JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 94, NO. B10, PAGES 14,215–14,225, OCTOBER 10, 1989

Some Comments on Magnetotelluric Response Function Estimation

ALAN D. CHAVE AND DAVID J. THOMSON

AT&T Bell Laboratories, Murray Hill, New Jersey

A new set of computational procedures are proposed for estimating the magnetotelluric response functions from time series of natural source electromagnetic field variations. These combine the remote reference method, which is effective at minimizing bias errors in the response, with robust processing, which is useful for removing contamination by outliers and other departures from Gauss-Markov optimality on regression estimates. In addition, a nonparametric jackknife estimator for the confidence limits on the response functions is introduced. The jackknife offers many advantages over conventional approaches, including robustness to heterogeneity of residual variance, relative insensitivity to correlations induced by the spectral analysis of finite data sequences, and computational simplicity. These techniques are illustrated using long-period magnetotelluric data from the EMSLAB Lincoln line. The paper concludes with a cautionary note about leverage effects by high power events in the dependent variables that are not necessarily removable by any robust method based on regression residuals.

INTRODUCTION

Magnetotelluric and geomagnetic depth sounding theory is based on the assumption that the function space representing the external sources has low dimensionality. This is usually achieved by specifying a form for the external source electric currents; the plane wave or zero wavenumber model is the most common example. A variety of studies have shown that this implies the existence of a set of linear relations between the electromagnetic field components observed at the Earth's surface [e.g., *Berdichevsky and Zhdanov*, 1984; *Egbert and Booker*, this issue]. In the absence of noise, these linear relationships for magnetotellurics may be written

$$\mathbf{E} = \mathbf{Z}\,\mathbf{B} \tag{1}$$

where $\mathbf{E}$ and $\mathbf{B}$ are frequency domain, two vectors of the horizontal electric and magnetic field components at a single site and frequency and $\mathbf{Z}$ is the second rank response or transfer tensor connecting them. The solution of (1) is

$$\mathbf{Z} = (\mathbf{E}\,\mathbf{B}^{\mathrm{H}})(\mathbf{B}\,\mathbf{B}^{\mathrm{H}})^{-1} \tag{2}$$

where the superscript H denotes the Hermitian transpose and the terms in parentheses are the exact cross-power and autopower spectra. Similar response functions may be obtained between the vertical and horizontal magnetic fields or quantities derived from them.

A variety of methods have been proposed for the numerical computation of electromagnetic response functions and their associated errors from a finite realization of the induction process and in the presence of noise. Most of these are based on classical spectral analysis procedures and least squares regression, and are ultimately derived from simple, wide-sense stationary, Gaussian models. It is generally recognized that natural source electromagnetic data exhibit gross departures from this simple situation, including nonstationary phenomena such as geomagnetic storms and outliers caused by both measurement errors and source field inhomogeneity, and it is equally well-known that these disturbances can destroy conventional spectral estimates. This has motivated the development of methods which are robust, in the sense of being relatively insensitive to a moderate amount of nonstationarity and outliers or to small inadequacies of the model, and which react gradually rather than abruptly to perturbations of either. These methods are sensitive only to outliers in the dependent variable (usually $\mathbf{E}$ in a magnetotelluric context), but response functions computed with real data will also be biased downward by noise in the independent variables. This led to the remote reference method of *Gamble et al.* [1979] in which auxiliary observations from a second location, usually of the horizontal magnetic field, are used to minimize bias effects. The success of remote reference methodology is attested to by its nearly universal adoption by magnetotelluric practitioners. While a vast improvement over conventional approaches, outliers in the electric field or noise coherence between the local and reference fields can cause it to fail.

An essential feature of any statistical method is the provision of both an estimate and a measure of its accuracy. The traditional accuracy estimates, or confidence intervals, on spectral quantities are obtained using explicit statistical models which are ultimately based on a Gaussian distribution. Despite this simplifying assumption, the probability distributions associated with spectra are complicated, especially for multivariate problems. These procedures also require auxiliary information such as the correct number of degrees of freedom that is often difficult to obtain due to estimator-induced correlations or the presence of nonstationary signals. These (and other) problems with distribution-based error estimates on variables with complicated properties have led to the development of nonparametric estimators which require few restrictive conditions. The most widely used of the nonparametric estimators is the jackknife which is reviewed in this work.

In this paper, new procedures for the analysis of electromagnetic induction data are proposed which combine some of the best features of the robust and remote reference methods and use the jackknife to provide error estimates. In the next section, a brief review of the principles of robust statistics and its application to response function estimation is given. This is followed by a description of a hybrid robust remote reference estimator and its implementation. A review of the jackknife and its use for the computation of confidence limits on the response functions is then presented. Examples from the long-period EMSLAB magnetotel-

Paper number 88JB03910.
0148-0227/89/88JB-03910$05.00

luric data are used to illustrate the advantages of this approach. The paper concludes with a cautionary note about the effect of leverage by high power events in the magnetic field that may not be removable by robust estimation, along with a suggested diagnostic to detect its presence.

ROBUST RESPONSE FUNCTION ESTIMATION

When **E** and **B** are actual measurements, (1) and (2) do not hold exactly due to the presence of noise from sampling errors, violations of the induction model assumptions, and variability produced by a finite realization of an infinite process. It becomes necessary to estimate the response functions **Z** from imperfect data, and the problem becomes statistical.

In the sequel, it is assumed that simultaneous, finite time sequences of the electric and magnetic field from one or more sites are available. It is also presumed that the time series are collected without aliasing and that any necessary preprocessing to remove trends and gross data errors has been performed correctly. This may include the removal of periodic phenomena like ocean tides in the case of seafloor data, and robust least squares methods are essential for this step in the data preparation. The time series may optionally be prewhitened using an autoregressive or differentiation filter. A subset size is then chosen based on the lowest frequency of interest and a target value for the final degrees of freedom. Each subset is tapered with a data window, Fourier-transformed, and stored; the subsets may be overlapped by an amount that depends on the correlation properties of the data window. The superiority of the discrete prolate spheroidal sequences as data windows is now well-documented. See *Thomson* [1977] for a thorough discussion of the properties of prolate data windows and their use. In the remainder of this paper, the Fourier transforms of windowed subsets of the electric and magnetic field will be referred to as the data. The implementation of robust processing requires the use of the overlapped section-averaging method outlined here and is not amenable to a more conventional, straight band-averaging approach. The subsets need not be contiguous, so that sections of poor quality data may be excluded if necessary. However, it is sometimes desirable to combine band and section averaging to raise the effective degrees of freedom or produce estimates that are evenly spaced in frequency.

There is a large body of literature on response function computation; in addition to standard references on spectral analysis and regression, it includes the papers by *Sims et al.* [1971], *Gamble et al.* [1979], *Larsen* [1980], and *Egbert and Booker* [1986], which explicitly address the magnetotelluric problem. These procedures are based on least squares methods in which the tensor equation (1) is replaced by an equivalent matrix form

$$\mathbf{e} = \mathbf{b}\,\mathbf{z} + \mathbf{r} \tag{3}$$

where there are N observations (i.e., the number of data sections times the number of frequencies in each section is N) so that **e** and **r** are N-vectors, **b** is an $N\times 2$ matrix, and **z** is a two vector. The residual power in (3) is then minimized, yielding

$$\mathbf{z} = (\mathbf{b}^{\mathrm{H}}\,\mathbf{b})^{-1}(\mathbf{b}^{\mathrm{H}}\,\mathbf{e}) \tag{4}$$

The elements of $(\mathbf{b}^{\mathrm{H}}\,\mathbf{b})$ and $(\mathbf{b}^{\mathrm{H}}\,\mathbf{e})$ are the averaged autopower and cross-power spectral estimates based on the available data.

The advantages of least squares include simplicity and the optimality properties established by the Gauss-Markov theorem [e.g., *Kendall and Stuart*, 1977, chapter 19]. For example, linear regression yields the best linear unbiased estimate when the errors in (3) are uncorrelated and share a common variance; this holds independent of any assumptions about their statistical distribution except that it must have a variance. In addition, if the residuals are drawn from a multivariate normal probability distribution, then the least squares result is a maximum likelihood, fully efficient, minimum variance estimate. It should be noted that the latter condition is essential to the computation of confidence limits using a distribution-based approach but is not necessary to obtain the response functions themselves.

With natural source electromagnetic data, the Gauss-Markov assumptions about the error structure are rarely tenable. First, it is often true that the error variance depends on the signal power over at least part of a data series. A large portion of the misfit of the data to the model in (3) is due to source field complications. It is well-known that the energetic early phases of magnetic storms correspond to times of source field complexity; other active events may be produced by morphologically complex, small spatial scale current systems. Second, finite duration, transient features in the geomagnetic field cause outliers to occur in patches, violating the Gauss-Markov independence condition. Finally, the requirement that the errors be normally distributed is often untenable. Due to marked nonstationarity, departures from the model that produce very large residuals are likely, and such outliers are poorly described by a Gaussian model. This situation virtually guarantees problems with conventional least squares. As a consequence, data analysts have adopted a variety of screening techniques ranging from inspection of the data for outliers to ad hoc coherence weighting or high power event rejection in an attempt to alleviate the limitations of conventional approaches. True robust methods can accomplish this under more general circumstances and in a more automatic fashion.

The least squares or Gaussian maximum likelihood solution (4) to (3) was obtained by minimizing the L_2 norm $\mathbf{r}^{\mathrm{H}}\cdot\mathbf{r}$ of the residuals. Some of the problems with the L_2 norm treatment of electromagnetic data have already been noted. The first of these, the occurrence of unequal error variance, is easily detected by plotting the residual power against the power in the magnetic field and noting any correlations. It can be treated using weighted least squares where the rows of (3) are scaled by the inverse of the total power in the same row of **b**. This is discussed in the context of geomagnetic depth sounding by *Egbert and Booker* [1986], but has rarely been found by the authors to suffice for magnetotelluric data, probably because the error variance structure is itself nonstationary. In addition, the existence of a fraction of residuals whose distribution is anomalous compared to the remainder seems to be ubiquitous and requires additional treatment to achieve robustness. However, in some instances it may be necessary to prescale the data to equalize the error variance; it will be assumed that such preprocessing is applied when required prior to utilizing the robust estimators described in this paper.

The least absolute deviations or L_1 norm, where the sum of the absolute values of the elements of **r** is minimized, is a commonly used robust measure. This has led to the suggestion to substitute it for least squares in many geophysical applications. Such a course of action is not recommended for reasons discussed by *Chave et al.* [1987]. There are several classes of more general robust estimates in current use; the most widely applied are the M estimates which are motivated by analogy to the statistical concept of maximum likelihood. For the present application, M estimation is similar to least squares in that it minimizes a norm of the residuals, but the misfit measure is chosen so that a few extreme values cannot dominate the answer. The M estimate is obtained by solv-

ing $\min\{\mathbf{R}^H \cdot \mathbf{R}\}$, where $\mathbf{R}$ is an N vector whose ith entry is $(\rho(r_i/d))^{1/2}$, d is a scale factor, and $\rho(x)$ is called a loss function. For standard least squares, $\rho(x)=x^2/2$, while for the L_1 estimator, $\rho(x)=|x|$. In general, if $\rho(x)$ is chosen to be $-\log f(x)$, where $f(x)$ is the true probability density function of the regression residuals, then the M estimator is maximum likelihood. In practice, $f(x)$ cannot be estimated accurately from finite samples, and the loss function is chosen on theoretical or empirical grounds. Performing the minimization yields

$$\mathbf{b}^H \Psi = 0 \tag{5}$$

where Ψ is an N vector whose ith entry is the influence function value $\psi(r_i/d)$ and $\psi(x)=\partial_x \rho(x)$. Expression (5) corresponds to the familiar normal equations of least squares.

Many methods exist to solve such equations, but it is easiest to rewrite them as weighted least squares problems and iterate to get a linear approximation. The weighted least squares form of (5) is

$$\mathbf{b}^H \mathbf{w} \mathbf{r} = 0 \tag{6}$$

where $\mathbf{r}$ is the residual in (3) and $\mathbf{w}$ is an $N \times N$ diagonal matrix of weights whose ith element is $w_i=\psi(r_i/d)/r_i$. The weights are computed based on the residuals and scale estimate from the previous iteration to linearize the problem; they are initialized using the residuals and scale from ordinary least squares. Note that since the weights are chosen to minimize the influence of data corresponding to large residuals, the M estimator is data adaptive. The solution of (6) is

$$\mathbf{z} = (\mathbf{b}^H \mathbf{w} \mathbf{b})^{-1} (\mathbf{b}^H \mathbf{w} \mathbf{e}) \tag{7}$$

and the terms in parentheses are identified as the weighted autospectra and cross-spectra analogous to those in (4).

There are a multitude of possible forms for the weight matrix used in (6)–(7). The most widely used is the Huber [1964] weight and is based on the least favorable model for the residual distribution that turns out to be Gaussian in the center and Laplacian in the tails

$$w_i = 1 \quad |x_i| \le a$$
$$w_i = \frac{a}{|x_i|} \quad |x_i| > a \tag{8}$$

A value of $a=1.5$ gives better than 95% efficiency with outlier-free Gaussian data. Downweighting of the data with (8) begins when $x_i=r_i/d=a$, so the scale factor d determines which of the residuals are considered large. It is necessary because (5) is not scale invariant without it, in the sense that multiplication of the data by a constant will not produce an affine change in the solution. The scale estimate must be chosen robustly, and L_2 norm parameters like the standard error are not suitable due to serious sensitivity to outliers. *Chave et al.* [1987] and *Rousseeuw and Leroy* [1987] discuss possible forms for d and show that either

$$d_1 = \frac{s_{IQ}}{\sigma_{IQ}} \tag{9}$$

or

$$d_2 = \frac{s_{MAD}}{\sigma_{MAD}} \tag{10}$$

are suitable, where s and σ refer to the sample and theoretical values of the interquartile distance (IQ) and median absolute deviation from the median (MAD), respectively. The sample interquartile distance or MAD are easily obtained by sorting the residuals. If N real residuals $\{r_i\}$ are converted to order statistics by placing them in the ascending order $r_{(1)} \le r_{(2)} \le \cdots \le r_{(N)}$, then the sample interquartile distance is

$$s_{IQ} = r_{(3N/4)} - r_{(N/4)} \tag{11}$$

while the sample MAD is

$$s_{MAD} = \text{median}\{|r_i - r_{(N/2)}|\} \tag{12}$$

where the median is the $N/2$th order statistic.

Either (9) or (10) require a target distribution for the residuals to get the theoretical entities, and while the Gaussian is usually used as a reference for real data, the complex Gaussian is not necessarily the optimal choice with complex data. It is preferable to measure residual size by the magnitude of the complex residuals because this is rotationally (i.e., phase) invariant. Because the real and imaginary parts of a complex Fourier transform are independent, its magnitude or absolute value is Rayleigh, and this distribution will be used exclusively in the present paper. The real and imaginary parts of the data are identically constrained with the Rayleigh approach, whereas weights based on the Gaussian treat the real and imaginary parts separately, altering the phase of the data. The theoretical Rayleigh interquartile distance and MAD are derived by *Chave et al.* [1987].

Because the weights (8) fall off slowly for large residuals and never descend to zero, they provide inadequate protection against the severe outliers that are observed in magnetotelluric data. However, the loss function corresponding to (8) is convex so that convergence to a local rather than a global minimum cannot occur in (6). A common practice is to use (8) for a few iterations to get a solution near the optimal one while computing a good estimate for the scale d, fix the scale, and switch to the more severe weight function proposed by *Thomson* [1977]

$$w_i = \exp\{-e^{\alpha(|x_i|-\alpha)}\} \tag{13}$$

where the parameter α determines the residual size at which the downweighting begins and is analogous to a in (8). *Chave et al.* [1987] discuss a choice for α which depends explicitly on the number of data, yielding a nearly automatic, data-adaptive result.

To summarize, robust processing of magnetotelluric data begins with data editing and the computation of the windowed Fourier transforms of subsets of a longer time sequence. At each frequency of interest, an initial solution (4) is obtained with ordinary least squares and used to compute the residuals $\mathbf{r}$ and an initial scale estimate using (9) or (10). Several iterations are then performed using weighted least squares (6)–(7) and the Huber weight function (8), with the weights and scale computed at each stage using the residuals from the previous iteration. This is terminated when the weighted residual power $\mathbf{r}^H \mathbf{w} \mathbf{r}$ does not vary above a threshold value. The scale is then fixed at the final Huber value, and several iterations are carried out using the weight function (13). This stops when the weighted residual power does not change appreciably, yielding the final robust estimate of the response functions.

ROBUST REMOTE REFERENCE METHOD

The remote reference method was introduced to magnetotellurics by *Goubau et al.* [1978] and developed by *Gamble et al* [1979]. It is based on classical principles but utilizes additional observations of the electromagnetic field at a site some distance from the station of interest to reduce the influence of noise. The necessary separation of the two measurements depends on the

local geology, the source field behavior, and the instrument noise and will not be considered here; for a relevant experimental study, see *Goubau et al.* [1984].

It is instructive to consider the remote reference method in the context of norm minimization used earlier. Let the local electric (**e**) and magnetic (**b**) fields and the remote magnetic ($\mathbf{b}_R$) field be observed over the same time interval and processed into frequency domain data in an identical manner. At the first site, the response functions **z** satisfy the linear regression

$$\mathbf{e} = \mathbf{b}\,\mathbf{z} + \mathbf{r} \tag{14}$$

while at the second site a similar relationship is

$$\mathbf{e} = \mathbf{b}_R(\mathbf{z}+\delta\mathbf{z}) + \mathbf{r}_R \tag{15}$$

where the dimensions of the vectors and matrices are as for (3). The term $\delta\mathbf{z}$ accounts for any changes in the response function using the remote magnetic field caused by the separation of the electric and magnetic field measurements. The remote reference response functions are computed by minimizing the magnitude of the covariance of the residuals, or the norm $|\,\mathbf{r}_R^H \cdot \mathbf{r}\,|$, yielding

$$\mathbf{z} = (\mathbf{b}_R^H\,\mathbf{b})^{-1}(\mathbf{b}_R^H\,\mathbf{e}) \tag{16}$$

which is the remote reference response function given by *Gamble et al.* [1979]. The equivalence of remote reference estimates with minimization of the absolute covariance of the residuals was also recognized by *J.C. Larsen* (private communication, 1987).

The advantages of the remote reference method are well-known. Since the crosspowers of **b** and $\mathbf{b}_R$ appear in the denominator of (16) rather than the autopower of **b** as in (4), remote reference estimates are in principle not biased downward by noise in the local magnetic field **b**. The remote reference method effectively eliminates this bias if the noise is not correlated between the two sites. It also has some effect in reducing local electric field noise if it is uncorrelated with the remote magnetic field. However, the remote reference method is not robust; if there is correlated noise on all of the channels, then it will break down like least squares. Such noise is obviously not instrumental but must be related to inhomogeneous source structure; see *Egbert and Booker* [this issue] for an example. It should also be noted that the remote reference method is not data adaptive since there is no explicit weighting based on data quality, so that severe contamination of the electromagnetic field at one of the sites cannot necessarily be removed.

The remote reference technique is equivalent to the method of instrumental variables for solving errors-in-variables problems as introduced by *Reiersol* [1945] and is used in the field of econometrics [e.g., *Malinvaud*, 1970]. The errors-in-variables model allows for measurement error in both **e** and the components of **b**, and (3) becomes

$$\bar{\mathbf{e}} = \bar{\mathbf{b}}\,\mathbf{z} + \mathbf{r} \tag{17}$$

where $\bar{\mathbf{e}}$ and $\bar{\mathbf{b}}$ are the true electric and magnetic fields which are related to the measured ones by

$$\begin{aligned} \mathbf{e} &= \bar{\mathbf{e}} + \mathbf{f} \\ \mathbf{b} &= \bar{\mathbf{b}} + \mathbf{g} \end{aligned} \tag{18}$$

and **f** and **g** represent the errors. A least squares solution is valid only if $\mathbf{r} - \mathbf{f} + \mathbf{g}\,\mathbf{z}$ meets the Gauss-Markov conditions. It is not difficult to show that the complete solution of (17)–(18) requires additional information such as the ratio of the noise powers in **f** and **g**; see *Miller* [1986, chapter 5] for details. Accurate estimation of such additional quantities is often difficult, yet errors in their values will bias the solution **z**. The method of instrumental variables instead utilizes a set of auxiliary measurements to overcome the indeterminate form of (17)–(18). Under the conditions that the instrumental variables (e.g., the remote reference magnetic field) are uncorrelated with the noise **g** and that no linear combination of them is orthogonal to **b**, this gives a solution identical to the remote reference result (16). However, it can be shown that the answer is not precise unless the coherences between the reference field, $\bar{\mathbf{b}}$, and $\bar{\mathbf{e}}$ are high and that correlation of the reference measurements with the errors in (18) will produce substantial bias. The instrumental variable estimate is not minimum variance, and confidence limits on the result will in general exceed those for more conventional ones. For details, see *Fuller* [1987].

The effectiveness of robust *M* estimation in dealing with outliers in the electric field and the ability of the remote reference method to remove bias induced by local magnetic field noise suggests that a hybrid approach could be especially fruitful. The implementation of such a combination is conceptually straightforward. The remote reference solution to the weighted least squares problem analogous to (7) may be written

$$\mathbf{z} = (\mathbf{b}_R^H\,\mathbf{w}\,\mathbf{b})^{-1}(\mathbf{b}_R^H\,\mathbf{w}\,\mathbf{e}) \tag{19}$$

where $\mathbf{b}_R$ is the $N\times 2$ matrix containing the remote field data and **w** is the $N\times N$ diagonal matrix containing the robust weights of (7). By hypothesis, the remote and local magnetic fields are identical except for a possible coordinate rotation and uncorrelated noise, so the distribution of the absolute value of the residuals will still be approximately Rayleigh in the absence of outliers. This means that the scale and weight parameters outlined in the last section are applicable. In all other respects, the robust remote reference procedure is like the regular robust one described in the last section.

THE JACKKNIFE

By use of a loose form of the central limit theorem, it is often claimed that Fourier transforms of time series are nearly Gaussian even if the time domain data are not. While this argument may be true for sequences that are drawn from a single statistical distribution, it fails in the presence of mixtures of distributions, nonstationarity, and outliers, particularly when the underlying process has a complicated spectrum. However, assuming that the Gauss-Markov conditions are approximately met, even non-Gaussian frequency domain data can be analyzed with least squares to get reasonably accurate response functions. This does not hold for distribution-based estimates of the errors on the response functions; these are extremely sensitive to departures from normality.

Even for stationary Gaussian processes, the probability distribution for the response functions is a complicated multivariate form of the *t* distribution [*Brillinger*, 1981]. Because of the difficulties in working directly with this distribution, asymptotic or Taylor series approximations are in standard use. The range of validity of this type of expression is often difficult to establish a priori, and its undetected breakdown can lead to erroneous inferences. In addition, distribution-based confidence limits are characterized by a degrees-of-freedom parameter whose estimation must be performed accurately. While this is frequently done by counting Fourier frequency bins and attributing two degrees of freedom to each, the presence of correlations caused by the use of data windows (especially the low-bias types) and section overlapping can alter the answer substantially, and must be corrected for.

Quantitative removal of these effects is not straightforward; see *Thomson and Chave* [1989] for an extended discussion. Finally, the presence of spectral leakage due to data window failure or nonstationarity and outliers can dramatically alter the effective degrees of freedom. It should be noted that all of these complications reduce the true number of degrees of freedom, and unless properly accounted for, the standard distribution-based confidence limits will be too small.

Because they depend explicitly on hypotheses about the probability distribution and auxiliary quantities, the standard confidence limit estimators are parametric. In the field of statistical inference, parametric estimators have been displaced by newer nonparametric types due to the latter's simplicity and better performance in complicated situations. The most frequently applied nonparametric method is the jackknife. For an introduction to the jackknife and related methods, see *Efron* [1982] or *Efron and Gong* [1983]. The jackknife has recently been applied to a variety of spectral analysis problems by *Thomson and Chave* [1989].

Let $\{x_i\}$ be a sample of N observations drawn from some distribution characterized by a statistical parameter θ which is to be computed. For the moment, only a scalar-valued parameter θ will be considered. Denote an estimate of θ based on all of the data by $\hat{\theta}$. The data are then divided into N groups of size $N-1$ each by deleting an entry in turn from the whole set. Let the estimate of θ based on the ith subset, where the ith datum has been removed, be $\hat{\theta}_{-i}$. The pseudovalues are

$$\hat{\Phi}_i = N\hat{\theta} - (N-1)\hat{\theta}_{-i} \tag{20}$$

and serve as substitute jackknife data in standard statistical procedures. The jackknife mean is just the arithmetic average of the pseudovalues

$$\begin{aligned}\tilde{\theta} &= \frac{1}{N}\sum_{i=1}^{N}\hat{\Phi}_i \\ &= N\hat{\theta} - \frac{N-1}{N}\sum_{i=1}^{N}\hat{\theta}_{-i}\end{aligned} \tag{21}$$

>From (21), it is clear that the pseudovalues (20) are not really necessary, and the delete-one estimates can be used directly. However, the pseudovalue form pervades the jackknife literature. The quantity (21) was originally introduced as a lower bias replacement for the regular mean $\hat{\theta}$; see *Efron* [1982] for details. For a statistic θ which is linear in the data, the jackknife value $\tilde{\theta}$ and the conventional one $\hat{\theta}$ will be equivalent. Several studies have shown that the variability of the jackknife mean can be large for some statistics, and $\tilde{\theta}$ should be used as a substitute for $\hat{\theta}$ when it is distinct only with caution.

A more important application for the jackknife is in the nonparametric estimation of the variance of an arbitrary statistic. The jackknife variance is just the standard sample variance of the pseudovalues, and may conveniently be written in terms either of the pseudovalues or of the delete-one estimates as

$$\begin{aligned}\tilde{s}^2 &= \frac{1}{N(N-1)}\sum_{i=1}^{N}(\hat{\Phi}_i - \tilde{\theta})^2 \\ &= \frac{N-1}{N}\sum_{i=1}^{N}(\hat{\theta}_{-i} - \bar{\theta})^2\end{aligned} \tag{22}$$

where

$$\bar{\theta} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_{-i}$$

An important advantage of the jackknife is computational simplicity with complicated statistics; as long as a rule relating the data to θ is available, it can be applied. This should be contrasted to parametric approaches where θ is typically contained in a complicated distribution function which is reduced by maximum likelihood, often yielding a highly nonlinear equation. Another important property of the jackknife variance estimate is its conservatism: *Efron and Stein* [1981] have shown that the expected value of the jackknife variance always slightly exceeds the true value even when the data are not identically distributed.

Under very general conditions, it can be shown that $(\hat{\theta}-\theta)/\tilde{s}$ is asymptotically normally distributed, allowing approximate double-sided $100(1-\gamma)\%$ confidence limits to be placed on θ in the usual way

$$\hat{\theta} - t_\nu(1-\frac{\gamma}{2})\tilde{s} \le \theta < \hat{\theta} + t_\nu(1-\frac{\gamma}{2})\tilde{s} \tag{23}$$

where $t_\nu(1-\gamma/2)$ is a value from Student's t distribution with ν degrees of freedom. When ν exceeds 50, the t distribution can be approximated by the normal one. Note also that the use of an incorrect degrees-of-freedom estimate leads to small errors in (23) for $\nu>10$; this is generally not true for the more complicated distributions used in parametric estimates on spectra. However, asymptotic behavior does not guarantee correct results for finite samples, and *Hinkley* [1977*a*] has shown that substantial errors can accrue if (23) is used blindly on markedly non-normal statistics with small samples. This can be corrected using transformations on $\hat{\theta}$ to get a more Gaussian form. In a magnetotelluric context, the response functions are roughly Gaussian without modification, and nonnormality is not a critical consideration. However, jackknifing of coherence estimates requires an inverse hyperbolic tangent transformation; see *Thomson and Chave* [1989] for details.

Some further complications occur when the jackknife is applied to the regression problems (3) or (6). First, the response functions $\mathbf{z}$ are vector-valued, so the scalar variance (22) becomes a covariance matrix. Second, regression is an unbalanced operation in the sense that the overall sample sizes and possibly the sample variances are different for the vector $\mathbf{e}$ and the matrix $\mathbf{b}$. To apply the balanced jackknife (22) to (3) or (6), (4) or (7) can be used with a row removed in turn to get the delete-one estimates, and (22) can be applied directly after modification to allow for the vector form of θ and replacement of $N-1$ with $N-p$, where p is the number of columns in $\mathbf{b}$, to correct the sample bias. However, *Hinkley* [1977*b*] examined the small sample properties of the balanced jackknife variance applied to regression and suggested that it produces a consistently biased result. He proposed the use of a weighted pseudovalue to eliminate the error. Rewriting (20) in vector form and including the weight gives

$$\mathbf{P}_i = (N(1-h_i)+1)\,\hat{\Theta} - N(1-h_i)\,\hat{\Theta}_{-i} \tag{24}$$

where $\hat{\Theta}$ is the vector statistical parameter replacing $\hat{\theta}$ and the $\{h_i\}$ are the diagonal elements of the hat matrix. Since the hat matrix is important to several points contained in the remainder of this paper, a digression to examine its properties is in order.

The hat matrix is defined by considering the residual $\mathbf{r}$ in (3) as the difference between the observed electric field $\mathbf{e}$ and that predicted by the regression $\hat{\mathbf{e}}$, so that

$$\mathbf{r} = (\mathbf{I} - \mathbf{H})\,\mathbf{e} \tag{25}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{H}$ is the hat matrix given explicitly by

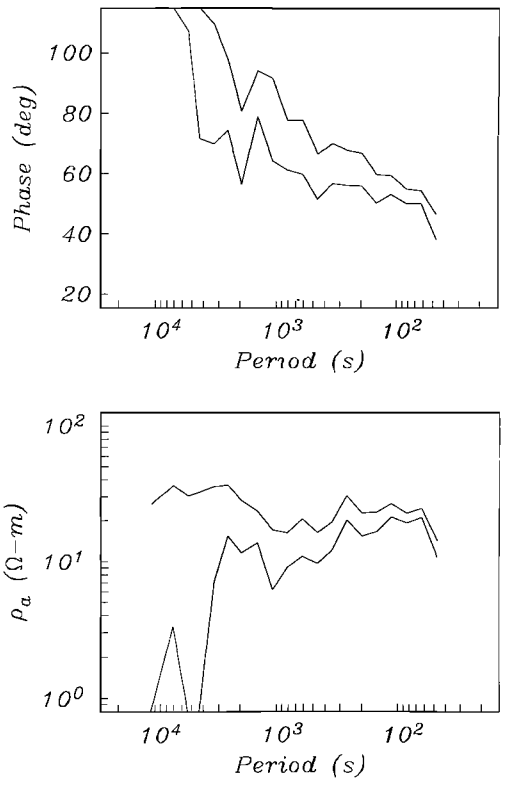

Fig. 1. The phase (top panel) and apparent resistivity (bottom panel) for the TE mode response function at Lincoln line site 4 remote referenced to site 1. The TE mode has the electric field polarized geographic north-south. The bands are the double-sided 95% confidence limits computed using the jackknife estimates of standard error and (29)–(30). Note the variability of the response function and the large error bars; the uncertainty in the phase is typically 15° and for the apparent resistivity it often exceeds 50%.

$$\mathbf{H} = \mathbf{b}(\mathbf{b}^H \mathbf{b})^{-1}\mathbf{b}^H \tag{26}$$

For robust regression using (6)–(7), (26) must be modified by replacing $\mathbf{b}^H$ with $\mathbf{b}^H\mathbf{w}$ to account for the weights. The hat matrix is a projection matrix which maps $\mathbf{e}$ onto $\hat{\mathbf{e}}$. The lack of balance in a regression problem is reflected in its diagonal. *Hoaglin and Welsch* [1978] summarize some of the properties of the hat matrix diagonal that are essential to its interpretation. First, $\mathbf{H}$ is a projection matrix, hence is symmetric and idempotent ($\mathbf{H}^2=\mathbf{H}$). These characteristics can be used to show that $0 \le h_t \le 1$. Second, the eigenvalues of a projection matrix are either 0 or 1, and the number of nonzero eigenvalues is equal to its rank. Since rank($\mathbf{H}$)=rank($\mathbf{b}$)=p, where p is the number of columns in $\mathbf{b}$, the trace of $\mathbf{H}$ is p. This suggests that the average size of h_t is p/N. Finally, the two endpoint values for h_t have special meaning. If $h_t=0$, then the ith predicted value is fixed at zero and not affected by any datum in $\mathbf{e}$. If $h_t=1$, then the predicted and observed values of e_t are identical, and the model fits the data exactly.

Some additional considerations become important for remote reference regression. In this instance, the counterpart to the hat matrix is

$$\mathbf{H}_R = \mathbf{b}(\mathbf{b}_R^H \mathbf{b})^{-1}\mathbf{b}_R^H \tag{27}$$

and the robust version is (27) with $\mathbf{b}_R^H$ replaced by $\mathbf{b}_R^H\mathbf{w}$. This matrix is idempotent but does not possess Hermitian symmetry and is not formally a projection matrix. The diagonal elements of (27) are not necessarily real, unlike those of (26). It is not difficult to show that the magnitudes of the diagonal elements of (27) are bounded by 0 and 1. Furthermore, numerical simulations show that the diagonal elements are nearly real when $\mathbf{b}$ and $\mathbf{b}_R$ are highly correlated and behave like those of (26). In addition, the equivalence of (26) and (27) can be verified using the errors-in-variables model (17)–(18). While these are not rigorous arguments, they can be justified heuristically for the applications considered here. It will be assumed that the magnitude of the diagonal of (27) can be used interchangeably with the diagonal of (26) in applications used in the remainder of this paper.

When $h_t=p/N$ in (24), the balanced pseudovalue is obtained; since this is actually the expected value of h_t, the unbalanced pseudovalue is in general different for finite samples. The jackknife estimate of the regression covariance matrix is

$$\tilde{\mathbf{S}} = \frac{1}{N(N-p)}\sum_{t=1}^{N}[\tilde{\Theta}-\mathbf{P}_t][\tilde{\Theta}-\mathbf{P}_t]^H \tag{28}$$

where $\tilde{\Theta}$ is the arithmetic average of the $\{\mathbf{P}_t\}$ as in (20). The diagonal terms of (28) give the variances of the corresponding terms in $\mathbf{z}$, while the off-diagonal terms are the covariances; note that the errors on the real and imaginary parts are identical. An important property of the jackknife regression variance is robustness in the presence of inhomogeneity of error variance, in contrast to parametric estimators [*Hinkley*, 1977*b*]. Note also that (22) or (28) do not require an explicit accounting for the effective degrees of freedom. In the presence of the inevitable correlations

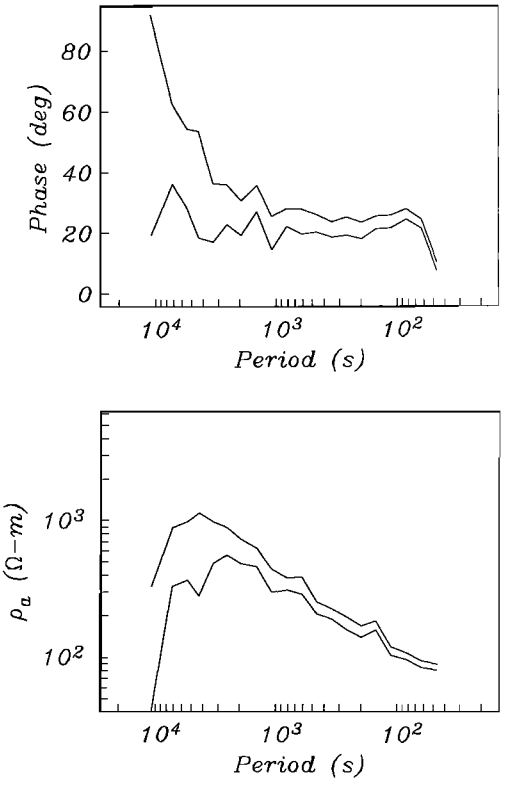

Fig. 2. The phase (top panel) and apparent resistivity (bottom panel) for the TM mode response function at Lincoln line site 4 remote referenced to site 1. The TM mode has the electric field polarized geographic east-west. The bands are the double-sided 95% confidence limits computed using the jackknife estimates of standard error and (29)–(30). While these values are smoother than for the TE mode in Figure 1, the phase error is typically 3°-5°, and the apparent resistivity uncertainty is typically 15%.

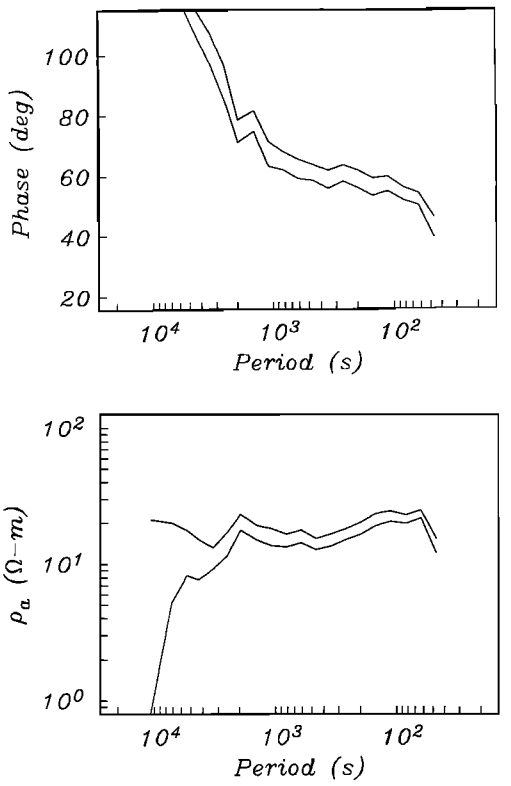

Fig. 3. Robust remote reference results for the data used in Figure 1. Note the much smoother response and smaller error bars. The uncertainty in the phase is typically 3°, and for the apparent resistivity it is under 10%. The anomalous kink near 2000 s is discussed in the text.

induced by a data window with finite length sequences, the actual degrees of freedom will be lower than the standard value of twice the number of raw frequency estimates, and the usual jackknife assumption of independent data is violated. *Thomson and Chave* [1989] have shown that (22) and (28) will give asymptotically reliable results in spite of this with section-averaged spectra.

To summarize, estimation of the jackknife confidence limits on the response functions first requires the delete-one estimates of **z**. These can be derived for either the conventional (4), robust (7), or robust remote reference (19) estimators by deleting a row from **e** and **b** in turn and solving the problem. Significant computational advantages accrue if (3) or its counterpart for (6) are solved by QR or Cholesky decomposition since standard methods for removing a row by downdating the upper triangular factor are available, avoiding the need to completely solve (4) or (7) N additional times. For the remote reference method, this is precluded by the lack of Hermitian symmetry in (19). The magnitude of the diagonal of the hat matrix (27) is also required. The jackknife estimate of the covariance matrix follows from (28). Estimates of the standard errors on the response functions are given by the square roots of the diagonal terms. This may be converted to a confidence level using (23).

The additional computational burden imposed by using the jackknife over conventional, parametric error estimates is sometimes a source of complaint. However, the authors feel that getting the correct answer is more important than reducing the computational load. For readers that are still in doubt, examples with order-of-magnitude differences between Gaussian-based and jackknife confidence limits on spectra are given by *Thomson and Chave* [1989].

DISCUSSION AND EXAMPLES

The examples in this section are typical of the results achieved on the long-period EMSLAB data and have not been selected to accentuate the capabilities of either robust processing or the jackknife. Only the remote reference and robust remote reference methods are treated here; for a more thorough comparison of a variety of commonly used processing schemes, see *Jones et al.* [this issue].

The response function examples were computed for Lincoln line site 4 (44°51.2'N, 123°31.6'W) using the site 1 (44°54.2'N, 123°55.1'W) horizontal magnetic field as a reference. The time series at both locations were collected by the Geological Survey of Canada under the direction of A. G. Jones and are among their highest-quality data from the Lincoln line; see *Wannamaker et al.* [this issue] for details. A 1-week long segment beginning at 0400 UT on August 11, 1985, was selected for processing. This interval displayed moderate activity with no obvious problems. Aside from interpolating across a few very short data gaps, no preliminary editing of the data was deemed necessary on the basis of visual inspection. Six-hour sections of the time series overlapped by 70% with adjacent ones were then tapered by a prolate data window of time-bandwidth product 4, converted to frequency domain data, corrected for the response of antialiasing and high-pass filters, and rotated to a geographic coordinate system. Target center frequencies were selected to yield eight values per decade evenly spaced on a logarithmic scale, and a combination of section and band averaging was applied to get the remote reference response functions (16) and their robust counterparts (19). The unbalanced jackknife (28) was used to get the response variances.

The response functions will be shown in geographic coordinates because the dominant geologic features are believed to

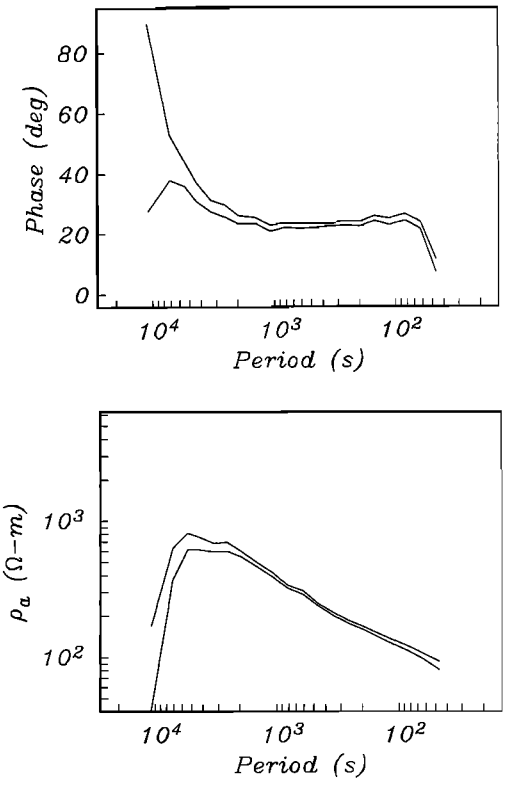

Fig. 4. Robust remote reference results for the data of Figure 2. Note the very smooth response and consistent error bars. The uncertainty in the phase is typically 1°, and for the apparent resistivity it is under 5%.

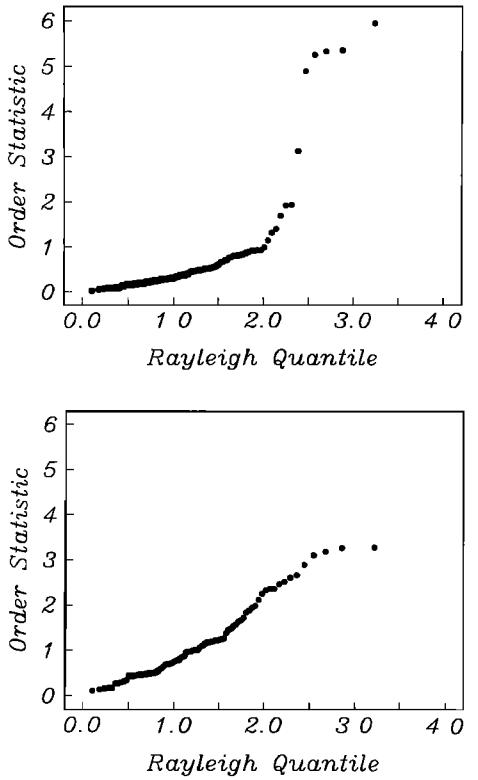

Fig. 5. Quantile-quantile plots for the results of Figures 2 and 4 at a period of 5400 s. The top panel shows the original q-q plot for the ordinary remote reference estimate. The residual distribution is quite long-tailed. The bottom panel shows the final q-q plot for the robustly weighted residuals and is approximately Gaussian.

strike north-south in the EMSLAB region. The TE mode is the response function component determined primarily by the north electric field, while the TM mode is the conjugate one where the electric field is east-west. The remote reference values for the two modes appear in Figures 1 and 2. The robust remote reference values are in Figures 3 and 4. For display purposes, approximate standard errors on the apparent resistivity and phase were computed by expanding them in first-order Taylor series to get

$$\delta\rho \approx 0.4T\,|\,z\,|\,|\,\delta z\,| \tag{29}$$

and

$$\delta\phi \approx \sin^{-1}\left(\frac{|\,\delta z\,|}{|\,z\,|}\right) \tag{30}$$

where T is the period in seconds and z and δz are the relevant components of the complex response function and its standard error. Approximate double-sided 95% confidence limits follow from (29)–(30) by multiplying the standard error by 1.96 and are shown in the figures. The apparent resistivity and phase from standard remote reference processing for both polarizations show a variety of unphysical kinks and have relatively large error bounds. By contrast, the robust remote reference results in Figures 3 and 4 are relatively smooth and possess much smaller errors. This is especially true of the TM mode result in Figure 4. The robust TE mode response still contains a sharp peak near 2000 s that is neither physical nor caused by outlier contamination and will be examined in detail later. Note that parametric error bounds for the response functions of Figures 1 and 2 are only slightly larger than those in Figures 3 and 4 and would clearly be too small to be consistent with the estimates. Since there are 20 frequencies in the responses, for Gaussian data, one estimate is expected to lie outside a smooth 95% confidence band. This is observed for Figures 3 and 4, but more departures are seen in Figures 1 and 2. This attests to the non-Gaussian form of the least squares residuals in the absence of robust processing. The unusual behavior of the response at low frequencies is caused by topographic effects; the geomagnetic north-south electric field line was placed over ground with substantial relief. In addition, the turn-down in the apparent resistivity and phase at short periods is due to inadequate knowledge of the system response.

It is instructive to apply some diagnostic tools that can provide insight to the statistical nature of robust estimation. The first of these is the quantile-quantile (q-q) plot which gives a qualitative picture of the residual distribution. The quantiles are the theoretical entities that divide the area under a probability density function into $N+1$ pieces of equal area and are easily computed for a specified distribution. The order statistics are obtained by ranking the data and provide the empirical quantiles. By producing a q-q plot of the quantiles against the order statistics, a qualitative impression of the fit to a given distribution can quickly be obtained. If the residuals are drawn from the model distribution, then the q-q plot will approximate a straight-line segment whose slope is proportional to the scale. Outliers typically appear as sharp departures from a straight line at the distribution extremes.

Figures 5 and 6 show q-q plots for the TM mode response functions of Figures 2 and 4 at periods of 5400 and 617 s respectively. Ordered residual plots for the TE mode are similar in appearance. Since the residuals from a regression like (3) are complex, the order statistics are computed from their absolute magnitudes, and the appropriate theoretical distribution is Ray-

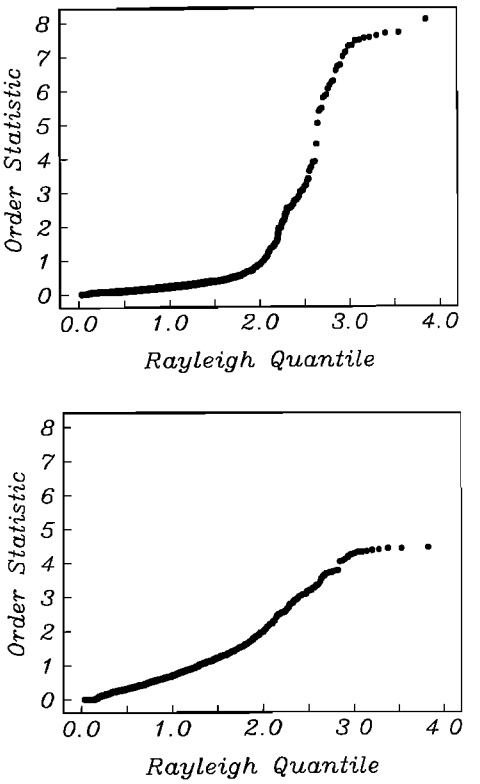

Fig. 6. Quantile-quantile plots for the results in Figures 2 and 4 at a period of 617 s. The ordinary remote reference result (top panel) is markedly long-tailed, while robust weighting (bottom panel) has reduced the residual distribution to one which is roughly Gaussian.

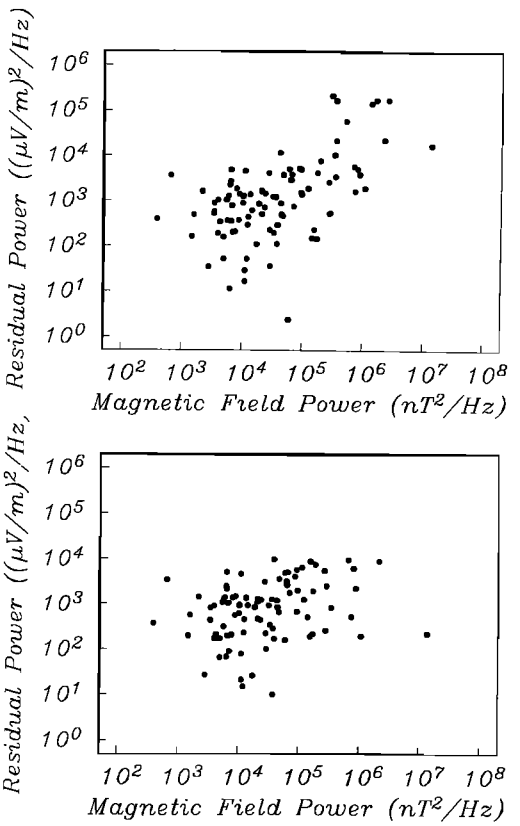

Fig. 7. Scatter plots of the power in the magnetic field against the residual power for each data section and frequency for the TM mode result of Figures 2 and 4 at a period of 5400 s. The remote reference result (top panel) indicates a slight correlation of the residual variance with the magnetic field power. This is effectively removed by robust weighting (bottom panel).

leigh. The order statistics have been scaled so that their sample variance is 2, the expected Rayleigh value. For both period bands, the original q-q plot displays markedly long-tailed behavior; residuals appear which are much too large to have come from a Gaussian-derived distribution. The robust q-q plot, in which the order statistics are derived from the weighted residuals $\mathbf{w}^{1/2}\mathbf{r}$, approximates a straight line much more closely. The departure in slope seen for the largest quantiles is due to the shape of the weight function (13) and is not of concern because of the minimal number of terms involved and the small size of their deviation. However, the remaining residuals show a slight curvature and approximate two line segments of different slope with a break between 1.5 and 2 on the x axis. This suggests a mixture situation in which the original residuals are drawn from a combination of two Gaussian distributions with different variances plus a few extreme outliers. The proportion of the populations is about two to one with the lower variance group being larger. To determine if a heterogeneous residual distribution of this type could bias the results of Figures 3 and 4, the parameter α in (13) was reduced enough to eliminate the higher variance population. The apparent resistivity and phase were almost unchanged, while their errors were reduced slightly. This suggests that while least squares-based estimators are very sensitive to extreme outliers, they are relatively immune to slight mixture situations. However, the presence of a mixture does require an increase in the size of the errors on the response functions. While the jackknife accommodates this automatically, as observed when one population was eliminated, parametric confidence limit estimators cannot detect heterogeneity and will generally be too small in its presence.

The use of summary tests such as the χ^2 goodness-of-fit type to assess the fit of the residuals to a given distribution is sometimes suggested in preference to the use of q-q plots. While this may prove useful under some circumstances, such tests are binary (in the sense that they either accept or reject a goodness of fit at some confidence level) and cannot be used to understand the form of the residual distribution and guide subsequent corrective action. The potential importance of this is seen in Figures 3 and 4. In any case, the χ^2 test involves arbitrary binning and is most sensitive to misfit near the middle of the distribution where problems are least likely to occur. If summary goodness-of-fit statistics are desired, the Anderson-Darling test [*Anderson and Darling*, 1952, 1954] is preferred because it does not group the residuals arbitrarily and is preferentially sensitive to discrepancies at the tails of their distribution.

Another commonly used diagnostic is a plot of residual versus input power. Figures 7 and 8 show the residual power against the power in the corresponding row of $\mathbf{b}$ for the TM mode response using the same period bands as in Figures 5 and 6. At 5400 s, there is a weak correlation of residual and magnetic field power that disappears when the data corresponding to large residuals are robustly downweighted. The relationship of residual and input power is less pronounced at 617 s and is again reduced by robust processing. Attempts to process the data with a priori scaling of the rows of $\mathbf{w}$ by the inverse power in the input magnetic field

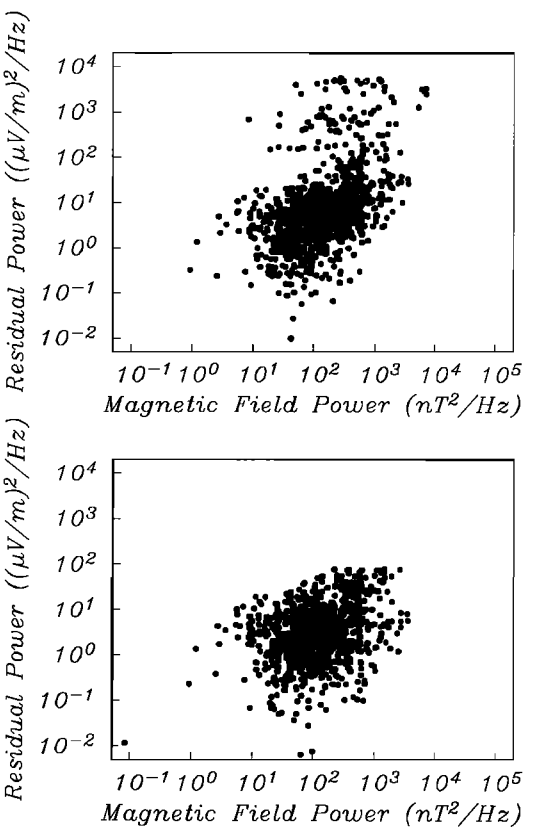

Fig. 8. Scatter plots of the power in the magnetic field against the residual power for each data section and frequency for the TM mode result of Figures 2 and 4 at a period of 617 s. The remote reference result (top panel) indicates a weak association of residual and magnetic field power that is greatly reduced by the robust weighting (bottom panel).

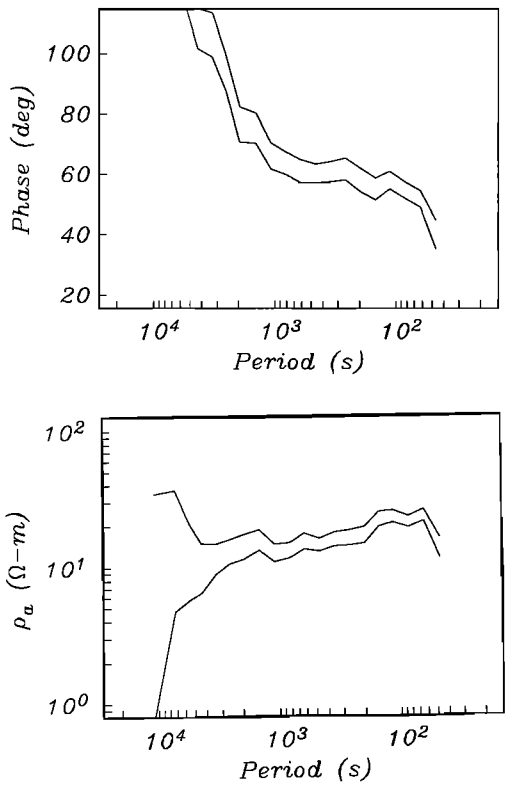

Fig. 9. The robust remote reference TE mode response functions of Figure 3 recomputed so that leverage effects are reduced. Data corresponding to hat matrix residuals exceeding $8/N$, where N is the number of estimates in a given frequency band, have been discarded. Note the marked reduction in the kink near 2000 s and the much smoother response compared to Figure 3.

substantially degraded the quality of the response functions. While this will certainly reduce the influence of the highest power events, it may inadvertently accentuate noise in low power intervals. *Egbert and Booker* [1986] suggest scaling by the inverse of the sum of the magnetic field power plus a constant to set a minimum magnetic field power level below which the weighting is stopped. In the authors' experience, this alleviates the difficulties in scaling (7) or (19) only slightly, probably because of nonstationarity of the residual power relationship. In any case, Figures 7 and 8 suggest that slight correlations of the residuals with power are readily eliminated by standard robust processing in this example, and further treatment is not called for. The same observations hold for the TE mode data in Figure 3.

The final effect to be considered is leverage. Leverage is the extent to which a given row of **b** is extraordinary compared to its remaining rows. The diagonal of the hat matrix (26) is a standard statistical measure of leverage. To illustrate, consider (25) when the ith element of **e** is replaced by e_i+c so that the corresponding predicted value becomes $\hat{e}_i+h_i c$, and a large diagonal element h_i can produce a substantial change in the predicted electric field, the residual, and ultimately the answer. This pulling effect is reminiscent of the influence which large regression residuals can exert on ordinary least squares. Since the hat matrix (26) is a function only of the input variables **b**, this is an effect that may not be detected and eliminated by robust procedures that use residual-dependent weighting. While this characterizes the normal behavior of h_i, some method to detect anomalous values is still required. When the rows of **b** are independently Gaussian, then $(N-p)[h_i-1/N]/[(p-1)(1-h_i)]$ is F distributed with $p-1$ and $N-p$ degrees of freedom [*Hoaglin and Welsch*, 1978]. Since $p=2$ and $N>10$ for most magnetotelluric problems, this gives a critical value of $h_i \approx 2.8p/N$ at the 95% level, which compares favorably to the statistical rule of thumb that data corresponding to $h_i > 2p/N$ are leverage points deserving of attention.

With these properties in mind, it is instructive to examine the effect of leverage points on the response functions in Figures 3 and 4. A crude attempt to account for leverage in solving (19) was made by iterating with Huber weights (8) to find the scale and an approximate robust solution in the usual way and computing the hat matrix (27) using the final weights. The robust processing then proceeded using the weight (13) in the standard manner, but data corresponding to hat matrix diagonal values whose magnitude exceeded a threshold were discarded. The threshold was a free parameter, but typically a few times p/N. It should be noted that the approach considered here is distinct from weighting the rows of **w** in (7) or (19) by the inverse of the power in the independent variables **b**; in the present case, only extreme values are affected, whereas weighting influences all of the data and accentuates noise in weak power intervals. Figure 9 shows the apparent resistivity and phase for the TE mode corresponding to Figure 3 but with leverage points removed. Both the apparent resistivity and phase are smoother and have slightly larger error

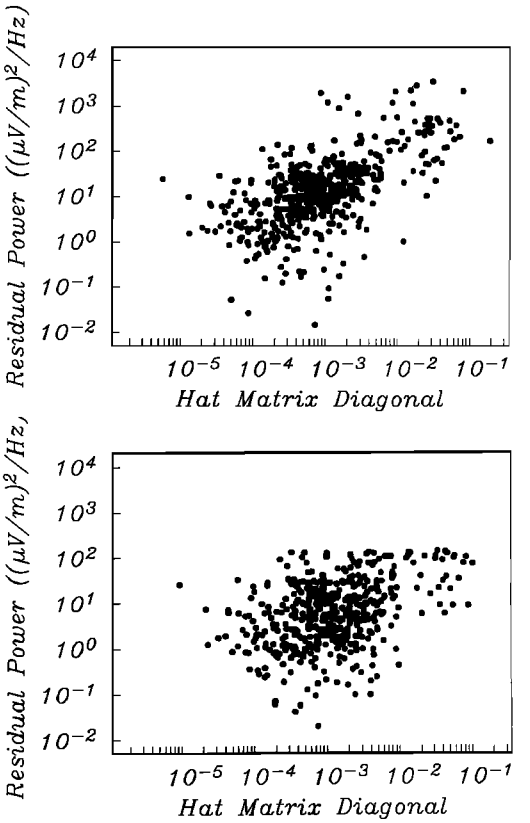

Fig. 10. Scatter plots of the residual power against the magnitude of the hat matrix diagonal for the TE mode response of Figure 3 at a period of 1440 s, near the peak of the anomalous kink in the apparent resistivity and phase. There are 460 estimates in this frequency band, so the expected value of the hat matrix diagonal is about 0.0043 and leverage points correspond to values exceeding about 0.009. The extreme leverage points seen in both panels are as much as 40 times the expected value in size. Note that robust weighting based on the regression residuals does not eliminate leverage points.

bounds, but the unphysical kink near 2000 s is greatly reduced in amplitude. The TM mode result with leverage point rejection is virtually indistinguishable from Figure 4.

To understand the behavior seen in Figure 9, it is instructive to examine a scatter plot of the residual power on the magnitude of the hat matrix diagonal. This is shown in Figure 10 at a period of 1440 s, near the peak of the kink in Figure 3. An association of residual power with the magnitude of the hat matrix diagonal is obvious for the nonrobust solution, and is just the slight correlation with magnetic field power seen in Figures 7 and 8. There are 470 data in the plot, so the mean value of the hat matrix diagonal is ≈ 0.004 and data corresponding to those exceeding ≈ 0.01 are leverage points. It is clear that robust weighting does not eliminate the leverage points in this example and that many remain in association with small to moderate residuals. These leverage points dominate the solution (19) in a similar manner to the way large residuals affect (16), and produce the non-physical kink in the response seen in Figure 3. This example is not pathological; similar effects have been observed in many magnetotelluric data series.

The leverage point rejection scheme used here is a simple type of generalized *M* or *GM* estimator, and is not recommended for routine use. While conventional *M* estimators are reasonably well understood, *GM* estimators are a topic of ongoing research in applied statistics, and each contribution seems to pose more new questions than it answers old ones. For a comprehensive discussion of *GM* estimates, see *Huber* [1983]. The increased complexity of *GM* estimation is obvious: while data corresponding to large regression residuals can be regarded as spurious, only leverage points that occur in association with anomalous residuals should be rejected. Quantifying this statement is difficult, and detecting leverage points in a manner that preserves efficiency, yet yields an unbiased result is a delicate business. However, the example presented here, as well as many similar trials, suggests that serious attention to leverage points is required.

REFERENCES

Anderson, T.W., and D.A. Darling, Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes, *Ann. Math. Stat., 23*, 193–212, 1952.

Anderson, T.W., and D.A. Darling, A test of goodness-of-fit, *J. Am. Stat. Assoc., 49*, 765–769, 1954.

Berdichevsky, M.N., and M.S. Zhdanov, *Advanced Theory of Deep Geomagnetic Sounding*, Elsevier, Amsterdam, 1984.

Brillinger, D.R., *Time Series: Data Analysis and Theory*, Holden-Day, San Francisco, 1981.

Chave, A.D., D.J. Thomson, and M.E. Ander, On the robust estimation of power spectra, coherences, and transfer functions, *J. Geophys. Res., 92*, 633–648, 1987.

Efron, B., *The Jackknife, the Bootstrap, and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia, 1982.

Efron, B., and G. Gong, A leisurely look at the bootstrap, the jackknife, and cross-validation, *Am. Stat., 37*, 36–48, 1983.

Efron, B., and C. Stein, The jackknife estimate of variance, *Ann. Stat., 9*, 586–596, 1981.

Egbert, G.D., and J.R. Booker, Robust estimation of geomagnetic transfer functions, *Geophys. J. R. astron. Soc., 87*, 173–194, 1986.

Egbert, G.D., and J.R. Booker, Multivariate analysis of geomagnetic array data, 1, The response space, *J. Geophys Res.*, this issue.

Fuller, W.A., *Measurement Error Models*, John Wiley, New York, 1987.

Gamble, T.D., W.M. Goubau, and J. Clarke, Magnetotellurics with a remote magnetic reference, *Geophysics, 44*, 53–68, 1979.

Goubau, W.M., T.D. Gamble, and J. Clarke, Magnetotelluric data analysis: Removal of bias, *Geophysics, 43*, 1157–1162, 1978.

Goubau, W.M., P.M. Maxton, R.H. Koch, and J. Clarke, Noise correlation lengths in remote reference magnetotellurics, *Geophysics, 49*, 433–438, 1984.

Hinkley, D.V., Jackknife confidence limits using Student *t* approximations, *Biometrika, 64*, 21–28, 1977*a*.

Hinkley, D.V., Jackknifing in unbalanced situations, *Technometrics, 19*, 285–292, 1977*b*.

Hoaglin, D.C., and R.E. Welsch, The hat matrix in regression and ANOVA, *Am. Stat., 32*, 17–22, 1978.

Huber, P.J., Robust estimation of a location parameter, *Ann. Math. Stat., 35*, 73–101, 1964.

Huber, P.J., Minimax aspects of bounded-influence regression, *J. Am. Stat Assoc., 78*, 66–80, 1983.

Jones, A.G., A.D. Chave, G. Egbert, D. Auld, and K. Bahr, A comparison of techniques for magnetotelluric response function estimation, *J. Geophys. Res.*, this issue.

Kendall, M., and A. Stuart, *The Advanced Theory of Statistics*, 2 vols., MacMillan, New York, 1977.

Larsen, J.C., Electromagnetic response functions from interrupted and noisy data, *J Geomagn. Geoelectr., 32, Supp. I*, SI89–SI103, 1980.

Malinvaud, E., *Statistical Methods of Econometrics*, Elsevier, New York, 1970.

Miller, R.G., *Beyond ANOVA, Basics of Applied Statistics*, John Wiley, New York, 1986.

Reiersol, O., Confluence analysis by means of instrumental sets of variables, *Ark Mat., Astron. Fys., 32*, 1–119, 1945.

Rousseeuw, P.J., and A.M. Leroy, *Robust Regression and Outlier Detection*, John Wiley, New York, 1987.

Sims, W.E., F.X. Bostick Jr., and H.W. Smith, Estimation of magnetotelluric impedance tensor elements from measured data, *Geophysics, 36*, 938–942, 1971.

Thomson, D.J., Spectrum estimation techniques for characterization and development of WT4 waveguide, I, *Bell Syst Tech. J., 56*, 1769–1815, 1977.

Thomson, D.J., and A.D. Chave, Jackknife error estimates for spectra, coherences, and transfer functions, in *Advances in Spectral Analysis and Array Processing*, edited by S. Haykin, Prentice-Hall, Englewood Cliffs, NJ, in press, 1989.

Wannamaker, P.E., et al., Magnetotelluric observations across the Juan de Fuca subduction system in the EMSLAB project, *J. Geophys. Res.*, this issue.

A.D. Chave and D.J. Thomson, AT&T Bell Laboratories, 600 Mountain Ave., Murray Hill, NJ 07974.

(Received March 17, 1988;
revised October 3, 1988;
accepted October 18, 1988.)