66b

LBL-12203
Preprint

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

EARTH SCIENCES DIVISION

Submitted to Geophysical Prospecting

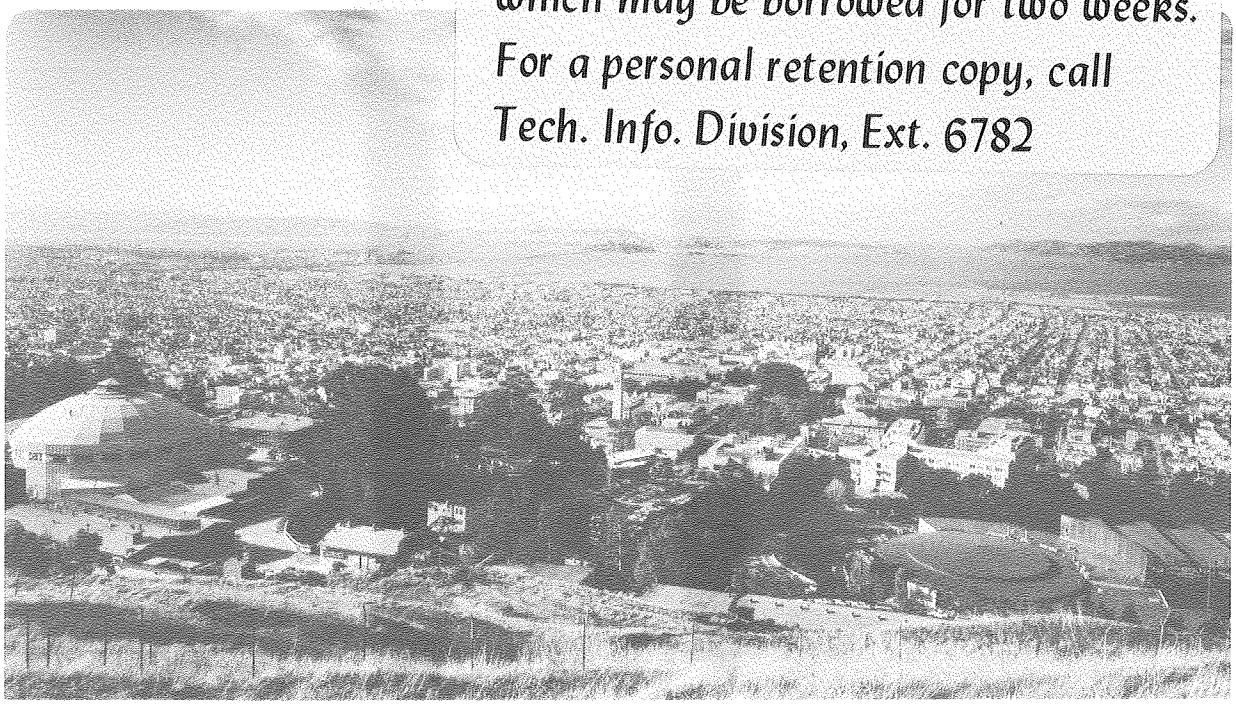
COMPARISON OF FIVE LEAST-SQUARES INVERSION TECHNIQUES IN RESISTIVITY SOUNDING

G.M. Hoversten, A. Dey, and H.F. Morrison

July 1981

TWO-WEEK LOAN COPY

This is a Library Circulating Copy which may be borrowed for two weeks. For a personal retention copy, call Tech. Info. Division, Ext. 6782



LBL-12203 c.2

Prepared for the U.S. Department of Energy under Contract W-7405-ENG-48

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

LBL-12203

COMPARISON OF FIVE LEAST-SQUARES INVERSION TECHNIQUES IN RESISTIVITY SOUNDING

G. M. Hoversten, A. Dey,* and H. F. Morrison

Earth Sciences Division
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

and

Engineering Geosciences
Material Science and Mineral Engineering
University of California
Berkeley, California 94720

August, 1981

This work was supported by the Assistant Secretary for Conservation and Renewable Energy, Office of Renewable Technology, Division of Geothermal and Hydropower Technologies of the U. S. Department of Energy under Contract No. W-7405-ENG-48.

*Now with Chevron Resources, Inc., San Francisco, California 94119

LBL-12203

COMPARISON OF FIVE LEAST-SQUARES INVERSION TECHNIQUES IN RESISTIVITY SOUNDING

G. M. Hoversten, A. Dey,* and H. F. Morrison

Earth Sciences Division
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

and

Engineering Geosciences
Material Science and Mineral Engineering
University of California
Berkeley, California 94720

August, 1981

ABSTRACT

A brief history of the development of the inverse problem in resistivity sounding is presented with the development of the equations governing the Least Squares Inverse. Five algorithms for finding the minimum of least square problem are described and their speed of convergence is compared on data from two planar earth models. Of the five algorithms studied, the ridge-regression algorithm required the fewest numbers of forward problem evaluations to reach a desired minimum.

---

*Now with Chevron Resources, Inc., San Francisco, California 94119

Solution space statistics, including 1) parameter started errors, 2) parameter correlation coefficients, 3) model parameter eigenvectors and 4) data eigenvectors are discussed. The type of weighting applied to the data affects these statistical parameters. Weighting the data by taking $\log_{10}$ of the observed and calculated values is comparable to weighting by the inverse of a constant data error. The most reliable parameter standard errors are obtained by weighting by the inverse of observed data errors. All other solution statistics, such as data-parameter eigenvector pairs, have more physical significance when inverse data error weighting is used.

INTRODUCTION

Interpretation of resistivity soundings has been a topic of research since the early 1900's. Contribution to the recent geophysical literature have dealt with the development and application of a wide variety of one-dimensional inversion techniques and approaches for estimating uncertainties in the resulting parameters. This paper presents a comparison of five least squares minimization algorithms. The comparison is made in terms of the number of forward problem evaluations required by each technique to reach a residual minimum. Three weighting schemes which affect the estimated parameter errors are compared at the minimum reached for a specific model.

Until the advent of fast digital computers, the interpreter relied primarily on curve matching procedures, where albums of theoretical curves (Compagnie Generale de Geophysique 1955, 1963, Mooney and Wetzel 1965, Flathe 1955, Orellana and Mooney 1966, Rijkswaterstaat 1969) are used alone or in conjunction with the auxiliary point method of partial curve matching (Kalenov 1957, Orellana and Mooney 1966, Zohdy 1965). This method, while undoubtedly the most convenient and simple, suffers from the drawback that the published curves cover only a limited number of cases.

HISTORICAL DEVELOPMENT OF INVERSE METHODS

The forward problem expressed in terms of integral expression for potential and apparent resistivity for the Schlumberger electrode array was developed by Stefanesco, Schlumberger and Schlumberger (1930). With

these and the associated kernel function, development began on interpretation based on determining earth parameters by fitting a theoretical apparent resisitivity curve or kernel function to its observed counterpart. Similar expressions for apparent resistivity have been developed for many other electrode arrays (Roy and Apparao 1971, Alpin 1966, Keller and Frischknecht 1966). This general approach remains a very popular method of inversion today. Another popular inversion method, which will not be considered here, involves the use of Dar Zarrouk parameters. This is well described by Zohdy (1965, 1968, 1974a, 1974b, 1974c).

Stefanesco <u>et al</u>. (1930) derived the integral expressions

$$V(r) = \frac{\rho_1 I}{2\pi}\left[\frac{1}{r} + 2\int_0^\infty \theta(\lambda, k, t)\, J_0(\lambda r)\, d\lambda\right] , \qquad (1)$$

$$\rho_s(r) = \rho_1\left[1 + 2r^2 \int_0^\infty \theta(\lambda, k, t)\, J_1(\lambda r)\, d\lambda\right] ,$$

I = current applied to the earth

r = current electrode spacing

$V(r)$ = the electrode potential at $r = AB/2$

$\rho_s(r)$ = Schlumberger apparent resistivity

ρ_1 = top layer resistivity

J_0, J_1 = zero and first order Bessel functions, respectively

$\theta(\lambda,k,t)$ = Stefanesco kernel function

λ = integration variable

k = resistivity reflection coefficient

t = layer thickness

The Slichter kernel (Vozoff 1958), is defined by

$$T(\lambda,k,t) = \rho_1[1 + 2\theta(\lambda,k,t)] \quad . \tag{3}$$

This kernel can also be related to apparent resistivity through the inverse Hankel transformation of (2);

$$T(\lambda,k,t) = \int_0^\infty \frac{\rho_s(r)}{r} J_1(\lambda r)\, dr \quad . \tag{4}$$

The Slichter kernel can also be expressed in a closed nonintegrable recursive form for an arbitrary number of n layers of thickness t_i (Sunde 1949), where,

$$T_{1,2\ldots n} = \frac{1 - \mu_{1,2\ldots n} e^{-2\lambda t_1}}{1 + \mu_{1,2\ldots n} e^{-2\lambda t_1}} \quad ,$$

where

$$\mu_{1,2\ldots n} = \frac{\rho_1 - \rho_2 \; T_{2,3\ldots n}}{\rho_1 + \rho_2 \; T_{2,3\ldots n}} \quad ,$$

$$T_{(m-1)m\ldots m} = \frac{1 - \mu_{(m-1)m\ldots n} e^{-2\lambda t_{m-1}}}{1 + \mu_{(m-1)m\ldots n} e^{-2\lambda t_{m-1}}} \quad ,$$

$$\mu_{(m-1)m\ldots n} = \frac{\rho_{m-1} - \rho_m T_{m(m-1)\ldots n}}{\rho_{m-1} + \rho_m T_{m(m-1)\ldots n}} \quad , \tag{5}$$

and

$$\mu_{(m-1)m\ldots n} = \frac{\rho_{n-1} - \rho_n}{\rho_{n-1} + \rho_n}$$

The first general approach mentioned above consists of an algorithm that uses (2) or some similar expression to calculate the forward problem and then to vary model parameters until the calculated apparent resistivity, ρ_A matches the observed data to some specified tolerance. Alternatively, the kernel function is derived from the apparent resistivity data via (4), and (5) is used as the forward problem in a least squares method. This latter approach has the advantage that (5) is much faster to calculate than the forward problem represented by (2).

The inverse problem was first approached in the kernel domain by Slichter (1933) using a solution for surface potentials developed by Langer (1933) for a one dimensional resistivity function which varied continuously with depth. Slichter's procedure was to determine the kernel function from the apparent resistivity and then solve for the conductivity profile from the kernel function. Stevenson (1934) modified the approach to accommodate a step-wise resisitivity function of depth. A partially graphic, partially numerical method was developed by Pekeris (1940) which also used the Slichter kernel. The ease with which the kernel function can be calculated encouraged many workers to concentrate on interpretation in the kernel domain. Slichter used a power series representation for the kernel and later recursion formulae

(i.e., equation (5)) were developed (Sunde 1949, Flathe 1955, Vanyan, Morozova and Lozhemitina 1962, Kunetz 1966, Meinardus 1967).

More recently, Ghosh (1971) developed a set of coefficients for transforming Schlumberger and Wenner sounding curves into their corresponding Slichter kernels, using linear filter theory rather than numerical integration. Koefoed (1965, 1966, 1968) used the raised kernel $H(\lambda)$. Many other authors have also done recent work concerning inversion in the kernel domain. These include Crous (1971), Meinardus (1967, 1970), Onodera (1960), Pekeris (1940), Vozoff (1958), and Ginzberg, Loewenthal, Shoham (1976).

Whether inversion is performed with kernel functions or measured data, fitting the calculated to the observed data is usually carried out in a least-squares sense. The least-squares fitting is in turn performed on a linearized version of the governing equations.

The most common method for generating this system of equations is to expand the calculated functions, such as (2) and (5) in a Taylor series about an initial estimate P^0 in the parameter space. The series, neglecting second and higher order terms, is

$$O(X,P)_i = C(x,P^0)_i + \sum_{j=1}^{N} \frac{\partial}{\partial P_j} [C(x,P^0)_i] (P_j - P^0{}_j) \quad , \qquad (6)$$

with $j = 1, M$ and $i = 1, N$,

where

$O(X,P)_i$ = the ith observation or kernel as a function of X, the system parameter vector

$C(X,P^0)_i$ = the ith calculated function value, (i.e., (2) or (5)), for X and P^0 the current estimate of the unknown parameter set.

$(P_j - P_j^0)$ = the linear estimate of the correction needed in the jth unknown parameter.

Rewriting (6) in matrix notation, neglecting second and high order terms,

$$\Delta G = A \Delta P \quad , \tag{7}$$

where

$$\Delta G_i = O(x,P)_i - C(x,P^0)_i$$

$$A_{ij} = \left. \frac{\partial C(x,P)_i}{\partial P_j} \right|_{\substack{x = x_i \\ p = p^0}} \quad \text{(The System Matrix)} \quad . \tag{8}$$

$$\Delta P_j = P_j - P^0{}_j$$

Equation (7) represents a system of N linear equation in M unknowns and since this is a non-linear problem, the solution of (7) will provide only a linear estimate of ΔP, the correction needed to reach the minimum. The classic least-squares inverse is

$$\Delta P = (A^T A)^{-1} A^T \Delta G \quad . \tag{9}$$

The solution must be iterated, beginning each time from the current estimate of the parameter set, P^0.

Vozoff (1958) calculated the kernel by inverse Hankel transformation and then used least-squares minimization of

$$\sum_{i=1}^{N} [K_{1,2,\ldots n}(\lambda_1) - K_f(\lambda_i)] \equiv \Delta G \quad , \tag{10}$$

where

$K_{1,2,\ldots n}$ = kernel calculated for a particular parameter set

K_f = kernel integrated from field data

Vozoff (1958) considered two methods based on gradients of the system matrix to minimize (10). The first was a functional iteration method and the second, a steepest descent method. The functional iteration procedure follows Newton's technique for finding roots of a non-linear equation of one unknown parameter. The procedure is discussed by Hildebrand (1949). It consists essentially of an iterative Taylor series technique. An initial estimate is made for the values of the

parameters, assuming that the estimate is quite close to the true values. The difference or the objective function ΔG is approximated by the first two terms of its Taylor series expansion in terms of the unknown parameters. The expansion is then used to calculate the necessary changes in the initial estimate. The procedure is iterated until no further changes are needed.

The steepest descent method deals directly with the difference function ΔG. If one maps ΔG in its m-parameter space, an m-dimensional least-squares surface is generated. Any choice of a parameter vector not at the minimum gives a point on the least-squares surface where its m-dimensional gradient is calculated. The parameter vector is changed so that the value of ΔG descends along the steepest initial gradient. Because the surface is not truly described by its first order gradient alone, the initial direction will probably not extend to the true minimum. Therefore the parameter vector is changed until the minimum of the least-squares surface ΔG is found in that particular direction. At this point a new gradient is calculated and a new descent is made in the direction of steepest gradient. This procedure is iterated until a point of zero gradient is reached or until ΔG becomes less than some predetermined value. These two methods of minimizing a function of sums of squares are forerunners of newer and more advanced methods which will be dealt with in this paper. The general format of a least-squares approach is now widely used in both the kernel domain and the apparent resistivity domain since the speed of new large computers make this time

consuming method more feasible and, more importantly, because it lends itself well to statistical evaluation of the resulting parameters.

Many modern least-squares techniques are centered on the method of the generalized inverse. The generalized inverse has been thoroughly discussed for the inversion of surface wave and free oscillations data by Wiggins (1972), and by Jackson (1972). Jupp and Vozoff (1975), Vozoff and Jupp (1975), Inman, Ryu and Ward (1973) found that in many resistivity problems, when model parameters are strongly correlated, the system matrix (8) was nearly singular. The nonorthogonality this caused was not satisfactorily dealt with by the generalized inverse. Hoerl and Kennard (1970a and 1970b) developed a theory which shows that linear estimation from nonorthogonal data could be improved by the use of biased estimators. Hoerl gave the name "ridge regression" to this method. Marquardt (1970) discussed the relationship between generalized inverse and ridge regression and summarized his finding in these words:

> The Ridge and Generalized Inverse estimators share many properties. Both are superior to least-squares for ill conditioned problems. The generalized inverse solution is especially relevant for precisely zero eigenvalues (in the system matrix). The ridge solution is computationally simpler and seems better suited to coping with very small, but non-zero, eigenvalues.

Most practical resistivity problems, owing to the many highly correlated parameters that can result, involve small but non-zero eigenvalues. For this reason Inman (1975) used ridge regression, with good results, for the inversion of resistivity data. The ridge

regression approach has also been used recently by Glenn and Ward (1976), Rijo, Palton, Feitosa and Ward (1977) and Petrick, Pelton and Ward (1977).

FIVE LEAST-SQUARES MINIMIZATION ALGORITHMS

The following descriptions of the algorithms closely follow the published literature. Some notation has been changed from the original publications for consistency within this paper.

1) The Simplex Method

The method is designed for the minimization of a function of M variables without constraints.

Let Q_0, Q_1, Q_m be the (m+1) points in the m-dimensional parameter space which define the current least-squares surface, or simplex. Each point Q_k has its own parameter vector of length m associated with it. Let ΔG_k be the function value at Q_k and define

h as the suffix such that $\Delta G_h = \max(\Delta G_k)$

ℓ as the suffix such that $\Delta G_\ell = \min(\Delta G_k)$

Next we define Q as the centroid of the points with $k \neq h$, (Q_1, Q_j) represents the distance from P_i to P_j. At each stage in the process one of three operations -- reflection, contraction or expansion -- is used to replace Q_h with a new point in the simplex. These true possible replacement points are defined as follows: the reflection of

Q_h is denoted Q^* and its co-ordinates are defined by the relation

$$Q^* = (1+\alpha)\,\overline{Q} - \alpha Q_h \quad ,$$

where α, the reflection coefficient, is an arbitrary positive constant. This expression indicates that Q^* is $\overline{Q}$ on a line joining $\overline{Q_h}$ and Q on the far side of $\overline{Q}$ from $\overline{Q_h}$ with $(Q^*,\overline{Q}) = \alpha\,(Q_h\overline{Q})$. If Q^* which corresponds to ΔG^* falls between ΔG_h and ΔG_ℓ, the Q_h is replaced by Q^* and the next iteration proceeds with the newly defined simplex.

If the reflection produced a new minimum, $\Delta G^* < G_\ell$, then a step is taken to try and find a further minimum. The expanded point Q^{**} is given by the relation

$$Q^{**} = Q^* = (1-\gamma)\overline{Q} \quad ,$$

where the expansion coefficient, γ, is the ratio of the distance $(Q^*, {}^*\overline{Q})$ to $(Q, {}^*\overline{Q})$. If $\Delta G^* < G_{\ell\rho}$, Q_h is replaced by Q^{**}, and the process begins again from the start. However, if $\Delta G^{**} > \Delta G_\ell$ it is called a "failed expansion," and Q_h is replaced by Q^* before restarting.

The third operation of contraction is used if, on reflecting Q to Q^*, the condition $\Delta G^* > \Delta G_\ell$ exist for all $k \neq h$, i.e., that replacing Q_h by Q^* leaves ΔG^* the maximum, then a new Q_h is defined to be either the old Q_h or Q^*, whichever has the lower ΔG value. The contracted point is defined by

$$Q^{***} = \beta Q_h = (1-\beta)\overline{Q} \quad .$$

The contraction coefficient β lies between 0 and 1 and is defined by the ratio of distance $[Q^{***},\overline{Q}]$ to $[Q_h,\overline{Q}]$. Q^{***} is used in place in Q_h unless $\Delta G^{***} > \min(\Delta G_h, \Delta G^*)$, i.e., the contracted point Q^{***} is worse than the better of Q_h and Q^*. In this case of a "failed contraction", all the Q_k's are replaced by $(Q_k + Q_\ell)/2$ and the process is restarted.

The iteration process is set to run until ΔG reaches or surpasses some preset minimum value. The application of the simplex method to the minimizing of a function of many variables is discussed by Nelder and Mead (1965).

2) <u>Uncontrained Global Optimization</u> (Bremermann)

This routine was initially developed for the solution of systems of non-linear equations of up to 100 variables (Bremermann, 1970). The routine has also been used for finding maxima or minima of sums of squares, sums of exponentials, and curve fitting. The method is briefly described as follows:

a) ΔG is evaluated for the initial estimate of the parameter set, p^0

b) A random direction r is chosen. The probability distribution of the r is an N dimensional Gaussian with $\sigma_1=\sigma_2=..=\sigma_n=1$

c) On the line determined by P^0 and r the restriction of ΔG to this line is approximated by five point Lagrangian interpolation, centered at P^0 and equidistant with distance h, which is a parameter of the method.

d) The Lagrangian interpolation of the restriction of ΔG is a fourth degree polynomial in parameter λ describing the line $P^0 + \lambda r$. The five coefficients of the Langrangian interpolation polynomial are determined.

e) The derivative of the interpolation polynomial is a third degree polynomial with one or three real roots. The roots are computed by Cordan's formula.

f) If there is one root λ_0 the procedure is iterated from the point with a new random direction provided that $\Delta G(P^0 + \lambda_0 r) \leq \Delta G(P^0)$. If this inequality does not hold, the method is iterated from P^0 with a new random direction.

g) When there are three real roots, $\lambda_1, \lambda_2, \lambda_3$, then the polynomial is elevated at $P^0 + \lambda_1 r$, $P^0 + \lambda_2 r$ and $P^0 + \lambda_3 r$. Also, considering the value of P^0, the procedure is iterated from the point where ΔG has the smallest value. If ΔG has a minimum value at more than one point, the algorithm chooses one of them.

h) Iteration continues until a predetermined number of interactions has been run or until a prescribed minimum has been reached.

3) Peckham's method

This method, Peckham (1970), was specifically designed for minimizing a sum of squares of non-linear functions without calculating the gradients which make up the matrix A.

The work of Spendley, Hext and Himsworth (1962), Nelder and Mead (1965) in developing methods for minimizing functions in which the function is evaluated at (N+1), or more, points forming a simplex in N-dimensional space suggested to Peckham, that, for problems where the function is a sum of squares, the function values at a set of (N+1), or more, points might be used to estimate values for the coefficients C_i and A_{ij} in (8). These could then be used in (9) for a linear estimate of the minimum. One iteration consists of replacing the point of the set with the highest function value by the linear estimate of the minimum position.

Assume that there are function values $O_{k\ell}$ for a set of p points with parameters $P_{j\ell}$ where $p \geq N+1$ and $\ell = 1,2,\ldots p$. Now consider the minimization at each of the points in the M dimensional hyperspace. The linear approximation is obtained by picking C and A to minimize the N expressions.

$$\sum_{\ell=1}^{P} \omega_\ell^2 \left(C_i + \sum_{j=1}^{M} A_{ij}P_{j\ell} - O_{k\ell}\right)^2 , \qquad (11)$$

where $i = 1,2\ldots N$ and ω_ℓ is a weighting factor to determine the relative importance of each point in the simplex. The weighted mean is chosen as the origin for P_j so that

$$\sum_{\rho=1}^{P} \omega_\ell^2 P_{j\ell} = 0 \quad .$$

With the change of variables $P'_{j\ell} = \omega_\ell P_{j\ell}$ and $O'_{k\ell} = \omega_\ell O_{k\ell}$ the values of C and A which minimize ΔG are given by

$$P'P^T A^T = P'O'^T \quad ,$$

where

$$\Omega = \sum_{\ell=1} \omega_\ell^2 \quad \text{and} \quad C = \frac{1}{\Omega} O'\omega \quad .$$

If these values are substituted into (9) the linear estimation of the parameter set at the minimum, is

$$P_E = \frac{-1}{\Omega} (P'P'^T)(P'O'^T)(P'O'^T O'\omega) \quad . \tag{12}$$

In order to solve equation (12) it is rewritten as

$$P_E = \frac{-1}{\Omega} P'P'^T Z \quad , \tag{13}$$

where

$$(P'O'^T O'P'^T)Z = P'O'^T O'\omega \quad . \tag{14}$$

Equation (13) is the normal equation of a linear least-square problem, i.e., the Euclidean norm $\|O'_\omega - O'P'^T Z\|$ has a minimum when Z satisfies (14).

The solution is obtained by use of orthogonal transformation, Golub (1965). Peckman used the ALGOL procedure "Orthlin 2" for his solution. In addition the values of ω_ℓ were chosen to give function values near the minimum more weight in determining C and A

$$\omega_\ell = \frac{1}{\Delta G_\ell} \quad .$$

4) Ridge Regression

Rather than using function values in a M dimension hyperspace to estimate values for C_i and A_{ij}, it is very popular to use direct forward problem calculation for C_i and finite differencing of these values to calculate A_{ij}. The simplest approach is to solve (7), neglecting weighting, for ΔP by calculating the least-squares inverse $(A^T\underline{A})^{-1}A^T$. However, Hoerl and Kennard (1970a) show that when (A^TA) is nearly singular, as it can be in many geophysical problems, the average difference between ΔP estimated and ΔP true becomes very large. The ridge regression method (Levenberg 1944; Foster 1961; Marquardt 1963, 1966) seeks to reduce this difference during the iteration process by damping the diagonal terms of (A^TA). The ridge regression estimate of ΔP_{RR} is

$$\Delta P_{RR} = (A^TA + KI)^{-1} A^T\Delta G \quad , \tag{15}$$

where I is the density matrix and $K \geq 0$. According to Inman (1975), "The eigenvalues of $(A^TA + KI)$ are $(\lambda_j^2 + K)$, where λ_j^2 are the eigenvalues of A^TA. Any very small eigenvalues of the least-squares estimator will be increased in the ridge regression

estimator by a factor K. Hence the inversion of the matrix $(A^TA + KI)$ will be more stable. Increasing the size of all the eigenvalues results in a significant decrease of, a) the mean of the squared length between ΔP true and ΔP_{RR} and b) the variance of the estimated solution."

The basic concept of the technique is that the best direction for finding a reduced sum of squares lies somewhere between the direction given by the Taylor series increment and the direction of steepest descent. In (15), when $K = 0$, ΔP approaches the Taylor series direction. This process insures second convergence near the minimum. The effect of decreasing K in each iteration on the resulting intermediate model is that the model will at first fit the broad, low frequency aspects of the data with higher frequency components being fitted as K decreases. Hoerl and Kennard (1970) give an excellent example of the comparison of the ridge regression inverse and the classic least-squares inverse. A more detailed discussion of ridge regression and its relationship to the generalized inverse is covered by Marquardt (1970). The algorithm tested is routine ZXSSQ for the IMSL computer library.

5) <u>The Spiral Algorithm</u>

The Spiral algorithm and its comparison to methods by Marquardt and Powell is described in a paper by Jones (1970) for a number of models whose derivatives with respect to the parameter (i.e. the elements of A) are analytic. Only the general concepts of the algorithm, following Jones (1970), will be given here.

The main principal of the algorithm is that a reduction in ΔG can always be found in parameter space plane defined by the Taylor series estimate of the minimum, the steepest descent estimate of the minimum and the starting point for that iteration (see Fig. (1). Figure 2 represents the plane ODT in parameter space, where O is the starting point for an iteration, T is the Taylor series point and D is the steepest descent point, chosen so the distance OD is equal to OT. The basic strategy is that the next starting point should be as far away from O as possible while keeping the number of evaluations of the least-squares surface to a minimum.

Within an iteration the first point checked is the Taylor series point T, which is generated by Marquardt's method. If $\Delta G_T < \Delta G_O$, this point is accepted as the new minimum and a new Taylor series point is calculated from there. If $\Delta G_T \geq \Delta G_O$, then the linear approximation of the model at O is not valid at T. This implies that the sum of squares valley must curve in one of the two directions, indicated by the dashed lines in Figure 2. In order to try and intercept the valley the spiral OST is searched. The curve moves out from T at angle β into the search area OTD and return to O along a tangent to line OD. The optimum equation for the spiral found by Jones (expressed in polar co-ordinates with origin at O) is

$$r = r_0(1-\theta \cos \beta-(1-\gamma \cos \beta)\ (\theta/\gamma)^2) \quad ,$$

where r is the distance OS and r_0 is the distance OT.

The points S that are checked on the spiral are determined from a sequence of points L which are generated on the segment TD. The points L divide the segment TD by the ratio: $(1-\mu)$ where μ is computed from the recurrence relation

$$\mu_{n+1} = 2\mu_n/(1 + \mu_n) \quad .$$

This relation was chosen to ensure that the points L become closer together as they near D.

The coordinates (a,θ) of the points L are given by the relations

$$\theta = \tan^{-1}\left[\frac{\mu \sin \gamma}{1-\mu + \mu \sin \gamma}\right] ,$$

and

$$a = \frac{r_0 \mu \sin \gamma}{\sin \theta} \quad .$$

With the starting point O as origin, the coordinates of the point S in parameter space are given in terms of t and d, the coordinates of T and D respectively by the relation

$$S = \frac{r}{a} \mu d + (1 - \mu)t \quad . \tag{16}$$

Equation (16) is the main operating equation of the Spiral algorithm aside from the Marquardt technique for generating T. If $\Delta G_T > \Delta G_0$, each successive search point is derived as a weighted sum of two parameter space vectors t and d.

The entire algorithm is more complicated since it contains a provision for dealing with spurious local minima. Interpolations are also performed when three consecutive search points yield reduced sums of squares, the interpolated minimum being checked to speed convergence.

PARAMETER STATISTICS

The simplest view of the inverse problem considered here is that of an automated "curve fitting" procedure. A model is derived by finding a parameter distribution that will produce a theoretical function, either ρ_a or kernel function, that fits the observed data in some least-squares sense. Weighted least-squares is usually used so that the data can be fitted on one of two ways: 1) data are fitted uniformly when the percentage data error is equal for all data (e.g., resistivity data is usually assmed to have a constant percent error at all electrode spacings) and (2) data are selectively fitted when data errors are variable (e.g., in EM soundings data error is usually a function of frequency).

An ideal weighting scheme for the first case is a $\log_{10}$ Rijo et al. (1973) such that the function to be minimized is

$$\Delta G = \sum_{i=1}^{N} \left[\log_{10}(\rho_{observed})_i - \log_{10}(\rho_{calculated})_i\right] . \qquad (17)$$

Since the data errors are a constant percentage, this scheme weights the date equally and eliminates the requirements for a weight vector to be included.

In the case of variable data error it would seem appropriate to weight each data point in inverse proportions to its error. The following equation (18), represents this type of weighting.

$$(\Delta G^W)^2 = \sum_{i=1}^{N} \frac{(\rho_{obs} - \rho_{cal})^2_i}{\hat{\sigma}_i^2} , \qquad (18)$$

where

$\hat{\sigma}_i$ = the standard error for the ith data point.

For the least squares inverse the weights enter as a matrix and (9) becomes $\Delta P = (A^T W^{-1} A)^{-1} A^T W^{-1} \Delta G$ where W is the diagonal matrix of data variance σ_i^2. The choice of a weighting scheme does not appreciably affect the speed of a minimization. However, it can drastically affect the position of the minimum in parameter space and the parameter statistics.

The effects of weighting, represented by (17) and (18) along with no weighting, on statistical parameter such as standard errors, correlation coefficients and eigenvectors, will be discussed after these statistical parameters have been examined.

The statistical parameters which are useful in characterizing our models are: 1) parameter standard errors, and 2) parameter correlation coefficient. In addition to these statistical parameters, the parameter and data eigenvectors with their associated eigenvalues can yield great insight into the relations between individual model parameter's and specific data.

Parameter standard errors and correlations are derived from the covariance matrix V, evaluated at the minimum.

$$V = \sigma^2 (A^T w^{-1} A)^{-1} \quad , \tag{19}$$

where

$$\sigma^2 = \frac{\Delta G^T w^{-1} \Delta G}{N-M} \quad ,$$

$$w_{ij} := \begin{cases} \sigma_i^2 & i=j \\ 0 & i \neq j \end{cases} \quad ,$$

and

$$(\Delta G)^2 = \sum_{i=1}^{N} [\text{Observed}_i - \text{Calculated}_i]^2 \quad ,$$

If $\log_{10}$ weighting is being used as in Equation (17), then w is replaced by taking $\log_{10}$ of the observed and calculated values in ΔG.

The parameter standard errors are defined by the square root of the diagonal term of V(e.g., $\sqrt{V_{11}}$ equals the standard error for parameter number 1).

The correlation matrix is the diagonally normalized covariance matrix. Its terms are the correlation coefficients, which are measures of the linear dependence between parameters. The correlation matrix C is given by Jenkins and Watts (1968):

$$C_{ij} \equiv \frac{V_{ij}}{(V_{ii})^{1/2}\,(V_{jj})^{1/2}} \quad . \tag{20}$$

If an element C_{ij} is near ±1 then the ith and j^{th} parameters are strongly linearly dependent. For example, if i represented the thickness t, and j the resistivity, ρ, of a layer (i.e., C_{ij} represents the correlation between the thickness and resistivity of a layer), then only the ratio t/ρ is well determined by the data if $C_{ij} \simeq 1$. This case is true for layers that are highly conductive relative to their surroundings. If $C_{ij} \simeq -1$, then only the product ρt is well determined, as is the case for the relatively resistive layers. This is the familiar equivalence problem discussed, for example, by Sunde (1949).

The relationship between parameter correlations and parameter standard errors is well explained by Inman (1975) and we will paraphrase him here: {If the correlations are small, then the standard errors, given by the square roots of the diagonals of (19), are a good measure of the uncertainty of each parameter. If, however, two parameters are highly correlated, $C_{ij} \simeq \pm 1$, then the standard deviations will be larger than the actual uncertainties. Figure 3 illustrates this fact with a generalized slice of solution space. The two coordinates axes correspond to two parameters of the estimated layered earth model. The ellipse indicates a confidence region within which the residual sum of squares is expected to lie for a certain percent of the repeated experiments. This region also defines the values of the parameter ρ_2 (resistivity) and t_2 (thickness) which will give a residual sum of

squares within the contour. The origin is defined by the parameter value at the final solution. The tilt of the axis of the ellipse is a measure of the degree of correlation between the two parameters. If the standard errors from (19) are taken to be the true deviation estimates, then the ellipse is enclosed by a box whose sides are defined by the standard deviation. The box, which ignores parameter correlation, represents a much larger confidence region than the ellipse. By using the standard deviation implied by the box one obtains a very conservative estimate of the parameter confidence interval for correlated parameters.}

Therefore, by considering the standard deviations in conjunction with parameter correlations a more realistic parameter standard deviation can be arrived at which is always less than or equal to the standard deviation computed from (19). Two models, one described by Inman (1975) and one of our choosing, called model 3, are considered for comparison of the inversion routines and to illustrate some concepts of the parameter statistics. See Fig. 11 and 12.

A common misconception about parameter standard errors is that a single model parameter can be varied by its estimated standard error with no significant change resulting in the calculated forward problem. In fact, the parameter standard errors and correlation coefficients must be viewed as representing a complex interactive system that describes combinations of parameter changes which can be made without a significant change in the estimated least-squares residual. For example, consider model 3 with its conductive middle layer. The eigenvectors, eigenvalues, correlation coefficients, and parameter standard errors,

calculated using data errors as weights in Equation (18), are shown in Figure 8. Note the high positive correlation between ρ_2 and t_2 indicating a linear relation between these two parameters, $S_2 = t_2/\rho_2$. Figure 4 shows the sounding curve for model 3 along with error envelopes generated by varying t_2 by its standard error. Similarly, Figure 5 shows the error envelope generated by varying ρ_2 by its standard error. Clearly the change in ρ_A curve is much larger than the 1% error assumed in the data. However, if the ratio t_2/ρ_2 is varied (see Figure 6), as indicated by their correlation, simultaneously by their standard errors, the change observed in ρ_A is on the order of 1%, the assumed error level, and thus is not a statistically significant change.

The parameter eigenvectors and their associated eigenvalues are also very useful in defining the relation between parameters and their overall effect on the data generated from a particular model.

Lanczos (1961) factored the system matrix A into its row (parameter) and column (observation) eigenvectors. The generalized inverse of A is defined in terms of these eigenvectors and eigenvalues as

$$H = A^{-1} = V\Lambda^{-1}U^T \quad .$$

The matrix U consists of q (q is the rank of A) eigenvectors u_i of length N associated with the columns (data) of A. V is made up of the q eigenvectors V_i of length M associated with the rows (parameters) of A. The matrix Λ^{-1} is the inverse of the diagonal matrix comprised of the eigenvalues of A. Figures 7 through 9 show these quantities for model 3 and will be discussed later.

The q parameter eigenvectors comprising V are a new parametrization of the model, they are the q specific linear combinations of the parameters that can be uniquely determined by the data. Similarly, the q data eigenvectors are the linear combination of data which are tied, through the assumed model, to the linear combination of parameters formed in V.

A useful analogy is to consider the eigenvectors as spectral components of the input and output of the linearized system. Then the decomposition of the matrix A is similar to the decomposition of the impulse response of an ordinary linear filter in terms of sinusoids (eigenvectors) of various amplitudes (eigenvalues). For a linear filter the amplitude response at a particular frequency determines how the filter will amplify the corresponding spectral component of the output.

Similarly, if we think of the matrix A as a filter relating the input, (parameters), to the output, (calculated data), then the eigenvalues are the amplification coefficients which determine the magnitude of the effect of the linear combination of parameters, v_i, on the linear combination of data u_i. Small eigenvalues and their associated eigenvector represent the spectral components which are poorly transferred through the earth model. By considering these eigenvector eigenvalue decompositions, one can optimize data sets to contain the maximum information related to a model parameter of particular interest. The subject of experiment design by this method is discussed by Glenn _et al_. (1976).

Figure 8 presents, (1) the parameter eigenvectors, (columns of V), (2) eigenvalues, (diagonal elements of Λ), (3) the data eigenvectors, (columns of U) whose numbering refers to the data points labelled in Figure 12, (4) the parameter correlation coefficients, and (5) the model parameter value with estimated standard errors. Some insight into the physical significance of this eigenvector decomposition can be gained by considering the effects of varying parameters on the sounding curve. Figure 13 shows the variation in ρ_A caused by changing ρ_1 by approximately 50%. Similarly, Figure 14 shows the variation in ρ_A caused by a 50% change in t_1. Note that the variation in ρ_A occurs from data points 1 to data points 10 or 11. Compare this with the second and third eigenvector pairs of Figure 8, both of whose parameter eigenvectors are composed of ρ_1, t_1 components. Their corresponding data eigenvectors have components from position 1 to 10. The same relation holds for the other eigenvector pairs. Compare Figure 4, produced by changing, ρ_2, with the first eigenvector pair of Figure 8. The correspondence between the eigenvector pair and changes induced in ρ_A by varying a particular parameter are not as clear for the fourth and fifth vector pairs of Figure 8. However, if the two are lumped together (the eigenvalues are of the same order of magnitude) the effects of varying t_2 or ρ_2 manifests itself in data from positions 9 to 21 as expected.

Effects of Weighting

By comparing Figures 8 and 9, it is obvious (as proposed by Rijo et al. 1977) that taking $\log_{10}$ of the observed and calculated data is the same as weighting by a constant percent error. The only difference appears in the estimated standard errors where those in Figure 8 used $\sigma^2 = 1$ in Equation (19) and those in Figure 9 used

$$\sigma^2 = \frac{\Delta G^T \log_{10} \Delta G \log_{10}}{N-M}$$

in Equation (19). The standard errors estimated in Figure 8 encompass the true parameter errors in all cases, whereas the standard errors shown in Figure 9 are consistently too small.

The eigenvector pairs, correlation, and parameter standard errors calculated with no data weighting are shown in Figure 7: the eigenvectors are essentially the same as those in Figures 8 and 9 with one major exception. The data eigenvectors associated with parameter ρ_2, t_2, and ρ_3 have their components shifted to larger $AB/_2$ spacings. This bias occurs because the large ρ_A values toward larger $AB/_2$ dominate the sum of squares. In effect, each data point is weighted by its own magnitude. In addition a much higher degree of parameter correlation is found when no data weighting is used and estimated parameter standard errors are much larger than those calculated by weighted schemes. For an example of this compare the correlation coefficients for ρ_3 ρ_2 and ρ_3 t_2 in Figures 7 and 8.

Weighting the data by their standard errors as indicated in Equation (12) seems to be the most advisable because it is the most flexible to varying data errors and yields the smallest standard errors, consistent with the true error, as shown in Figures 7 through 9.

The parameter eigenvectors and correlation coefficients for Inman (1975) model are given in Figure 10 and present a good example for their interpretation. The first three eigenvalues are all of about the same order of magnitude with the last two very much smaller. The linear combination of parameters represented by the first three eigenvectors have the greatest effect on the sounding curve. In these three eigenvectors and ρ_1 and t_1 elements have opposite signs while the ρ_2 and t_2 elements have the same sign. This indicates that if ρ_2 and t_2 are both changed in the same direction, either positive or negative, the effect on the sounding curve (Figure 11) will be larger compared to the effect of similar changes on other parameters. In addition, if ρ_1 increases and t_1 decreases, or vice versa, the sounding curve will also change. The eigenvector associated with $\lambda_4^2 = 0.081$ indicates, since λ_4 is small, that increasing or decreasing ρ_1 and t_1 together will have little effect on Figure 11. In other words, the ratio t_1/ρ_1 is the combination of these parameters which effects the sounding curve the most (note the correlation coefficient between ρ_1 and t_1 is +0.86). The eigenvector associated with $\lambda_5^2 = 0.0097$ indicates that increasing ρ_2 and decreasing t_2 or vice versa has little effect on the sounding curve, (i.e., only the product $\rho_2\ t_2$ affects Figure 12). Again, this is also indicated by considering the correlation coefficient between ρ_2 and t_2, which equals -0.988.

By considering the parameter eigenvector and parameter correlation coefficients of this model it can be seen that the products $\rho_2\ t_2$ are better determined by, and have greater effect on, this data than either ρ_2 or t_2 separately. This is also true to a lesser extent of the parameters in the ratio t_1/ρ_1.

Comparison of the Inversion Algorithms

The five methods discussed here were compared on two three-layer planar earth models (Figures 11 and 12). The data was generated by a numerical evaluation of Equation (2). Each of the five routines was modified to use log parameters to ensure determination of physically meaningful parameter sets (i.e., no negative parameter values) Rijo et al. (1977). Each routine minimizes the same ΔG (Equation (17) and begins its process from the same initial guess of the parameter vector.

The comparison can be made in terms of the number of forward problem evaluations required to reach a desired minimum and by considering the accuracy of the determined parameters. The criterion for defining the degree of fit between the observed and calculated apparent resistivity is the data variance, σ^2, given by Hamilton (1964) as:

$$\sigma^2 = \frac{\sum_{i=1}^{N} [\rho_0 - \rho_c]_i^2}{N - M} \quad . \tag{18}$$

where (N-M) are the degrees of freedom of the system.

Note that the measure of the fit is defined in real space and not in log or weighted space.

The results of the five inversions on both models are given in Table 1. The number of forward problem evaluations, N, required to reach a data variance estimate of σ^2 and the resulting parameters are shown. The table also contains the linear combination of parameters for the middle layer, which are best determined by the data. For Inman's model (resistive middle layer) this is t_2/ρ_2.

The initial guess for Inman's model was that used by Inman (1975).

ρ_1 = 8 ohm.m $\quad$ t_1 = 15 m

ρ_2 = 500 ohm.m $\quad$ t_2 = 150 m

ρ_3 = 5 ohm.m

The initial guess for model 3 was:

ρ_1 = 80 ohm.m $\quad$ t_1 = 100 m

ρ_2 = 20 ohm.m $\quad$ t_2 = 50 m

ρ_3 = 500 ohm.m

The true model parameter for Inman's model and model 3 are given in Figures 11 and 12, respectively.

A general convergence criterion of σ = 0.05 was used. The values for σ can become very much less than σ (e.g., on iteration 30, σ = 0.36 for Marquardt's method, an iteration 31, σ = 0.026). Iterations were stopped if the number of function calls was excessive compared with the other routines; this only occurred for Bremermann's method, and the simplex method.

In a comparison of this type, where we are using the ideal data and requiring a close fit, all the parameters are just about equally resolved, given a low σ. For both models, the three (Table 1), Ridge

Regression, Peckham's and Spiral, all reached fairly close values and very similar parameter estimates. For this reason, a comparison can be made simply in terms of N, the number of function calls. The iterative progression of parameters in different routines is of interest since it can give an indication of how a particular routine reaches a model.

For Inman's model the ranking in terms of N is 1) Ridge Regression, 2) Spiral, 3) Peckham, 4) Bremermann, 5) Simplex. Considering the accuracy of parameters ρ_2 and t_2, it is interesting to note that Peckham's method has more accurate estimates for ρ_2 and t_2 than Ridge Regression or Spiral, although the additional 50 plus function calls outweighs the increased accuracy. The Bremmerman and Simplex estimates for t_2 are very poor, but it is interesting to note in the case of both Simplex and Bremmerman that, although ρ_2 and t_2 are in error by as much as 100%, their product $\rho_2 t_2$ is only 2% low.

The ranking in terms of N for model 3 is identical to the ranking for Inman's model. All of the routines gave good estimates for ρ_1, t_1, ρ_3. All five also gave good estimates of the ratio t_2/ρ_2, even though the individual parameters were either both high (i.e. Peckham, Ridge Regression, Bremermann, Simplex, or low (i.e., Sprial).

The procession of conductive middle layer parameters, ρ_2, t_2, and t_2/ρ_2 are plotted as a function of forward problem evaluation in figures (15), (16), and (17) respectively for the three fastest routines. The behavior of these two parameters, ρ_2 and t_2, are characteristic of the other, unplotted, parameters.

It is immediately obvious that parameter procession for Peckham's method is characterized by large oscillations from one forward problem evaluation to another as contrasted with rather smooth transitions of parameter values for Spiral and Marquardt's methods. The oscillations in Peckham's method are due to using a Simplex to define the forward problem function space, since each new set of parameter estimates is obtained through an arbitrary operation such as reflection of contraction. This is to be contrasted with the almost monotonous procession of parameters in the Spiral and Marquardt routines, which essentially use Taylor series increments to define the forward problem function space. It is also worth noting that Peckham's parameter estimates develop an oscillation in values near the true solution; whereas the other two routines converge much more rapidly to a solution.

One final property of the solutions that should be noted in the expression of the equivalence principal as seen by the procession of the longitudinal conductance $S_2 = t_2/\rho_2$. For all three routines the longitudinal conductance S_2 reached the true value much faster than either individual parameter; and even when both parameter ρ_2 and t_2 are in error as they are both Spiral and Peckham, their ratio t_2/ρ_2 has been accurately determined. This is a realization of the fact that for a thin conductive layer, S_2 is the quantity best determined by the data.

Considering both models, the ranking in terms of reaching the lowest σ^2 with the fewest number of forward problem evaluations is:

1) Ridge Regression
2) Spiral
3) Peckham's method
4) Bremmerman
5) Simplex

It should be noted that, for Bremermann's method, the number of function evaluations used in the iterative process is independent of the number of parameters describing the system, while the other four routines require more function evaluations as the number of parameters increases. For this reason, the Bremerman method would compare more favorably for models with a large number of parameters.

Conclusion

It has been demonstrated that parameter statistics such as parameter standard errors, parameter correlation, and associated eigenvectors can be greatly affected by choice of data weighting. The use of inverse data error weighting is flexible and yields the most reliable parameter standard error estimates. In addition, the relationships between parameter and data eigenvectors is physically correct and not biased as in the case where no weighting is used.

In the comparison of the five least squares minimization algorithm, the Ridge Regression algorithm proved to require the fewest number of forward problem evaluation to reach a desired fit. The ranking of the five algorithms is the same for both models tested; indicating that the relative speeds of the algorithms is, at least to some degree, model independent.

Acknowledgement

This work was supported by the Assistant Secretary for Conservation and Renewable Energy, Office of Renewable Technology, Division of Geothermal and Hydropower Technologies of the U. S. Department of Energy under Contract No. W-7405-ENG-48. A Fortran IV version of Spiral was obtained for testing through the courtesy of Shell Research Limited.

References

Alpin, L.M., 1966, "Dipole methods for measuring earth conductivity," tran. by G. V. Keller, Consultants Bureau.

Bremermann, 1970, A method of unconstrained global optimization, Mathematical Biosciences, 9, 1-15.

Compagnie Generale de Geophysique, 1955, Abaques de sondage electrique: Geophysical Prospecting, 3, supplement 3.

_____, 1963, Abaques de sondage electrique; The Hague, E.A.E.G.

Crous, C. M., 1971, Computer-assisted interpretation of electrical soundings: Colorado School of Mines, M.S. Thesis, 108 pp.

Flathe, H., 1955, A practical method of calculating geoelectric model graphs for horizontally stratified media: Geophysical Prospecting 3, 268-294.

Foster, M., (1961), An application of the Wiener-Kolmogorov smoothing theory to matrix inversion, Journal of the Society for Industrial and Applied Mathematics, 9, 387-392.

Ghosh, D. P., 1971, The application of linear filter theory to the direct interpretation of geolectric resistivity measurements: Geophysical Prospecting, 19, 192-217.

Glen, W. E., Ward, S. H., 1976, Statistical evaluation of electrical sounding methods. Part 1: Experiment design: Geophysics, 41, 6A, 1207-1222.

Golub, G., 1965, Numerical methods for solving linear least squares problems, Numerische Mathematik, 7, 206-216.

Ginzburg, A., Loewenthal, D., Shohan, Y., 1976, On the automated interpretation of direct current resistivity, Pageoph. 114, 983-995.

Hamilton, W. C., 1964, Statistics in physical sciences, estimation, hypothesis testing, and least squares: New York, Ronald Press Co.

Hildebrand, F. B., 1949, Advanced calculus for engineers: New York, Prentice Hall.

Hoerl, A. E. and Kennard, R. W., 1970a, Ridge regression: Biased estimation for nonorthogonal problems, Technometrics, 12, 55-67.

_____, 1970b, Ridge regression: Applications to nonorthogonal problems, Technometrics 12, 69-82.

Inman, J. R., Ryu, J. and Ward, S. H., 1973, Resistivity inversion, Geophysics, 38, 1088-1108.

Inman, J. R., 1975, Resistivity inversion with ridge regression, Geophysics, 40, 798-817.

Jackson, D. D., 1972, Interpretation of inaccurate, insufficient, and inconsistent data, Geophysical Journal of the Royal Astronomical Society, 28, 97-109.

Jenkins, G. M. and Watts, D. G., 1968, Spectral analysis and its application: San Francisco, Holden-Day Inc.

Jones, A., 1970, Spiral - a new algorithm for non-linear parameter estimation using least squares, The Computer Journal 12, 301-308.

Jupp, D. L. B., Vozoff, K., 1975, Stable iterative methods for the inversion of geophysical data, Geophysical Journal of the Royal Astronomical Society, 42, 952-976.

Kalenov, E. N., 1957, Interpretation of vertical electrical sounding curves, Moscow, Gostoptekhizdat, 431 pp.

Keller, G. V. and Frischknecht, F. C., 1966, Electrical methods in geophysical prospecting: New York, Pergamon Press, 517 pp.

Koefoed, O., 1965, Direct methods of interpreting resistivity observation: Geophysical Prospecting, 13, 568-591.

_____, 1966, The direct interpretation of resistivity observations made with a Wenner electrode configuration; Geophysical Prospecting, 14, 71-79.

_____, 1968, The application of the kernel function in interpreting geoelectrical measurements: Berlin, Gebruder Borntraeger, 111.

Kunetz, G., 1968, Principals of direct-current resistivity prospecting, Berlin, Gebrüder Bormtraeger, 103.

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

Langer, R. E., 1933, An inverse problem in differential equations, Bull, American Mathematical Society, Series 2, 29, 814-820.

Levenberg, K., 1944, A method for the solution of certain nonlinear problems in least squares: Quart. Appl. Math., 2, 164-168.

Marquardt, D. W., 1963, An algorithm for least-squares estimation of non-linear parameters, Journal of the Society for Industrial and Applied Mathematics, 11, 431-441.

_____, 1970, Generalized inverses, ridge regression, biased linear estimation, and non-linear estimation, Technometrics, 12, 591-612.

Meinardus, H. A., 1967, The kernel function in direct-current resistivity sounding, Colorado School Mines, M.S. Thesis, 151 pp.

_____, 1970, Numerical interpretation of resistivity soundings over horizontal beds, Geophysical Prospecting, 18, 415-433.

Mooney, H. M. and Wetzel, W. W., 1956, The potentials about a point electrode and apparent resistivity curves, Minneapolis, University of Minnesota Press.

Nelder, J. A. and Mead, R., 1965, A simplex method for function minimization, The Computer Journal, 308-313.

Onodera, Seibe, 1960, The kernel function in the multiple-layer resistivity problem, Journal of Geophysical Research, 65, 3787-3794.

Orellana, E. and Mooney, H. M., 1966, Master tables and curves for vertical electrical sounding over layered structures, Madrid, Interciencia, 150 pp., 66 tables.

Peckham, G., 1970, A new method for minimizing a sum of squares without calculations, gradients, The Computer Journal, 13, 418-420.

Pekeris, C. K., 1940, Direct method of interpretation in resistivity prospecting, Geophysics 31-42.

Petrick, W. R., Pelton, W. H., and Ward, S. H., 1977, Ridge regression inversion applied to crustal resistivity sounding data from South Africa: Geophysics, 42, 5, 995-1006.

Powell, M. J. D., 1964, An efficient method for finding the minimum of a function of several variables without calculating derivatives: The Computer Journal, 7, 155-162.

Rijo, L., Pelton, W. H., Feitosa, E. C. and Ward, S. H., 1977, Interpretation of apparent resistivity data from Apodi Valley, Rio Grande de Norte, Brasil, Geophysics, 92, 9, 811-822.

Roy, A., Apparao, A., 1971, Depth of investigation indirect current methods, Geophysics, 36, 5, 943-959.

Rijkswaterstaat, 1969, Standard graphs for resistivity prospecting, EAEG, The Hague.

Slichter, L. B., 1933, Interpretation of resistivity prospecting for horizontal structures: Physics, 4, 307-322; also erratum, p. 407.

Spendley, W., Hext, G. R., and Himsworth, F. R., 1962, Sequential application of simplex design in optimization and evolutionary operation, Technometrics, 4, 441-461.

Stefanesco, S. S., Schlumberger, C., Schlumberger, M., 1930, Sur la distribution electrique potentielle auour d'une prix de terre ponctuelle dans un terrain a couches horizontals, homogenes, et isotropes, Journal de Physique et le Radium, 1, 132-140.

Stevenson, A. F., 1934, On the theoretical determination of earth resistance from surface potential measurements, Physics, 5, 114-124.

Sunde, E. O., 1949, Earth conductor effects in transmissions systems, New York, Doyer Publications, Inc..

Vanyan, L. L., Morozova, G. M. and Lozhenitina, L., 1962, On the calculation of theoretical electrical sounding curves, Prikladnaya Geofizika, 34, 135-144 (in Russian).

Vozoff, K., 1958, Numerical resistivity analysis - Horizontal layers, Geophysics 28, 536-556.

Vozoff, K., and Jupp, D. L. B., 1975, Joint inversion of Geophysical Data, Geophysical Journal of the Royal Astronomical Society, 42, 977-991.

Wiggins, R. A., 1972, The generalized inverse problem, Reviews of Geophysics and Space Physics 10, 251-286.

Zohdy, A. A. R., 1965, The auxiliary point method of electrical sounding interpretation, and its relationship to the Dar Zarrouk parameters, Geophysics 30, 644-660.

_____, 1968, The effect of current leakage and electrode spacing errors on resistivity measurements, in Geological Survey Research, U.S. Geological Survey Professional Paper 600-D, p. D258-D264.

_____, 1974a, Use of Dor Zarrouk curves in the interpretation of vertical electrical sounding data, U.S. Geological Survey Bull., 1313-D, 41 pp.

_____, 1974b, A computer program for the calculation of Schlumberger sounding curves by convolution, U.S. Geological Survey Report GD-74-010, PB-232 056.

_____, 1974c, A computer program for the automatic interpretation of Schlumberger sounding curves over horizontally stratified media, U.S. Geological Survey Report GD-74-017, PB-232-703.

Table 1

Model 3

	σ	N	ρ_1	t_1	ρ_2	t_2	ρ_3	t_2/ρ_2
Peckham	0.44	75	100.20	49.7	3.6	121.2	998	33.66
Marquardt	0.188	28	101.01	49.6	3.97	133.3	1005	33.57
Bremmerman	212.0	192	101.53	50.2	3.92	133.4	1107	34.0
Simplex	0.034	332	100.02	49.6	3.6	120.0	1002	33.33
Spiral	0.048	45	100.01	50.1	2.7	91.7	997	33.96
	True Values		100	50	3	100	1000	33

Model 4

	σ	N	ρ_1	t_1	ρ_2	t_2	ρ_3	$t_2 \cdot \rho_2$
Peckham	0.065	100	9.99	9.98	385.2	252.3	10.0	97185.0
Marquardt	0.019	29	9.99	9.96	382.8	254.7	9.99	97499.0
Bremmerman	1.56	67	10.41	11.29	524.23	181.87	10.05	95342.0
Simplex	9.03	299	10.00	10.3	594.8	162.0	10.09	96357.0
Spiral	0.041	43	10.00	9.89	380.3	256.4	10.0	97508.0
	True Values		10	10	390	250	10	97500

σ = data variance estimate
N = the number of calculation of the matrix A

List of Figures

1) Contours of constant sum of squares, illustrating different directions toward the minimum sum of squares.

2) Geometric picture of the operation of the spiral algorithm.

3) Standard error ellipse (after Inman, 1975).

4) Model 3 sounding curve with 1% variation of ρ_2.

5) Model 3 sounding curve with 1% variation of t_2.

6) Model 3 sounding curve with 1% variation of t_2/ρ_2.

7) Parameters and data eigenvectors with associated eigenvalues, parameter correlations, and best fit model parameters for minimization with no. weighting.

8) Parameter and data eigenvectors with associated eigenvalues, parameter correlations, and best fit model parameters for minimization with data weighted by their standard errors (standard error 1% of data value).

9) Parameters and data eigenvectors, parameter correlations and best fit model parameters for weighting by taking $\log_{10}$ of the observed and calculated data.

10) Parameter eigenvectors and eigenvalues, and parameter correlations for Inman's 1975 model.

11) Schlumberger sounding curve for Inman's model.

12) Model 3 Schlumberger sounding curve.

13) Variation of ρ_A for Model 3 with 10% change in ρ_1.

14) Variation of ρ_A for Model 3 with 10% change in t_1.

15) Progression of the estimate of ρ_2 for Model 3 as a function of the number of forward problem evaluations.

16) Progression of the estimate of t_2 for Model 3 as a function of the number of forward problem evaluations.

17) Progression of the estimate of t_2/ρ_2 for Model 3 as a function of number of forward problem evaluations.

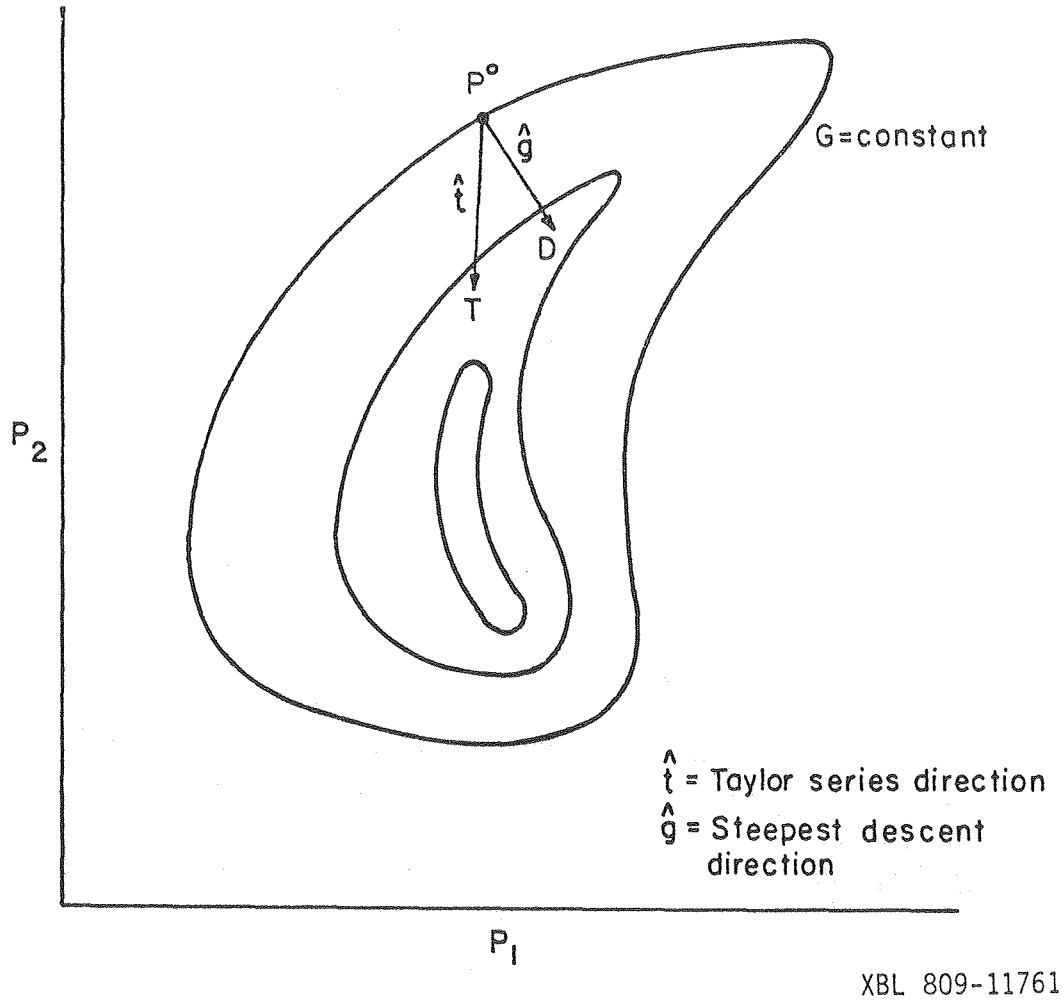
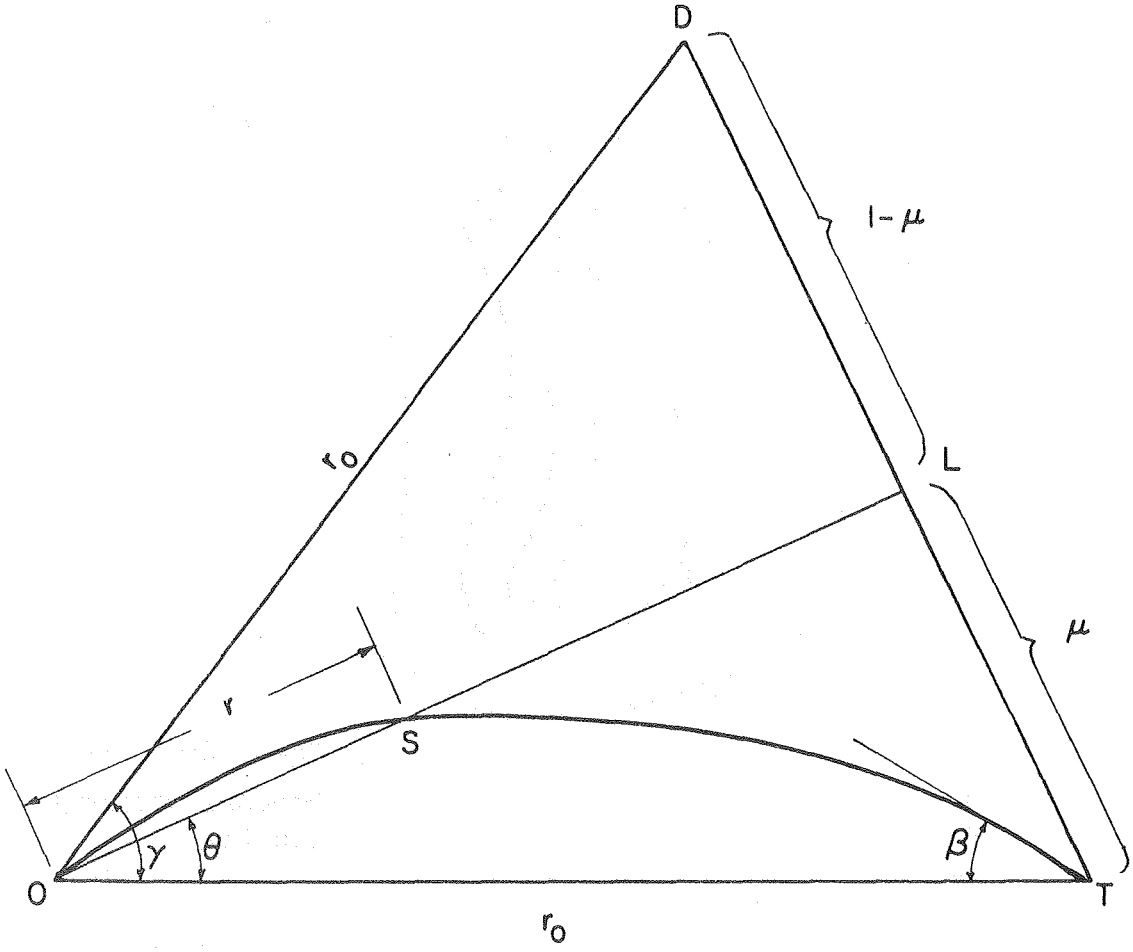

P°
ĝ
t̂
D
T
G=constant
P2
P1
t̂ = Taylor series direction
ĝ = Steepest descent direction
XBL 809-11761

Fig. 1



XBL 809-11760

Fig. 2 Geometry of SPIRAL Algorithm

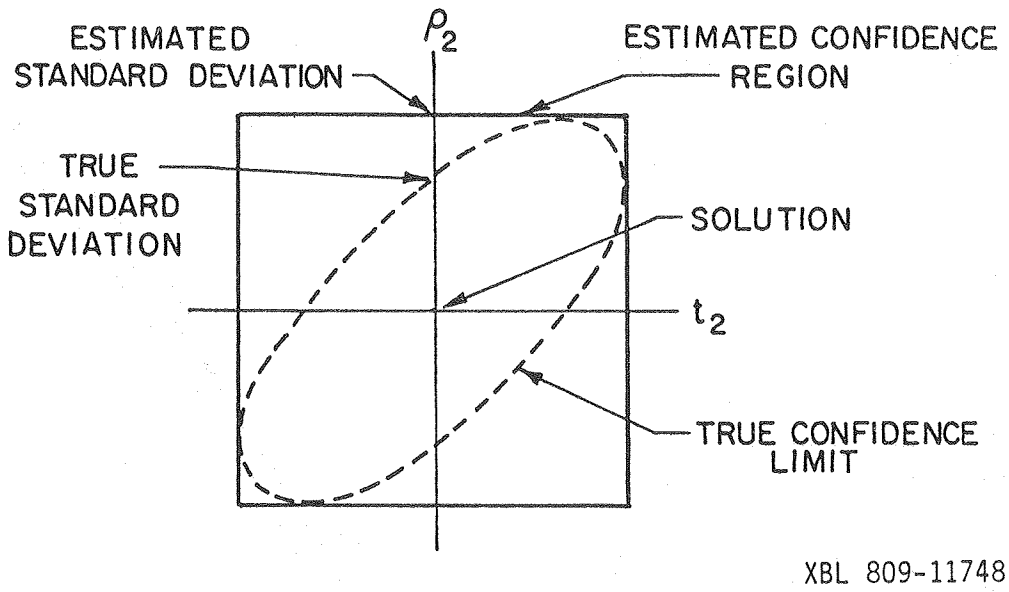
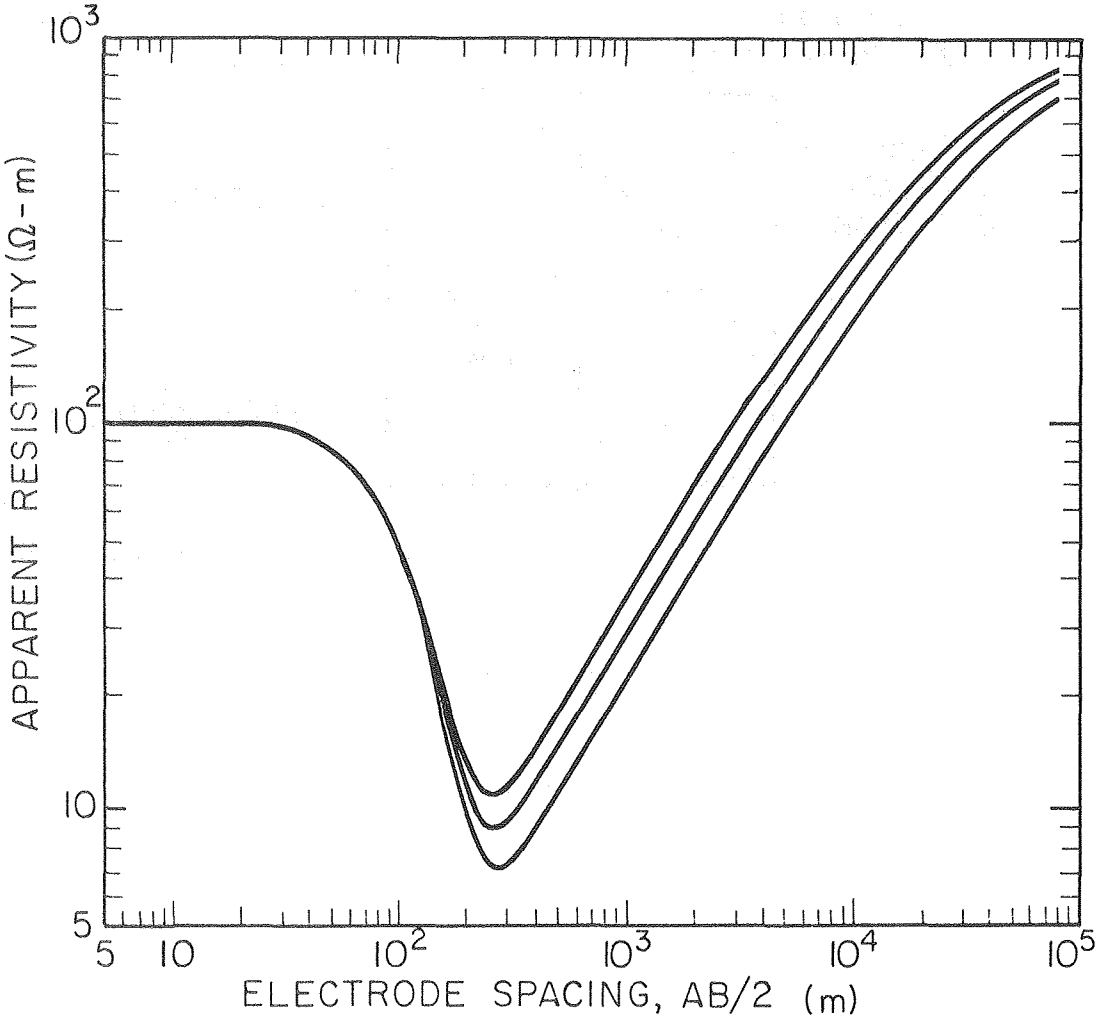
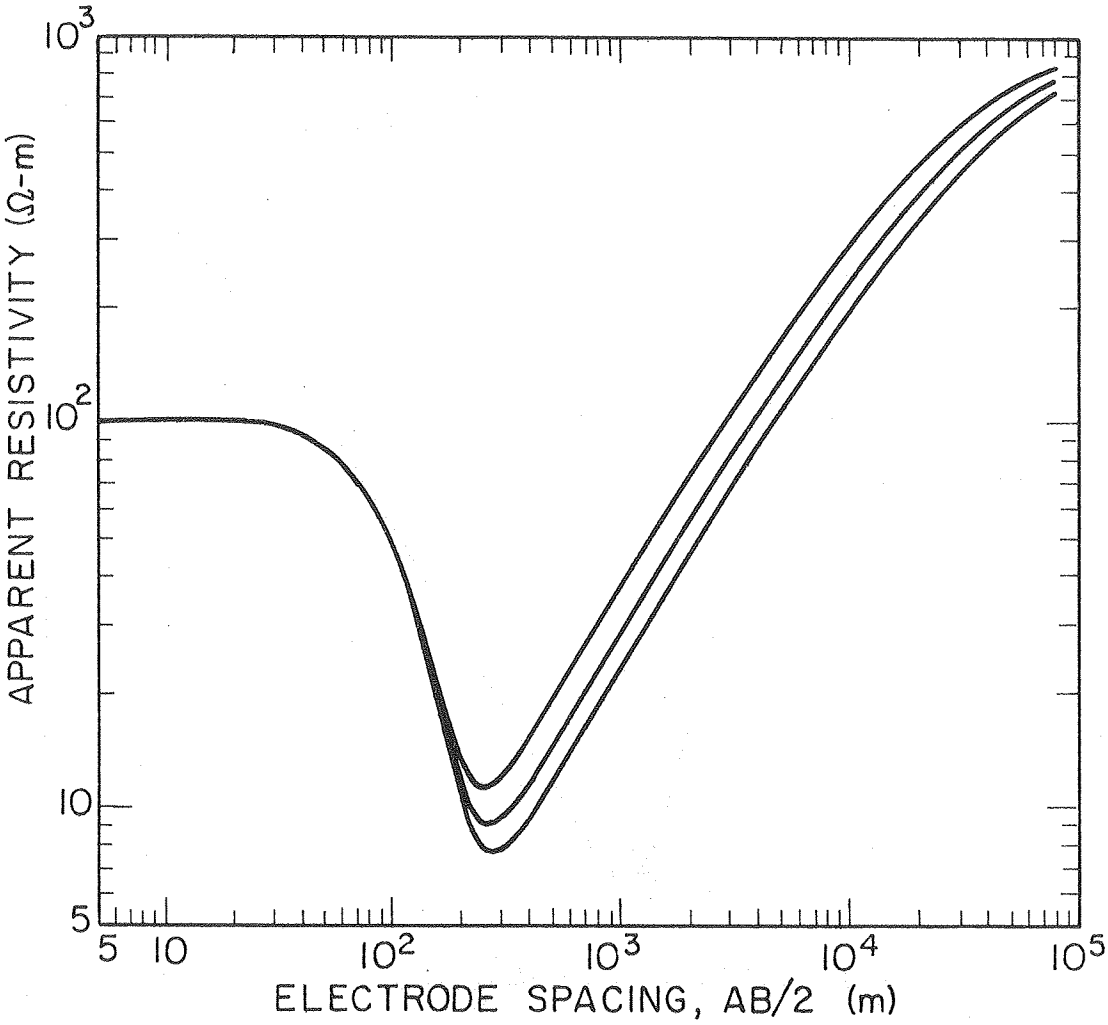

ESTIMATED
STANDARD DEVIATION
ρ_2
ESTIMATED CONFIDENCE
REGION
TRUE
STANDARD
DEVIATION
SOLUTION
t_2
TRUE CONFIDENCE
LIMIT
XBL 809-11748

Fig. 3



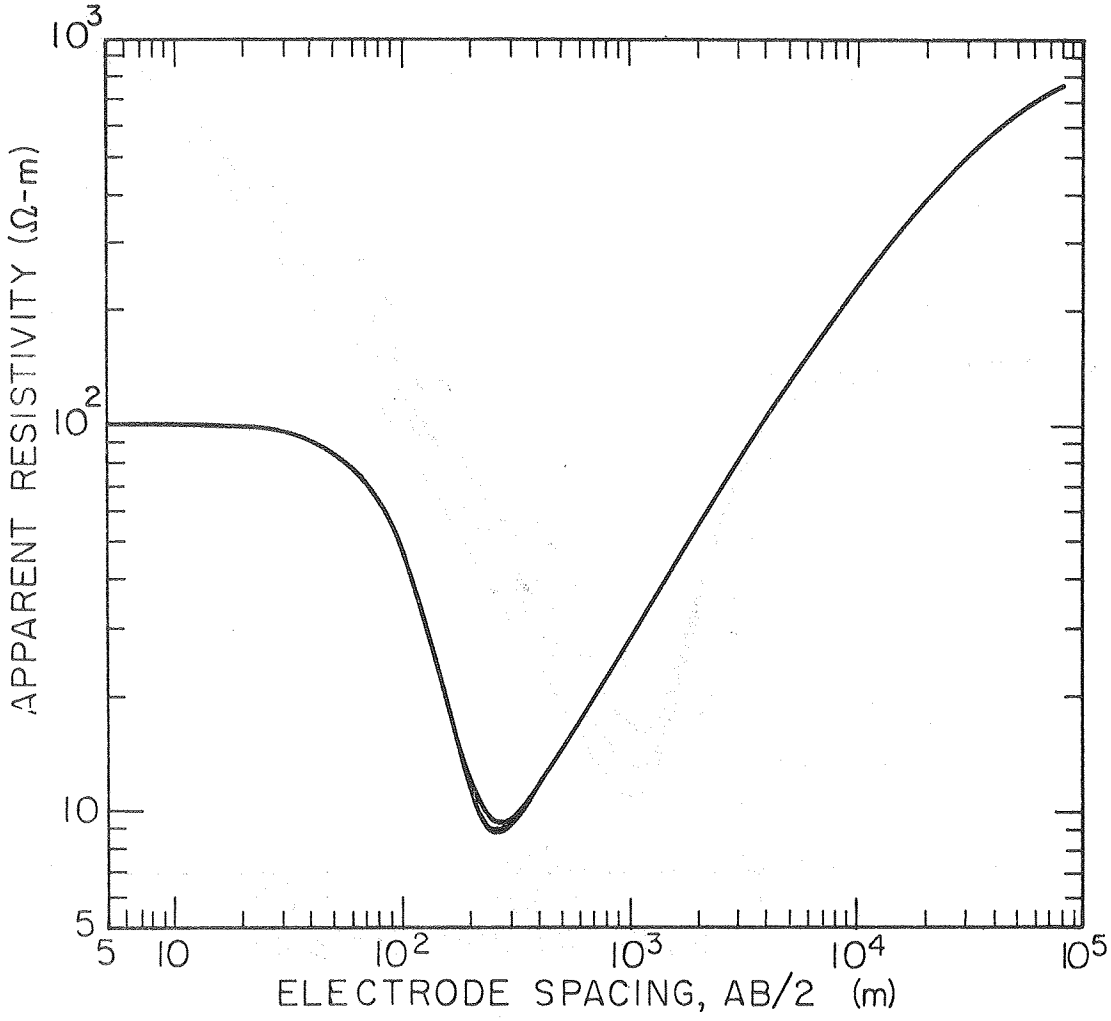
XBL 809-11745

Fig. 4



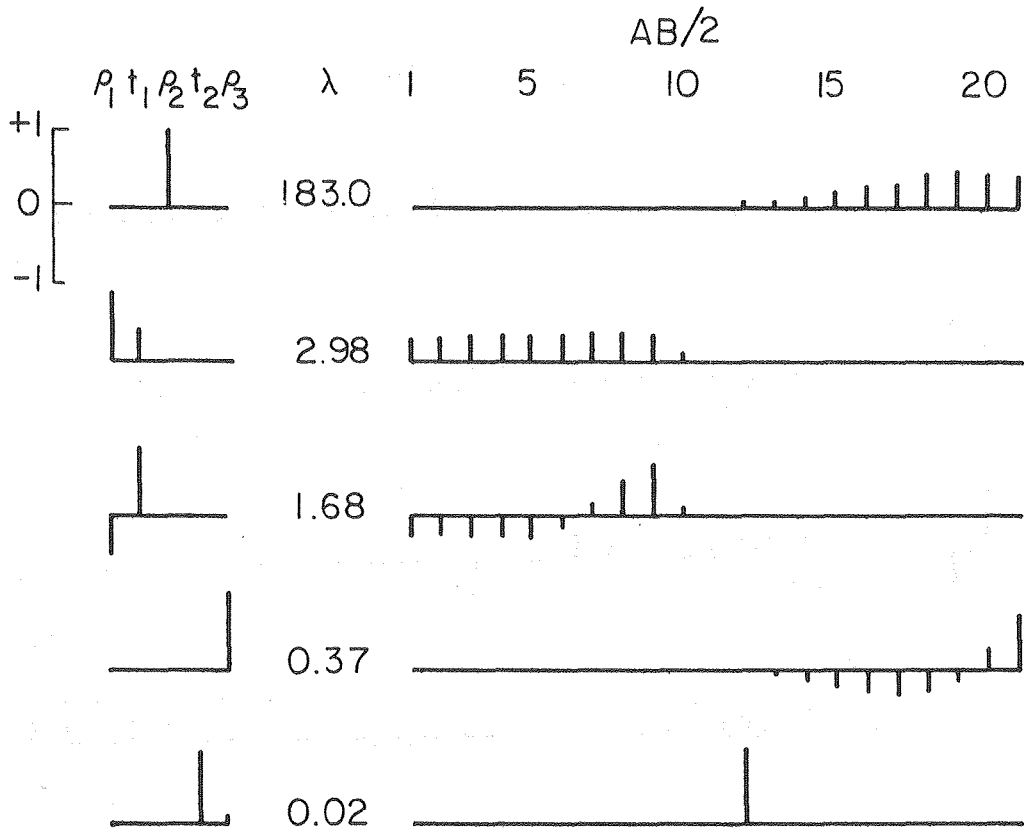
XBL809-11746

Fig. 5



XBL809-11747

Fig. 6

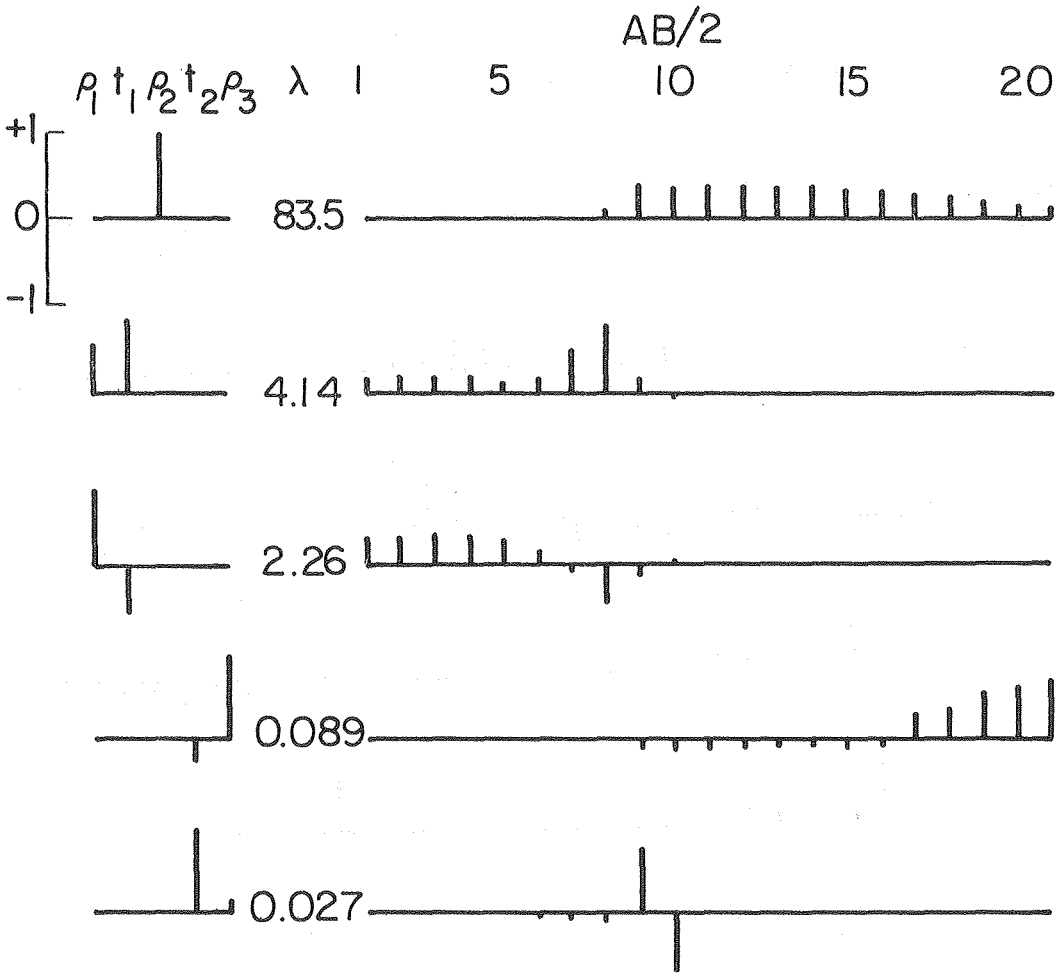


Correlation Matrix

	ρ_1	t_1	ρ_2	t_2	ρ_3	True Model	Best Fit	Standard Error
ρ_1	1.0					100.0	ρ_1 = 99.99	±0.21
t_1	-.27	1.0				50.0	t_1 = 49.85	±1.053
ρ_2	.17	-.96	1.0			3.0	ρ_2 = 3.31	±2.17
t_2	.16	-.96	1.0	1.0		100.0	t_2 = 110.81	±73.26
ρ_3	.16	-.90	.94	.94	1.0	1000.0	ρ_3 = 1005.7	±4.48

XBL 809-11759

Fig. 7

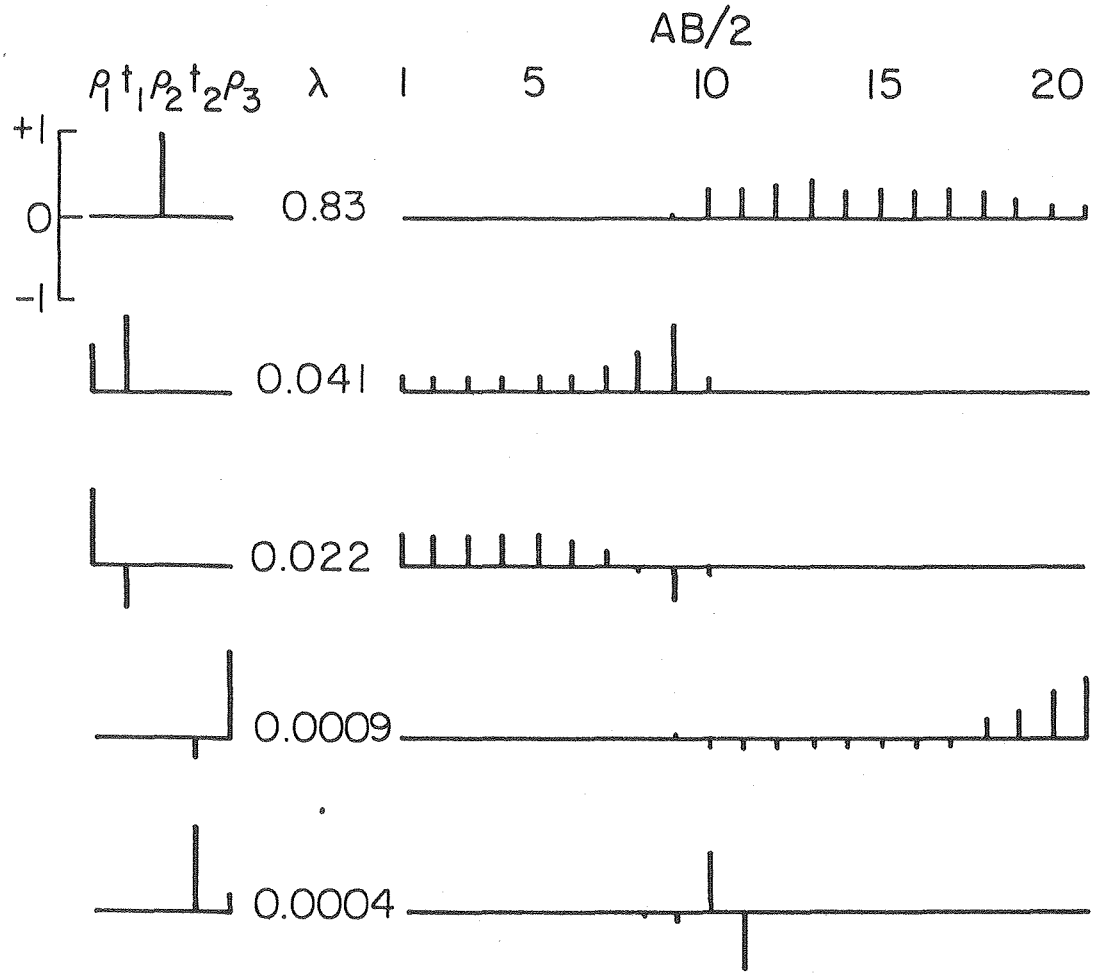


Correlation Matrix

	ρ_1	t_1	ρ_2	t_2	ρ_3	True Model	Best Fit	Standard Error
ρ_1	1.0					100.0	ρ_1 = 99.99	±0.39
t_1	-.41	1.0				50.0	t_1 = -49.85	±0.56
ρ_2	.15	-.83	1.0			3.0	ρ_2 = -3.31	±0.81
t_2	.15	-.83	.99	1.0		100.0	t_2 = -110.81	±27.80
ρ_3	.03	-.27	.36	.36	1.0	1000.0	ρ_3 = 1005.7	±11.90

XBL809-11758

Fig. 8

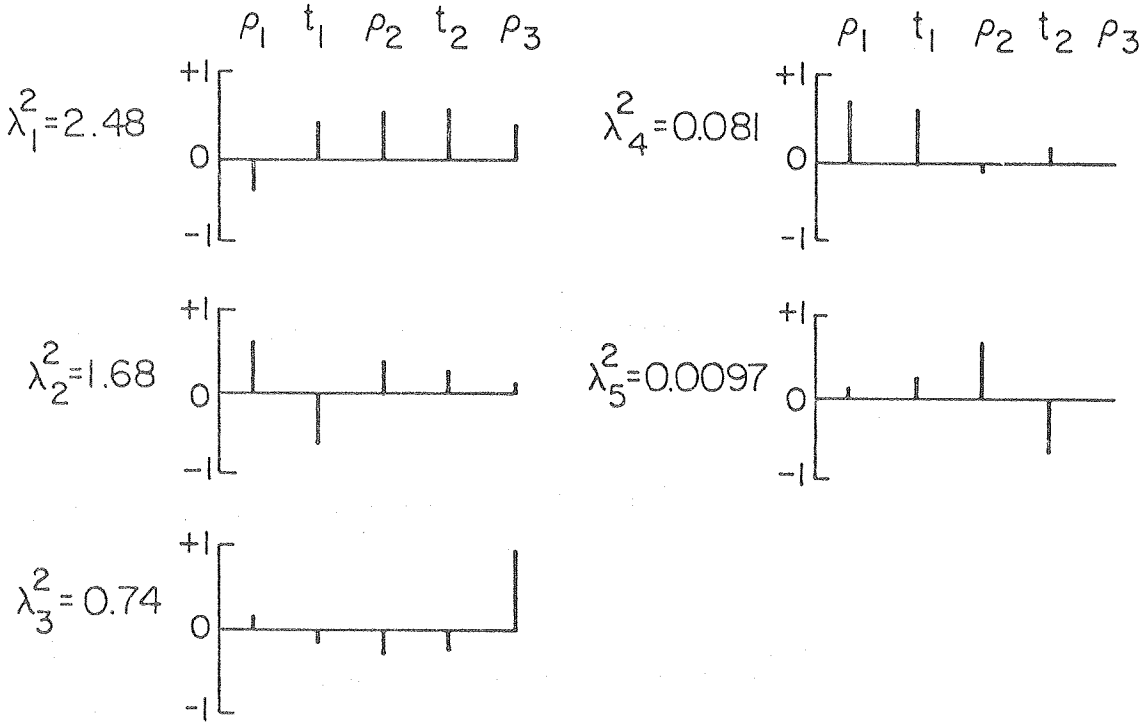


Correlation Matrix

	ρ_1	t_1	ρ_2	t_2	ρ_3	True Model	Best Fit	Standard Error
ρ_1	1.0					100.0	ρ_1 = 99.99	±0.04
t_1	-.41	1.0				50.0	t_1 = 49.85	±0.06
ρ_2	.15	-.83	1.0			3.0	ρ_2 = 3.31	±0.09
t_2	.15	-.83	.99	1.0		100.0	t_2 = 110.81	±3.10
ρ_3	.02	-.21	.28	.29	1.0	1000.0	ρ_3 = 1005.7	±1.30

XBL 809-11757

Fig. 9



CORRELATION MATRIX

	ρ_1	t_1	ρ_2	t_2	ρ_3
ρ_1	1.0	.86	.21	-.22	.11
t_1		1.0	.57	-.58	.16
ρ_2			1.0	-.988	.24
t_2				1.0	-.29
ρ_3					1.0

XBL 809-11756

Fig. 10

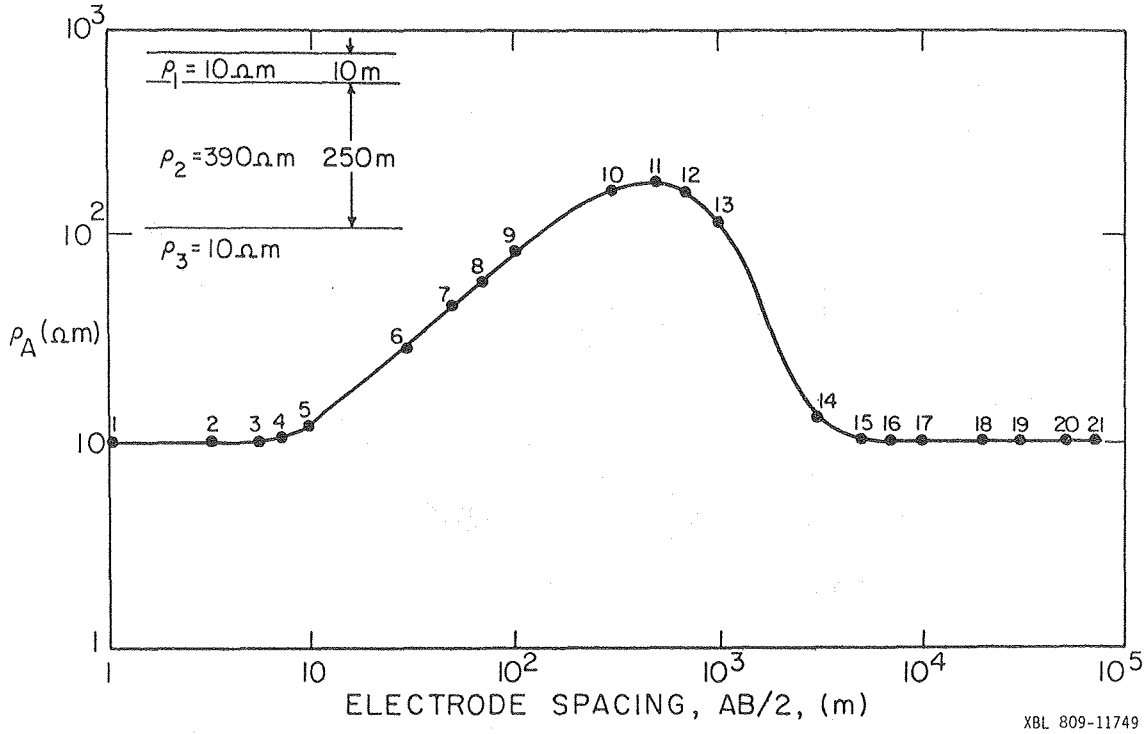

10^3
10^2
10
1
ρ_A (Ωm)
ρ_1 = 10Ωm
10m
ρ_2 = 390Ωm
250m
ρ_3 = 10Ωm
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
1
10
10^2
10^3
10^4
10^5
ELECTRODE SPACING, AB/2, (m)
XBL 809-11749

Fig. 11

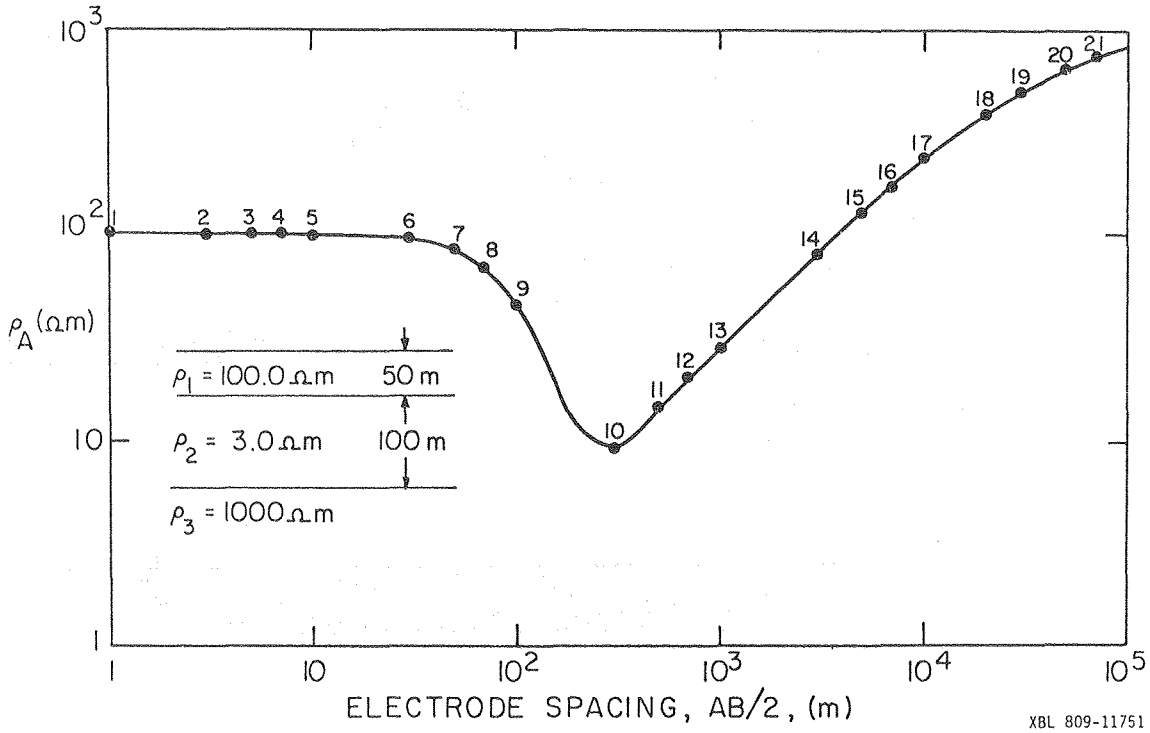
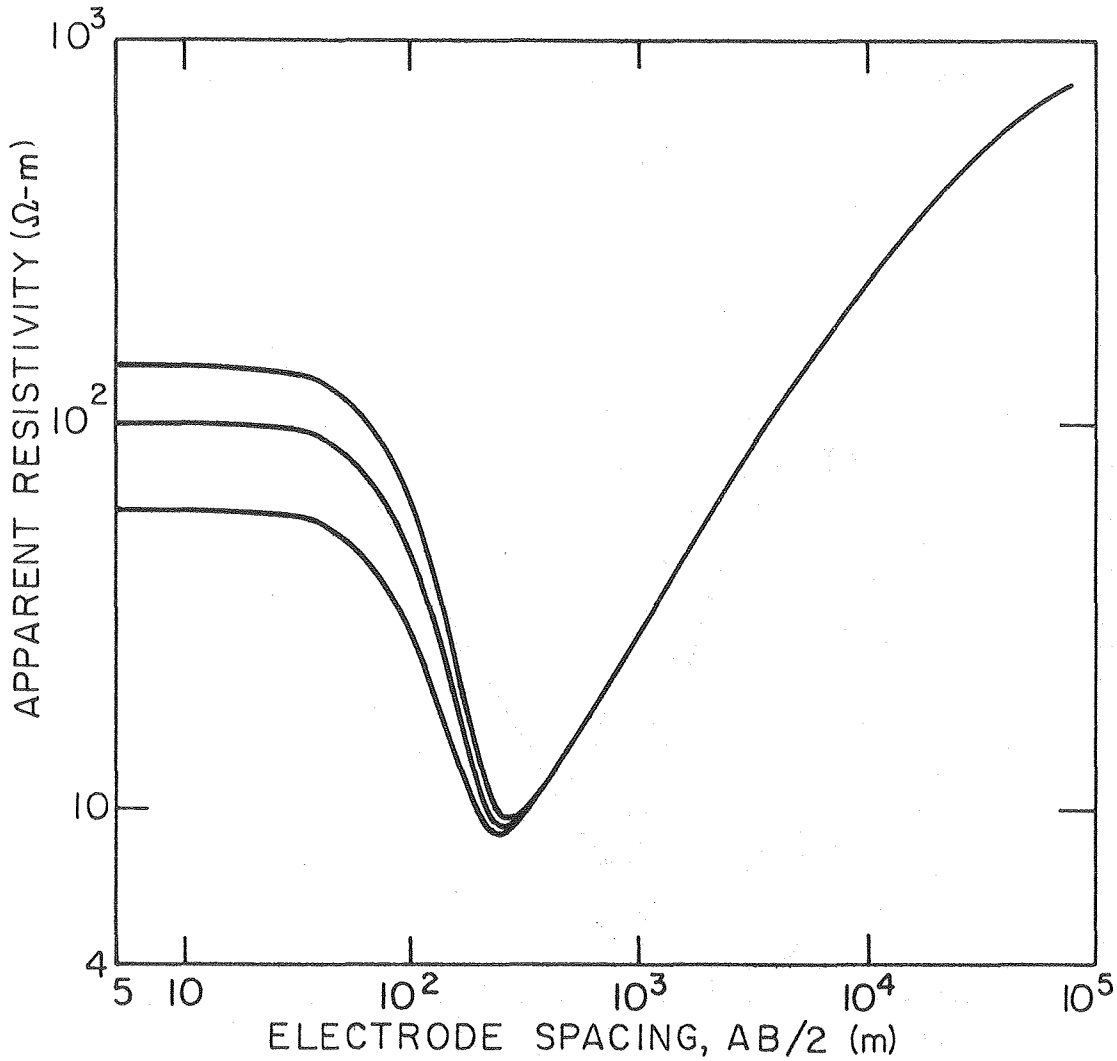
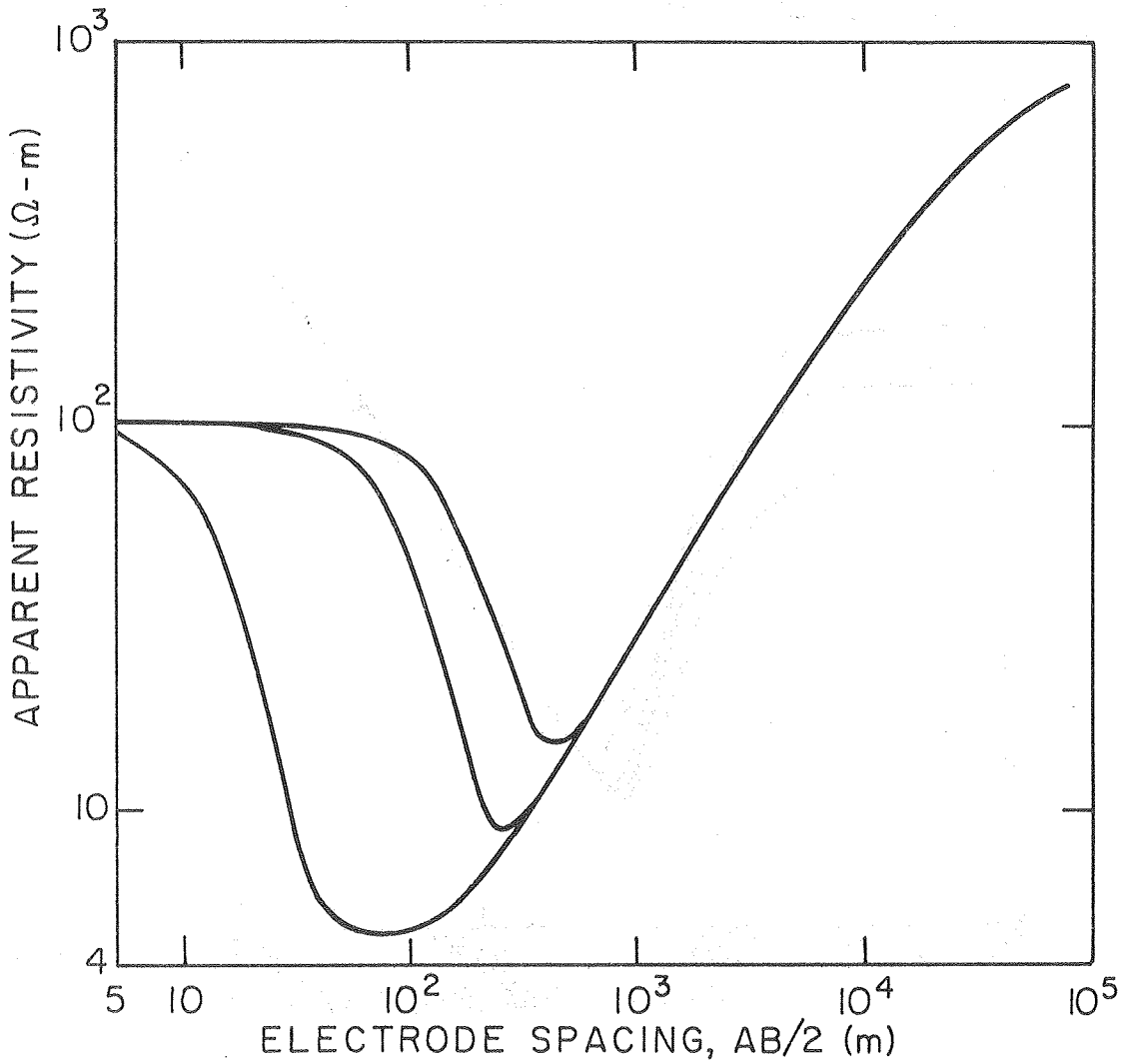

10^3
10^2
10
1
ρ_A (Ωm)
ρ_1 = 100.0 Ωm
50 m
ρ_2 = 3.0 Ωm
100 m
ρ_3 = 1000 Ωm
1
10
10^2
10^3
10^4
10^5
ELECTRODE SPACING, AB/2, (m)
XBL 809-11751

Fig. 12



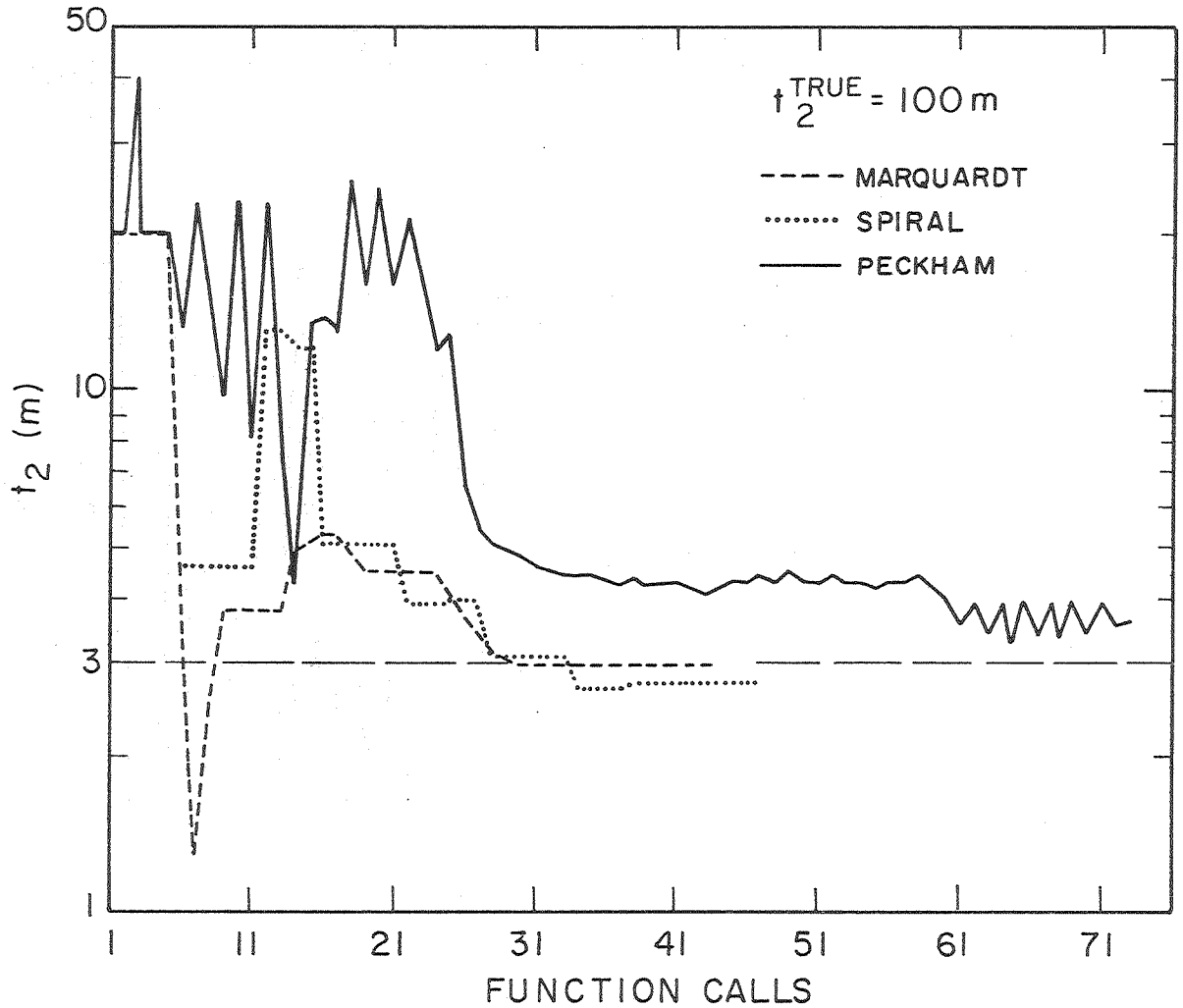
XBL 809-11750

Fig. 13



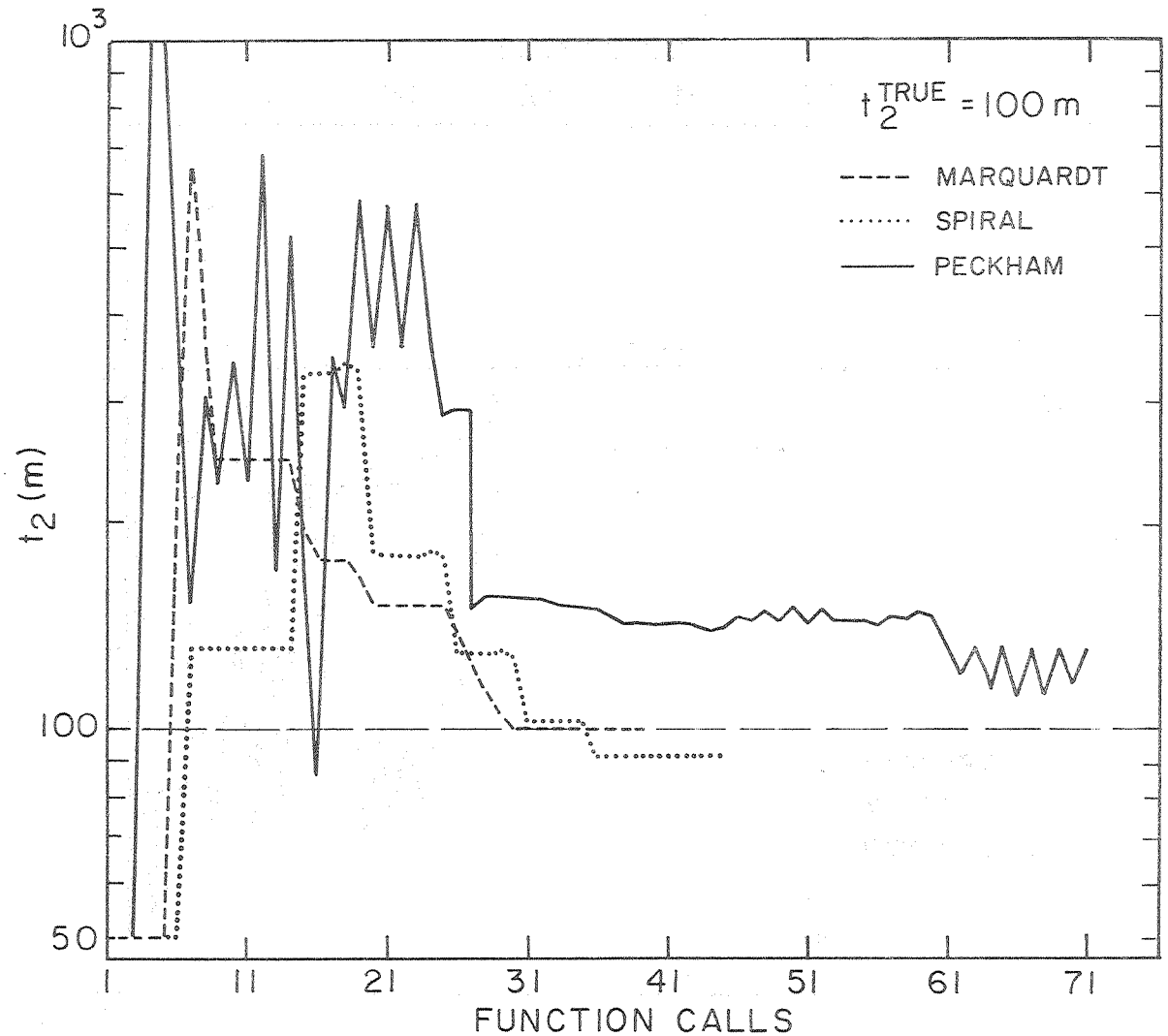
XBL 809-11752

Fig. 14



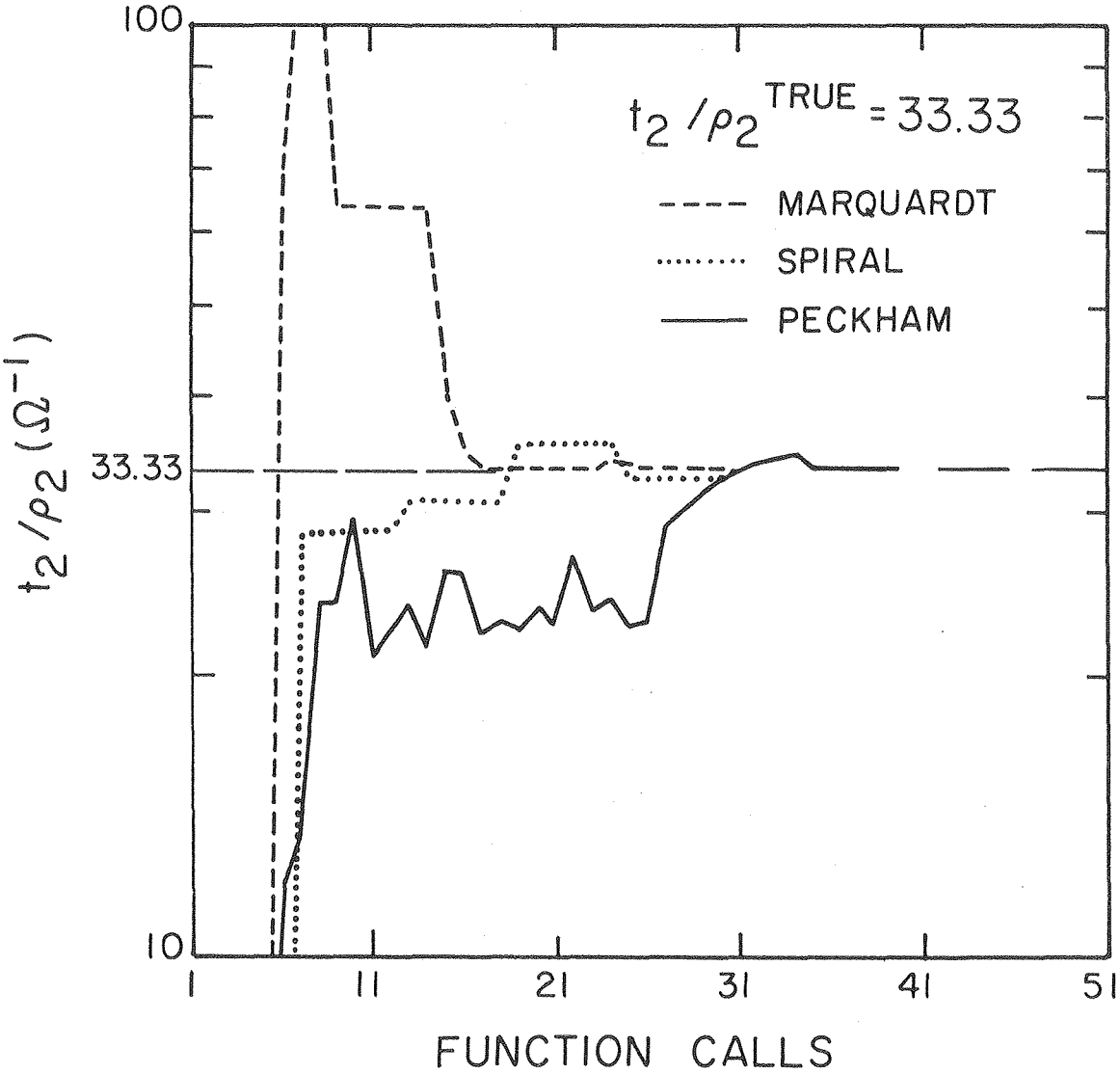
XBL 809-11753

Fig. 15



XBL809-11755

Fig. 16



XBL 809-11754

Fig. 17