

Magnetotelluric Inversion for minimum structure

J. Torquil Smith* and John R. Booker*

ABSTRACT

Structure can be measured in terms of a norm of the derivative of a model with respect to a function of depth $f(z)$, where the model $m(z)$ is either the conductivity σ or $\log \sigma$. An iterative linearized algorithm can find models that minimize norms of this form for chosen levels of chi-squared misfit. The models found may very well be global minima of these norms, since they are not observed to depend on the starting model. Overfitting data causes extraneous structure. Some choices of the depth function result in systematic overfitting of high frequencies, a "blue" fit, and extraneous shallow structure.

Others result in systematic overfitting of low frequencies, a "red" fit, and extraneous deep structure. A robust statistic is used to test for whiteness; the fit can be made acceptably white by varying the depth function $f(z)$ which defines the norm. An optimum norm produces an inversion which does not introduce false structure and which approaches the true structure in a reasonable way as data errors decrease. Linearization errors are often so small that models of σ (but not $\log \sigma$) may be reasonably interpreted as the true conductivity averaged through known resolution functions.

INTRODUCTION

One-dimensional (1-D) inversion remains an important tool for interpreting magnetotelluric (MT) data. There are many instances, particularly at very low frequencies, when multidimensional effects may be approximated by a frequency-independent static distortion and only a 1-D interpretation is necessary (Weidelt, 1972; Larson, 1977). Also, 1-D inversions are routinely performed to constrain starting models for 2-D or 3-D modeling or inversion.

In solving any inverse problem, one seeks not merely a model which fits a given set of data, but also knowledge of what features in that model are required by the data and are not merely incidental to the manner in which the model was obtained. This is particularly important in 1-D models intended as starting points for 2-D or 3-D models, since unconstrained details may persist in later iterations and be mistakenly interpreted as significant structure.

Evaluating what features are resolved has been well studied for the linear inverse problem. Backus and Gilbert (1968) show how to construct averages of models that are uniquely determined by the data. These averages are the truth viewed through peaked resolution functions, whose locations may be varied. Knowledge of the resolution functions and the variances of the averages allows critical evaluation of details in the structure.

The same methods have been applied to nonlinear prob-

lems, like the inversion of MT data, by linearization about models fitting the data (e.g., Parker, 1970; Oldenburg, 1979). Unfortunately, the averages are unique only for models close to the models about which the linearization was made. Oldenburg reports quite different averages for different models of $\log \sigma$ (where σ is the conductivity in S/m) fitting the same MT data when the averaging functions are centered in low-conductivity areas. This result casts doubt on the uniqueness of his averages in the better resolved high-conductivity zones. A later study by Oldenburg (1981), using averages of $\log \sigma$ determined by his linearized $\log \sigma$ model-construction algorithm, reached conclusions regarding the conductivity beneath the Pacific plate which later had to be recanted (Oldenburg et al., 1984).

Given the uncertainties surrounding nonlinear effects in MT inversion, we argue that one should seek models that have the minimum structure possible for some tolerable level of misfit to the data. If a minimum-structure model exhibits a particular feature, we have confidence that that feature is required. Conversely, if a minimum-structure model does not exhibit a particular feature, then that feature certainly is not required by the data.

We also show that the nonlinear errors (i.e., those resulting from linearization) made in interpreting minimum-structure models of σ as averages of the true conductivity through reso-

Manuscript received by the Editor December 1, 1986; revised manuscript received June 9, 1988.

*Geophysics Program, University of Washington, Seattle, WA 98195.

©1988 Society of Exploration Geophysicists. All rights reserved.

lution functions can be quite small, allowing the investigator the use of resolution functions as a means of quantifying the resolution of a data set. In contrast we will show that errors are not small for models of $\log \sigma$ (nor of resistivity ρ), which may explain some of the erroneous conclusions of Oldenburg (1981).

All real data have measurement errors, so that it is generally neither possible nor desirable to fit the data exactly. The chi-squared statistic

$$\chi^2 = \sum_{i=1}^{2N} \left[\frac{\Delta\gamma_i}{\varepsilon_i} \right]^2, \quad (1)$$

where $\Delta\gamma_i$ are the data residuals and ε_i are the data standard errors, is a common measure of the misfit between a model and the data. For 1-D data with independent Gaussian errors, the χ^2 misfit of the data to the truth is distributed as the standard χ^2 for which probabilities are given in most books on statistics. The expected value of χ^2 for the misfit of the data to the truth is $2N$ for $2N$ data points. Parker (1980) shows that when no model fits MT data exactly, the model which minimizes χ^2 (which he calls D^+) consists of delta functions with finite conductance but locally infinite conductivity. Other types of models that approach the same level of misfit develop oscillations. As χ^2 decreases, the oscillations increase as they try to mimic the delta functions of D^+ . Thus, if one seeks models with minimum structure, it is a bad idea to demand that χ^2 be close to its minimum possible value or be much less than the expected value $2N$. In fact, minimum-structure models with greater amounts of misfit (such as the 90 percent or 95 percent confidence limit values of χ^2) may be desired to place more conservative bounds on the amount of structure required.

The χ^2 statistic does not give a complete picture of the misfit. We call a fit which distributes the normalized residuals uniformly across the frequency spectrum a white fit, one that overfits low-frequency data a red fit, and one that overfits high-frequency data a blue fit. It is important that an inversion not systematically overfit some frequency ranges and underfit others. We show that a red fit results in more structure than required at depth for a given χ^2 and less structure than required in the shallow part of the model. We use a robust statistic to test for whiteness and show how to make the fit acceptably white by tailoring the norm that defines the minimum-structure model. Using artificial data, we show that the optimum norm produces an inversion which does not introduce false structure and which approaches the true structure in a reasonable way as the data errors decrease.

We restrict our examples to inversions of artificial 1-D data with Gaussian, zero-mean independent errors of known scale, so that we can compare to the truth and test statistically the residuals to compare different inversions. Considerable caution must be used in interpreting statistical tests made on the residuals left upon inverting real data, since the distributions and scales of the errors may be poorly known and the 1-D assumption is at best an approximation.

CHOOSING THE MODEL VARIABLE AND RESPONSE FUNCTION

Three possible model variables are conductivity σ , resistivity $\rho = 1/\sigma$, and log conductivity $\log \sigma = -\log \rho$. To interpret

a model as a linearly filtered version of the truth, it is essential that errors associated with linearization be small. This cannot be the case for ρ because adding a thin layer of infinite resistance (zero conductance) has no effect on the response, but can produce a vastly different filtered model. The same is true for $\log \rho$ and $\log \sigma$ but not for σ . A physical argument in favor of σ is that it is large in conductors where MT gives the most information and small in resistors where MT gives the least information. Thus, a filtered σ is dominated by regions where we know the most, while a filtered ρ is dominated by regions where we know the least. However, despite being more nonlinear than σ models, $\log \sigma$ models reduce the masking of structure in resistive zones through side-band leakage from conductive zones because they are less variable. Using $\log \sigma$ is somewhat akin to prewhitening in time series analysis. Modeling $\log \sigma$ also ensures that σ will be positive.

Having chosen σ as the model variable for which the inverse problem is most linear, we select an appropriate response to measure based on a heuristic argument. In the 1-D MT problem, assuming a time dependence $\exp(-i\omega t)$ and a piecewise-continuous conductivity $\sigma(z)$, the governing equation for the horizontal electric field E is

$$E'' = -i\omega\mu_0\sigma E, \quad (2)$$

where μ_0 is the permeability and the left side is differentiated twice with respect to the vertical coordinate z . The boundary conditions at the surface and great depth are $E(0) = E_0$ and $E'(z) = 0$, respectively. Integrating once and normalizing by the surface field, we get

$$\frac{E'(0)}{E(0)} = \int_0^{\infty} i\omega\mu_0\sigma(z) \frac{E(z)}{E(0)} dz. \quad (3)$$

We define the complex response

$$\gamma(\omega) = \frac{E'(0, \omega)}{E(0, \omega)} = \frac{i\omega B_y(0, \omega)}{E_x(0, \omega)}, \quad (4)$$

where B_y and E_x are the magnetic and electric fields in orthogonal horizontal directions. (Note that in all other sections of this paper γ has been normalized by dividing by the standard errors of the measurements of γ .) Since γ would be linear in σ if $E(z)$ were independent of σ , γ may be more linear in σ than the response $c = -1/\gamma$ used by Weidelt (1972) and Parker (1970). This motivates our choice. Other choices could be made, but we are doubtful they would give linearization errors as small as those we have obtained.

FINDING MINIMUM-STRUCTURE MODELS

A convenient way to minimize structure is to minimize a norm of a derivative of the model. Models minimizing the first derivative are commonly called "flattest." We define the flattest model as the one that minimizes

$$F(m, f) = \int_0^c \left[\frac{dm}{df(z)} \right]^2 df(z) \quad (5)$$

for a given value of χ^2 , where m is either σ or $\log \sigma$ and the function f controls the norm. The choice of f has effects somewhat similar to the choice of layer thicknesses in the fitting of

layered models. The simplest choice is $f = z$. However, this choice is likely to lead to a red fit with unnecessary structure at depth, because the resolution of MT data generally decreases with depth. The deeper structure required to fit low-frequency data typically has a longer length scale and contributes less to $F(m, z)$. Thus, low-frequency data will be easier to fit and will end up with smaller residuals. To compensate for this effect, one can contract the effective scale of the derivative at depth by choosing $f(z)$ such that

$$\frac{df(z)}{dz} = (z + z_0)^\eta \quad (6)$$

for some η and $z_0 > 0$. Equation (6) is a useful parameterization for f , since it includes the obvious choices of $f = z$ and $f = \log(z + z_0)$. Below, we compare models using $f = z$, $f = \log(z + z_0)$, and $f = -1/(z + z_0)$, corresponding to $\eta = 0, -1$, and -2 .

The constant z_0 in the definitions of f ensures that the integration of dm/df to recover m is not singular. Physically, z_0 is required because the resolution length approaches a constant at the Earth's surface rather than approaching zero. We somewhat arbitrarily choose z_0 equal to half the penetration depth $\text{Re}(c)$ (Weidelt, 1972) for the highest frequency in the data, since we cannot hope to resolve structure much shallower than this. Conceivably, one could adjust the fit of middle frequencies, as compared to high frequencies, by varying z_0 .

Marchisio (1985) (see also Marchisio and Parker, 1984) presents a fully nonlinear inversion which minimizes a quantity that is a bound on $F(\log \sigma, z)$ when the model is close to a uniform slab over an infinitely conducting half-space. While a significant advance in nonlinear inverse theory, Marchisio's solution is not necessarily the flattest and is likely to produce a red fit with structure at depth that is not required by the data. Whittall and Oldenburg (1986) also present several nonlinear inversions which minimize various norms of the impulse response of the model rather than norms of the model itself. This is another step in the right direction, but still falls short of finding truly minimum-structure models.

Constable et al. (1987) present a many-layered, linearized inversion that minimizes the sum of the squared first differences (or second differences) of adjacent layers of their models, for a given misfit. Their inversion minimizing the first differences should give very similar results to one minimizing F , in the limit of vanishing layer thicknesses and a sufficiently deep final layer. Since Constable et al. weight all differences equally, their choice of layer thicknesses (as a function of depth) plays the role of the function $f(z)$ in controlling the "color" of the fit to the data.

We minimize structure directly by minimizing $F(m, f)$ in a stable linearized scheme. Let $m_0(z)$ be the starting model of σ_0 or $\log \sigma_0$ for the current step and $m_1 = m_0 + \Delta m$ be the model considered for the next step. Let γ_i for $i = 1$ to N be the real part and for $i = N + 1$ to $2N$ be the imaginary part of the measured data normalized by their standard errors ϵ_i . Similarly, let γ_{0i} and γ_{1i} be the data predicted by m_0 and m_1 normalized by the standard errors. The normalized misfits $e_{0i} = \gamma_i - \gamma_{0i}$ and $e_{1i} = \gamma_i - \gamma_{1i}$ have total squared misfits χ_0^2 and χ_1^2 , respectively.

If Δm is small, perturbing equation (3) and neglecting second-order terms in Δm gives

$$\gamma_{1i} - \gamma_{0i} = \int_0^\infty g_i(z) \Delta m(z) dz. \quad (7)$$

For $m = \sigma$ and $i = 1$ to N ,

$$g_i(z) = \frac{1}{\epsilon_i} \text{Re} \left\{ i \mu_0 \omega_i \left[\frac{E(\sigma_0, z, \omega_i)}{E(\sigma_0, 0, \omega_i)} \right]^2 \right\} \quad (8)$$

(see Oldenburg, 1979). When $i = N + 1$ to $2N$, one takes the imaginary part and when $m = \log \sigma$, g_i is replaced by $\sigma_0(z)g_i(z)$. Letting

$$\Gamma_i = \int_0^\infty g_i(z)m_0(z) dz, \quad (9)$$

we can write

$$\gamma_{1i} - \gamma_{0i} + \Gamma_i = \int_0^\infty g_i(z)m_1(z) dz. \quad (10)$$

Integrating by parts,

$$\gamma_{1i} - \gamma_{0i} + \Gamma_i + G_i(0)m_1(0) = \int_0^\infty G_i m' dz, \quad (11)$$

where

$$G_i(z) \equiv \int_z^\infty g_i(x) dx \quad (12)$$

and

$$m' = dm/dz. \quad (13)$$

If our goal were to fit the γ_{1i} s to the γ_i s exactly and $m_1(0)$ were known, replacing γ_{1i} with γ_i , equation (11) would provide $2N$ constraints to the minimization of F . However, since our goal is to fit the γ_{1i} s to the γ_i s only to some prescribed χ_i^2 , we replace γ_{1i} by $\gamma_i - e_{1i}$, rewrite equation (11) as

$$\gamma_i - \gamma_{0i} + \Gamma_i - G_i(0)m_1(0) = \int_0^\infty G_i m' dz + e_{1i}, \quad (14)$$

and minimize

$$W(m_1, \chi_i^2, \beta_i) = F(m_1, f) + \beta_i \chi_i^2 \quad (15)$$

with the linearized constraints (14).

In the Appendix we show how to choose β_i so that minimizing $W(m_1, \chi_i^2, \beta_i)$ results in the smallest F for a specified value of χ_i^2 when the linearization inherent in equation (7) is valid. If $m_1(0)$ is also unknown, we solve simultaneously for the $m_1(0)$ which minimizes $W(m_1, \chi_i^2, \beta_i)$.

Our algorithm is a method of keeping the change to the model small enough at each iteration so that the linearization is valid, yet large enough so that the flattest model with the desired χ^2 is arrived at quickly, without an excessive number of forward calculations. The process involves choosing the target $\chi_i^2 < \chi_0^2$, calculating m_1 by minimizing $W(m_1, \chi_i^2, \beta_i)$ using the linearization, and then forward modeling to compute χ_a^2 , the actual χ^2 attained by the model $m_0 + a\Delta m$, and $W(m_0 + a\Delta m, \chi_a^2, \beta_i)$, where $0 < a \leq 1$. If $a\Delta m$ is small enough, the linearization will hold; and $W(m_0 + a\Delta m, \chi_a^2, \beta_i)$ will be smaller than $W(m_0, \chi_0^2, \beta_i)$, its value for the previous model. We then begin another iteration, further reducing the

target, until χ_a^2 reaches our ultimate goal. However, the value of W may increase at any step because $a\Delta m$ is too large for the linearization to hold. The remedy depends on whether the large $a\Delta m$ is due to trying to flatten the model too much in a single step or attempting to decrease the misfit too much. To determine which is the case, we use the linearization to find the model m_f which minimizes $W(m_f, \chi_f^2, \beta_f)$ with β_f selected so that $\chi_f^2 = \chi_0^2$. This produces a Δm_f which flattens the model without reducing the misfit. We then compare the size of Δm_f to $\Delta m - \Delta m_f$. For simplicity, we compare using the maximum of the absolute value of the functions (ℓ_∞ norm). If Δm_f is similar in size to $\Delta m - \Delta m_f$, too much flattening is to blame, and we reduce a by a factor of 2. If Δm_f is much smaller than $\Delta m - \Delta m_f$, too large an attempted decrease in χ^2 is to blame. We must then repeat the minimization with a smaller decrease in the target χ_f^2 . The change $a\Delta m$ can always be made small enough for the linearization to hold, and W will decrease. Then $m_0 + a\Delta m$ is used as the starting model for the next iteration. To avoid unnecessary failed steps, we never choose χ_f^2 less than $0.1\chi_0^2$, nor do we change a by more than a factor of 2 between tries. To be certain that we reach a minimum of W and F , we must iterate until Δm_f is negligible. All the necessary decisions in the process can be made automatically and our algorithm typically converges from a half-space to a model with the expected χ^2 in about eight iterations. These iterations are quite rapid, since only one or two forward calculations are generally needed to find an m_1 which reduces W . (When modeling $\log \sigma$, starting models with unnecessary structure often increase the number of iterations required.)

It is difficult to be sure that we have found the *global* minimum of F . We have tried starting models ranging from half-spaces to smoothed versions of D^+ and have never found a case where the final model depended on the starting model. Thus it seems likely that the $\log \sigma$ models found are globally the flattest. In a previous version of the algorithm, the decision to accept a model $m_0 + a\Delta m$ for use as the starting point for the next iteration was made solely on the basis of improved χ^2 rather than W . With that criterion, the algorithm was occasionally trapped in local minima when the model variable was σ and the starting model was very far from one fitting the data. These minima were easily recognized; χ^2 was very large and the model had large negative values of σ . Since we changed the criterion, this trapping has not recurred. Other convergence problems may occur when the magnitude of a σ model approaches zero with increasing depth. In this case, the magnitude of $|a\Delta m(\infty)|$ is typically found to be of the order of $|\sigma(\infty)|$ or smaller. As $|\sigma(\infty)|$ decreases in successive iterations, the algorithm successively decreases a to keep $|a\Delta m(\infty)| < |\sigma(\infty)|$ and Δm_f may never become negligible. This case occurs when the lowest frequency data are overfit and the best fitting D^+ model ends in a resistor. [Two of the models presented in the results (Figures 7a and 7b) suffer this problem. In these cases we have let the inversion continue until $|\sigma(\infty)| < 10^{-8}$ S/m.] Fortunately, choosing a norm, such as $F(\sigma, -1/z + z_0)$, which does not overfit the lowest frequency data circumvents the problem. Except for these cases, we have never found final σ models which depended on the starting model.

RESULTS AND DISCUSSION

Level of misfit

Requiring too small a misfit requires large oscillations mimicking the best fitting model (D^+). Ideally, we should aim for the misfit that our data has with respect to the true Earth response. However, since we do not know the true Earth's structure, the best we can do is to aim for the expected value, $E(\chi^2) = 2N$. Even this level is not always desirable or possible, because the D^+ misfit may approach or exceed it, particularly when the frequencies are very closely spaced and the misfit of the truth itself is larger than $E(\chi^2)$. Instead one may want to find the flattest model with some higher level of misfit, such as the 95 percent confidence level. Then one can be more confident that the structures which remain in the model are required to fit the data. To illustrate the dangers of overfitting data and other points, we generated 11 frequencies of synthetic data from 3.2×10^{-3} Hz to 1.6×10^3 Hz. Since the disastrous effects of overfitting the data are evident only if there are errors in the data, we added 1 percent Gaussian noise to the synthetic data (Table 1).

In Figure 1, we plot models of $\log \sigma$ which are flattest with respect to $\log(z + z_0)$ (a) with the expected misfit $\chi^2 = 22$ (model 1a), (b) with a much smaller misfit $\chi^2 = 4.73$ (model 1b), and (c) with the 95 percent confidence limit $\chi^2 = 33.9$ (model 1c). (We refer to models by the number of the figure in which they are shown, and the letter of their trace in the figure, e.g., model 1a is shown in trace (a) of Figure 1.) For comparison, in Figure 2 we have plotted the true model, and the locations of the conductances of the best fitting D^+ model ($\chi^2 = 3.75$). The values of the D^+ conductances have been scaled into conductivities by dividing them by the distance between the midpoints to the adjacent spikes. These are the conductivities that would result from redistributing the conductances into uniform layers extending between the midpoints. These scaled conductances have values very close to the true model; this is consistent with the fact that inversion for conductance is well posed (cf., Weidelt, 1985). The true model makes no attempt to fit the errors in the data and has a misfit of 25.6. In this case, fitting to the expected χ^2 (model 1a) recovers essentially all the structure of the true model with the exception of the resistive zone between 1 km and 1.6 km. Fitting only to the 95 percent confidence limit χ^2 (model 1c)

Table 1. Synthetic data expressed in terms of $c = -1/\gamma$, generated from model (a) of Figure 2, with 1 percent Gaussian noise added.

| Frequency (Hz) | Re (c) (m) | Imag (c) (m) | 1 std error (m) |
|-------------------|---------------|-----------------|--------------------|
| 0.0031835 | 39577. | -62879. | 526. |
| 0.0159155 | 10587. | -22403. | 175. |
| 0.0954930 | 2726. | -5866. | 46. |
| 0.3183099 | 1378. | -2359. | 19. |
| 0.9549297 | 722.4 | -1132.3 | 9.5 |
| 3.183099 | 427.4 | -470.8 | 4.5 |
| 6.366198 | 350.1 | -303.1 | 3.3 |
| 15.91549 | 246.2 | -203.2 | 2.3 |
| 63.66198 | 138.82 | -110.77 | 1.27 |
| 159.1549 | 93.94 | -76.43 | 0.86 |
| 1591.549 | 28.34 | -28.40 | 0.28 |

loses resolution but reduces the sensitivity to data errors. Requiring a misfit close to the minimum possible χ^2 requires the false structures in model 1b to fit the noise in the data. Note that most of the extraneous peaks in model 1b correspond to spikes in the D^+ model 2b. This correspondence increases as the misfit of a minimum-structure log σ model approaches the minimum possible.

In comparing misfit statistics, it should be noted that we measure normalized misfit in terms of $\gamma_i = -1/c_i/\epsilon_i$. Parker's D^+ minimizes the squared misfit in terms of $c_i/\tilde{\epsilon}_i$ where $\tilde{\epsilon}_i$ is the estimated standard error in c_i , so D^+ does not necessarily obtain the smallest squared misfit expressed in terms of γ_i . When the relative misfit at each frequency is small (e.g., ≤ 5 percent), the squared misfit is very nearly identical expressed in γ or c . When the relative misfit is larger, the squared misfit of D^+ expressed in terms of γ may be somewhat larger than the minimum possible and may be different from the squared misfit of D^+ expressed in terms of c .

Whiteness of fit

The choice of what norm is minimized can affect the color of the fit significantly. Changing the norm by decreasing η in equation (6) penalizes structure at depth and typically increases the size of the low-frequency residuals relative to the high-frequency residuals, making for a bluer fit of the model to the data. In Figure 3, we compare the truth (model 3d) to flattest models of log σ with respect to z , $\log(z + z_0)$, and $-1/(z + z_0)$, (models 3a, 3b, 3c) corresponding to $\eta = 0, -1, -2$, respectively. All these flattest models have χ^2 misfits equal to its expected value of 22. We plot the normalized residuals associated with the three models and the truth in Figure 4. The model flattest with respect to $\log(z + z_0)$, model 3b, shows the true structure most clearly of the three inversions. The model flattest with respect to z , model 3a, shows fluctuations at depth which are not present in the true model, and the structure is less clearly defined near the surface. Model 3a clearly fits the high frequencies systematically better than the low frequencies (Figure 4). The model flattest with respect to $-1/(z + z_0)$ shows less detail at depth, more fluctuations near the surface, and systematically overfits the high frequencies (Figure 4). Overfitting the low frequencies demands the oscillations at depth, whereas underfitting the high frequencies loses resolution near the surface. The model flattest with respect to $\log(z + z_0)$, model 3b, achieves a fairly even fit, resulting in more accurate detail and fewer extraneous oscillations.

To quantify the color of the fit, we use Spearman's statistic D . This robust statistic is used to test the significance of a trend (cf., Bickel and Doksum, 1977, p. 365-369) and is based on the ranks R_i and S_i of two variables. The samples of a variable are arranged by size and the ordered samples are numbered (e.g., 1, 2, 3, ...). Then the rank of each sample is simply the number of its place in the ordered set. In our case, we let R_i be the ranks of the sums of the squares of the real and imaginary parts of the residuals (normalized by their standard errors) and let S_i be the ranks of the corresponding frequencies. Spearman's statistic D is

$$D = \sum_{i=1}^N (S_i - R_i)^2 \tag{16}$$

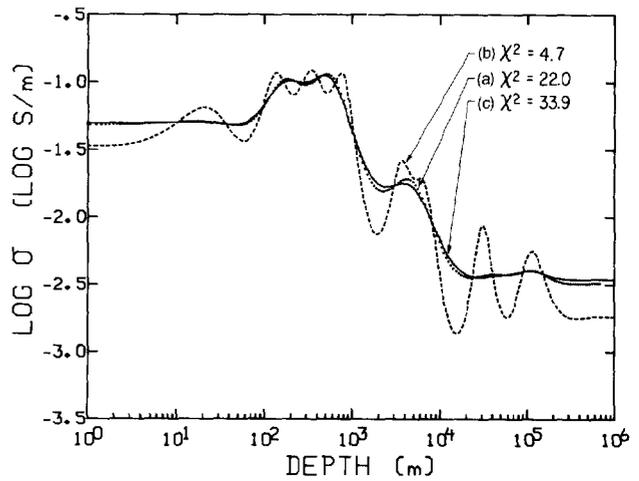


FIG. 1. Models minimizing $F[\log \sigma, \log(z + z_0)]$ fit to data of Table 1, with misfits (a) $\chi^2 = E(\chi^2) = 22.0$, (b) $\chi^2 = 4.73$, and (c) $\chi^2 = 33.9$.

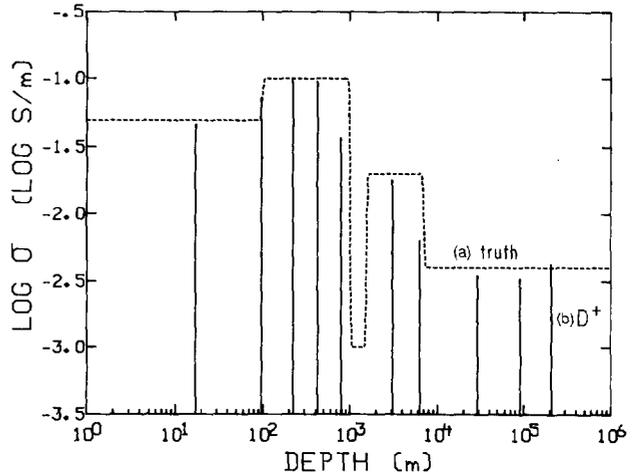


FIG. 2. (a) Model from which data of Table 1 were generated; $\chi^2 = 25.6$. (b) Conductances of best-fitting D^+ model scaled into conductivities by dividing by the midpoint distances between the conductance spikes; $\chi^2 = 3.75$.

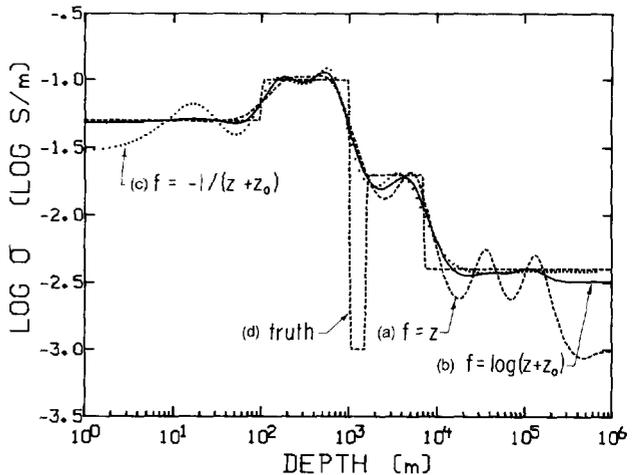


FIG. 3. Models all with misfit $\chi^2 = 22.0$ fit to the data of Table 1, minimizing (a) $F(\log \sigma, z)$, (b) $F(\log \sigma, \log(z + z_0))$, and (c) $F(\log \sigma, -1/(z + z_0))$. (d) Model from which data were generated; $\chi^2 = 25.6$.

and is equivalent to a correlation coefficient between the two sets of ranks. Low values of D correspond to positive correlations and high values to negative correlations. A statistic based on the ranks is more robust than one based on the actual values, because it does not depend on distributional assumptions (such as the errors being Gaussian) and is invariant to transformations of either variable as long as the transformations conserve order. (Our use of D does, however, require that the errors in the data be independent.)

Standard tables exist for the distribution of D (Lehman, 1975, p. 433). For no correlation between R and S , the distribution of D is symmetric about its expected value $E(D)$, which is $(N^3 - N)/6$, and has variance $\text{var}(D) = N^2(N+1)^2(N-1)/36$. For large N , the distribution of $[D - E(D)]/[\text{var}(D)]^{1/2}$ is approximately normal for no correlation between R and S . In the case of 11 frequencies, $P(D \leq 102 \text{ or } 338 \leq D) = 0.094$, since $P(D \leq 102) = 0.047$ and $E(D) = 220$. Also $P(D \leq 84 \text{ or } 356 \leq D) = 0.048$. We may conclude with a 90 percent confidence level that there is a trend when $D \leq 102$ or $D \geq 338$, and with a 95 percent confidence level when $D \leq 84$ or $D \geq 356$.

After we have normalized the observations to have unit variance, the actual squared data errors should be randomly distributed and uncorrelated with frequency. The presence of a trend in the residuals may indicate a frequency-dependent misestimation of the errors in the data, a failure of the 1-D assumption, or a failure to model the data adequately due to a systematic bias in an inversion routine. (By bias we mean the tendency to fit some frequency ranges better than others.) One certainly should rule out the last possibility before invoking either of the first two as probable causes of an observed trend.

Since D may vary by approximately $(\text{var}(D))^{1/2}$ from its expected value when no trend exists, as it does for the true model (3d), we do not require that $D = E(D)$ exactly for a model to be acceptable. Values of D for the residuals of the models of Figure 3 are listed in Table 2. Either model 3b or 3c is acceptable at a 90 percent confidence level on the basis of D . With synthetic data, we can compare the obtained values of D to the value from the residuals to the truth, to check for biases of an inversion algorithm. The value of D for the truth lies about 0.4 of the way between those given by using $f(z) = \log \sigma$ and $f(z) = -1/(z + z_0)$. Thus an η of about -1.4 should give the least biased fit to this data set. With real data one should check that the values obtained are within a range that has a reasonable chance of occurring (e.g., $P \geq 0.1$). If D indicates a trend, one then can evaluate the bias of the inversion algorithm by inverting synthetic data at the same frequencies, with the same scale errors as the real data, generated from a model that at least roughly fits the data.

Our experience is that models minimizing $F[\log \sigma, \log(z + z_0)]$ tend to have values of D close to the value given

Table 2. Spearman's D for the residuals left by the models shown in Figures 3 and 7, showing the effect of choice of depth function $f(z)$. For 11 frequencies, $E(D) = 220 \pm 118$ (90 percent confidence limits), if no trend is present.

| $f(z)$ | z | $\log(z + z_0)$ | $-1/(z + z_0)$ | Truth | Variable |
|----------|-----|-----------------|----------------|-------|----------------|
| Figure 3 | 86 | 256 | 322 | 284 | $\log(\sigma)$ |
| Figure 7 | 54 | 110 | 300 | 284 | σ . |

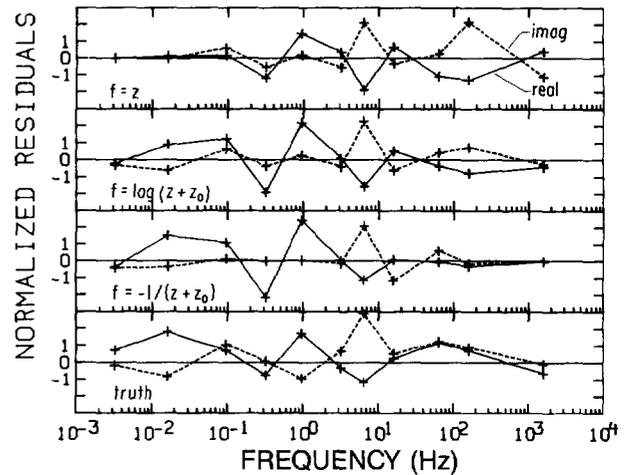


FIG. 4. Residuals of models of Figure 3, from top to bottom (3a)–(3d), normalized by the standard errors of the data. (—, real part; ---, imaginary part.)

by the truth, so this is a very good choice of F . Our experiments used logarithmically spaced data and fairly uniform error estimates; for less uniformly distributed data, this may not be as good a choice. For models minimizing $F(\sigma, f)$, such as those shown in Figure 7, there does not appear to be a single best choice of $f(z)$ independent of the true conductivity. For models which are resistive at depth, using $F[\sigma, \log(z + z_0)]$ tends to overfit the low frequencies, since it does not penalize structure in resistive regions as much as using $F[\log \sigma, \log(z + z_0)]$ does. Minimizing $F[\sigma, \log(z + z_0)]$ fits the data more uniformly when the true conductivity is more uniform.

To avoid unnecessary structure at some depths and insufficient structure at other depths, we must reject any models for which Spearman's statistic indicates a red or blue fit, regardless of how the models are obtained. At the very least, we must exercise caution in interpreting the deeper portions of models for which D takes on small values, since they may contain large oscillations due to fitting the errors in the data. One must also realize that these models may not have the neces-

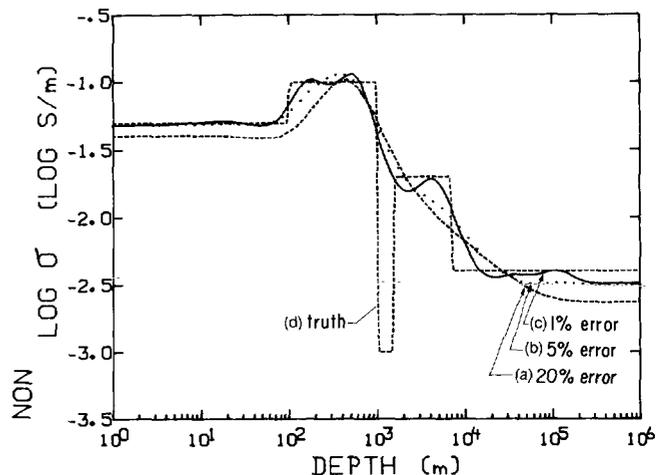


FIG. 5. Models minimizing $F[\log \sigma, \log(z + z_0)]$ for data sets generated from model 2a (Figure 2), each with $\chi^2 = E(\chi^2) = 22.0$, for three different levels of error: (a) 20 percent error, (b) 5 percent error, and (c) 1 percent error. (d) Model from which data were generated; $\chi^2 = 25.6$.

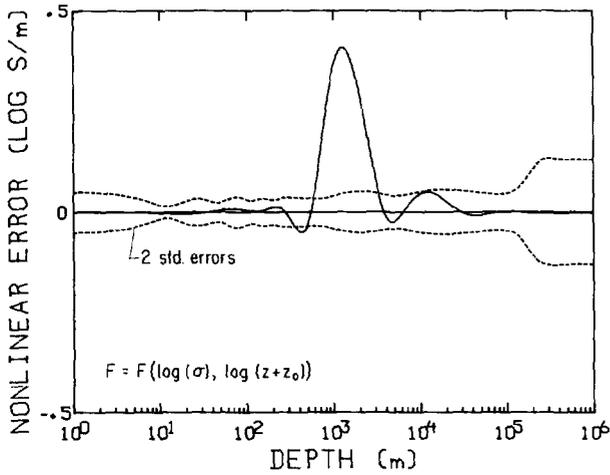


FIG. 6. Nonlinear error versus depth for model 3b (Figure 3) for which the model variable is $\log \sigma$, plotted with an envelope of ± 2 linear standard errors.

sary shallow structure to fit the high-frequency data adequately. Similarly, for models with large values of D , the shallow portions may contain large oscillations, due to overfitting the errors in the high-frequency data, and deep structure, which may not adequately fit the lower frequency data.

Effect of error level on resultant models

As the errors in MT data decrease, flattest inversions reproduce the true structure with increasing fidelity. We have generated three sets of synthetic data with 20 percent, 5 percent, and 1 percent errors added. The frequencies and the true model are the same as for our first set of data. In Figure 5, we plot models minimizing $F[\log \sigma, \log(z + z_0)]$ for the three data sets, fitting each model to the expected χ^2 . As expected, we resolve more details of the true conductivity as the level of errors in the data decreases.

In inverting data, it is essential that the estimates of the errors in the data be accurate. If the estimates of the errors are too large, then the estimated χ^2 misfit [equation (1)] will be too small; fitting to the expected χ^2 may underfit the data, losing resolution. Worse yet, if the estimated variances are unrealistically small, even fitting only to the 95 percent confidence limit χ^2 may be overfitting the data, and may require false structures to fit the noise in the data. Egbert and Booker (1986) have shown that GDS transfer function estimates found by conventional nonrobust methods often have unrealistically small error estimates due to violations of the assumptions of uncorrelated Gaussian errors implicit in the standard methods. Similar results are likely to hold for MT impedance error estimates, so robust transfer function estimation methods such as those used by Egbert and Booker (1986) or Chave et al. (1987) should be used.

Nonlinear error in flattest models considered as averages of the truth

The flattest model $m(z)$ can be shown to be the truth $M(z)$ smoothed through a resolution function, plus a nonlinear error and a stochastic error. The term neglected in writing equation (7) for the change between the flattest model and the truth is

$$e_{L,i} = \gamma_{M_i} - \gamma_{m_i} - \int_0^x g_i(z, m)[M - m] dz, \quad (17)$$

where γ_{M_i} and γ_{m_i} are the normalized data predicted by the truth and the flattest model. We call $e_{L,i}$ the linearization error.

In vector form, if M is the true earth, then our measured data γ are the sum of the true data γ_M and the data errors e_d ,

$$\gamma = \gamma_M + e_d. \quad (18)$$

Letting $\Delta\gamma$ be the residual from fitting the model m , we also have that

$$\gamma = \gamma_M + \Delta\gamma. \quad (19)$$

Assume that our iterative inversion process has converged so that Δm is negligibly small. Then the starting model for a step m_0 and the resulting model m are identical. Using equations (18) and (19) with equations (7), (9), and (A-12), we get

$$m(z) = \int_0^z \mathbf{B}(z)' \mathbf{g}(z_0, m) M(z_0) dz_0 + \mathbf{B}'e_L + \mathbf{B}'e_d, \quad (20)$$

where

$$\mathbf{B}' \equiv \mathbf{A}' + \alpha' - \mathbf{A}'\mathbf{G}(0) \alpha'. \quad (21)$$

(\mathbf{A}' and α' are defined in the Appendix.) Equation (20) characterizes the flattest models as the true earth M smoothed through the resolution function $\mathbf{B}(z)'\mathbf{g}(z_0)$ plus the nonlinear error $\mathbf{B}'e_L$ and the stochastic error $\mathbf{B}'e_d$. The nonlinear error made in interpreting the model as a filtered version of the truth is given by propagating e_L in our model estimates in exactly the same way that random errors e_d propagate; i.e., $\mathbf{B}'e_L$. This procedure was used effectively in the seismic travel-time problem by Pavlis and Booker (1983). The nonlinear error is just the difference between the flattest model and the truth smoothed through the resolution kernels of the flattest model, with a correction for the differences that the linear theory predicts should be due to the differing responses of the two models [cf., equation (17)]. We emphasize that nonlinear error and linearization error are only relevant to interpreting flattest models as averages, and that minimization of W does not depend on having small linearization errors.

The resolution kernels may be used to display the inherent resolution limitations of a data set. In addition to this, given a flattest model and a set of resolution kernels, one might be tempted to try to deconvolve the flattest model to obtain the truth. This is not possible even assuming that both error terms are negligible ($e_d = e_L = 0$). In this case, both the truth and the flattest model averaged through the resolution kernels yield the same flattest model, so a unique deconvolution is impossible. Since resolution kernels have been presented for MT inversions previously (see for example Parker, 1970, or Oldenburg, 1979), we will consider only the nonlinear errors inherent in their use.

When the truth contains large variations not resolved by the data, such as the resistive zone between 1.0 km and 1.6 km included in our test case, the magnitude of the linearization error inherent in interpreting models of $\log \sigma$ to be averages of the true $\log \sigma$ is large. The nonlinear errors $\mathbf{B}'e_L$ for $\log \sigma$ models 3a, 3b, and 3c are extremely similar, so we plot only one (3b) in Figure 6. For comparison we have plotted an envelope of ± 2 standard errors (linear stochastic error) of the

Table 3. Squared linearization error for various models, comparing the effects of the choice of model variable. (Parenthetic values omit the lowest frequency.)

| Model | 3a | 3b | 3c | 7a | 7b | 7c | 10a |
|-----------|---------------|---------------|---------------|-----------------|-----------------|----------|----------|
| Variable | $\log \sigma$ | $\log \sigma$ | $\log \sigma$ | σ | σ | σ | σ |
| $ e_L ^2$ | 3746. | 2488. | 1724. | $> 10^6$ (3.02) | $> 10^8$ (1.52) | 0.467 | 1.72 |

model interpreted as averages through resolution functions. The nonlinear error is greatest near the depth of the unresolved layer, where it is much larger than the uncertainties in the averages due to random noise in the data. The squared magnitude of the linearization errors $|e_L|^2$, listed in Table 3, is large for all the $\log \sigma$ models. Each has large squared errors (> 40) at each of the seven lowest frequencies. These errors are due principally to the unresolved resistive layer.

The nonlinear error made in interpreting models of σ as averages of the true conductivity need not be so large. Figure 7 shows models minimizing $F(\sigma, f)$ for comparison with the $\log \sigma$ models in Figure 3. In Figure 8 we plot the nonlinear error as a function of depth with envelopes of ± 2 standard (stochastic) errors for σ models 7a and 7b. For model 7c, the nonlinear errors (not shown) are less than 0.5 standard errors at all depths. For the models 7a and 7b the nonlinear errors are smaller than 1 standard error at all depths above 500 km. The magnitude of the nonlinear error increases below 500 km for models 7a and 7b, reflecting the fact that the data no longer constrain the model enough for the Frechet kernels g_i to be similar for different models fitting the data. The increase in nonlinear error at depth reflects huge linearization errors at the lowest frequency ($> 10^6$ for models 7a and 7b). The sums of the squared linearization errors of all frequencies except the lowest are only 3.02 and 1.52 for models 7a and 7b, respectively. Even with very large linearization errors at the lowest frequency in these two examples, the nonlinear errors in interpreting the σ models as averages are insignificant at all depths of interest.

We suspect that nonlinear errors should be largest when the integrated conductivity of a flattest model (as a function of

depth) is furthest in some sense (or norm) from that of the truth. The integrated conductivities of two admissible profiles may easily differ at great depth, where the data no longer constrain the conductivity in any way. Parker (1981) has shown that one can often find models terminating in an infinite conductance which still fit the data within an acceptable χ^2 misfit. For these models, the conductivity below the infinite conductance has no effect on the data. Parker calls the shallowest level where one can place the infinite conductance, while still having χ^2 less than or equal to the 95 percent confidence limit of χ^2 the "maximum depth of inference." Models with an infinite conductance can never be linearly close to any model lacking the infinite conductance, since the

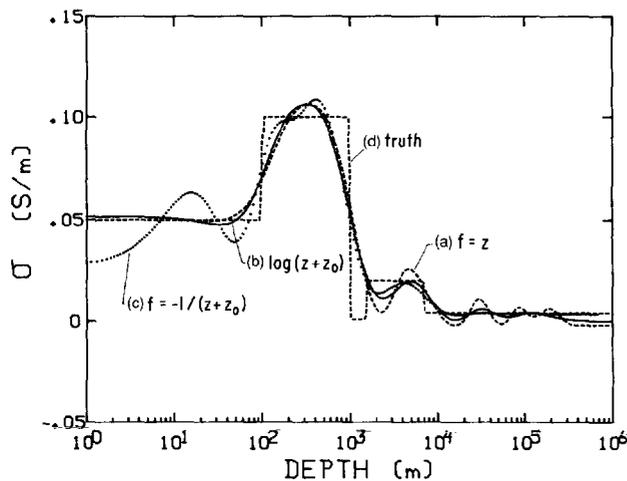


FIG. 7. (a, b, c) Models all with misfit $\chi^2 = 22.0$ fit to data of Table 1: (a) minimizing $F(\sigma, z)$, (b) $F[\sigma, \log(z + z_0)]$, and (c) $F[\sigma, -1/(z + z_0)]$. (d) Model from which data were generated; $\chi^2 = 25.6$.

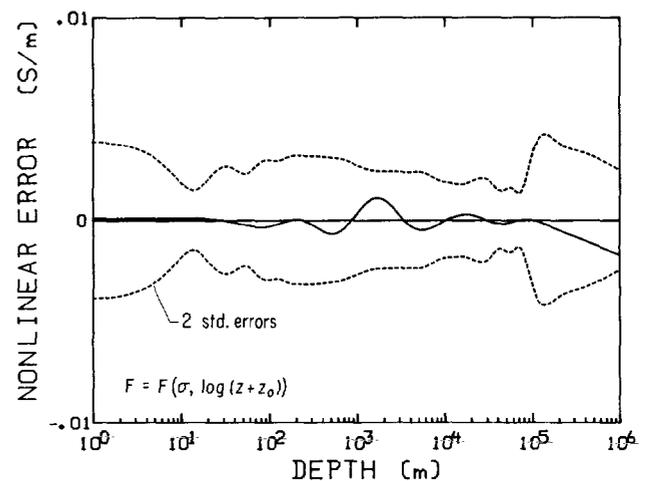
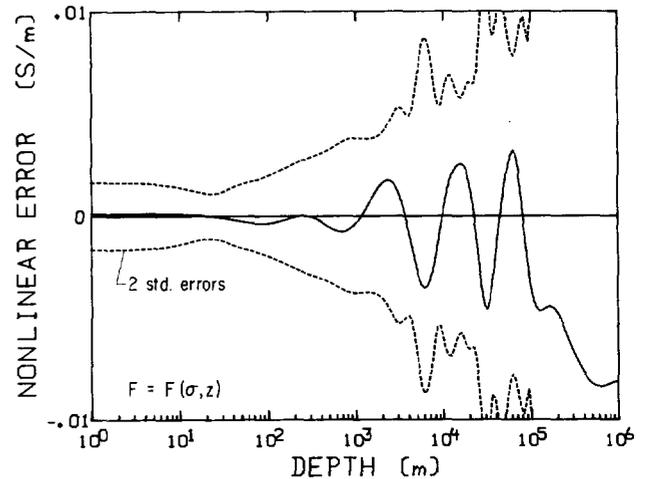


FIG. 8. (a, b) Nonlinear error versus depth for models 7a, and 7b (Figure 7), for which the model variable is σ , plotted within an envelope of ± 2 linear standard errors.

latter have Frechet kernels g_i that are nonzero at all finite depths. As finite conductivities at the bottom of a model are increased, the electric fields (from which the Frechet kernels are calculated) are excluded from the high-conductivity region and the data are less affected by the change in conductivity than the linearized theory would predict. An example of increased nonlinear errors due to this effect follows at the end of the next section. The practical meaning is that the kernels obtained by linearization may tend to overestimate the effects of a large increase in conductivity at great depth. In light of this, we cannot use the averages through the resolution kernels to exclude the possibility of large increases in conductivity near or below the maximum depth of inference.

A more fundamental concern remains: Within the depth range for which a data set contains information, how large an effect can variations of integrated conductivity have on nonlinear errors? This concern would be best addressed by considering the nonlinear errors that would be indicated if the true conductivity were one of the D^+ models minimizing or maximizing conductance for a given level of χ^2 (Weidelt, 1987). Since the code to compute these models is not widely distributed, we consider instead the nonlinear errors that would be indicated if the best-fitting D^+ model were actually the true conductivity. The results of this exercise for model 7b are shown in Figure 9. The nonlinear error is substantially larger than two standard errors in many parts of the model. Despite this, D^+ smoothed through the resolution kernels of the model still bears a strong resemblance to the model, indicating that in this case the data may constrain the integrated conductivity well enough for resolution kernels to remain a worthwhile means of expressing the resolution properties of the data. Although in each case, this exercise considers the nonlinear errors indicated for only one of the infinite number of possible candidates for the true conductivity, it provides an indication of how poor the linearization may be if the conductance of the truth is distributed in as uneven a manner as that of D^+ . We have not experimented much with this exercise, but it seems probable that more reasonable candidates for the truth may be expected to yield smaller linearization errors.

Log σ models recast as σ models

Models minimizing a derivative of $\log \sigma$ are formulated as averages of $\log \sigma$, not as averages of σ , so interpreting them as averages is subject to large nonlinear errors. By reformulating the problem slightly, it is possible to find averages of σ that share the desirable characteristics of nonnegativity and reduced variability in resistive zones that averages of $\log \sigma$ have, but which avoid the nonlinearity of $\log \sigma$. Instead of minimizing $F[\log \sigma, f(z)]$, we minimize

$$F_2 \left[\sigma, 1/\sigma_0, f(z) \right] \equiv \int_0^\infty \left[\frac{1}{\sigma_0(z)} \frac{d\sigma}{df(z)} \right]^2 df(z), \quad (22)$$

where, at any iteration, $\sigma_0(z)$ is the conductivity profile of the previous iteration or alternatively the conductivity profile found by directly minimizing the derivative of $\log \sigma$. These minimizations would be equivalent to minimizing the derivative of $\log \sigma$, except that the weight function is held constant with respect to variations $\delta\sigma$ in any single iteration of the inversion. Treating $1/\sigma_0(z)$ as a weight function in equation

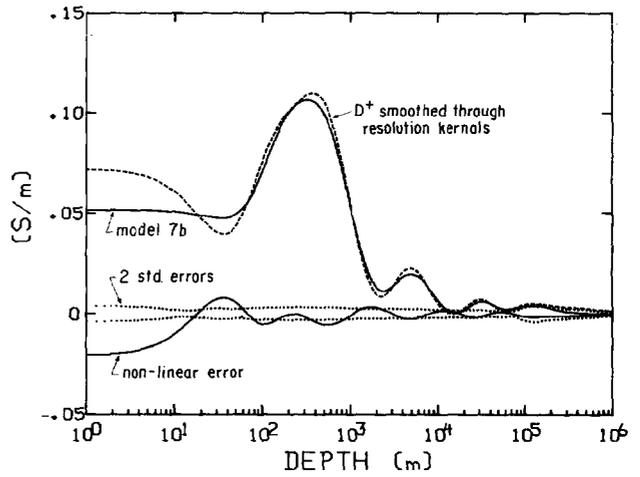


FIG. 9. (a) Model 7b. (b) D^+ model 2b smoothed through resolution kernels of model 7b. (c) Nonlinear error for model 7b with D^+ model 2b considered as true model. (d) ± 2 linear standard errors of model 7b.

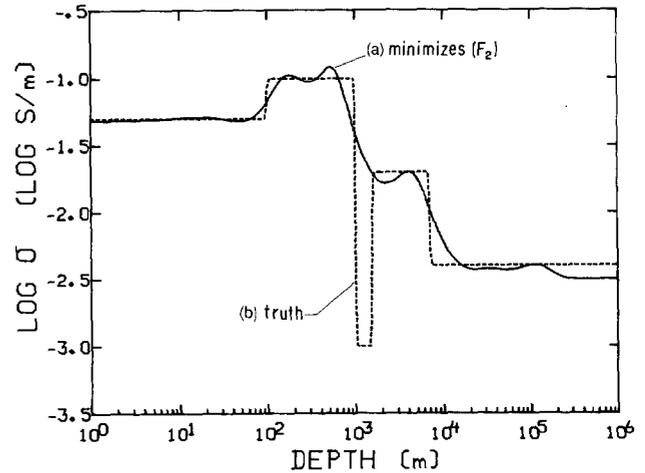


FIG. 10. (a) Logarithm of model minimizing $F_2[\sigma, 1/\sigma_0, \log(z + z_0)]$ with $\chi^2 = 22.0$, and where σ_0 is the conductivity at the previous iteration. (b) True model.

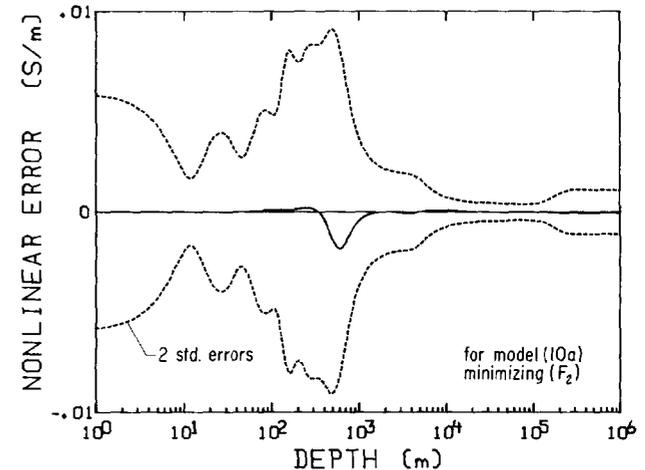


FIG. 11. Nonlinear error of model 10a, plotted with an envelope of ± 2 linear standard errors.

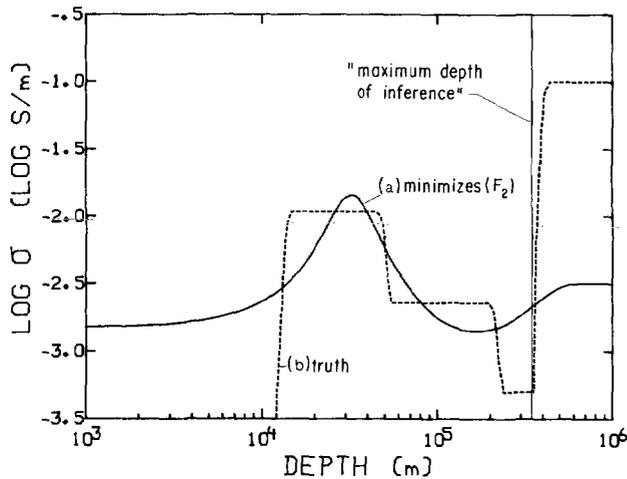


FIG. 12. (a) Logarithm of model minimizing $F_2[\sigma, 1/\sigma_0, \log(z + z_0)]$ with $\chi^2 = 43.8$, and where σ_0 is the conductivity at the previous iteration. (b) True model.

(5), equation (20) lets us interpret models minimizing this weighted norm as averages of the true conductivity through resolution functions. In Figure 10 we plot the logarithm of the conductivity model minimizing this norm with $f = \log(z + z_0)$ and σ_0 as the model of the previous iteration. The log of this model is almost identical to model 3b, so we may be certain that differences in $|e_L|^2$ are due to our choice of model variable (σ or $\log \sigma$), not due to differences in the models themselves. For the σ model 10a, $|e_L|^2$ is only 1.72, negligible compared to the data errors and three orders of magnitude smaller than $|e_L|^2$ for the comparable $\log \sigma$ model 3b. In Figure 11 we plot the nonlinear error inside an envelope of ± 2 standard (linear) errors of the averages. As expected, the nonlinear error is negligible compared to the stochastic uncertainties.

A final example (Figures 12 and 13) demonstrates the increase in linearization error for the case when the true conductivity increases greatly below the maximum depth of inference. We have generated 15 frequencies of artificial data from model 12b, a slightly smoothed version of a uniform slab model provided by Parker (1983). The frequencies are the same as for the COPROD data set (cf., Parker, 1983), ranging between 5.099×10^{-4} Hz to 3.509×10^{-2} Hz. We have added 15 percent Gaussian errors to the data at the lowest three frequencies and the highest frequency and 5 percent Gaussian errors to the other data, approximating the error levels estimated in the COPROD data. The maximum depth of inference for this data set is 336 km, which is shallower than the rise in conductivity centered at 396 km. In Figure 12, we also plot the logarithm of the σ model that minimizes $F_2[\sigma, 1/\sigma_0, \log(z + z_0)]$, fit to the 95 percent confidence limit of $\chi^2 = 43.8$. (The minimum possible misfit for this data is 30.2, which is greater than $E(\chi^2) = 30$, so we only fit to the 95 percent confidence level.) The linearization error is 2010, which is much larger than the random errors in the data. The nonlinear errors (Figure 13) are largest at the depths of the final good conductor and still exceed the stochastic errors at shallower depths. When the same test is made by inverting synthetic data from a model (not shown) similar to model 11b but with a less conducting final layer of 0.01 S/m, the linearization errors of the resultant

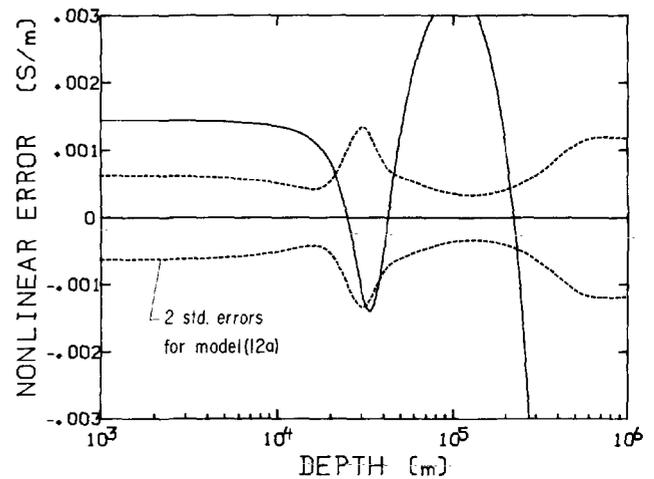


FIG. 13. Nonlinear error of model 12a (Figure 12), plotted with an envelope of ± 2 standard errors.

model (not shown) are reduced to 14.6, and the nonlinear error attains a value of just twice the stochastic standard error in the final layer and is less than the stochastic error elsewhere. The much larger nonlinear errors in the first of these examples are evidently due to the large discrepancies in conductance at great depth. An explanation is clear: side bands of the resolution kernels for the flattest model extend into the conducting region, and the large conductivities there have a large effect on the averages through the resolution kernels, even when the amplitudes of the kernels are very small at depth. Despite large nonlinear errors, both of these models appear to give reasonable averages of the conductivity at depths above the maximum depth of inference, the caveat being that the averages are less affected by conductance near or below the maximum depth of inference than the averaging kernels indicate.

CONCLUSIONS

Tests with synthetic data show that norm minimization may be highly successful in recovering the large-scale features of the true conductivity, even in cases where nonlinear effects may be very large, such as in modeling $\log \sigma$. Features which are not resolved by the flattest models fitting a data set are not necessary, and their existence cannot be determined from the data. Flattest models of conductivity have the further advantage that nonlinear effects are often so small that model values may be interpreted reasonably as the true conductivity averaged through known resolution functions.

No matter how a model is obtained, it is essential that the model fit high-frequency and low-frequency data equally well (a white fit). Failure to assure whiteness results in models with unnecessary structure in some depth ranges and possibly inadequate structure in other depth ranges. We have proposed use of Spearman's statistic D to test against selective overfitting or underfitting of data from either end of the spectrum, while making minimum assumptions about the functional form of any relationship between frequency and residual size. We find that minimizing some norms results in systematic overfitting of low-frequency data, whereas minimizing others does not.

Which norms result in white fits may be somewhat data-dependent (particularly for σ models), so the test of Spearman's D should be made for every inversion to protect against off-white fits.

ACKNOWLEDGMENTS

We wish to thank Gary Egbert for the suggestion of Spearman's statistic and for many useful discussions and S. Constable and an anonymous reviewer for thoughtful review of an earlier version of this paper. Parts of the research were supported by the Department of Energy under Grant DE-FG06-86ER13472 and the National Science Foundation under Grant EAR-8500248.

REFERENCES

Backus, G. E., and Gilbert, J. F., 1968, The resolving power of gross earth data: *Geophys. J. Roy. Astr. Soc.*, **16**, 169–205.
 Bickel, P. J., and Doksum, K. A., 1977, *Mathematical statistics: basic ideas and selected topics*: Holden-Day, Inc.
 Chave, A. D., Thomson, D. J., and Ander, M. E., 1987, On the robust estimation of power spectra, coherences, and transfer functions: *J. Geophys. Res.*, **92**, 633–648.
 Constable, S. C., Parker, R. L., and Constable, C. G., 1987, Occam's inversion: a practical algorithm for generating smooth models from EM sounding data: *Geophysics*, **52**, 289–300.
 Egbert, G. D., and Booker, J. R., 1986, Robust estimation of geomagnetic transfer functions: *Geophys. J. Roy. Astr. Soc.*, **87**, 173–194.
 Larson, J. C., 1977, Removal of local surface conductivity effects from low frequency mantle response curves: *Acta Geodact., Geophys. et*

Montanist. Acad. Sci. Hung., **43**, 897–907.
 Lehman, E., 1975, *Nonparametrics: statistical methods based on ranks*: Holden-Day, Inc.
 Marchisio, G. B., 1985, Exact non-linear inversion of electromagnetic induction soundings: Ph. D. thesis, Univ. of Calif. at San Diego.
 Marchisio, G. B., and Parker, R. L., 1984, Exact nonlinear inversion of electromagnetic induction soundings (abstract): *EOS*, **64**, 692.
 Oldenburg, D. W., 1979, One-dimensional inversion of natural source magnetotelluric observations: *Geophysics*, **44**, 1218–1244.
 ——— 1981, Conductivity structure of the oceanic upper mantle beneath the Pacific plate: *Geophys. J. Roy. Astr. Soc.*, **65**, 359–394.
 Oldenburg, D. W., Whittall, K. P., and Parker, R. L., 1984, Inversion of ocean bottom magnetotelluric data revisited: *J. Geophys. Res.*, **89**, 1829–1833.
 Parker, R. L., 1970, The inverse problem of electrical conductivity in the mantle: *Geophys. J. Roy. Astr. Soc.*, **22**, 121–138.
 ——— 1980, The inverse problem of electromagnetic induction: Existence and construction of solutions based upon incomplete data: *J. Geophys. Res.*, **85**, 4421–4428.
 ——— 1981, The existence of a region inaccessible to magnetotelluric sounding: *Geophys. J. Roy. Astr. Soc.*, **68**, 165–170.
 ——— 1983, The magnetotelluric inverse problem: *Geophys. Surveys*, **46**, 5763–5783.
 Pavlis, G. L., and Booker, J. R., 1983, A study of the importance of nonlinearity in the inversion of earthquake arrival time data for velocity structure: *J. Geophys. Res.*, **88**, 5047–5055.
 Weidelt, P., 1972, The inverse problem of geomagnetic induction: *Z. Geophys.*, **38**, 257–289.
 ——— 1985, Construction of conductance bounds from magnetotelluric impedances: *J. Geophys.*, **57**, 191–206.
 ——— 1987, Bounds on spatial conductivity averages from MT impedances: Presented at the XIX General Assembly, Internat. Union Geod. Geophys., Vancouver.
 Whittall, K. P., and Oldenburg, D. W., 1986, Inversion of magnetotelluric data using a practical inverse scattering formulation: *Geophysics*, **51**, 383–395.

APPENDIX

SOLUTION OF THE LINEARIZED EQUATIONS

The side conditions (14) are most easily applied to the minimization of W when rewritten (in vector form) as

$$\Gamma + \Delta\gamma \int_0^\infty \mathbf{K} \frac{1}{(f')^{1/2}} [m'_0 + \Delta m'] dx + \mathbf{e}, \quad (\text{A-1})$$

where we have dropped the subscript from \mathbf{e}_1 , and where

$$\Delta\gamma' \equiv \Delta\gamma - [m_0(0) + \Delta m(0)]\mathbf{G}(0), \quad (\text{A-2})$$

$$\Delta\gamma \equiv \gamma - \gamma_0, \quad (\text{A-3})$$

and

$$\mathbf{K} \equiv (f')^{1/2}\mathbf{G}. \quad (\text{A-4})$$

For now we treat the surface value $m_0(0) + \Delta m(0)$ as a fixed parameter. Define

$$\mathbf{H} \equiv \int_0^\infty \mathbf{K} \mathbf{K}' dz, \quad (\text{A-5})$$

where the superscript $'$ denotes transpose. \mathbf{H} is symmetric (and positive semidefinite) and may be diagonalized by an orthogonal transformation \mathbf{Q} . Let

$$\mathbf{H} = \mathbf{Q}\boldsymbol{\lambda}\mathbf{Q}'. \quad (\text{A-6})$$

Then, by minimizing $W(m, |\mathbf{e}|^2, \beta)$ with respect to perturbations $\delta\mathbf{e}$ and $\delta[m'/(f')^{1/2}]$, one finds that

$$\frac{m'}{(f')^{1/2}} \equiv \mathbf{K}' \mathbf{Q}[\boldsymbol{\lambda}']^{-1} \mathbf{Q}' [\Gamma + \Delta\gamma'], \quad (\text{A-7})$$

where

$$\boldsymbol{\lambda}' \equiv \boldsymbol{\lambda} + \frac{1}{\beta} \mathbf{I}, \quad (\text{A-8})$$

which may be integrated for m . "Squaring" the misfit vector (which is not displayed here), one gets the squared misfit $|\mathbf{e}|^2$,

$$\mathbf{e}'\mathbf{e} = \frac{1}{\beta^2} [\Gamma + \Delta\gamma']' \mathbf{Q}[\boldsymbol{\lambda}']^{-2} \mathbf{Q}' [\Gamma + \Delta\gamma'], \quad (\text{A-9})$$

where $\boldsymbol{\lambda}'$ depends on β through equation (A-8) and $\Delta\gamma'$ depends on $\Delta m(0)$. Since $|\mathbf{e}|^2$ is a monotonic function of β for $\beta > 0$, this may be solved for β numerically using Newton's method.

The above holds for any choice of $\Delta m(0)$. We use equation (A-7) to form an expression for W and minimize W with respect to changes in $\Delta m(0)$, yielding

$$m_0(0) + \Delta m(0) = \alpha' [\Gamma + \Delta\gamma'], \quad (\text{A-10})$$

where

$$\alpha' \equiv \frac{\mathbf{G}(0)' \mathbf{Q}[\boldsymbol{\lambda}']^{-1} \mathbf{Q}'}{\mathbf{G}(0)' \mathbf{Q}[\boldsymbol{\lambda}']^{-1} \mathbf{Q}' \mathbf{G}(0)}. \quad (\text{A-11})$$

Then combining equations (A-2), (A-7), and (A-10), we have the model which minimizes W for a given choice of β :

$$m(z) = [\mathbf{A}' + \boldsymbol{\alpha}' - \mathbf{A}'\mathbf{G}(0)\boldsymbol{\alpha}'][\boldsymbol{\Gamma} + \Delta\boldsymbol{\gamma}], \quad (\text{A-12})$$

where

$$\mathbf{A} = \mathbf{Q}[\boldsymbol{\lambda}]^{-1}\mathbf{Q}' \int_0^z f' \mathbf{K}(z) dz. \quad (\text{A-13})$$

Equation (A-9) gives the value of β necessary to obtain a specific squared misfit, given a choice of $\Delta m(0)$. To obtain the

pair β and $\Delta m(0)$ that give the flattest model with a specific squared misfit, we solve for β with $\Delta m(0) = 0$ initially, obtain a new $\Delta m(0)$ from equation (A-10), and reiterate, solving for β with the improved $\Delta m(0)$ each time. In practice $\Delta m(0)$ and β converge rapidly, generally in less than five iterations. In the few cases where more than five iterations are needed, we continue iterating using weighted averages of the last two estimates to avoid cyclic repetition. Iterative solution of equations (A-9) and (A-10) is rapid, since it is not necessary to recompute \mathbf{H} , $\boldsymbol{\lambda}$, or \mathbf{Q} .