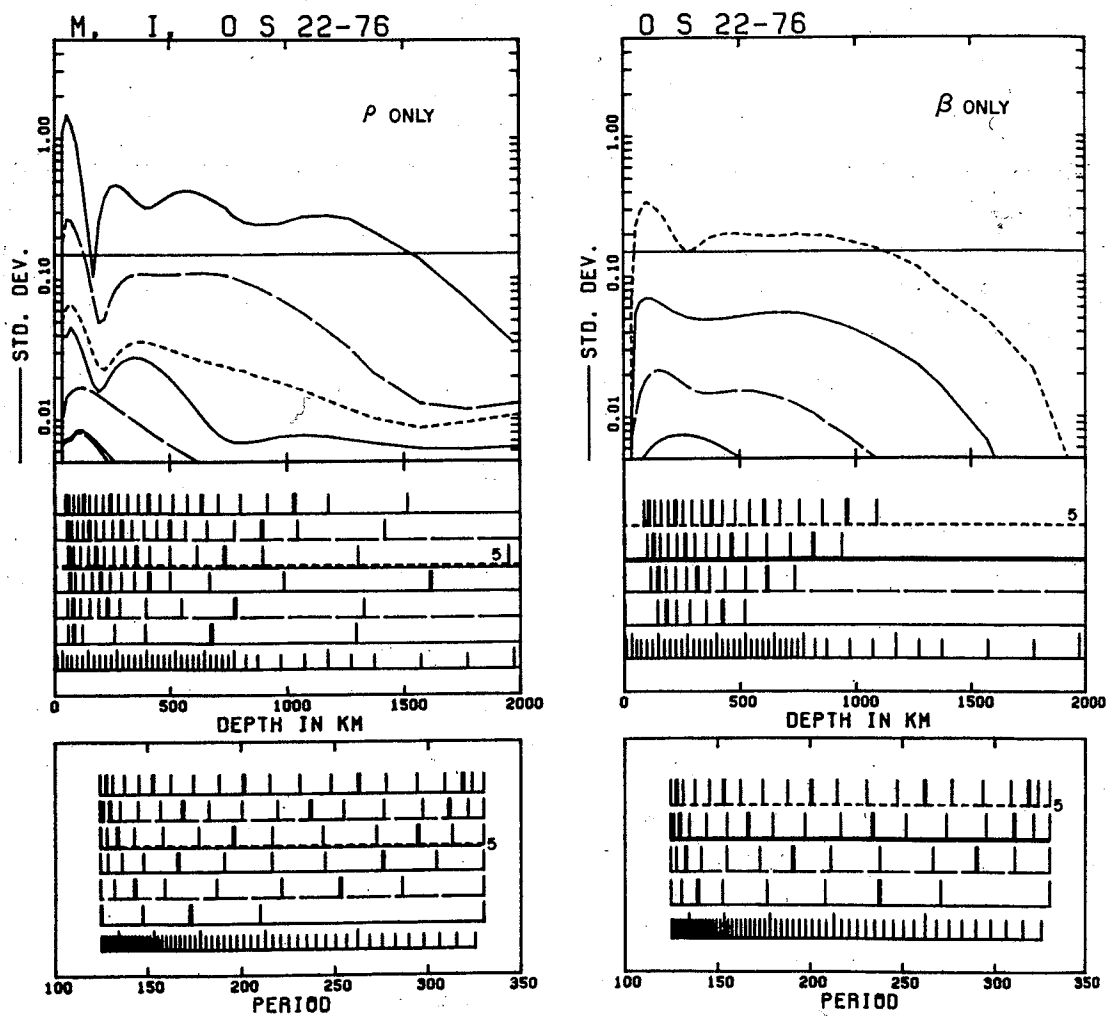

Fig. 7. Resolution diagrams for long-period Rayleigh waves (periods 125 to 300 sec) for inversion of β or ρ structure independently.

short-period Love waves are from *Anderson and Harkrider* [1968]. All the others are from *Wiggins* [1968].

My approach here is to display resolution diagrams for various subsets of the data. Almost all the information that is relevant to any particular inversion problem is contained in these diagrams. (See the last section for an explanation of the interpretation of these diagrams.)

Long-period Surface Waves

For the purpose of this paper, I define long-period surface waves to correspond to free oscillations between periods of 125 and 325 sec. These are the periods for which many recent observations of great-circle-path surface waves have been made. All the surface-wave observations are assumed to have independent standard deviations of 0.5%. The interpretations are made assuming a maximum allowable standard deviation for the parameter corrections of 0.15

km/sec for β and 0.15 g/cm^3 for ρ. If more accurate data are available, the same kind of analyses can be carried through by raising the standard deviation cutoff level in the same proportion as the accuracy is increased. All the major conclusions remain the same except the actual numbers of degrees of freedom found in each case.

We first consider long-period Love waves (Figure 6). Love waves in this period range offer very little information. If we invert β (shear-wave velocity) alone, we find 3 independent degrees of freedom or eigenvectors. If we invert ρ (density) alone, we find only 2 degrees of freedom. Further tests show that we would not be able to obtain a better fit to the observations by inverting both β and ρ at the same time.

Similar conclusions can be drawn about Rayleigh waves (Figure 7). They provide somewhat more information than do Love waves. If we invert β only, we find 4 degrees of freedom. If we invert ρ only, and include the mass M and moment of inertia I, we find 5 degrees of freedom. From the plot of information distribution, we see that only 3 of the degrees of freedom come from the Rayleigh

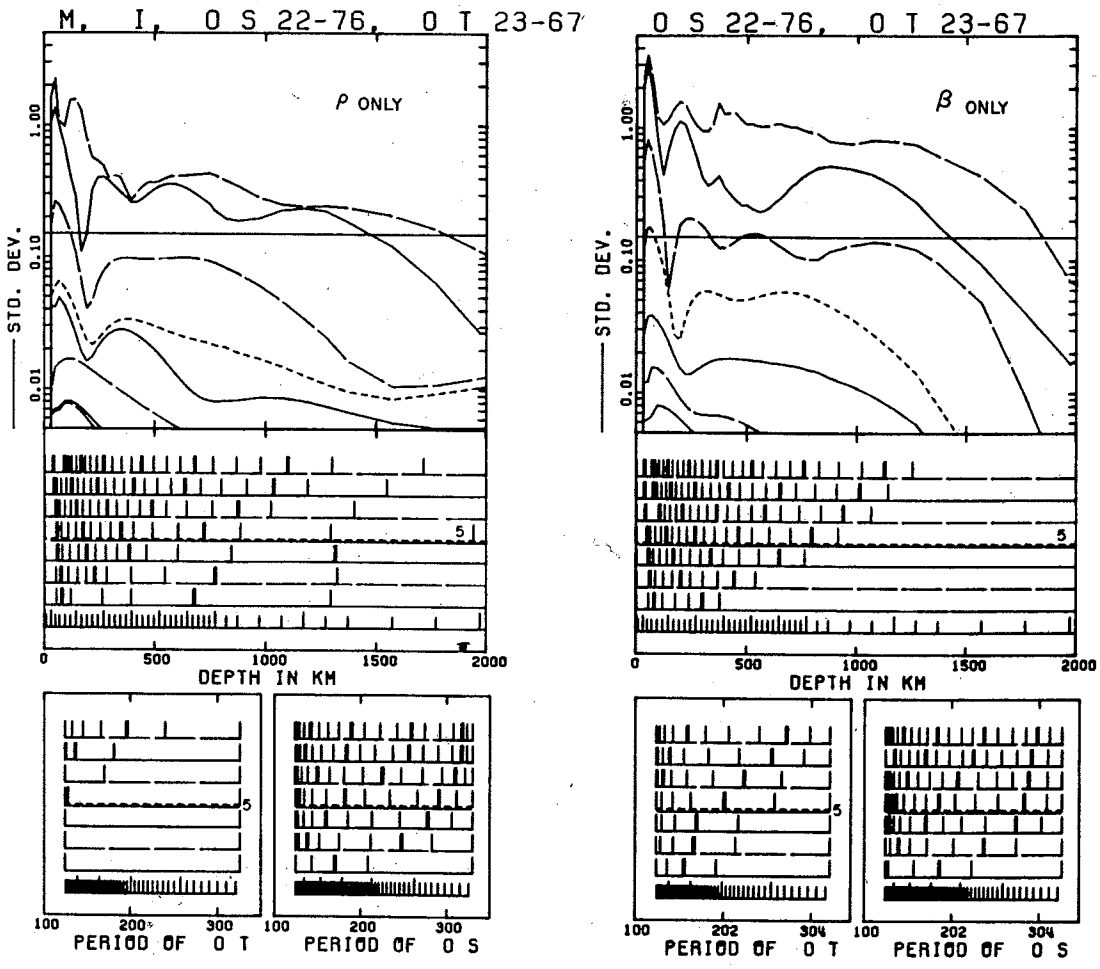

Fig. 8. Resolution diagrams for long-period Love and Rayleigh waves (periods 125 to 350 sec), mass, and moment of inertia for inversion of ρ or β structure independently.

waves. (There are only 3 heavy tick marks along the information distribution line for $k = 5$.) Again tests show that combining β and ρ in the inversion will not improve our ability to fit the observations.

The picture changes somewhat when we consider both Rayleigh and Love waves together (Figures 8 and 9). If we invert β only, we find 5 degrees of freedom. If we include M and I with the surface waves, we find 5 degrees of freedom when inverting ρ only. Figure 9 shows, however, that when we invert β and ρ at the same time, we find 8 degrees of freedom (2 of these degrees of freedom come from M and I). Thus, even though Love and Rayleigh waves individually do not offer information about both β and ρ structure, when taken together they do. This means that we may not be able to fit Rayleigh and Love wave observations simultaneously unless we vary both β and ρ in the inversion. Some investigators have in fact encountered difficulty when trying to simultaneously invert short-period Love and Rayleigh data, although this difficulty is generally ascribed to contamination of Love wave data by higher modes.

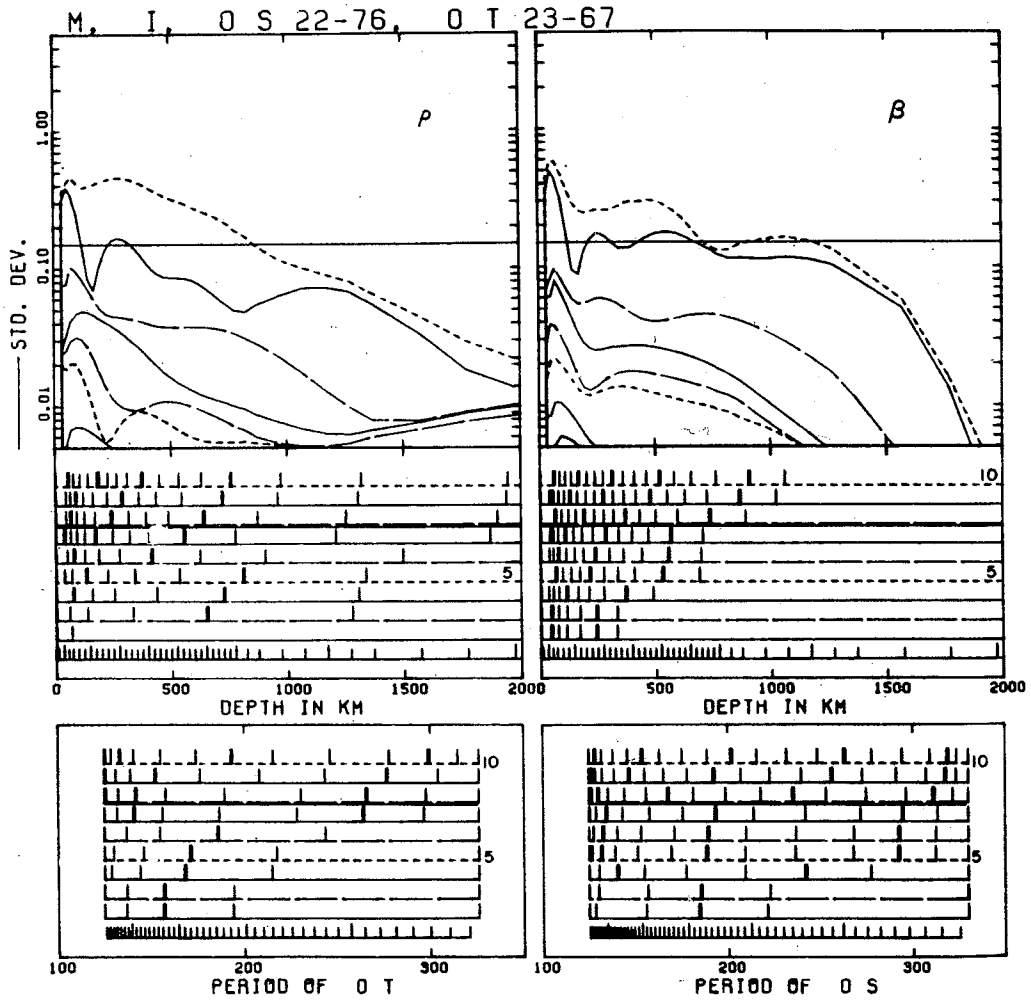

Fig. 9. Resolution diagram for long-period Love and Rayleigh waves (periods 125 to 350 sec), mass, and moment of inertia for joint inversion of β and ρ structure.

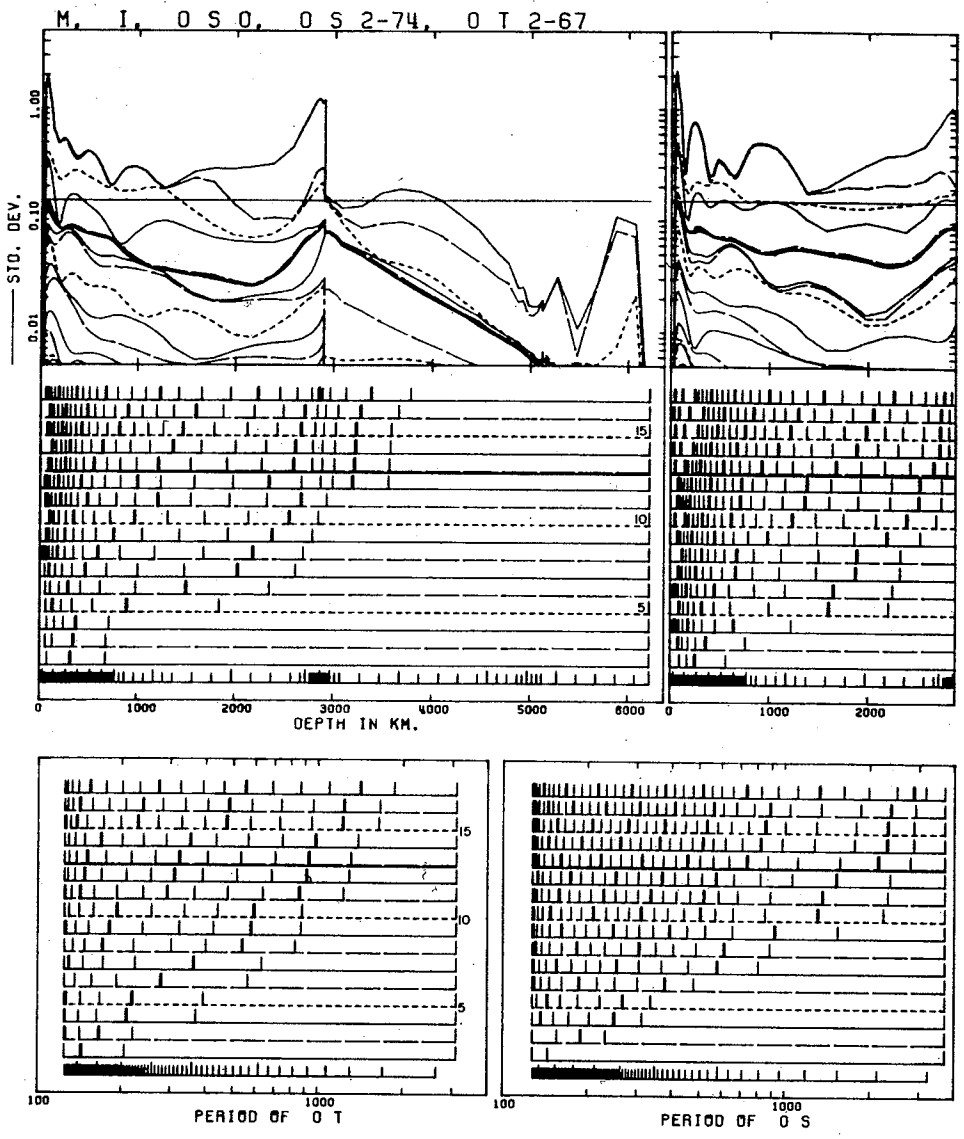

Fig. 10. Resolution diagram for mass, moment of inertia, and fundamental radial, toroidal, and spheroidal oscillations (periods greater than 125 sec) for the joint inversion of ρ (*left*) and β (*right*) structure. Dark lines indicate the case for which $\sigma_{max} < 0.15$.

Free Oscillations

We consider two cases of free-oscillation observations. The first case (Figure 10) includes only the fundamental spheroidal and toroidal modes, and the inversion is made with respect to β and ρ together. The relevant standard deviations of the observations are given in Table 1.

Figure 10 shows that for the assumed standard deviations (which, if anything, are optimistic) there are only 13 degrees of freedom. Of these, we can resolve only one density parameter in the core. Note that the greatest uncertainty and the greatest resolving power are located in the upper 100 km. There is also some higher resolution near the core-mantle boundary. This is, perhaps, a case

TABLE 1. Standard Deviations of Free-Oscillation Observations Used for Inversions Shown in Figures 10 and 11 (Per cent of observed values.)

Mode	σ
${}_0T_2$	0.50
${}_0T_{3-67}$	0.20
${}_1T_{1-15}$	0.40
${}_0S_0$	0.01
${}_{1-2}S_0$	0.05
${}_0S_{2-76}$	0.20
${}_{1-2}S_{1-15}$	0.40

Mass = 0.03; moment = 0.06.

where the standard deviations of the parameter corrections should be equalized by modifying the parameter covariance matrix **W**. Such an equalization would add 2 to 3 degrees of freedom to the solution but would not change the over-all resolution structure very much. We can tell how such an equalization would affect the resolution by examining the illustrated distributions for $k = 14$ or 15.

The standard-deviation-curve peaks in the inner core are fictitious. These peaks are caused by minor numerical errors in the computation of the partial derivatives near the center of the earth, where the motion amplitudes are very small. These instabilities appear only for cases that are well beyond achievable resolving power.

As in the case of Rayleigh and Love waves, the spheroidal oscillations have a much greater information content than do the toroidal oscillations. Beyond the information distribution shown, M, I, and ${}_0S_0$ (the radial mode) each provide independent information about earth structure.

Figure 11 illustrates the additional resolving power that can be achieved from observations of a limited number of overtones. For this case, we can resolve 17 parameter combinations, with three parameters in the core (standard-deviation equalization would raise this number to 19 or 20). As before, there are peaks in the standard deviation curves both near the surface and near the core-mantle boundary. The instability in the inner core also appears here; however, since the overtones have greater real energy penetration into the inner core than do the fundamentals, the instabilities are much less pronounced.

Notice that of the 4 additional degrees of freedom found for this case, 2 come from the overtones of the radial mode ${}_1S_0$ and ${}_2S_0$. The spheroidal overtones provide only 2 additional degrees of freedom, whereas the toroidal overtones provide practically no new information.

Short-Period Love Waves

Anderson and Harkrider [1968] have computed parameters for variations of phase velocity, group velocity, and amplitude of Love waves with respect to density and shear-wave-velocity perturbations of a model. These calculations were performed for the fundamental mode and the first three overtones in the

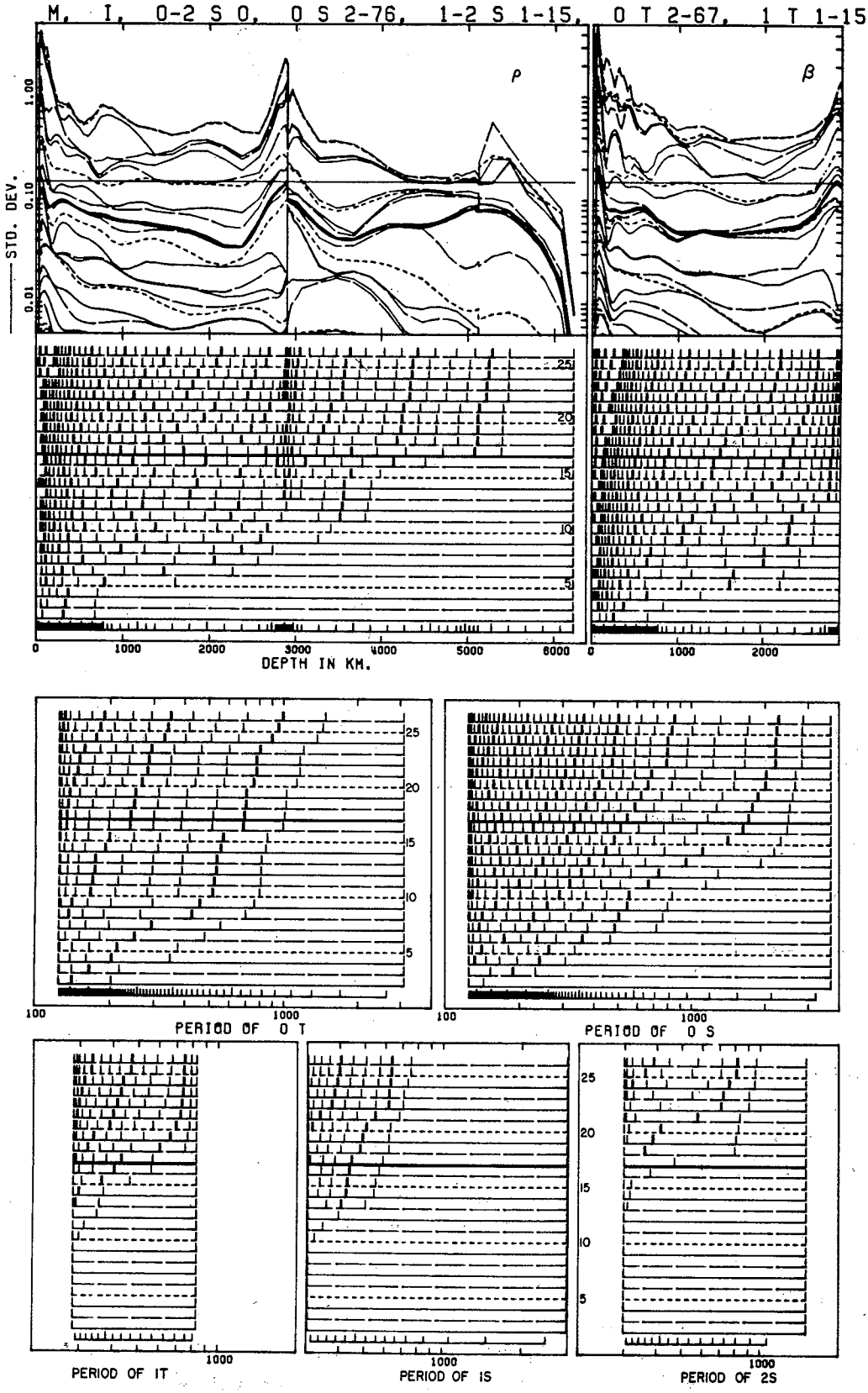

Fig. 11. Resolution diagram for the same case shown in Figure 10 with the addition of certain overtones.

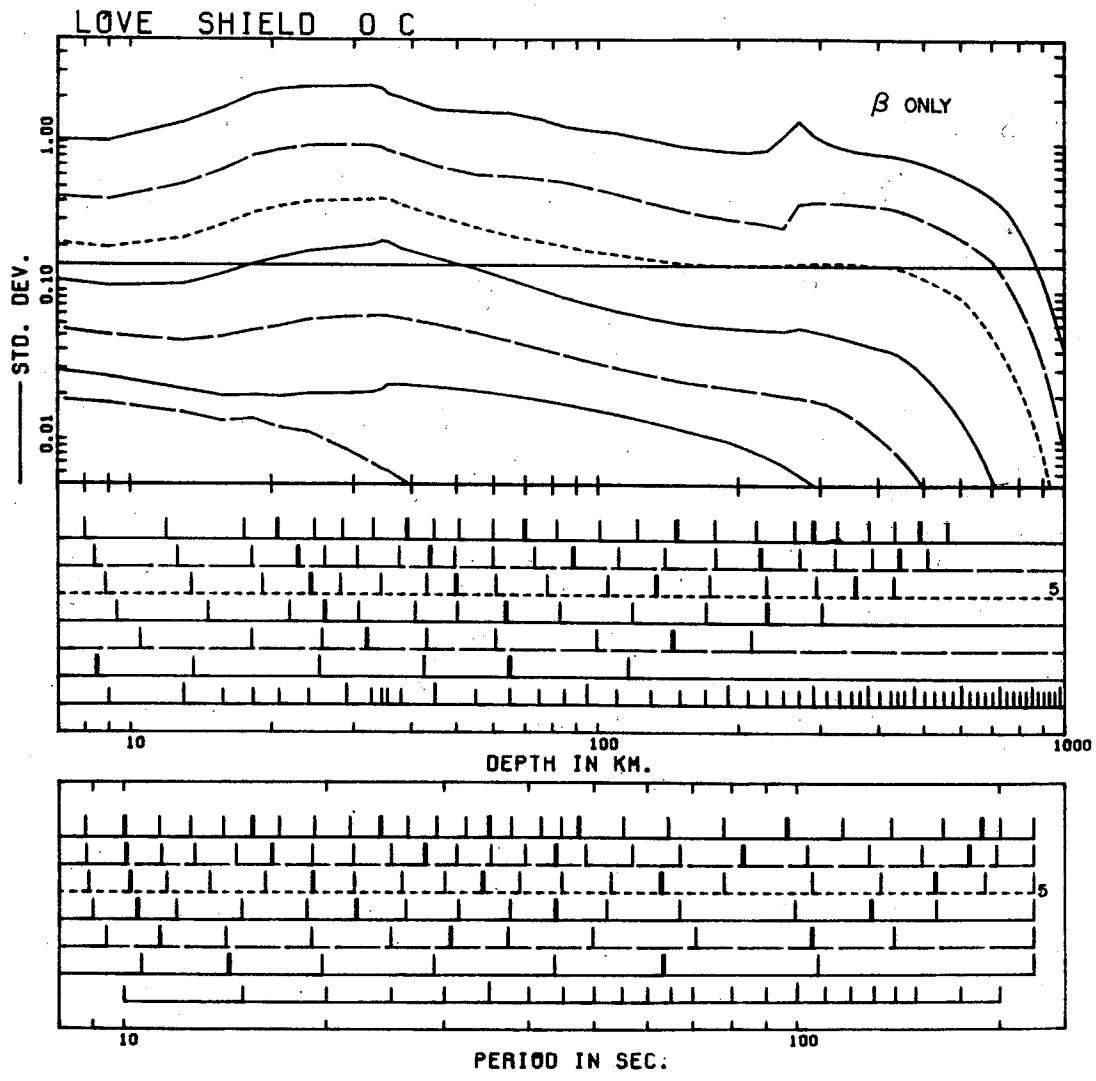

Fig. 12. Resolution diagram for fundamental short-period Love wave phase velocities for the shield model.

period range of 10 sec to 200 sec for both a 'shield' and an 'oceanic' model of the upper mantle and crust.

I shall present an analysis of a subset of these data that illustrates the principal features. Because of the insensitivity to density perturbations that was illustrated in the last section, I shall consider only perturbations of shear-wave velocities. *Der et al.* [1970] have also examined some of the properties of Love waves.

Figures 12–15 show resolution analyses for the fundamental mode. Figures 12, 14, and 15 show analyses for phase velocity C, group velocity U, and amplitude A for the shield model. Figure 13 shows an analysis of C for the oceanic model. The standard deviations assumed for the observations were 1% for C, 2% for U, and 10% for A.

Note that for the first 5 or 6 vectors, the resolution behavior is nearly

identical for the three types of observations. There are minor differences: the separation between curves is somewhat different, and C observations appear to resolve parameters that are slightly deeper than do U and A observations. These differences are not significant, however, compared to the similarities. Each type of observation provides basically the same information about the subsurface structure. Computational experiments with inverting C and U, C and A, or C, U, and A together show that the improvement in resolution comes from having more independent observations and not from basic differences in the type of information given.

The conclusion about the similarity of parameter resolution ceases to be valid if the density of observations is not greater than the information-distribution resolution. If the phase-velocity curve is not sampled sufficiently often, the phase-velocity observations do not provide enough information to establish the group-velocity behavior. Group-velocity observations at the same periods

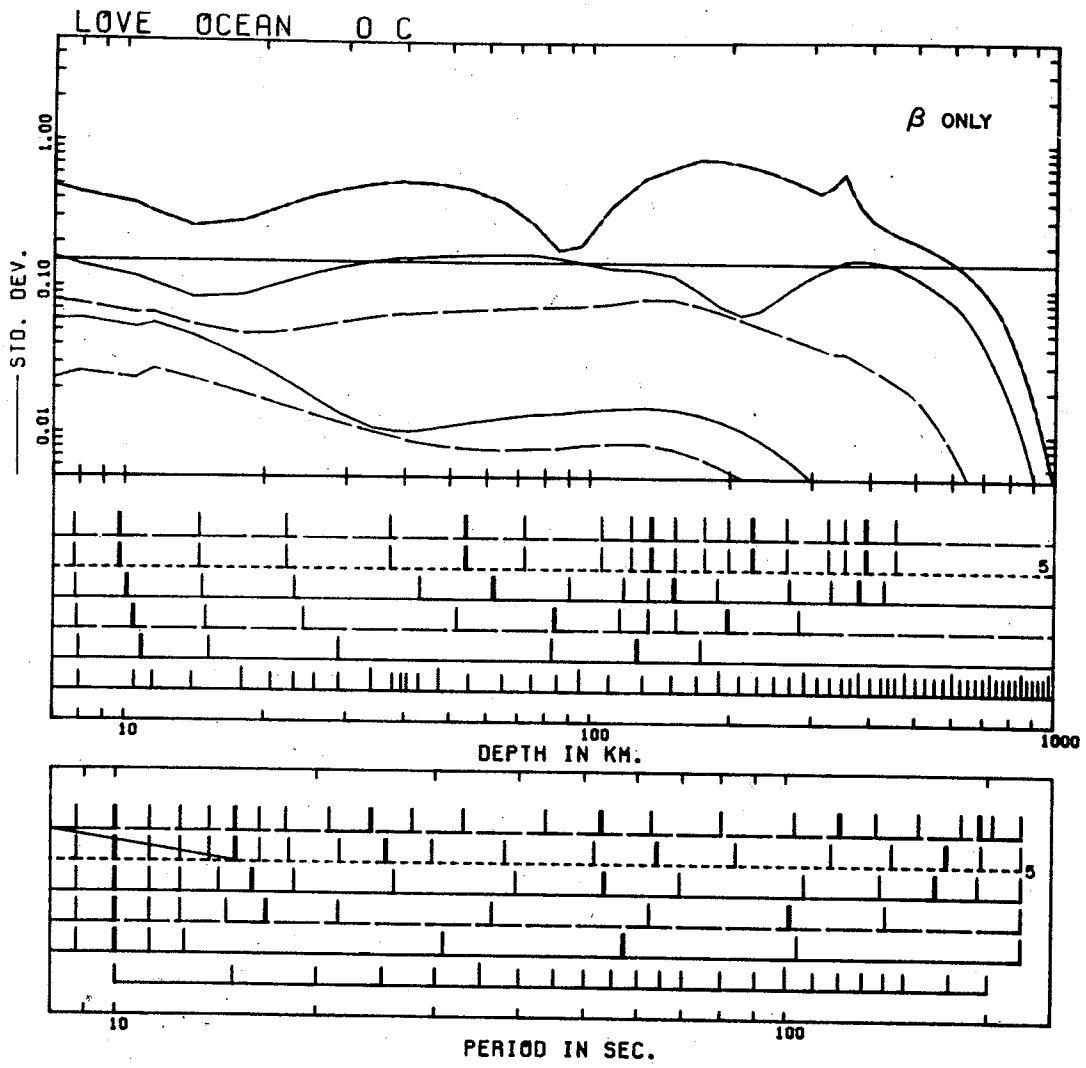

Fig. 13. Resolution diagram for fundamental short-period Love wave phase velocities for the oceanic model.

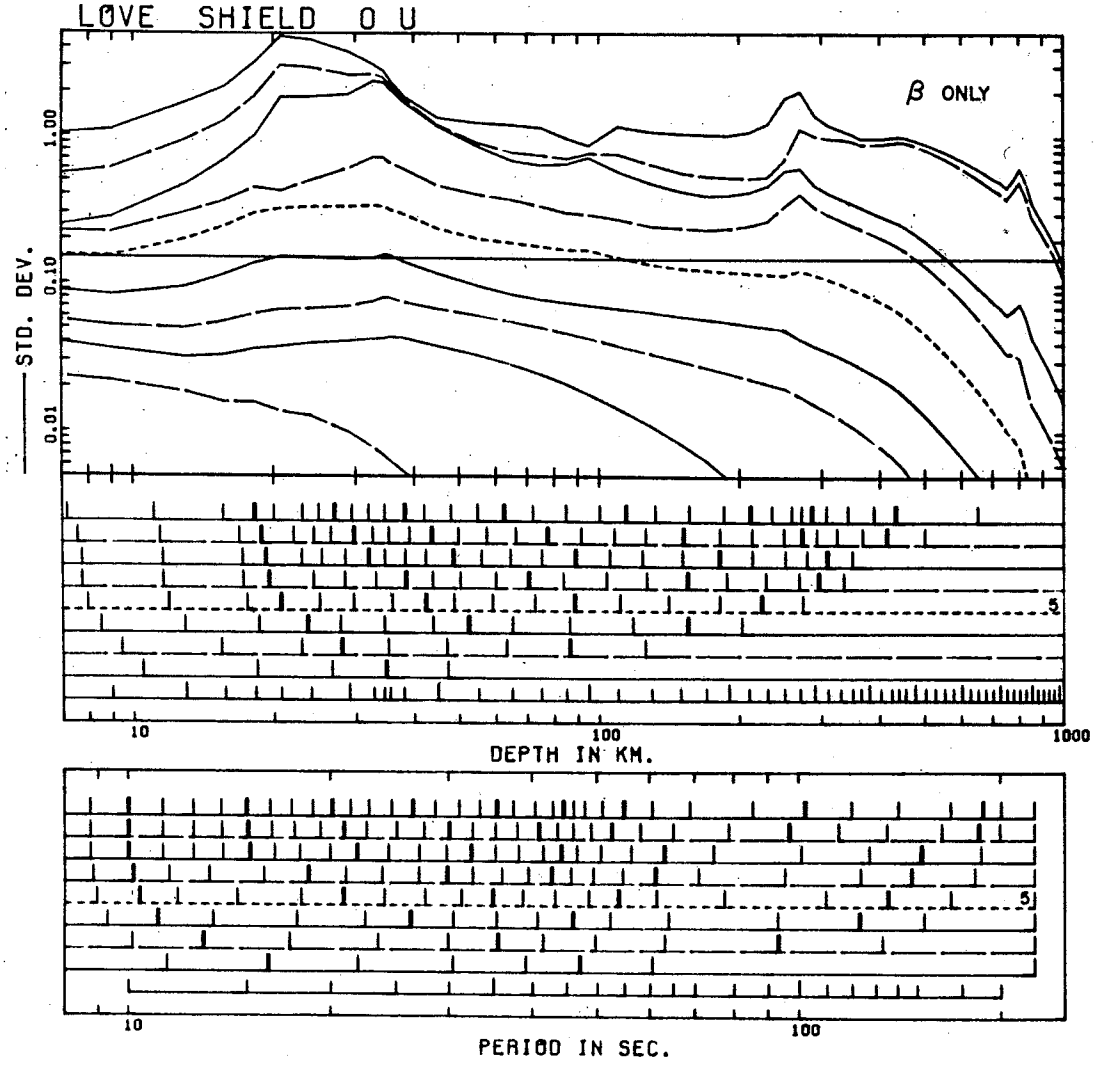

Fig. 14. Resolution diagram for fundamental short-period Love wave group velocities for the shield model.

would then exhibit different resolution properties than did the phase-velocity observations. The information-resolution diagram for U for $k = 7$, 8, and 9 indicates that the observation at 15 sec is completely independent from the observation at 10 sec. In order to properly represent the group-velocity curve in this range, we should have at least one more observation between 10 and 15 sec. Note that when the information-resolution width becomes nearly the same as the observation distribution, the shape of the standard-deviation curves change. If we had had several additional perturbation parameters between 10 and 30 sec, the shapes of the U and A standard-deviation curves would have remained nearly as parallel as those of the C curves.

Most of the kinks in the standard-deviation curves correspond to depths in the model for which there are discontinuous changes in the parameters. The small glitch at 800-km depth in Figure 14 is caused by mispunch, however.

Good resolution tends to be concentrated near the discontinuities for large

values of k. This behavior is similar to that noted for the free oscillations near the core-mantle boundary.

The dependence of resolution distribution on the specific model is well illustrated by comparing Figures 12 and 13. First, note that the oceanic phase-velocity curve was not sampled sufficiently often at short periods. This lack of sampling accounts for most of the erratic behavior of the standard-deviation curves. Even though the maximums of the standard-deviation curves are nearly the same for the two models, the resolution spread is quite different. The resolution in the oceanic model tends to be concentrated in the low-velocity zone.

Irregularities can also be seen in the information distribution. In all cases, most of the information comes from the periods between 10 and 50 sec. Periods greater than 50 sec correspond to an Airy phase, i.e., the group velocity is nearly constant. Thus, even though Airy phases are the easiest part of the phase-velocity curve to estimate, they provide relatively little information about the structure

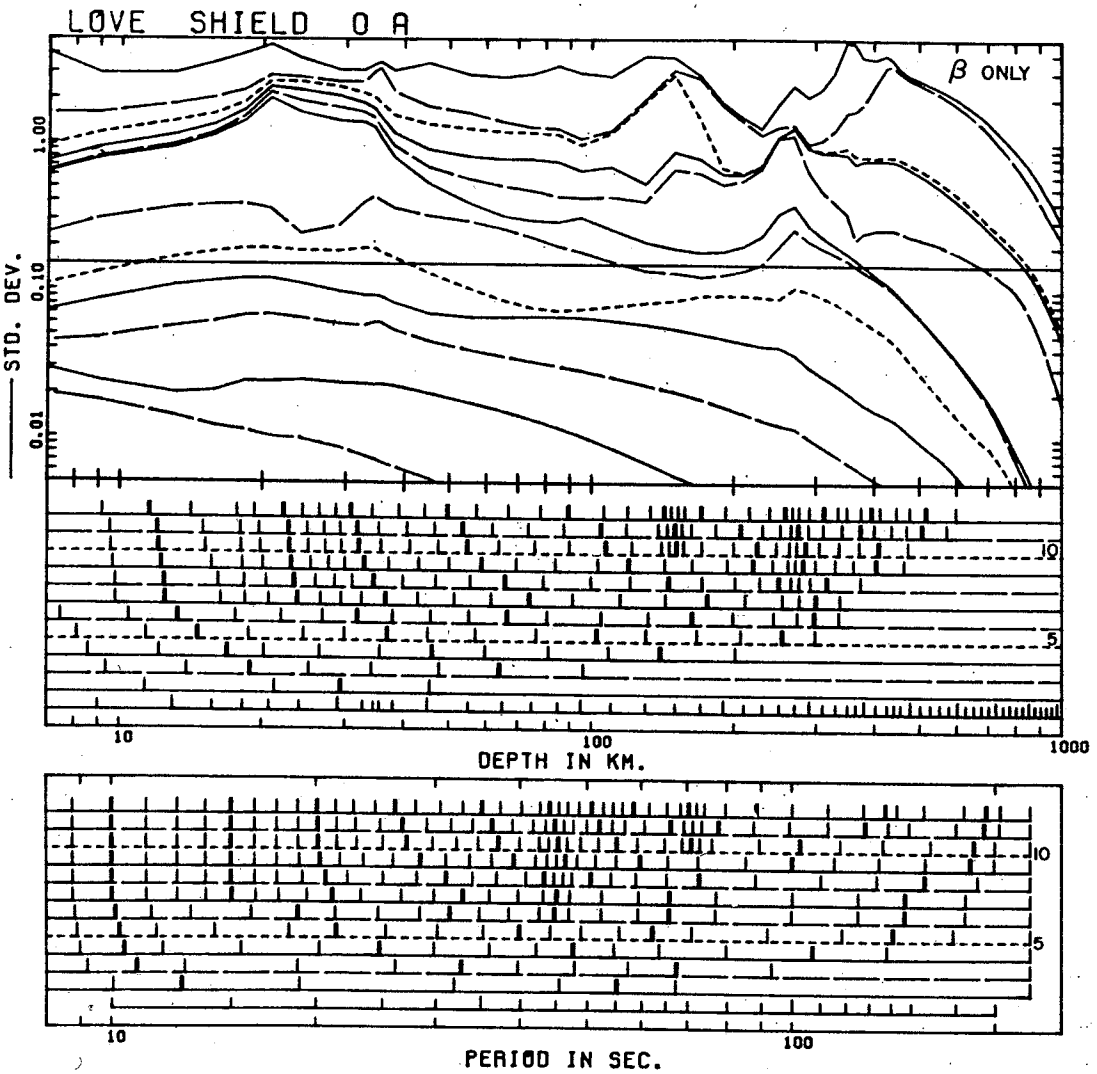

Fig. 15. Resolution diagram for fundamental short-period Love wave amplitudes for the shield model.

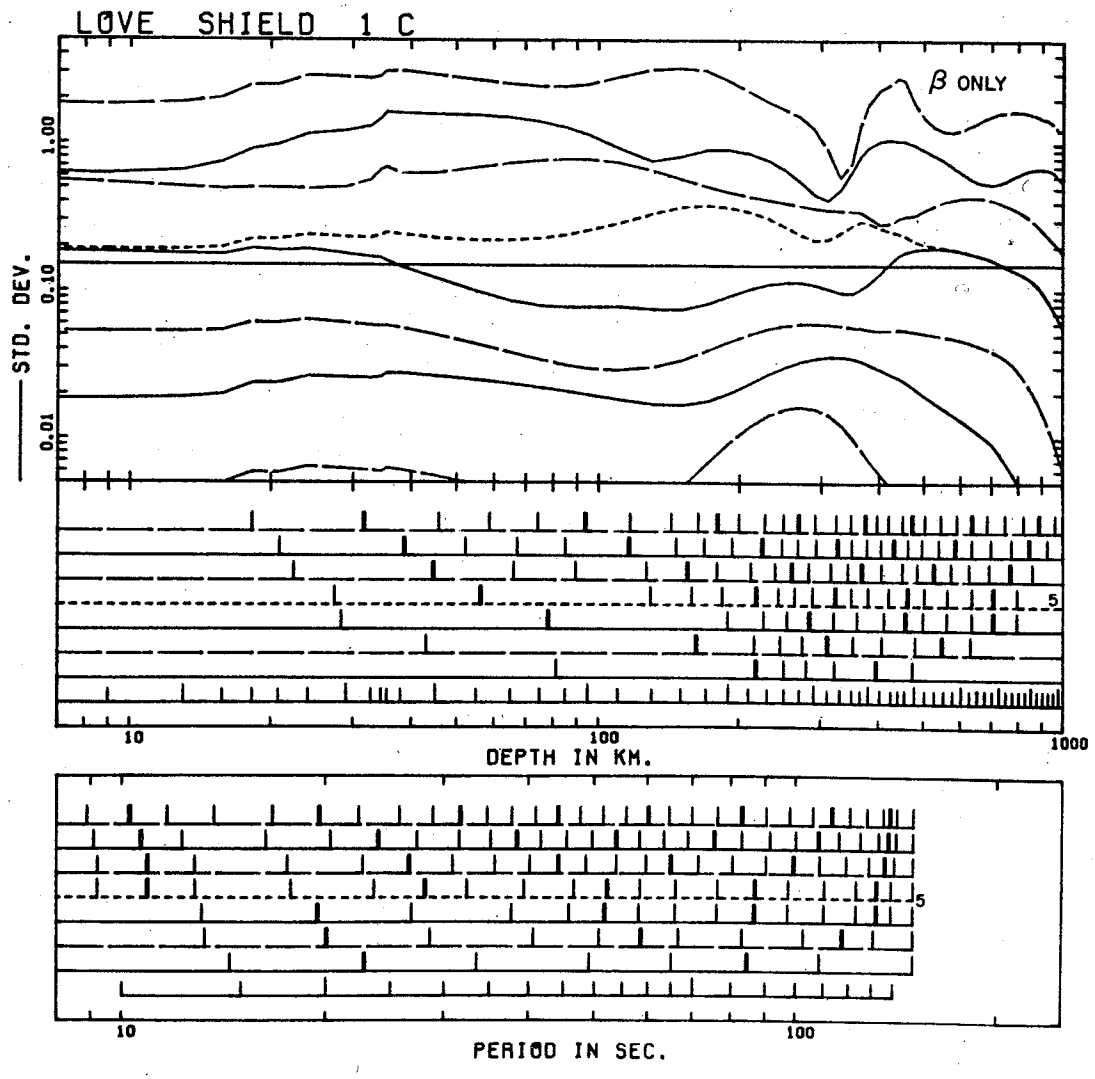

Fig. 16. Resolution diagram for first overtone short-period Love wave phase velocities for the shield model.

of the model. Alternatively, those parts of the curve that are changing very rapidly and consequently are more difficult to estimate have a more dense information content.

In contrast to the similarity of resolution information as determined from C, U, or A for the fundamental modes, the phase velocities for Love wave overtones provide distinctly different information about the model. Figure 16 shows a resolution plot for the first overtone for the shield model. Each higher overtone penetrates somewhat deeper for the same periods of observation. The lack of resolution near the surface is caused by the similarity of the displacement-function shapes for the period range sampled. Resolution tends to be concentrated at depths where the contrast between displacement shape is greatest. The overtones tend to be much more sensitive to the parameter discontinuities than was the fundamental mode.

The degree of orthogonality of information content between the fundamental

TABLE 2. Number of Degrees of Freedom k Allowed in Inversion of Shield Love Wave Phase-Velocity Observations for Various Combinations of the Fundamental Mode and Overtones

	Overtone Number						
σ_{max}	0	1	2	3	0–1	0–2	0–3
0.1	3	3	3	3–	5	7–	8
0.2	4	4	4	4–	6	9	10–
0.4	5	5	6	5	8	11–	12
0.8	5	6	7–	6	9	12	15–

Results for columns 1 and 2 can be read directly from Figures 12 and 15. A minus sign means that the standard-deviation curve rises slightly above the cutoff level.

and the overtones is outlined in Table 2, which shows the maximum number of degrees of freedom that would be obtained at various cutoff levels of the standard deviations for various combinations of modes. Each overtone adds distinctly new information at every level of observational precision.

CONCLUSIONS

In the absence of other constraints, surface-wave observations provide almost no information about the density structure of the earth unless observations for both Love and Rayleigh waves are available. When such joint interpretations are made, it is generally necessary to vary both the shear-wave velocity and the density structure independently to obtain a solution.

Phase-velocity, group-velocity, and amplitude observations provide almost exactly the same kinds of information about the structure of the earth. Since we know the analytical relationship between phase- and group-velocity curves, we can estimate a phase-velocity curve and its uncertainty on the basis of both types of observations. It would then be necessary to have only phase-velocity-perturbation parameters for inversion with no loss of generality. Similarly, if we had either an analytical or an empirical relationship between the amplitude and the phase-velocity behavior, we would not need to have amplitude partial derivatives. The fact that the resolution distribution for phase velocity and amplitude are so similar ensures that there must be some such relationship. The addition of group-velocity and amplitude estimates to phase-velocity observations serves to improve the precision of the observation. These estimates do not provide basically new kinds of information.

Overtones of Love waves or other types of observations, such as observations of Rayleigh waves, provide information about distinctly different parameter combinations, and thus the behavior of any single mode usually cannot be predicted from any of the other modes.

This brief summary of a few types of observations indicates the importance of resolution analysis in the study of inverse problems. From the simple computations described, one can gain a complete understanding of the role and interrelationship between observations and parameters. We not only gain a precise

picture of the limitations for particular inverse problems, but also obtain a better understanding of the physical behavior of the phenomena being studied. I believe that without such resolution analyses inversion studies are often misleading and always incomplete.

Acknowledgments. I am indebted to Dr. D. D. Jackson and Dr. J. Claerbout for their many helpful comments and suggestions when reviewing this paper.

The work was supported in part by NASA contract NGR-22-009-123 while the author was at MIT and in part by a National Research Council of Canada grant no. A7503 while the author was at the University of Toronto.

REFERENCES

Anderson, D. L., and D. G. Harkrider, Universal dispersion tables, 2, Variational parameters for amplitudes, phase velocity, and group velocity for first four Love modes for an oceanic and a continental earth model, *Bull. Seismol. Soc. Amer., 58,* 1407–1500, 1968.

Anderssen, R. C., The character of non-uniqueness in the conductivity modeling problem of the earth, *Pure Appl. Geophys. 80,* 238–259, 1970.

Backus, G. E., Inference from inadequate and inaccurate data, 1, *Proc. Nat. Acad. Sci., 65,* 1–7, 1970*a*.

Backus, G. E., Inference from inadequate and inaccurate data, 2, *Proc. Nat. Acad. Sci., 65,* 281–287, 1970*b*.

Backus, G. E., Inference from inadequate and inaccurate data, 3, *Proc. Nat. Acad. Sci., 67,* 282–289, 1970*c*.

Backus. G. E., and J. F. Gilbert. Numerical application of a formalism for geophysical inverse problems, *Geophys. J., 13,* 247–276, 1967.

Backus, G. E., and J. F. Gilbert, The resolving power of gross earth data, *Geophys. J., 16,* 169–205, 1968.

Backus, G. E., and J. F. Gilbert, Uniqueness in the inversion of inaccurate gross earth data, *Phil. Trans. Roy. Soc. London, Ser. A, 266,* 123–192, 1970.

Bullen, K. E., *An Introduction to the Theory of Seismology,* 3rd edition, 381 pp., Cambridge Univ. Press, London, 1965.

Der, A., R. Massé, and M. Landisman, Resolution of surface waves at intermediate distances, *J. Geophys. Res., 75,* 3399–3409, 1970.

Draper, N. R., and H. Smith, *Applied Regression Analysis,* 407 pp., John Wiley, New York, 1966.

Franklin, J. N., Well-posed stochastic extensions of ill-posed linear problems, *J. Math. Anal. Appl., 31,* 682–716, 1970.

Garland, G. D., *Introduction to Geophysics: Mantle, Core and Crust,* 420 pp., W. B. Saunders, Philadelphia, 1971.

Gilbert, F., Ranking and winnowing gross earth data for inversion and resolution, *Geophys. J., 23,* 125–128, 1971.

Johnston, J., *Econometric Methods,* 300 pp., McGraw-Hill, New York, 1963.

Jordan, T. H., and J. N. Franklin, Optimal solutions to a linear inverse problem in geophysics, *Proc. Nat. Acad. Sci., 68,* 291–293, 1971.

Kaula, W. M., *Theory of Satellite Geodesy,* 124 pp., Blaisdell, Waltham, Mass., 1966.

Keilis-Borok, V. I., and T. B. Yanovskaya, Inverse seismic problems (structural review), *Geophys. J., 13,* 223–233, 1967.

Lancsoz, C., *Linear Differential Operators,* 564 pp., D. Van Nostrand, London, 1961.

Moore, E. H., (untitled), *Bull. Amer. Math. Soc., 26,* 394–395, 1920.

Parker, R. L., The inverse problem of electrical conductivity in the mantle, *Geophys. J., 22,* 121–138, 1970.

Penrose, R., A generalized inverse for matrices, *Proc. Cambridge Phil. Soc., 51,* 406–413, 1955.

Press, F., Earth models obtained by Monte Carlo inversion, *J. Geophys. Res., 73,* 5223–5234, 1968.

Press, F., Earth models consistent with geophysical data, *Phys. Earth Planet. Interiors, 3,* 3–22, 1970*a*.

Press, F., Regionalized earth models, *J. Geophys. Res., 75,* 6575–6581, 1970*b*.

Smith, M. L., and J. F. Franklin, Geophysical application of generalized inverse theory, *J. Geophys. Res., 74,* 2783–2785, 1969.

Wiggins, R. A., Terrestrial variational tables for the periods and attenuation of the free oscillations, *Phys. Earth Planet. Interiors, 1,* 201–266, 1968.

Wiggins, R. A., Monte Carlo inversion of body-wave observations, *J. Geophys. Res., 74,* 3171–3181, 1969.

(Received May 5, 1971; revised September 24, 1971.)