Geophys. J. Int. (2004) **157,** 988–1006 doi: 10.1111/j.1365-246X.2004.02203.x

Bounded influence magnetotelluric response function estimation

Alan D. Chave[1] and David J. Thomson[2]

[1]*Deep Submergence Laboratory, Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA* 02543, *USA. E-mail: alan@whoi.edu*
[2]*Department of Mathematics and Statistics, Queens University, Kingston, Ontario* K7L 3N6, *Canada*

Accepted 2003 November 25. Received 2003 November 25; in original form 2002 April 22

SUMMARY

Robust magnetotelluric response function estimators are now in standard use in electromagnetic induction research. Properly devised and applied, these have the ability to reduce the influence of unusual data (outliers) in the response (electric field) variables, but are often not sensitive to exceptional predictor (magnetic field) data, which are termed leverage points. A bounded influence estimator is described which simultaneously limits the influence of both outliers and leverage points, and has proven to consistently yield more reliable magnetotelluric response function estimates than conventional robust approaches. The bounded influence estimator combines a standard robust M-estimator with leverage weighting based on the statistics of the hat matrix diagonal, which is a standard statistical measure of unusual predictors. Further extensions to magnetotelluric data analysis are proposed, including a generalization of the remote reference method which utilizes multiple sites instead of a single one and a two-stage bounded influence estimator which effectively removes correlated noise in the local electric and magnetic field variables using one or more uncontaminated remote references. These developments are illustrated using a variety of magnetotelluric data.

Key words: magnetotelluric impedance, robust processing.

1 INTRODUCTION

The magnetotelluric (MT) method utilizes naturally occurring fluctuations of electromagnetic fields observed at Earth's surface to map variations in subsurface electrical conductivity. The fundamental datum in MT is the site-specific, frequency-dependent tensor relationship between the measured electric and magnetic fields. Under general conditions on the morphology of the external source fields (e.g. Dmitriev & Berdichevsky 1979), in the absence of noise, and with precise data, this may be written

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

where $\mathbf{E}$ and $\mathbf{B}$ are two-vectors of the horizontal electric and magnetic field components at a specific site and frequency, and $\boldsymbol{Z}$ is the second rank, 2×2 MT response tensor connecting them. The solution to (1) at a given frequency is

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

where the superscript H denotes the Hermitian (complex conjugate) transpose and the terms in parentheses are the exact cross- and autopower spectra. Similar linear relationships can be derived between the vertical and horizontal magnetic field variations. Since the methodology is nearly identical, this paper will focus only on the MT case.

When $\mathbf{E}$ and $\mathbf{B}$ are actual measurements, (1) and (2) do not hold exactly due to the presence of noise, and it becomes necessary to estimate both $\boldsymbol{Z}$ and its uncertainty $\delta\boldsymbol{Z}$ in a statistical manner. Conventional MT data processing has historically been based on classical least-squares regression approaches and Gaussian statistical models (e.g. Sims *et al.* 1971) which are well known to be sensitive to small amounts of unusual data, uncorrelated noise in the magnetic field variables and inadequacies of the model description itself. This often leads to estimates of $\boldsymbol{Z}$ which are heavily biased and/or physically uninterpretable, and can affect estimates of $\delta\boldsymbol{Z}$ even more seriously.

In recent years this situation has been dramatically improved by three developments. First, the remote reference method (Gamble *et al.* 1979), in which the horizontal magnetic fields recorded simultaneously at an auxiliary site are combined with the fields at the local site, has proved to be quite effective at eliminating bias from uncorrelated local magnetic field noise. Second, a variety of data-adaptive robust weighting schemes, sometimes applied in conjunction with remote referencing, have been developed to eliminate the influence of data corresponding to large residuals which are usually called outliers (Egbert & Booker 1986; Chave *et al.* 1987; Chave & Thomson 1989; Larsen 1989; Sutarno & Vozoff 1989, 1991; Larsen *et al.* 1996; Egbert & Livelybrooks 1996; Egbert 1997; Oettinger *et al.* 2001; Smirnov 2003). The superior performance of robust methods was demonstrated in the comparison study of Jones *et al.* (1989), and these are now in general use in work on MT problems. Third, it is now recognized that conventional distribution-based estimates of the errors in the MT response tensor elements are often biased (e.g. Chave & Jones 1997), and a non-parametric jackknife method has

been introduced (Chave & Thomson 1989) and rigorously justified (Thomson & Chave 1991) to eliminate this problem.

However, there remain unusual data sets, typically due either to natural source field complexity or man-made noise, where even robust remote reference methods sometimes fail to yield physically interpretable MT responses. For example, Schultz *et al.* (1993) and Garcia *et al.* (1997) noted the serious bias that could be produced by auroral substorm source fields of a short spatial scale. MT tensor bias from cultural noise, especially that from DC electric rail systems, has been widely documented (Szarka 1988; Junge 1996; Larsen *et al.* 1996; Egbert *et al.* 2000). Some types of mid-latitude sources (notably those from Pc3 geomagnetic pulsations) have short and temporally variable spatial scales (Andersen *et al.* 1976; Lanzerotti *et al.* 1981) which can also alter MT responses (Egbert *et al.* 2000). For such situations, extensions to robust processing have been proposed. Schultz *et al.* (1993) and Garcia *et al.* (1997) applied a bounded influence estimator introduced by Chave & Thomson (1992) and further developed by Chave & Thomson (2003) and in this paper to minimize the influence of extreme magnetic field data which do not produce large residuals, and hence may be missed by robust estimators. Larsen *et al.* (1996) and Oettinger *et al.* (2001) proposed multiple transfer function methods to remove cultural noise using distant, uncontaminated sites as references. Both methods are special cases of the two-stage processing that is introduced in this paper. Egbert (1997) introduced a robust multivariate principal components approach to detect and characterize coherent cultural noise. This enables robust estimators to be tailored to the specific coherent noise characteristics derived from multiple-station data when these are available.

In a further attempt to deal with MT processing problems, this paper updates Chave & Thomson (1989) through extensions to robust remote reference MT data processing which can control problems from both outliers in the response (electric field) and extreme data (leverage points) in the predictor (magnetic field) variables, as well as eliminate correlated noise in the response and predictor data. The next section contains a brief discussion of spectral analysis principles directly relevant to this paper. Section 3 reviews robust estimation as applied to MT data. Section 4 discusses the role of the hat matrix as a leverage indicator. Section 5 introduces a generalization of the remote reference method utilizing multiple reference sites. Section 6 discusses bounded influence estimation as a means to simultaneously control both outlier and leverage effects. Section 7 describes bounded influence two-stage MT processing which can eliminate correlated noise in the local electromagnetic fields. Section 8 outlines some pre-processing steps that can improve MT processing results under special circumstances. Section 9 discusses statistical verification of robust or bounded influence procedures. Section 10 illustrates many of these concepts using a variety of field and synthetic MT data. Finally, Section 11 contains a discussion of a few remaining issues.

2 DISCOURSE ON SPECTRAL ANALYSIS

It is assumed that contemporaneous, finite time sequences of the electric and magnetic field variations from one or more sites are available. It is also presumed that the time-series have been digitized without aliasing and edited to remove gross errors like boxcar shifts or large spikes (although well-designed processing algorithms can often accommodate these problems). Finally, it is assumed that serious long-term trends have been removed by high-pass filtering or least-squares spline fitting procedures.

Because outlier detection is facilitated by comparing spectra estimated over distinct subsections in time, robust MT processing is typically based on the Welch overlapped section averaging (WOSA) approach (Welch 1967; see Percival & Walden 1993, Section 6.17, for a detailed discussion). The basic algorithm is computationally straightforward. After time-domain pre-whitening of each time-series and starting at the lowest frequency, a subset length that is of the order of a few over the frequency of interest is selected. The data sequences are then divided into subsets and each is tapered with a data window. Data sections may be overlapped to improve statistical efficiency. Fourier transforms are taken, pre-whitening is corrected for, and the instrument response is removed. In the context of this paper, the data are these Fourier transforms of the windowed data from each section at a single frequency. The section length is then repetitively reduced as higher frequencies are addressed. A variable section length WOSA analysis of this type is philosophically akin to wavelet analysis (e.g. Percival & Walden 2000) in that the timescale of the basis functions is changed along with the frequency scale to optimize resolution.

It is important to understand the bias (i.e. spectral leakage) and resolution properties of the selected data taper and design a frequency sampling strategy that accommodates both. Historically, data tapers have been chosen largely on the basis of computational simplicity or *ad hoc* criteria (e.g. Harris 1978). However, an optimal class of data tapers can be obtained by considering those finite time sequences (characterized by a length N) which have the largest possible concentration of energy in the frequency interval $[-W, W]$. The ensuing resolution bandwidth is $2W$. Work on the continuous-time version of this problem goes back to Chalk (1950), and culminated in a set of landmark papers by Slepian & Pollak (1961) and Landau & Pollak (1961, 1962). The discrete time problem was solved by Slepian (1978), and the solution is the zeroth-order discrete prolate spheroidal or Slepian sequence with a single parameter, the time–bandwidth product $\tau = NW$. Numerical solutions for the Slepian sequences are obtained by solving an eigenvalue problem; a numerically robust tridiagonal form is given by Slepian (1978) and a more accurate variant is described in Appendix B of Thomson (1990). The superiority of Slepian sequences as data tapers has been thoroughly documented (e.g. Thomson 1977, 1982; Percival & Walden 1993), and these are employed exclusively in the present work. The time–bandwidth product τ determines both the amount of bias protection outside the main lobe of the window and the main lobe half-width τ /N (Fig. 1). A useful range of τ for MT processing is typically 1 to 4. A $\tau = 1$ window provides limited (about 25 dB) bias protection but yields raw estimates that are essentially independent of the standard discrete Fourier transform (DFT) grid with frequencies separated by $1/N$. In this instance, pre-whitening is essential to avoid spectral bias. A $\tau = 4$ window provides over 100 dB of bias protection but yields raw estimates that are highly correlated over $[f - W, f + W]$, where f is the frequency of interest. As a rule of thumb, frequencies which are spaced τ apart on the DFT grid may be taken as independent; Thomson (1977) gives a more quantitative discussion of this point. It is also important to avoid frequencies within τ of the DC value, as these are typically contaminated by unresolved low frequency components. The amount that data sections can be overlapped without introducing data dependence also varies, ranging from 50 to 71 per cent as τ ranges from 1 to 4. Percival & Walden (1993, Section 6.17) give a quantitative treatment of this issue.

Pre-whitening is most easily achieved by filtering with a short autoregressive (AR) sequence fit to the time-series. Standard time-domain AR estimators are based on least-squares principles; for

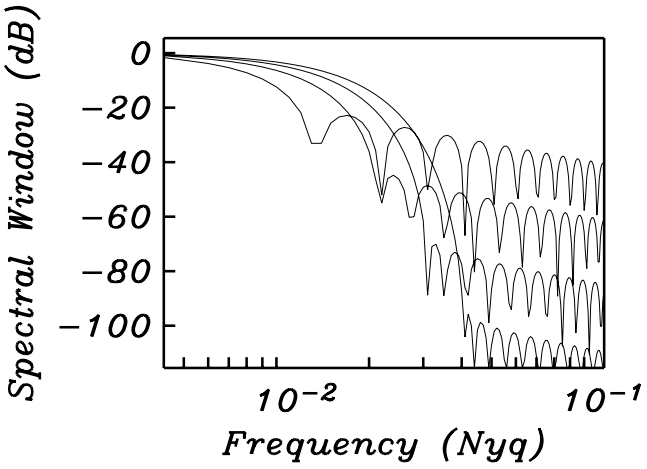

Figure 1. The spectral window (absolute square of the Fourier transform of the data taper) for 100 point Slepian sequences with time–bandwidth products of (from left to right) 1, 2, 3 and 4. As the time–bandwidth rises, the sidelobe suppression increases but the width of the mainlobe (and hence the correlation of nearby frequencies on the standard discrete Fourier transform grid) also rises.

example, the Yule–Walker equations (Yule 1927; Walker 1931) may be solved for the autocovariance sequence (acvs), from which the AR filter may be obtained using the Levinson–Durbin recursion (Levinson 1947; Durbin 1960). In the presence of extreme data, such least-squares estimators for the acvs are often seriously biased, yielding AR filters which may either be unstable (i.e. the zeros of the filter z-transform lie outside the unit circle) or highly persistent (i.e. the zeros of the filter z-transform lie inside but very close to the unit circle). In such instances, the filter may actually enhance spectral leakage by increasing the dynamic range of the data spectrum, which is opposite to the desired outcome. A robust approach to acvs estimation is essential to avoid these difficulties. A simple robust AR estimator may be obtained using acvs estimates obtained from the Fourier transform of a robust estimate of the power spectrum. The robust power spectrum may be computed using the WOSA method with a low bias (e.g. the Slepian sequence with $\tau = 4$) data taper and taking the frequency-by-frequency median rather than mean of the raw section estimates. The AR filter follows from the robust acvs using the Levinson–Durbin recursion. The importance of robust AR filter estimation for pre-whitening is illustrated in Section 10.

3 ROBUST RESPONSE FUNCTION ESTIMATION

In the standard linear regression model for the row-by-row solution of (1), the equivalent set of matrix equations is

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

where there are N observations (i.e. N Fourier transforms of N independent data sections at a given frequency), so that $\mathbf{e}$ is the response N-vector, $\mathbf{b}$ is the $N \times 2$, rank-2 predictor matrix, $\mathbf{z}$ is the solution 2-vector and ε is an N-vector of random errors. The error power $\varepsilon^{\mathrm{H}}\varepsilon$, or equivalently, the L_2 norm of the random errors, is then minimized in the usual way, yielding

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

The elements of $\mathbf{b}^{\mathrm{H}}\,\mathbf{b}$ and $\mathbf{b}^{\mathrm{H}}\mathbf{e}$ are the averaged estimates of the auto- and cross-power spectra based on the available data. The regression residuals $\mathbf{r}$ are the differences between the measured values of the response variable $\mathbf{e}$ and those predicted by the linear regression $\hat{\mathbf{e}} = \mathbf{b}\hat{\mathbf{z}}$, and serve as an estimate for the random errors ε.

The conditions on the variables in (3) and their moments that yield a least-squares solution (4) which is optimal in a well-defined sense are given by the Gauss–Markov theorem of classical statistics (e.g. Stuart *et al.* 1999 Chapter 29). The textbook version of the Gauss–Markov theorem applies when the predictor variables in (3) are fixed, but Shaffer (1991) has extended it to cover the cases where the joint distribution of $\mathbf{e}$ and $\mathbf{b}$ is multivariate normal with unknown parameters, the distribution of $\mathbf{b}$ is continuous and non-degenerate but otherwise unspecified, or under mild conditions, when $\mathbf{b}$ is a random sample from a finite population. It is as one of these cases that (3) will be considered in this paper. Under these circumstances, the linear regression solution (4) is the best linear unbiased estimate when the residuals $\mathbf{r}$ are uncorrelated and share a common variance independent of any assumptions about their statistical distribution except that the variance must exist. In addition, if the residuals are multivariate Gaussian, then the least-squares result is a maximum-likelihood, fully efficient, minimum-variance estimate. These optimality properties, along with the empirical observation that most physical data yield Gaussian residuals, explain the appeal of a least squares approach.

However, with natural source electromagnetic data, the Gauss–Markov assumptions about the data and their error structure are rarely tenable for at least three reasons. First, the residual variance is often dependent on the data variance, especially when energetic intervals coincide with source field complexity, as occurs for many classes of geomagnetic disturbance. Second, the finite duration of many geomagnetic events causes data anomalies to occur in patches, violating the independent residual requirement. Third, due to marked non-stationarity, frequent large residuals are much more common than would be expected for a Gaussian model, and hence the residual distribution is typically very long-tailed with a Gaussian centre. Any one of these issues can seriously bias the least-squares solution (4) in unpredictable and sometimes insidious ways; in the presence of all three, problems are virtually guaranteed.

These observations have led to the development of procedures which are robust, in the sense that they are relatively insensitive to a moderate amount of bad data or to inadequacies in the model, and that react gradually rather than abruptly to perturbations of either. Robust statistics is an active area of research (e.g. Hampel *et al.* 1986; Wilcox 1997), and many of its techniques have been adapted to spectral analysis beginning with the work of Thomson (1977). Chave *et al.* (1987) presented a tutorial review of robust methods in the context of geophysical data processing, and only an outline relevant to the MT problem will be presented here.

The most suitable type of robust estimator for MT data is the M-estimator which is motivated by analogy to the maximum-likelihood method of statistical inference (e.g. Stuart *et al.* 1999, Chapter 18). M-estimation is similar to least squares in that it minimizes a norm of the random errors in (3), but the misfit measure is chosen so that a few extreme values cannot dominate the result. The M-estimate is obtained computationally by minimizing $\mathbf{R}^{\mathrm{H}}\mathbf{R}$, where $\mathbf{R}$ is an N-vector whose ith entry is $\sqrt{\rho}(r_i/d)$, d is a scale factor and $\rho(x)$ is a loss function. The term 'loss function' originates in statistical decision theory (e.g. Stuart & Ord 1994, Section 26.52), and loosely speaking, yields a measure of the distance between the true and estimated values of a statistical parameter. For standard least squares, $\rho(x) = x^2/2$ and (4) is immediately recovered. More generally, if $\rho(x)$ is chosen to be $-\log f(x)$, where $f(x)$ is the true probability density function (pdf) of the regression residuals $\mathbf{r}$ which includes

extreme values, then the M-estimate is maximum likelihood. In practice, $f(x)$ cannot be determined reliably from finite samples, and the loss function must be chosen on theoretical or empirical grounds. Details may be found in Chave *et al.* (1987) and Wilcox (1997).

Minimizing $\mathbf{R}^{\mathrm{H}}\mathbf{R}$ gives the M-estimator analogue to the normal equations

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

where Ψ is an N-vector whose ith entry is the influence function $\psi(x) = \partial_x \rho(x)$ evaluated at $x = r_i/d$. The solution to (5) is obtained using iteratively reweighted least squares, so that at the kth iteration, Ψ is replaced by $\mathbf{v}^{[k]}\,\mathbf{r}^{[k]}$ where $\mathbf{v}^{[k]}$ is an $N \times N$ diagonal matrix whose ith entry is $v_{ii}^{[k]} = \Psi(r_i^{[k-1]}/d^{[k-1]})/(r_i^{[k-1]}/d^{[k-1]})$. The weights are computed using the residuals and scale from the previous iteration to linearize the problem; the solution is initialized using the residuals $r_i^{[0]}$ and scale $d_i^{[0]}$ from ordinary least squares, or for badly contaminated data, those from an L_1 estimate obtained using a linear programming algorithm. The weighted least-squares solution of (5) corresponding to (4) is

$$\hat{\mathbf{z}}_* = (\mathbf{b}^{\mathrm{H}}\mathbf{v}\mathbf{b})^{-1}(\mathbf{b}^{\mathrm{H}}\mathbf{v}\mathbf{e}) \tag{6}$$

Iteration ends when the solution does not change appreciably, typically after three to five steps. Since the weights are chosen to minimize the influence of data corresponding to large residuals in a well-defined sense, the M-estimator is data adaptive.

A simple form for the weights is due to Huber (1964). This has diagonal elements

$$\begin{aligned} v_{ii} &= 1 \qquad |x_i| \le a \\ v_{ii} &= \frac{a}{|x_i|} \quad |x_i| > a. \end{aligned} \tag{7}$$

A value of $a = 1.5$ gives better than 95 per cent efficiency with outlier-free Gaussian data. Downweighting of data with (7) begins when $|x_i| = |r_i/d| = a$, so that the scale factor d determines which of the residuals are to be regarded as large. It is necessary because the weighted least-squares problem is not scale invariant without it, in the sense that multiplication of the data by a constant will not produce a comparably affine change in the solution.

The scale estimate must itself be robust, and is typically chosen to be the ratio of the sample and theoretical values of some statistic based on a target distribution. An extremely robust scale statistic is the median absolute deviation from the median (MAD), whose sample value is

$$S_{\mathrm{MAD}} = |r - \tilde{r}|_{(\frac{N+1}{2})} \tag{8}$$

where the subscript (i) denotes the ith order statistic obtained by sorting the N values sequentially. The quantity $\tilde{r}$ is the middle order statistic or median of $\mathbf{r}$. The theoretical MAD is the solution σ_{MAD} of

$$F(\tilde{\mu} + \sigma_{\mathrm{MAD}}) - F(\tilde{\mu} - \sigma_{\mathrm{MAD}}) = 1/2 \tag{9}$$

where $\tilde{\mu}$ is the theoretical median and F denotes the target cumulative distribution function (cdf).

The solution of (5) and hence the choice of d requires a target distribution for the residuals $\mathbf{r}$. The data in MT processing are Fourier transforms and therefore complex, but the complex Gaussian is not necessarily the optimal choice. It is preferable to measure residual size using its magnitude since this is rotationally (phase) invariant, and Chave *et al.* (1987) showed that the appropriate distribution for the magnitude of a complex number is Rayleigh. The properties of the Rayleigh distribution are summarized in Johnson *et al.* (1994, Chapter 10). A Rayleigh statistical model for the residuals will be used exclusively in this paper.

Because the weights (7) fall off slowly for large residuals and never fully descend to zero, they provide inadequate protection against severe outliers. However, the loss function corresponding to (7) is convex and hence convergence to a global minimum is assured, making it safe for the initial iterations of a robust procedure to obtain a reliable estimate of the scale. Motivated by the form of the Gumbel extreme value distribution, Thomson (1977) and Chave & Thomson (1989) suggested using the more severe weight function

$$v_{ii} = \exp(e^{-\xi^2})\exp(-e^{\xi(|x_i|-\xi)}) \tag{10}$$

for the final estimate, where the parameter ξ determines the residual size at which downweighting begins and $v_{ii} = 1$ when $x_i = 0$. The weight (10) is essentially like a hard limit at $x_i = \xi$ except that the truncation function is continuous and continuously differentiable. Chave *et al.* (1987) advocated setting ξ to the Nth quantile of the Rayleigh distribution which automatically increases the allowed residual magnitude as the sample size rises. In an uncontaminated population, one expects the largest residual to increase slowly (approximately as $\sqrt{\log(N^2)}$), but if the fraction of contaminated points α is constant, a level at the $N(1-\alpha)$th quantile would be more appropriate.

To summarize, robust processing of MT data utilizes the Fourier transforms of comparatively short sections of a long data series, where the section length is chosen to be of order a few over the frequency of interest. Data sections may be overlapped depending on the data window characteristics. At each frequency, an initial least-squares solution is obtained from (4) and used to compute the residuals $\mathbf{r}$ in (3) as well as the scale d from the ratio of (8) and (9) using a Rayleigh model for the residual distribution. An iterative procedure is then applied using (6) with the Huber weights (7), where the residuals from the previous iteration are used to get the scale and weights. This continues until the weighted residual power $\mathbf{r}^{\mathrm{H}}\mathbf{v}\mathbf{r}$ does not change below a threshold value, typically 1–2 per cent. The scale is then fixed at the final Huber weight value and several iterations are performed using the more severe weight (10), again terminating when the weighted residual power does not change appreciably.

The algorithm outlined here is very similar to that introduced by Chave & Thomson (1989) except that the section length is variable and always maintained so that the frequency of interest is a few over the section length. It has been empirically shown that robust procedures are most effective at eliminating outliers without a serious reduction in statistical efficiency when this methodology is used. With variable section lengths, there is no need for band averaging, and it is recommended that its use be avoided to prevent errors on the confidence intervals of the response tensor caused by correlation. In addition, use of the interquartile distance for a scale estimate as discussed by Chave & Thomson (1989) has been discontinued since it has proven to be less robust than the MAD. Robust estimation as described here is capable of yielding reliable estimates of the MT response tensor in a semi-automatic manner for most data sets.

Finally, an essential feature of any method for computing MT responses is the provision of both an estimate and a measure of its accuracy. Traditionally, confidence intervals on the MT response are dependent on explicit statistical models which are ultimately based on a Gaussian distribution. In addition to problems in computing the correct number of degrees of freedom in the presence of correlated spectral estimates and the practical requirement for numerous approximations to yield a tractable result, these parametric methods

are even more sensitive to outliers or other data anomalies than the least-squares estimates themselves. For a more complete discussion of the limitations of parametric uncertainty estimates in a spectral analysis context, see Thomson & Chave (1991). This has led to the increasing use of non-parametric confidence interval estimators which require fewer distributional assumptions, among which the simplest is the jackknife. Thomson & Chave (1991) give a detailed description of its implementation and performance in spectral analysis, while Chave & Thomson (1989) describe its use for estimating confidence intervals on the MT response tensor. This requires only that N estimates of the response tensor elements obtained by deleting a single data section at a time from the ensemble be computed and saved for later use. In a regression context, the jackknife offers the advantage that it is insensitive to violations of the assumption of variance homogeneity which is implicit to parametric error estimates (Hinkley 1977). The jackknife also yields conservative confidence interval estimates in the sense that they will tend to be slightly too large rather than too small (Efron & Stein 1981), as documented for MT by Eisel & Egbert (2001). Finally, the jackknife offers a way to propagate error through either linear (e.g. rotation) or non-linear transformations of the MT response tensor which might otherwise be completely intractable; tensor decomposition to remove galvanic distortion as described in Chave & Smith (1994) is a prime example of the latter.

4 THE HAT MATRIX AND LEVERAGE

The hat or predictor matrix is an important auxiliary quantity in regression theory, and is widely used to detect unusually influential (i.e. high-leverage) data (Hoaglin & Welsch 1978; Belsley *et al.* 1980; Chatterjee & Hadi 1988). For the regression problem defined by (3), the $N \times N$ hat matrix is given by

$$\mathbf{H} = \mathbf{b}(\mathbf{b}^{\mathrm{H}}\mathbf{b})^{-1}\mathbf{b}^{\mathrm{H}} \tag{11}$$

The residual $\mathbf{r}$ is the difference between the observed electric field $\mathbf{e}$ and the value $\hat{\mathbf{e}}$ predicted from a linear regression on $\mathbf{b}$. It follows that

$$\hat{\mathbf{e}} = \mathbf{H}\mathbf{e} \tag{12}$$

and hence

$$\mathbf{r} = (\mathbf{I} - \mathbf{H})\mathbf{e} \tag{13}$$

where $\mathbf{I}$ is the identity matrix. The hat matrix is a projection matrix which maps $\mathbf{e}$ onto $\hat{\mathbf{e}}$. The diagonal elements of $\mathbf{H}$ (denoted by h_{ii}) are a measure of the amount of leverage which the ith value of the response variable e_i has in determining its corresponding predicted value $\hat{e}_i$ independent of the actual size of e_i, and this leverage depends only on the predictor variables $\mathbf{b}$. The inverse of h_{ii} is in effect the number of observations in $\mathbf{b}$ which determine $\hat{e}_i$.

Some relevant properties of the hat matrix are described by Hoaglin & Welsch (1978). First, because it is a projection matrix, $\mathbf{H}$ is Hermitian and idempotent ($\mathbf{H}^2 = \mathbf{H}$). These two characteristics can be used to show that the diagonal elements satisfy $0 \le h_{ii} \le 1$. Second, the eigenvalues of a projection matrix are either 0 or 1, and the number of nonzero eigenvalues equals its rank. Since rank($\mathbf{H}$) = rank($\mathbf{b}$) = p, where p is the number of columns in $\mathbf{b}$ (usually 2 for MT), the trace of $\mathbf{H}$ is p. Thus, the expected value of h_{ii} is p/N. Finally, the two extreme values 0 and 1 have special meaning. When $h_{ii} = 0$, then $\hat{e}_i$ is fixed at zero and not affected by the corresponding datum in $\mathbf{e}$. If $h_{ii} = 1$, then the predicted and observed values of the ith entry in $\mathbf{e}$ are identical, and the model fits the data exactly. In this instance, only the ith row of $\mathbf{b}$ has any influence on the regression problem, leading to the term high (or extreme) leverage to describe that point. More generally, if $h_{ii} \gg p/N$, the ith row of $\mathbf{b}$ will exert undue influence (or leverage) on the solution $\hat{\mathbf{z}}$ in an analogous manner to an outlier in $\mathbf{e}$. Thus, the hat matrix diagonal elements are a measure of the amount of leverage exerted by a predictor datum. The factor by which the hat matrix diagonal elements must exceed the expected value to be considered a leverage point is not well defined, but statistical lore suggests that values which are more than two to three times p/N are a concern (Hoaglin & Welsch 1978).

Table 1. Critical Points of $I_x(2, N-2)$ for $N \gg 2$.

$\eta = xN/2$	I_x
1	0.594
2	0.909
2.365	0.95
2.777	0.975
3.307	0.99
4.593	0.999
5.841	0.9999
7.064	0.99999

In Chave & Thomson (2003), the exact distribution of the diagonal elements of (11) for complex Gaussian data was derived from first principles, extending previous work which yielded only asymptotic forms for real data (e.g. Rao 1973; Belsley *et al.* 1980; Chatterjee & Hadi 1988). The result is the beta distribution $\beta\,(h_{ii}, p, N-p)$, where p is the number of predictor variables or columns in $\mathbf{b}$ and N is the number of data. The cumulative distribution function is the incomplete beta function ratio $I_x(p, N-p)$, and a series expression suitable for numerical solution is given in Chave & Thomson (2003). For the MT situation ($p = 2$) and when $N \gg 2$, some critical values η, where $x = \eta\,2/N$ and $I_x = \alpha$, are given in Table 1. The table shows that the probability that a given hat matrix diagonal element will exceed the expected value of $2/N$ by a factor of 4.6 is only 0.001. However, data whose corresponding hat matrix diagonal is only twice the expected value will occur with probability 0.091. Rejecting values at this level will typically remove a significant number of valid data unless N is fairly small. Chave & Thomson (2003) advocate rejecting data corresponding to hat matrix diagonal elements larger than the inverse of the beta distribution at a reasonable probability level. For example, selecting the 0.95 level will carry a 5 per cent penalty for Gaussian data, yet will effectively protect against large leverage points. In data sets whose hat matrix is longer tailed than beta (a common occurrence with MT data from auroral latitudes), larger cut-off values should be used.

It is important to note that using the hat matrix diagonal elements as leverage statistics differs considerably from a simpler approach based on some measure of the relative variance or power in the rows of $\mathbf{b}$. To see this, consider (11) with $p = 2$, as for a conventional single-site MT sounding. Let S^i_{jk} denote the power between the j, k variables in the ith row of $\mathbf{b}$ and let $\bar{S_{jk}}$ be the averaged cross spectrum of the j, k variables obtained from all N data in $\mathbf{b}$. The ith diagonal entry in (11) may be written

$$h_{ii} = \frac{1}{N(1-\gamma^2)}\left[\frac{S^i_{xx}}{\bar{S_{xx}}} + \frac{S^i_{yy}}{\bar{S_{yy}}} - 2\gamma^2 \mathrm{Re}\left(\frac{S^i_{xy}}{\bar{S_{xy}}}\right)\right] \tag{14}$$

where γ^2 is an estimate of the squared coherence between b_x and b_y derived from the averaged cross- and autopowers. The expected value of the right hand side of (14) is easily seen to be $2/N$. In the special case where $\gamma^2 \to 0$ so that b_x and b_y are uncorrelated, (14)

reduces to

$$h_{ii}^0 = \left(\frac{S_{xx}^i}{\sum_{i=1}^N S_{xx}^i} + \frac{S_{yy}^i}{\sum_{i=1}^N S_{yy}^i} \right) \tag{15}$$

which is just the sum of the ratios of the power in the ith row to the total power in each component of $\mathbf{b}$. As γ^2 increases, (14) may be rewritten as

$$h_{ii} = \frac{1}{1-\gamma^2} \left[h_{ii}^0 - 2\gamma^2 \mathrm{Re} \left(\frac{S_{xy}^i}{\sum_{i=1}^N S_{xy}^i} \right) \right]. \tag{16}$$

This quantity may be either larger or smaller than (15), depending on the size of γ^2 and the size and sign of the ratio of the ith row cross-power to the total cross-power. Thus, high leverage points may correspond to rows of $\mathbf{b}$ where γ^2 is small and the power in one or both of the predictor variables is much larger than the average power, or to rows of $\mathbf{b}$ where γ^2 is non-zero and the cross-power between b_x and b_y is much larger than the average cross-power. Intermediate situations with high leverage are also possible.

This discussion has centred on the diagonal elements of the hat matrix, but the off-diagonal entries also have a leverage interpretation. The i, j off-diagonal element of $\mathbf{H}$ gives the amount of leverage which the jth response variable e_j has on the ith predicted variable $\hat{e}_i$, and the leverage again depends only on the predictor variables $\mathbf{b}$. Thus, in the MT context of this paper, the off-diagonal elements of $\mathbf{H}$ give the amount of leverage exerted by magnetic field data which are non-local in time, and hence are expected to be small unless there is significant non-stationarity. The off-diagonal elements of $\mathbf{H}$ will not be considered further in this paper.

The hat matrix and its properties easily generalize to the robust algorithm of (5)–(6) using the definition in (11)

$$\mathbf{H} = \sqrt{\mathbf{v}}\mathbf{b}(\mathbf{b}^{\mathrm{H}}\mathbf{v}\mathbf{b})^{-1}\mathbf{b}^{\mathrm{H}}\sqrt{\mathbf{v}} \tag{17}$$

where $\sqrt{\mathbf{v}}$ is the diagonal matrix of robust weights with entries $\sqrt{v_{ii}}$.

5 A GENERALIZATION OF THE REMOTE REFERENCE METHOD

Consider the regression equation for the MT response (3), and assume that an auxiliary set of remote reference variables $\mathbf{Q}$ are available, where for a given frequency $\mathbf{Q}$ is $N \times q$ with $q \geq 2$. For standard remote reference MT data, $\mathbf{Q}$ would typically consist of horizontal magnetic field measurements collected some distance from the base site, and hence $q = 2$. For more than one remote reference site or where more than the auxiliary horizontal magnetic field is available at a single reference site, the rank q of $\mathbf{Q}$ may be larger. In this instance, the local magnetic field $\mathbf{b}$ may be regressed on $\mathbf{Q}$ in the usual way by solving

$$\mathbf{b} = \mathbf{Q}\boldsymbol{t} + \varepsilon \tag{18}$$

where $\boldsymbol{t}$ is a $q \times 2$ transfer tensor between the local horizontal magnetic field $\mathbf{b}$ and the remote reference variables in $\mathbf{Q}$ and ε is an $N \times 2$ random error matrix. In practice, (18) may be solved as a series of univariate regressions on the columns of $\mathbf{b}$, $\boldsymbol{t}$ and ε. By analogy to (12), the predicted local magnetic field is

$$\hat{\mathbf{b}} = \mathbf{Q}(\mathbf{Q}^{\mathrm{H}}\mathbf{Q})^{-1}\mathbf{Q}^{\mathrm{H}}\mathbf{b} \tag{19}$$

which is just a projection operation. Substituting $\hat{\mathbf{b}}$ for $\mathbf{b}$ in (3) and solving yields the generalized remote reference analogue to (4)

$$\hat{\mathbf{z}}^q = (\hat{\mathbf{b}}^{\mathrm{H}}\hat{\mathbf{b}})^{-1}\hat{\mathbf{b}}^{\mathrm{H}}\mathbf{e} \tag{20}$$

When $q = 2$ so that $\mathbf{Q}$ is the standard $N \times 2$ matrix of remote reference magnetic field variables $\mathbf{b}_{\mathrm{r}}$, (20) can be simplified using standard matrix identities to yield

$$\hat{\mathbf{z}}^{\mathrm{r}} = \left(\mathbf{b}_{\mathrm{r}}^{\mathrm{H}}\mathbf{b}\right)^{-1}\mathbf{b}_{\mathrm{r}}^{\mathrm{H}}\mathbf{e} \tag{21}$$

which is identical to the remote reference solution of Gamble *et al.* (1979), where the local auto- and cross-spectra are replaced by cross-spectra with the reference magnetic field. In fact, it is easy to show that $\hat{\mathbf{b}}^{\mathrm{H}}\hat{\mathbf{b}} = \hat{\mathbf{b}}^{\mathrm{H}}\mathbf{b}$, and hence (20) is not downward biased by uncorrelated noise because such noise is not present in $\hat{\mathbf{b}}$. Eq. (20) may be thought of as the remote reference solution (21) with the two-channel reference magnetic field $\mathbf{b}_{\mathrm{r}}$ replaced by the projection (19). Just as the remote reference method is equivalent to an econometric technique for solving errors-in-variables problems called the method of instrumental variables (Wald 1940; Reiersol 1941; Reiersol 1945; Geary 1949), the generalized remote reference method in (19)–(20) is identical to a related econometric technique called two-stage least squares (Mardia *et al.* 1979). Both methods arose to handle violations of the Gauss–Markov theorem where the predictor variables $\mathbf{b}$ in (3) are correlated with the residuals $\mathbf{r}$, in which case the solution (4) is not unbiased and statistically consistent.

From (20), the hat matrix for the two-stage least-squares solution is

$$\mathbf{H}_q = \hat{\mathbf{b}}(\hat{\mathbf{b}}^{\mathrm{H}}\hat{\mathbf{b}})^{-1}\hat{\mathbf{b}}^{\mathrm{H}} \tag{22}$$

which reduces to

$$\mathbf{H}_{\mathrm{r}} = \mathbf{b}_{\mathrm{r}}\left(\mathbf{b}_{\mathrm{r}}^{\mathrm{H}}\mathbf{b}_{\mathrm{r}}\right)^{-1}\mathbf{b}_{\mathrm{r}}^{\mathrm{H}} \tag{23}$$

when a single remote reference station is used. Eq. (23) differs from the mixed local and reference field form $\mathbf{H}_R$ proposed by Chave & Thomson (1989, see their Eq. 27) and further studied by Ghosh (1990). $\mathbf{H}_R$ is not formally a projection matrix, and hence shares the hat matrix properties outlined in Section 3 only approximately. The more natural version (23) should be adopted in remote reference applications.

Chave & Thomson (1989) combined a robust estimator like that described in Section 3 and (21) to yield a robust remote reference solution to the MT problem

$$\hat{\mathbf{z}}_*^{\mathrm{r}} = \left(\mathbf{b}_{\mathrm{r}}^{\mathrm{H}}\mathbf{v}\mathbf{b}\right)^{-1}\left(\mathbf{b}_{\mathrm{r}}^{\mathrm{H}}\mathbf{v}\mathbf{e}\right) \tag{24}$$

where the weights $\mathbf{v}$ are computed as for (6) based on the ordinary residuals $\mathbf{r}$ from (3). This is a straightforward extension of M-estimation that simultaneously eliminates downward bias in (6) from uncorrelated noise on the magnetic field channels and problems from outliers, and has been shown to perform well under general circumstances (e.g. Jones *et al.* 1989; Chave & Jones 1997). However, it provides limited protection from overly influential values in either the local or reference magnetic fields. Eq. (24) is easily generalized to more than two remote reference channels at a single station or to multiple reference sites by substituting $\hat{\mathbf{b}}$ for $\mathbf{b}_{\mathrm{r}}$ to give $\hat{\mathbf{z}}_*^q$. This can be very useful when remote reference sites contain uncorrelated noise with different frequency characteristics, as is demonstrated in Section 10.

6 BOUNDED INFLUENCE MT RESPONSE FUNCTION ESTIMATION

Two statistical concepts which are useful in understanding problems with least squares and M-estimators are the breakdown point and the influence function. Loosely speaking, the breakdown point is the smallest fraction of gross errors which can carry an estimate

beyond all statistical bounds that can be placed on it. Ordinary least squares has a breakdown point of $1/N$, as a single outlying data point can inflict unlimited change on the result. The influence function measures the effect of an additional observation on the estimate of a statistic given a large sample, and hence is the functional derivative of the statistic evaluated at an underlying distribution. Influence functions may be unbounded or bounded as the estimator is or is not sensitive to bad data. Further details may be found in Hampel *et al.* (1986, Section 1.3).

The procedures described in Sections 3 and 5 are capable in a semi-automatic fashion of eliminating bias due to uncorrelated noise in the local magnetic field and general contamination from data that produce outlying residuals. However, they do require that the reference channels in $\mathbf{Q}$ and the local magnetic field $\mathbf{b}$ be reasonably clear of extreme values or leverage points caused by instrument malfunctions, impulsive events in the natural source fields, or other natural or artificial phenomena. When this assumption is invalid, then any conventional robust M-estimator can be severely biased. As for least-squares estimators, M-estimators possess a breakdown point of only $1/N$, so that a single leverage point can completely dominate the ensuing estimate, and their influence functions are unbounded.

The discussion of Section 4 indicates that the hat matrix diagonal elements serve as indicators of leverage, and hence extensions of robust estimation that can detect both outliers based on the regression residuals and leverage points based on the hat matrix diagonal elements can be devised (e.g. Mallows 1975; Handschin *et al.* 1975; Krasker & Welsch 1982). When properly implemented, these algorithms have bounded influence functions, and are often termed bounded influence estimators as a result, although the term GM (for generalized M) estimator is also in common use. The bounded influence form of the normal equations analogous to (5) is

$$\mathbf{b}^{\mathrm{H}}\Psi\mathbf{w} = 0 \tag{25}$$

where $\mathbf{w}$ is a leverage weight. Two versions of (25) are in common use (Wilcox 1997). If $w_{ii} = \sqrt{1-h_{ii}}$ and $\Psi = \psi\,(r_i/d)$ as in Section 3, then the form is that originally suggested by Mallows (1975), in which residual and leverage weighting are decoupled and leverage points are gently downweighted according to the size of the hat matrix diagonal elements. If $\Psi = \psi\,[r_i/(w_{ii}d)]$, then the approach is that proposed by Handschin *et al.* (1975), and is more efficient than the Mallows solution since large leverage points corresponding to small residuals are not heavily penalized. However, Carroll & Welsh (1988) have shown under general conditions that the approach of Handschin *et al.* can lead to a solution which is not statistically consistent, and only the Mallows approach will be used here.

In practice, the cited estimators have proved to be less than satisfactory with MT measurements because downweighting of high-leverage data is mild and limited rather than aggressive; their breakdown point is only slightly larger than for conventional M-estimators. More capable, high breakdown point (up to 0.5) bounded influence estimators have been proposed (Rousseeuw 1984; Coakley & Hettmansperger 1993), but these typically entail a substantial increase in computational overhead which limits their applicability to the large data sets that occur in MT processing.

An alternative bounded influence estimator that combines high asymptotic efficiency for Gaussian data, high breakdown point performance with contaminated data, and computational simplicity that is suitable for large data sets was proposed by Chave & Thomson (1992, 2003) and is easily adapted to MT estimation. This is accomplished by replacing the weight matrix $\mathbf{v}$ in (6) or (24) with the product of two diagonal matrices $\mathbf{u} = \mathbf{vw}$, where $\mathbf{v}$ is the M-estimator weight matrix defined in Section 3 whose elements are based on the regression residuals and $\mathbf{w}$ is a leverage weight matrix whose elements depend on the size of the hat matrix diagonal. The bounded influence MT response obtained from (6) with $\mathbf{u}$ in place of $\mathbf{v}$ will be denoted by $\hat{\mathbf{z}}_{\#}$.

Based on numerous trials with actual data, the robust weights in (6) or (24) are frequently seen to both increase the amount of leverage exerted by existing leverage points or create entirely new ones, as will be demonstrated in Section 10. This occurs because M-estimators do not have bounded influence and probably accounts for many instances where robust procedures do not perform as well as might be expected. Even with Mallows weights in (25), $\mathbf{v}$ and $\mathbf{w}$ can interact to produce an unstable solution to the robust regression problem unless care is used in applying the leverage weights. This is especially true when the leverage weights are recomputed at each iteration step, which often results in oscillation between two incorrect 'solutions' each of which is dominated by a few extreme leverage points. Instead, the leverage weights are initialized with unity on the main diagonal, and at the kth iteration the ith diagonal element is given by

$$w_{ii}^{[k]} = w_{ii}^{[k-1]} \exp(e^{-\chi^2}) \exp(-e^{\chi(y_i-\chi)}) \tag{26}$$

where the statistic y_i is $Mh_{ii}^{[k]}/p$ with

$$h_{ii}^{[k]} = \sqrt{u_i^{[k-1]}}\, b_i \left(\mathbf{b}^{\mathrm{H}}\mathbf{u}^{[k-1]}\mathbf{b}\right)^{-1} b_i^{\mathrm{H}} \sqrt{u_i^{[k-1]}} \tag{27}$$

and M is the trace of $\mathbf{u}^{[k-1]}$ which is initially the number of data points N. χ is a free parameter which determines the point where leverage point rejection begins. If the usual statistical rule of thumb where rows of the regression (3) corresponding to hat matrix diagonal entries larger than twice the expected value are to be regarded as leverage points, then $\chi = 2$. A better, albeit empirical, choice of χ is the Nth quantile of the beta distribution or the value of its inverse corresponding to a chosen probability level. These choices will tend to automatically increase the allowed size of leverage points as N increases. To further ensure a stable solution, downweighting of leverage points is applied in half-decade stages beginning with χ set just below the largest normalized hat matrix diagonal element and ending with the final value, rather than all at once.

To summarize this section, bounded influence response function estimation for $\hat{\mathbf{z}}_{\#}$ without a remote reference begins with Fourier transform estimates of short data sections as described in Section 2. At each frequency, an initial least-squares solution is obtained from (4) and used to compute the residuals $\mathbf{r}$ as well as the scale d from (8) and (9) using a Rayleigh residual model and the hat matrix (11). A nested iterative procedure is then applied in which the outer loop steps over χ ranging from immediately below the largest hat matrix diagonal element to the selected minimum value χ_0 and the inner loop successively solves (6) using the composite weight matrix $\mathbf{u}$ computed at each step, where $\mathbf{v}$ is given by the Huber weights (7) using the residuals and scale from the previous inner iteration and the elements of $\mathbf{w}$ are given by (26). The inner loop terminates when the weighted residual power $\mathbf{r}^{\mathrm{H}}\mathbf{ur}$ does not change below a threshold value, typically 1–2 per cent, or the solution $\hat{\mathbf{z}}_{\#}$ does not change below a threshold value, typically a fraction of a standard error. After all of the outer loop iterations, a second pass is made with an identical outer loop and an inner one with the scale fixed at the final Huber value and the more severe robust weight (10), again terminating when the weighted residual power does not change appreciably. Extension to the ordinary remote reference solution (24) or its generalization is straightforward, and will be denoted by $\hat{\mathbf{z}}_{\#}^{r}$ and $\hat{\mathbf{z}}_{\#}^{q}$, respectively.

7 TWO-STAGE BOUNDED INFLUENCE MT RESPONSE FUNCTION ESTIMATION

Correlated noise of cultural origin in the local electric and magnetic fields is quite common, and hence extending processing methods to attack such problems is of considerable interest. Robust remote reference processing using (24) can be sensitive to correlated noise on the local electric and magnetic field channels even if the reference channels are noise-free. However, if a reference site which is not contaminated by the noise source is available, this can be remedied by a two-step procedure where the local and reference magnetic fields are first regressed using the bounded influence form of (18)–(19) to get a noise-free $\hat{\mathbf{b}}$, and then the bounded influence form of (6) is solved after substituting $\hat{\mathbf{b}}$ for $\mathbf{b}$. At each step, the absence of correlated noise in the predictor variables makes the noise which remains in the response variables appear like outliers so that it can be eliminated by the bounded influence weighting procedure.

The first stage is the straightforward bounded influence solution of (18) using the procedures described in Sections 2 and 5. The bounded influence estimate of the projected magnetic field is

$$\hat{\mathbf{b}} = \sqrt{\mathbf{u}_1}\mathbf{Q}(\mathbf{Q}^{\mathrm{H}}\mathbf{u}_1\mathbf{Q})^{-1}(\mathbf{Q}^{\mathrm{H}}\sqrt{\mathbf{u}_1}\mathbf{b}) \tag{28}$$

where $\mathbf{u}_1$ is an $N \times N$ diagonal matrix which contains the first stage bounded influence weights as defined in Section 6. The second stage in the solution is the bounded influence regression of $\hat{\mathbf{b}}$ on $\mathbf{e}$ to give

$${}^{\#}\hat{\mathbf{z}}^{q}_{\#} = \left(\mathbf{b}^{\mathrm{H}}\mathbf{u}\mathbf{b}\right)^{-1}\left(\mathbf{b}^{\mathrm{H}}\mathbf{u}\mathbf{e}\right) \tag{29}$$

where $\mathbf{u} = \mathbf{u}_1\,\mathbf{u}_2$, $\mathbf{u}_2$ is the second-stage bounded influence weight matrix computed as in Section 6 and $\mathbf{u}_1$ remains fixed at the final values from the first-stage solution.

8 PRE-SCREENING OF DATA

The procedures described in Sections 3–7 work well unless a large fraction of the data are quite noisy, as can occur when the natural signal level is comparable to or below the instrument noise level. This frequently occurs in the so-called dead bands between 0.5 and 5 Hz for MT and 1000 and 10 000 Hz for audiomagnetotellurics (AMT), where natural source electromagnetic spectra display relative minima. In this instance, robust or bounded influence estimators can introduce bias into the response tensor beyond that which would ensue from the use of ordinary least squares. This occurs because most of the data are noisy and the unusual values detected and removed by data-adaptive weighting are actually those rare ones containing useful information.

The solution to this difficulty is the addition of a pre-screening step which selects only those data which have an adequate signal-to-noise ratio applied prior to a robust or bounded influence estimator. Three approaches have been suggested, and others will undoubtedly evolve as future situations require. Egbert & Livelybrooks (1996) applied a coherence screening criterion to single-station MT dead-band data to select only those data segments with high electric to magnetic field multiple coherence for subsequent robust processing. Garcia & Jones (2002) pre-sorted AMT data based on the power in the magnetic field channels, selecting only those values which exceed the known instrument noise level by a specified amount. Jones & Spratt (2002) pre-selected data segments whose vertical field power was below a threshold value to minimize auroral source field bias in high-latitude MT data. All of these studies demonstrated substantially better performance for data-adaptive weighting schemes after pre-sorting. Similar pre-processing stages can be applied to data with single or multiple reference sites, although the number of data sets and hence the number of candidate selection criteria are increased.

9 STATISTICAL VERIFICATION

Quantitative tests for the presence of outliers and leverage points do not currently exist for realistic situations where anomalous data occur with multiplicity and intercorrelation. In addition, the difficulty of devising such tests rises rapidly with the number of available data and when the actual distribution of the outliers and leverage points is unknown. This means that *a priori* examination of a data set for bad values is not feasible, and hence it is necessary to construct *a posteriori* tests to validate the result of a bounded influence analysis. Further, due to the inherently non-linear form of bounded influence regression solutions like (24) and (29), and because they implicitly involve the elimination of a significant fraction of the original data set, it is prudent to devise methods for assessing the statistical veracity of the result. These can take many forms, as described in Belsley *et al.* (1980) and Chatterjee & Hadi (1988). In an MT context, the most useful have proved to be the multiple coherence suitably modified to allow for remote reference variables and adaptive weights, and quantile–quantile plotting of the regression residuals and hat matrix diagonal elements against their respective target distributions for Gaussian data.

In classical spectral analysis, the multiple coherence is defined as the proportion of the power in the response variable at a given frequency which is attributable to a linear regression with all of the predictor variables (e.g. Koopmans 1995, p. 155). In MT, this is just the coherence between the observed electric field $\mathbf{e}$ and its prediction $\hat{\mathbf{e}}$ given by (11). Care must be taken to incorporate the bounded influence weights for the first and second stages, as appropriate. Let the weighted cross-power between vector variables $\mathbf{x}$ and $\mathbf{y}$ be defined as

$$\mathbf{S}_{\mathbf{xy}} = \mathbf{x}^{\mathrm{H}}\mathbf{u}\mathbf{y}. \tag{30}$$

For two-stage processing, either $\mathbf{x}$ or $\mathbf{y}$ correspond to $\mathbf{b}$ and hence will also implicitly contain the corresponding first-stage weights. Using this notation, the generalized multiple squared coherence between the observed electric field $\mathbf{e}$ and its predicted value $\hat{\mathbf{e}}$ is

$$\gamma^2_{e\hat{e}} = \frac{\left|\mathbf{S}_{e\mathbf{b}}(\mathbf{S}_{\hat{\mathbf{b}}\mathbf{b}})^{-1}\mathbf{S}_{\hat{\mathbf{b}}e}\right|^2}{\mathbf{S}_{ee}\mathbf{S}^{\mathrm{H}}_{\hat{\mathbf{b}}e}\left(\mathbf{S}^{\mathrm{H}}_{\hat{\mathbf{b}}\mathbf{b}}\right)^{-1}\mathbf{S}_{\mathbf{bb}}(\mathbf{S}_{\hat{\mathbf{b}}\mathbf{b}})^{-1}\mathbf{S}_{\hat{\mathbf{b}}e}} \tag{31}$$

where $\hat{\mathbf{b}}$ is replaced by $\mathbf{b}_{\mathbf{r}}$ for single-site remote reference processing. In either instance, (31) is a complex quantity. Its amplitude is analogous to the standard multiple squared coherence and its phase is a measure of the similarity of the local and remote reference variables. In most instances, the phase should be very close to zero; a significant non-zero phase is indicative of source fields with spatial scales that are significant compared to the separation between the local and reference magnetic field sites. When $\hat{\mathbf{b}}$ is replaced by the local magnetic field $\mathbf{b}$, (31) reduces to the standard multiple squared coherence, and implicitly has zero phase.

A second and essential tool for analysing the results of a bounded influence MT analysis, either with or without two-stage processing, is comparison of the original and final quantile–quantile plots of the residuals and the hat matrix diagonal. Quantile–quantile or q–q plots compare the statistical distribution of an observed quantity with the corresponding theoretical entity, and hence provide a qualitative means for assessing the statistical outcome of robust processing. The N quantiles of a distribution divide the area under a pdf into

$N + 1$ equal-area pieces, and hence define equal probability intervals. They are easily obtained by solving

$$F(q_j) = \frac{j - \frac{1}{2}}{N} \tag{32}$$

for q_j, where $j = 1, \ldots N$, and $F(x)$ is the appropriate cdf. A q–q plot compares the quantiles with the order statistics obtained by ranking and sorting the data along with a suitable scaling to make them data unit independent. The advantage of q–q plots over some alternatives is that it emphasizes the distribution tails; most of a q–q plot covers only the last few per cent of the distribution range.

For residual q–q plots, it is most useful to compare the quantiles of the Rayleigh distribution with the order statistics of the residuals scaled so their second moment is 2, corresponding to that expected of a Rayleigh variate. For hat matrix diagonal q–q plots, the quantiles of the β $(p, N - p)$ distribution may be compared with the order statistics with both quantities scaled by p/N, so a value of 1 corresponds to the expected value.

The use of data-adaptive weighting which eliminates a fraction of the data requires that the quantiles be obtained from the truncated form of the original target distribution or else the result will inevitably appear to be short-tailed. The truncated distribution is easily obtained from the original one. Suppose data are censored in the process of robust or bounded influence weighting using an estimator such as those described in Sections 3, 5, 6 or 7. Let $f_X(X)$ be the pdf of a random variable X prior to censoring; this may be the Rayleigh distribution for the residuals or the beta distribution for the hat matrix diagonal. After truncation, the pdf of the censored random variable X' is

$$f_{x'}(x') = \frac{f_x(x')}{F_x(b) - F_x(a)} \tag{33}$$

where $a \leq x' \leq b$ and $F_X(X)$ is the cdf for $f_x(x)$. Let N be the original and M be the final number of data, so that $M = N - m_1 - m_2$, where m_1 and m_2 are the number of data censored from below and above, respectively. Suitable choices for a and b are the m_1th and $(N - m_2)$th quantiles of the original distribution $f_x(x)$. The M quantiles of the truncated distribution can then be computed from that of the original one using

$$F_x(q_j) = (F_x(b) - F_x(a))\frac{j - \frac{1}{2}}{M} + F_x(a) \tag{34}$$

where $j = 1, \ldots .M$.

Whether the data have been censored or not, a straight-line q–q plot indicates that the residual or hat matrix diagonal elements are drawn from the target distribution. Data which are inconsistent with the target distribution are suggested by departures from linearity which are usually manifest as sharp upward shifts in the order statistic. With MT data this is often extreme, and is diagnostic of long-tailed behaviour where a small fraction of the data occur at improbable distances from the distribution centre and hence will exert undue influence on the regression solution. The residual q–q plots from the output of a robust or bounded influence estimator should be approximately linear or slightly short tailed to be consistent with the optimality requirements of the Gauss–Markov theorem. However, the theorem does not prescribe a distribution for the predictor variables, and hence there is not a requirement for the hat matrix diagonal distribution to be β $(p, N - p)$ unless the predictors are actually Gaussian. Nevertheless, hat matrix diagonal q–q plots are useful in detecting extreme values, as will be shown in the next section.

Distributions and confidence limit estimates for the quantiles are given by David (1981). The qualitative use of q–q plots can be quantified by testing the significance of a straight line fit utilizing these results. Alternately, non-parametric tests for the goodness-of-fit to a target distribution of the Kolmogorov–Smirnov type can be applied.

10 EXAMPLES

The first example data set is taken from site 006 of a 150 km long, east–west wideband MT Lithoprobe (BC87) transect in British Columbia obtained in 1987 and described by Jones *et al.* (1988) and Jones (1993). These data have been used extensively in studies of distortion and its removal (see Chave & Jones (1997) for a summary). The data were sampled at a 12 Hz rate for about 17 hr and include a remote reference located 1 km from the local site. Each data series was processed into Fourier transform estimates at selected frequencies using variable section lengths (see Section 3) after pre-whitening with a five-term robust AR filter and tapering with a Slepian sequence with $\tau = 1$. Fig. 2 compares the xy (where x is north and y is east) apparent resistivity and phase estimates $\hat{\mathbf{z}}^r_*$ and $\hat{\mathbf{z}}^r_\#$ obtained using the ordinary robust (hereafter OR) and bounded influence (hereafter BI) remote reference estimators respectively, as described in Sections 5 and 6. The BI result was computed with the cut-off parameter χ in (26) taken as the 0.99999 (or upper 0.001 per cent) point of the beta distribution with $(2, N - 2)$ degrees of freedom. In general, the OR and BI results are similar except in the band between 2 and 20 s, where the OR estimates are biased relative to the BI ones and the OR jackknife confidence limits are sometimes dramatically larger, reflecting the effect of leverage. There are also more subtle but statistically significant differences between the two types of estimates at periods shorter than 2 s, as shown below. The differences seen in Fig. 2 cannot be removed by coherence thresholding; pre-processing of the data to eliminate all data sections with electric to magnetic field squared multiple coherence under 0.9 has no effect beyond a slight increase in the confidence limits due to a decrease in the effective degrees of freedom.

Fig. 3 shows an example of the severe bias that can ensue from pre-whitening with a least-squares AR filter rather than the robust AR filter described in Section 2. The data and BI methodology are those used for Fig. 2, and the only difference between the two results is in the pre-whitening. Note the erratic results at periods under 2 s (where the data power spectral density is weakest) obtained with the least-squares pre-whitening filter.

Fig. 4 shows a complex plane view of various response estimates for these data at a period of 5.3 s. This period is in the middle of the band where substantial differences between the confidence limits from the OR and BI estimators obtain. The ordinary least-squares (hereafter OLS) estimate given by (4) is compared with the OR and BI results; the latter are shown with the cut-off parameter χ in (26) taken as the 0.999, 0.9999, and 0.99999 (upper 0.1 per cent, 0.01 per cent and 0.001 per cent respectively) points of the beta distribution for $(2, N - 2)$ degrees of freedom, where $N = 4342$ is the number of estimates. There is no statistically significant difference between the estimates with substantially different values of the BI cut-off parameter, indicating that its selection is not critical. In Fig. 4, both the OLS and OR estimates display large uncertainties, reflecting heteroscedasticity that is not removed by data weighting based entirely on the size of the regression residuals despite the elimination of 7.2 per cent of the data. The BI estimator removes an additional 6.8, 8.5 and 11.5 per cent of the data as the cut-off varies

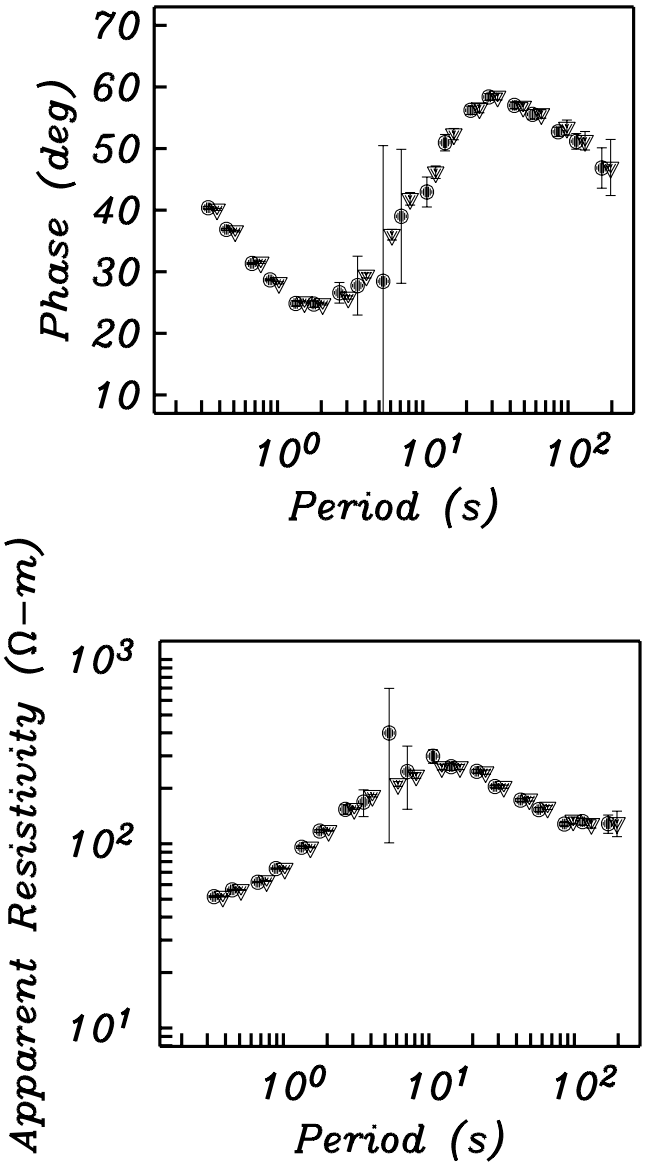

Figure 2. The apparent resistivity (bottom) and phase (top) as a function of period for the Z_{xy} component of the BC87 site 006 data described in the text. The error bars are double sided 95 per cent jackknife confidence limits. The solid circles denote the ordinary robust (OR) and the solid triangles denote the bounded influence (BI) remote reference results as described in the text. Spectra used for the computations have a TBW of 1 and were pre-whitened using a five-term robust AR filter. The BI estimates have been offset to slightly longer periods for clarity of presentation.

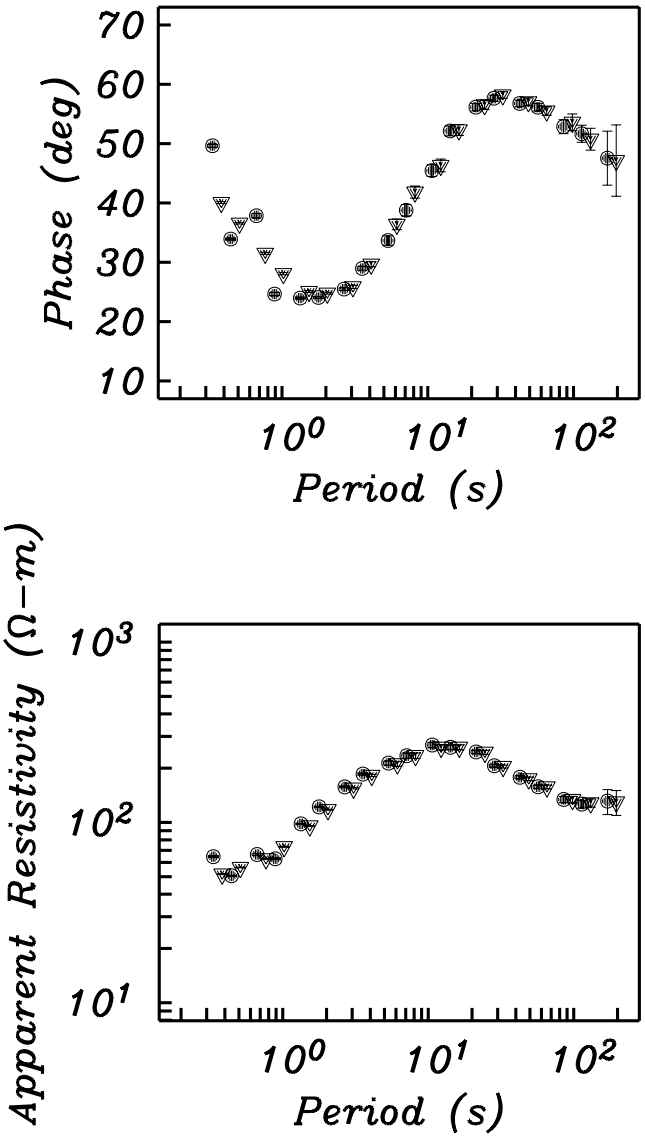

Figure 3. The apparent resistivity (bottom) and phase (top) as a function of period for the BI estimate shown in Fig. 2 computed using a conventional least-squares (solid circles) and robust (solid triangles) five-term autoregressive pre-whitening filter. The robust AR estimates have been offset to slightly longer periods for clarity of presentation. The variability with the least-squares approach is caused by AR filter instability which results in extreme spectral leakage where the data power spectral density is low.

between 0.99999 and 0.999, indicating that only a small fraction of the data is responsible for the heteroscedasticity. Further, the standard error is reduced by more than a factor of 34 over the OR result from bounding the influence.

Fig. 5 shows analogous complex plane responses from the region where more subtle response function differences are observed at a period of 0.89 s. There are large distinctions between both the OLS and OR and the OR and all of the BI estimates, although the relative standard error is less variable than for Fig. 4. In this instance, the *M*-estimator solution displays a substantially smaller standard error than both the OLS and BI values. It is also markedly biased compared with the BI result; the *M*-estimator solution differs from the BI one by more than six standard errors of the latter.

Figs 6 and 7 compare residual and hat matrix q–q plots for the OLS and 0.99999 BI estimates at 5.3 s shown in Fig. 4. The OLS q–q plots in Fig. 6 are both markedly long tailed, indicating the presence of severe outliers and leverage points, and in fact the hat matrix distribution appears to be more like a long-tailed version of log beta than long-tailed beta. The most serious outliers are about 50 standard deviations from the Rayleigh mean, while the most serious leverage points are over 1000 times the expected value of the hat matrix diagonal. In contrast, the BI residual q–q plot in Fig. 7 is only weakly long-tailed, and the upward concavity of the result is indicative of a residual distribution which is slightly but pervasively longer tailed than Rayleigh rather than the presence of significant outliers. The OR residual q–q plot is indistinguishable from that in Fig. 7, while the OR hat matrix q–q plot is very similar

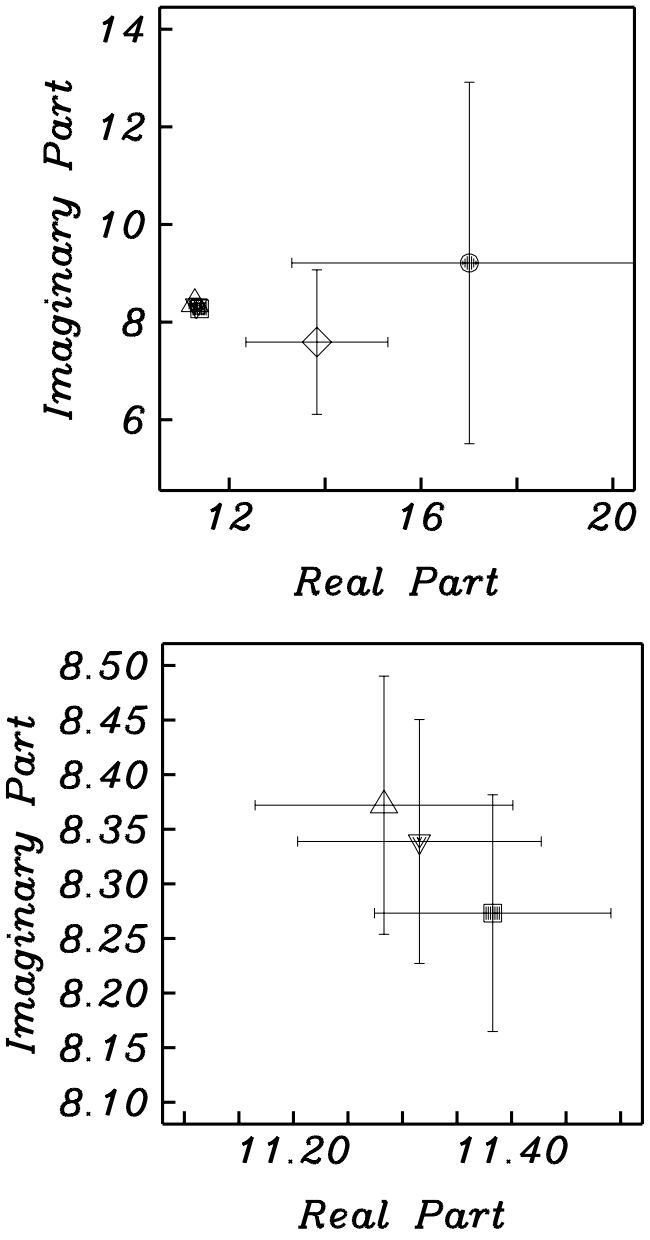

Figure 4. Complex plane view of the estimate in Fig. 2 at a period of 5.3 s. Each symbol is plotted with the jackknife standard error. The bottom panel is a magnified view with different *x*- and *y*-axis limits showing the three BI estimates at the left in the upper panel. The symbols correspond to OLS (open diamond), OR (solid circle) and BI estimates with a cut-off parameter χ in (26) at the 0.99999 (solid square), 0.9999 (solid triangle), and 0.999 (open triangle) points of the beta distribution with $(2, N-2)$ degrees of freedom, where N is the number of estimates. N is 4342 for the OLS estimate.

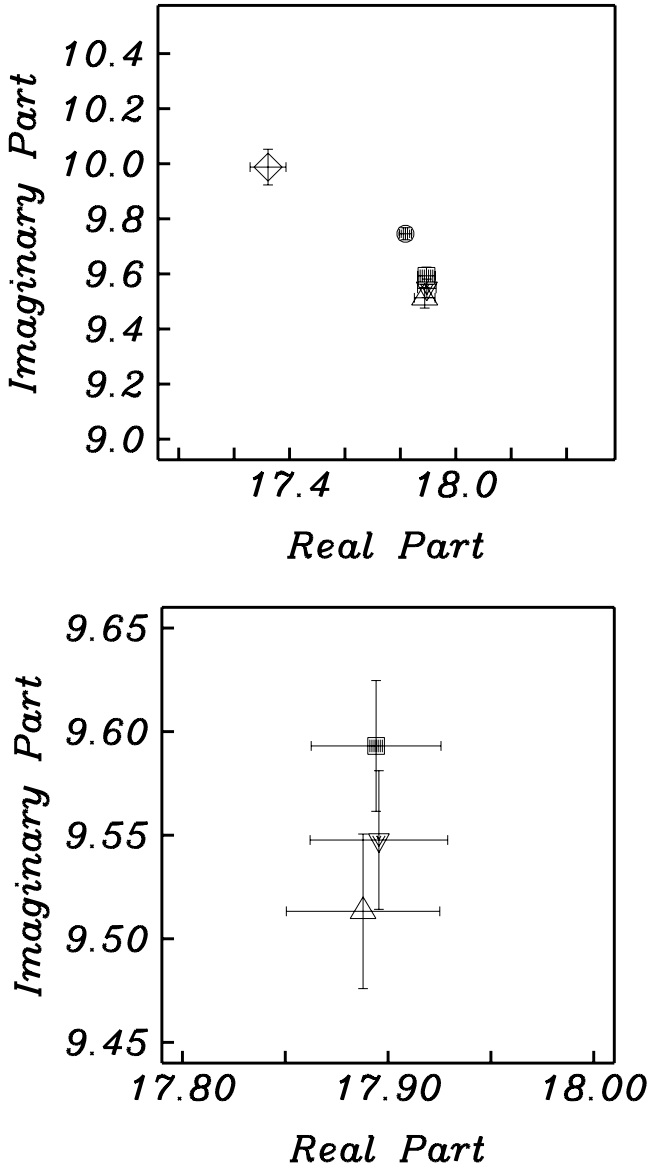

Figure 5. Complex plane view of the estimate in Fig. 2 at a period of 0.89 s. Each symbol is plotted with the jackknife standard error. The bottom panel is a magnified view with different *x*- and *y*-axis scales showing the three BI estimates at the bottom in the upper panel. The symbols correspond to OLS (open diamond), OR (solid circle) and BI estimates with a cut-off parameter χ in (26) at the 0.99999 (solid square), 0.9999 (solid triangle), and 0.999 (open triangle) points of the beta distribution with $(2, N-2)$ degrees of freedom, where N is the number of estimates. N is 34 788 for the OLS estimate.

to that in Fig. 6, indicating that robust weighting has little influence on leverage points for this data set. It is the failure to eliminate leverage effects which accounts for the large confidence limits on the OR estimate in Figs 2 and 4. The BI hat matrix q–q plot is approximately log beta with some smaller than expected values at the lower end, but the huge leverage points at the upper end of the distribution have been eliminated. It is easy to modify the bounded influence estimator to eliminate unusual data at both the lower and upper ends of the distribution, but the ensuing estimates are not changed from those shown in Fig. 4.

Fig. 8 compares OLS and OR q–q plots for the hat matrix diagonal corresponding to the 0.89 s estimates shown in Fig. 5. Comparison of the two shows that robust weighting has actually increased the size of the most extreme leverage points, possibly accounting for the difference between the OR and BI results in Fig. 5. This phenomenon is typical of OR procedures rather than unusual, and is not surprising given the form of (24), which is independent of any leverage measure. In fact, all of the short-period responses in Fig. 2, where the OR and BI estimates show large statistical differences, display this effect, indicating serious bias in the absence of leverage control.

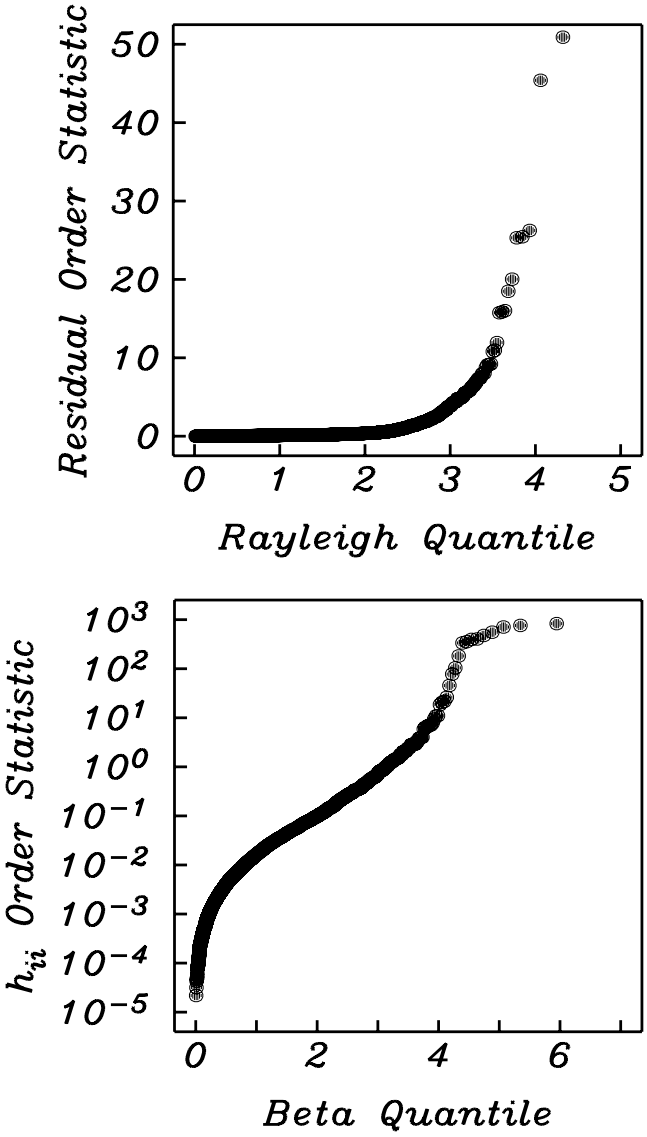

Figure 6. Quantile–quantile plots for the regression residuals (top) and hat matrix diagonal (bottom) for the OLS solution of Fig. 4. The absolute values of the regression residuals are scaled to have a variance of 2 and plotted against the Rayleigh quantiles in the top panel, while the hat matrix diagonal elements are scaled by N/p, converted to logarithms, and plotted against $\beta(p, N-p)$ quantiles, also scaled by N/p, in the bottom panel. Note the extremely long-tailed form of both quantities.

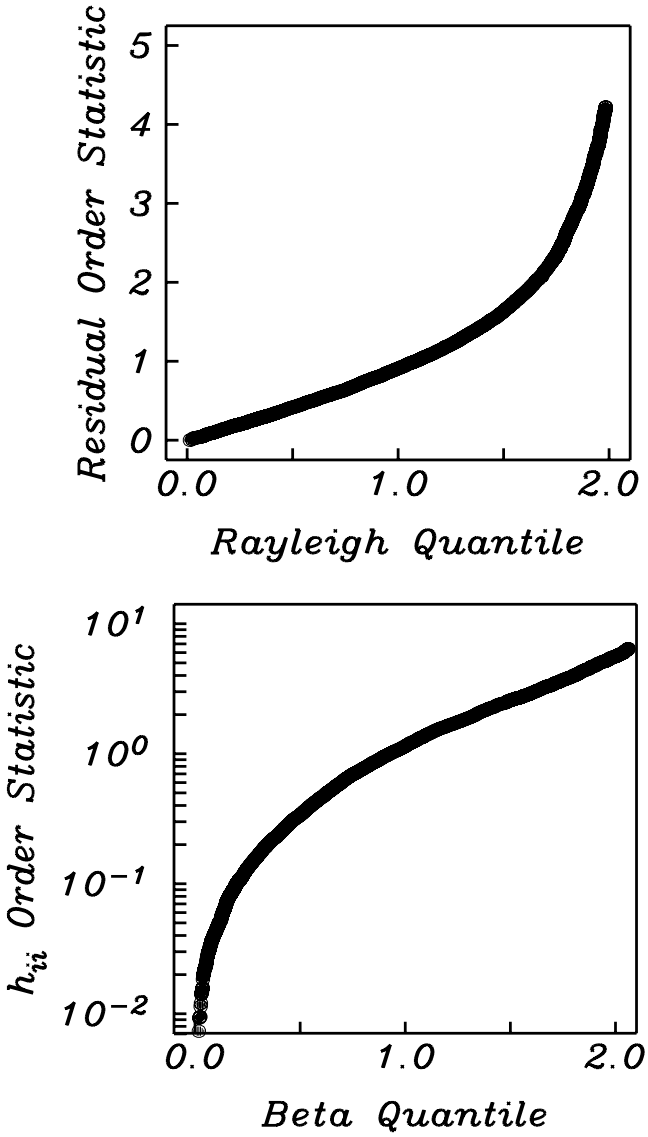

Figure 7. Quantile–quantile plots for the regression residuals (top) and hat matrix diagonal (bottom) for the 0.99999 BI solution in Fig. 4. See Fig. 6 caption for plotting details. The target distributions are the truncated forms of the Rayleigh and beta distributions, as discussed in Section 9. The residual distribution is slightly longer tailed than the Rayleigh, while the hat matrix diagonal is approximately log beta with some extreme values at the lower end.

The second example is taken from the Southeast Appalachians (SEA) transect extending from South Carolina to Tennessee collected in 1994 (Wannamaker *et al.* 1996). The data sets used here were part of the long-period MT effort described by Ogawa *et al.* (1996), and consist of 5 s sampled, five-component MT data collected contemporaneously at many sites. The local site will be taken as SEA335 (36°12′17″ N, 81°55′20″ W) which is located within the Blue Ridge Mountain belt, and the remote sites will be taken as SEA320 (36°38′52″ N, 82°19′05″ W) and SEA360 (34°53′08″ N, 80°16′30″ W), located 61 and 209 km to the northwest and southeast of the local site, respectively. The data have been rotated into an interpretation coordinate system, so that the x direction strikes N50E (or approximately along the strike direction for the Appalachian Mountains) and the y direction strikes N140E. All of the data are fairly clean and free of obvious problems, although the local site does display strong current channelling such that the Z_{xy} apparent resistivity is about a hundredth of the Z_{yx} apparent resistivity. In what follows, all spectral estimates were computed using a time-bandwidth three-Slepian sequence after pre-whitening with a five-term robust AR filter, and the fourth and sixth frequencies in each subsection were selected for processing to avoid contamination from unresolved low-frequency components.

In fact, these data are sufficiently clean that nearly identical processing results are obtained with most standard methods, yielding a baseline result for comparison with various artificially contaminated versions of the data. As a demonstration, Fig. 9 compares the yx component of the $\hat{\mathbf{z}}_{\#}^{r}$ apparent resistivities and phases using sites 320 and 360 separately as remotes. The results for different choices of remote references in Fig. 9 are indistinguishable. The $\hat{\mathbf{z}}_{\#}^{q}$ estimate computed with the two-stage formalism described in Sections 5 and

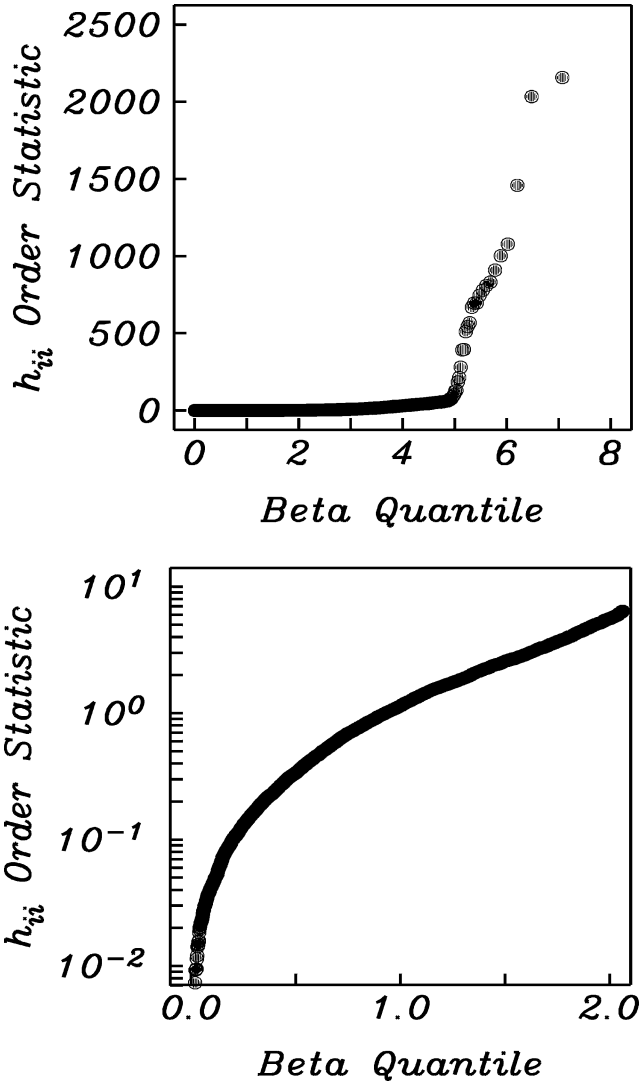

Figure 8. Quantile–quantile plots for the hat matrix diagonal elements for the OLS (top) and OR (bottom) estimates of Fig. 5. Note that robust weighting in the bottom panel has increased the size of extreme leverage points.

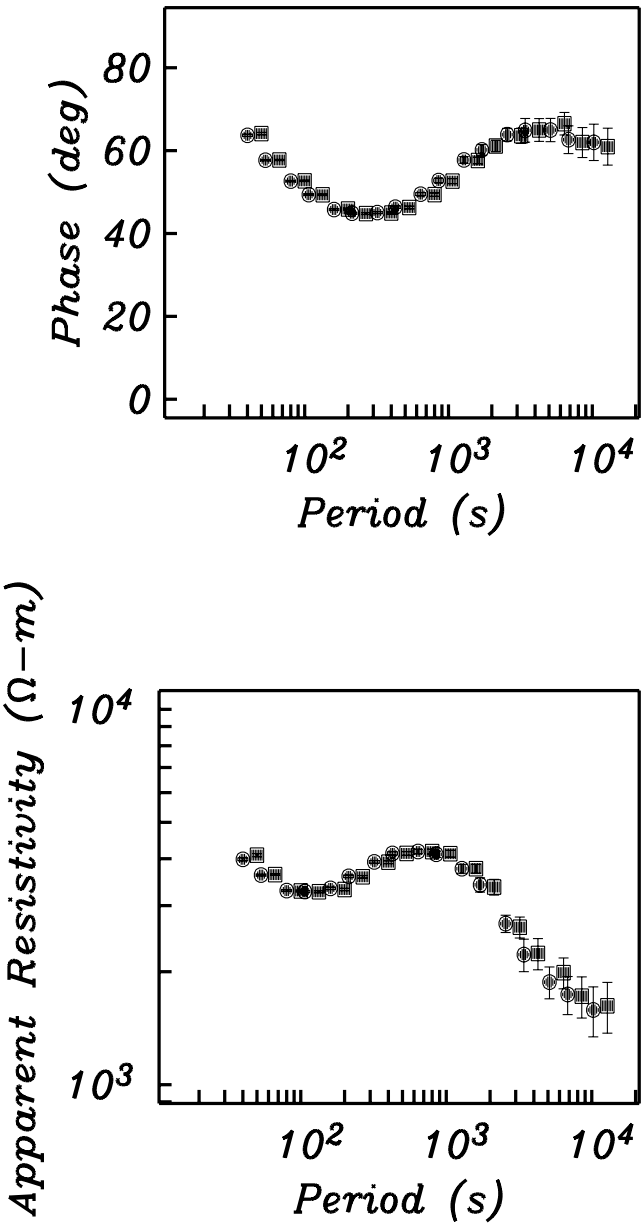

Figure 9. The BI Z_{yx} apparent resistivity and phase for SEA long-period MT transect site 335 computed using site 320 (solid circles) and site 360 (solid squares) as remote references. The site 360 result has been offset in period for clarity. The error bars are the double-sided 95 per cent confidence limits estimated using the jackknife.

7 using sites 320 and 360 simultaneously as remotes is essentially identical, reflecting the high quality of the data and the low environmental noise level. Residual and hat matrix q–q plots show the presence of only weak outliers. In fact, the OR estimates $\hat{\mathbf{z}}^r_*$ and $\hat{\mathbf{z}}^q_*$ are nearly identical to the BI estimates shown in Fig. 9, so that leverage effects in these data are minimal. The generalized coherence (31) is very high across the band while its phase remains near zero, reflecting the lack of source-field complications at this mid-latitude site. Because of these data features, little has been gained by using multiple reference sites. Comparable results are obtained for the *xy* component (not shown), although the longest period estimates have large uncertainties due to the very weak electric field in the *x* direction. The $\hat{\mathbf{z}}^r_\#$ response using site 320 as a remote will be used as a reference response in the sequel.

As a simulation of cultural effects, noise was added to both horizontal components of the site 360 magnetic field data. This consists of random samples drawn from a Cauchy distribution which have been filtered forward and backward with a fifth-order Chebyshev type-1 bandpass filter having cut-off points at about 200 and 1000 s. The MAD of the Cauchy noise was adjusted to be about one-third that of the data. The Cauchy distribution (Johnson *et al.* 1994, Chapter 16) is equivalent to Student's *t* distribution with one degree of freedom and is substantially longer tailed than the Gaussian. As a result, the additive noise appears to be quite impulsive, with peak values that exceed those of the original magnetic data by up to a factor of 10 000, and in fact the noise level is high enough to totally obscure the original data in a time-series plot.

Fig. 10 compares the Z_{yx} BI response using site 320 as a reference as in Fig. 9 with the OR estimates $\hat{\mathbf{z}}^r_*$ and $\hat{\mathbf{z}}^q_*$ using site 360 (with noise) and sites 320+360 (with noise) as remote references, respectively. In the 200–1000 s band, both the apparent resistivities and phases for the site 360 remote estimates are totally useless, being wildly biased by the high noise level in the reference data. However, the

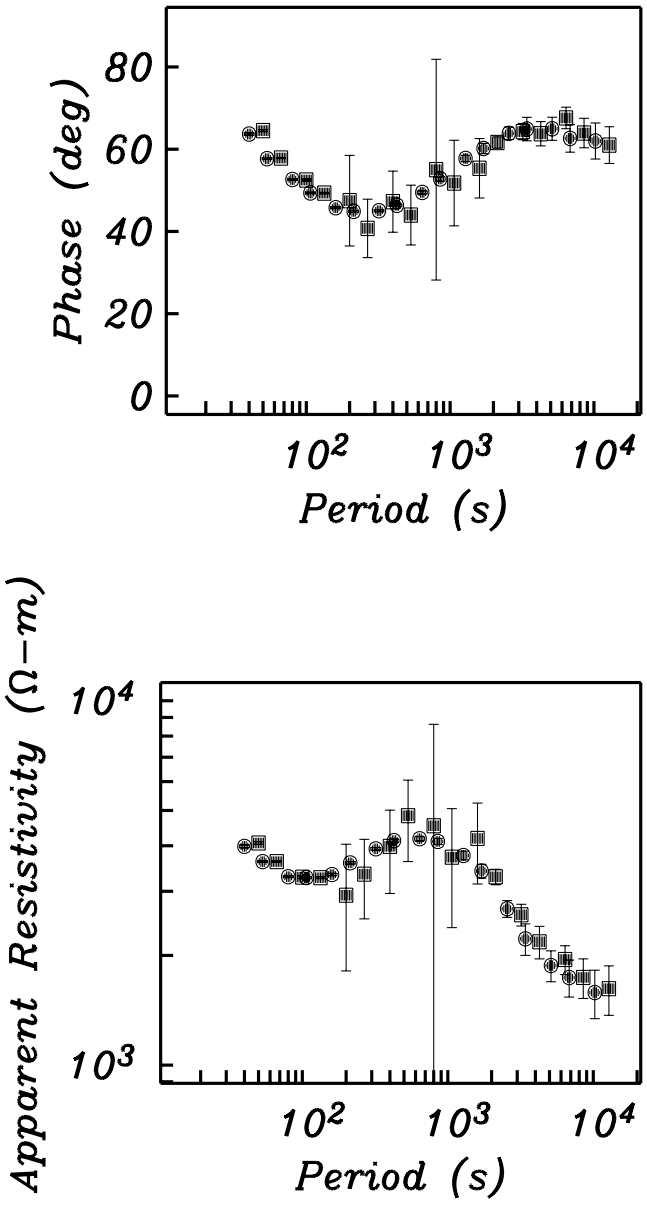

Figure 10. The BI Z_{yx} apparent resistivity and phase for SEA long-period MT transect site 335 computed using site 320 (solid circles) compared with the OR response using (panel A) site 360 (solid squares) and (panel B) sites 320 and 360 simultaneously (solid triangles) as remote references. The site 360 data include additive Cauchy-distributed noise as described in the text. The site 360 (A) and site 320+360 (B) results have been offset in period for clarity. The error bars are the double-sided 95 per cent confidence limits estimated using the jackknife.

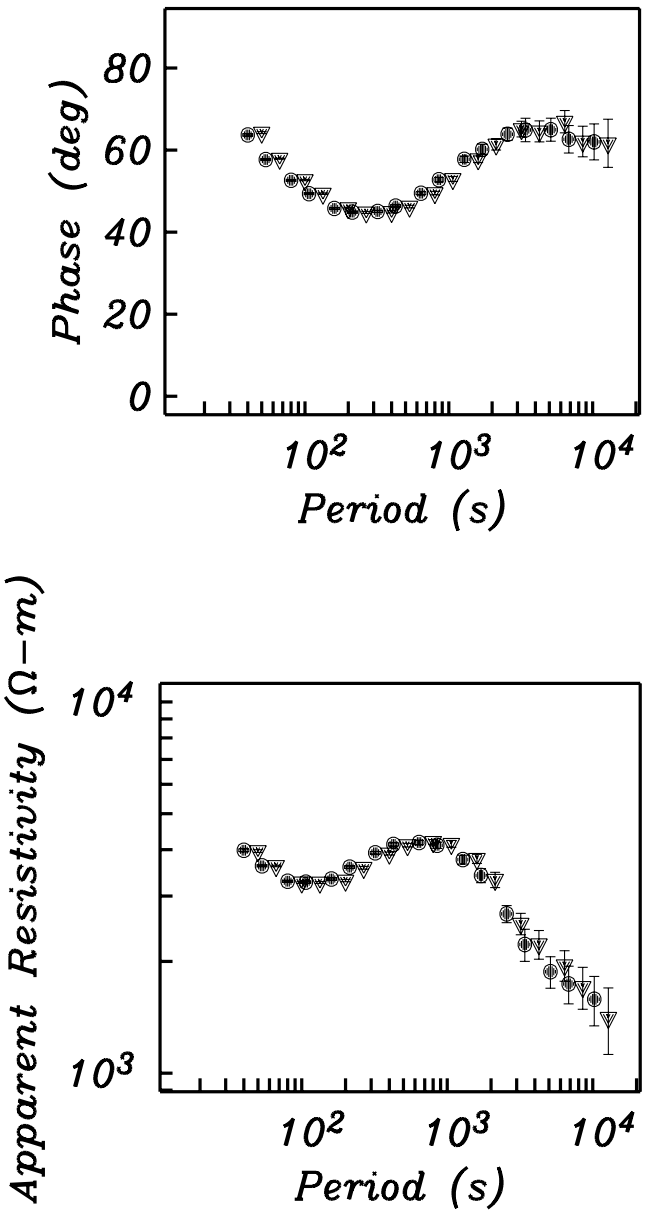

Figure 10. (*Continued*.)

dual remote reference method easily eliminates this effect, and is nearly identical to the result using clean remote reference data. This is a graphic illustration of the value of multiple remote references; good results will be obtained provided that at least one data set is uncontaminated at a given frequency.

Fig. 11 shows the BI counterpart to Fig. 10(a), indicating that control of leverage has substantially reduced both the estimate bias and the size of the confidence limits even with only the noisy site 360 as a remote. As for the OR approach, the dual remote estimate (not shown) is identical to that from uncontaminated data. While the multiple remote solution is clearly the preferred solution, bounding the influence of the remote data set can substantially improve the estimator performance and yield an interpretable response when the reference is noisy. However, the examples in Figs 10 and 11 suggest that multiple rather than single remote references do offer real advantages when one of the remote sites is noisy either over part of the period range of interest or part of the time interval. The two-stage regression approach which is implicit to the dual remote estimator automatically eliminates noisy frequency or time intervals, presuming that at least one clean remote is available.

Fig. 12 shows the magnitude of the t_{xx} component of the BI magnetic transfer tensor relating the x components of the local and remote magnetic fields from the first stage in this process for both noise-free and noisy site 360 data corresponding to Figs 9 and 11, respectively. With the original data, the site 360 transfer function is systematically smaller than its site 320 counterpart by up to a factor of 2 due to some combination of weaker coherence with or larger magnetic field variations at site 320. With the contaminated

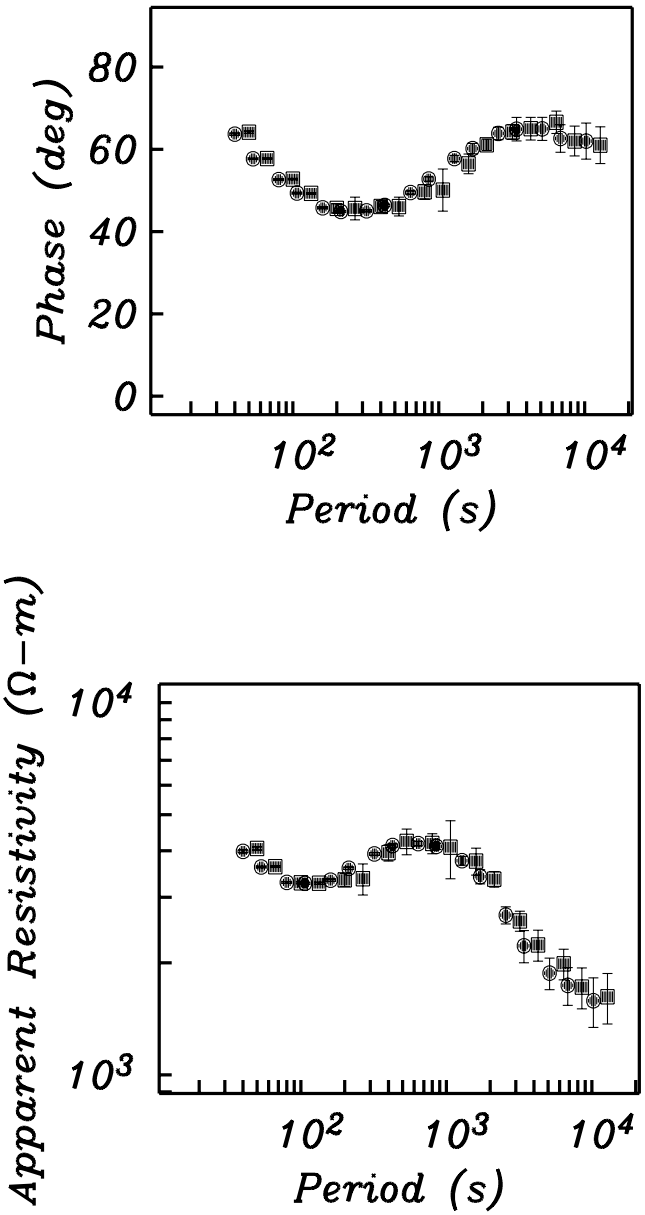

Figure 11. The BI Z_{yx} apparent resistivity and phase for SEA long-period MT transect site 335 computed using site 320 (solid circles) compared with the BI response using site 360 (solid squares) as a remote reference. The site 360 data include additive, band-limited (200–1000 s) Cauchy-distributed noise as described in the text. The site 360 results have been offset in period for clarity. The error bars are the double-sided 95 per cent confidence limits estimated using the jackknife.

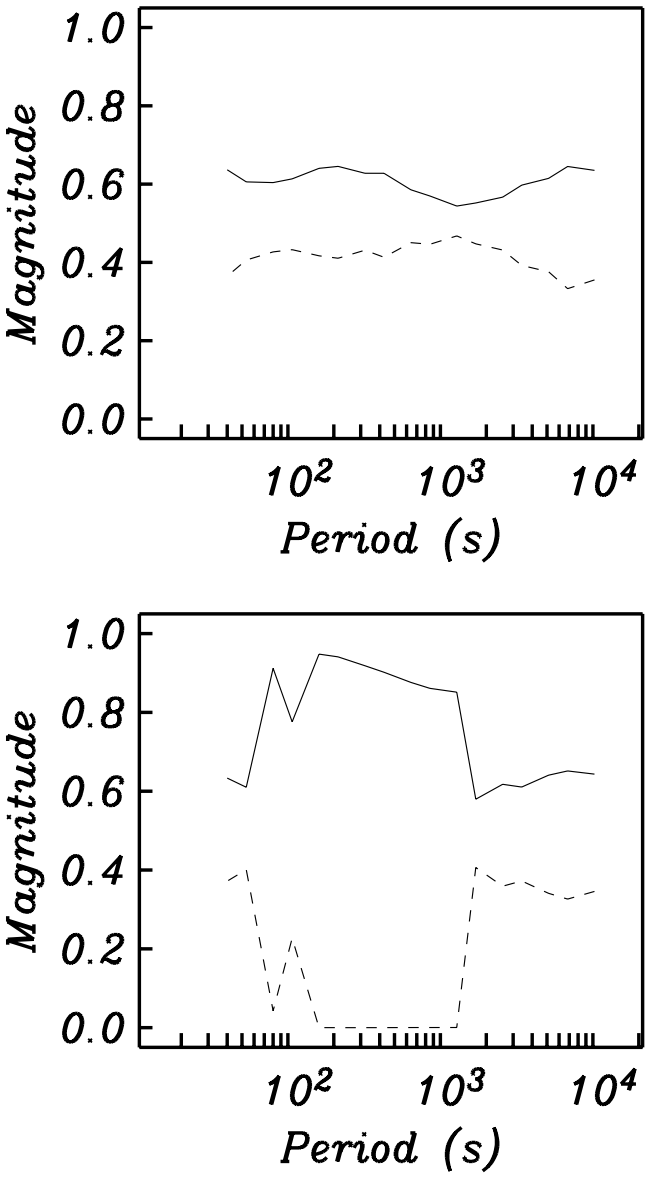

Figure 12. The magnitude of the t_{xx} component of the magnetic transfer tensor from the BI solution to (18) which relates the x component of the local magnetic field at site 335 to the x components of the remote magnetic fields at site 320 (solid line) and site 360 (dashed line). The upper panel utilizes the original remote magnetic data, while the lower panel utilizes the site 360 data with additive, band-limited (200–1000 s) Cauchy-distributed noise as described in the text.

site 360 data, the site 360 transfer function essentially vanishes in the 200–1000 s band.

As a simulation of correlated cultural contamination, bandpassed Cauchy-distributed noise filtered in the same way as for Figs 10–12 polarized 30° clockwise from the x-axis and with a MAD comparable to the original data was added to the local magnetic field variables at site 335. The non-causal impulse responses from the Z_{yx} MT impedances for the uncontaminated data were computed by minimizing a functional of the impulse response subject to fitting the MT impedance; a maximum smoothness constraint in log frequency was imposed as described by Egbert (1992). Simulated electric field noise data were produced by convolving the impulse response with the noise time-series. The size of the noise electric field was also varied by amplitude scaling. The resulting noise data were added to the site 335 data, varying the time duration of the noise from 0 to over 50 per cent of the total. Unlike for the earlier examples, the additive noise is now correlated between the predictor and response variables, and will result in rising bias in the response functions as the noise duration increases unless it is eliminated during processing. Fig. 13 compares BI processing on the original (clean) data with BI and two-stage processing on contaminated data with a noise fraction of 40 per cent and the noise electric field scaled so that its MAD is about 80 times larger than the background. The BI result is erratic over the affected period band since the estimator has difficulty distinguishing good from bad data. Leverage weighting is

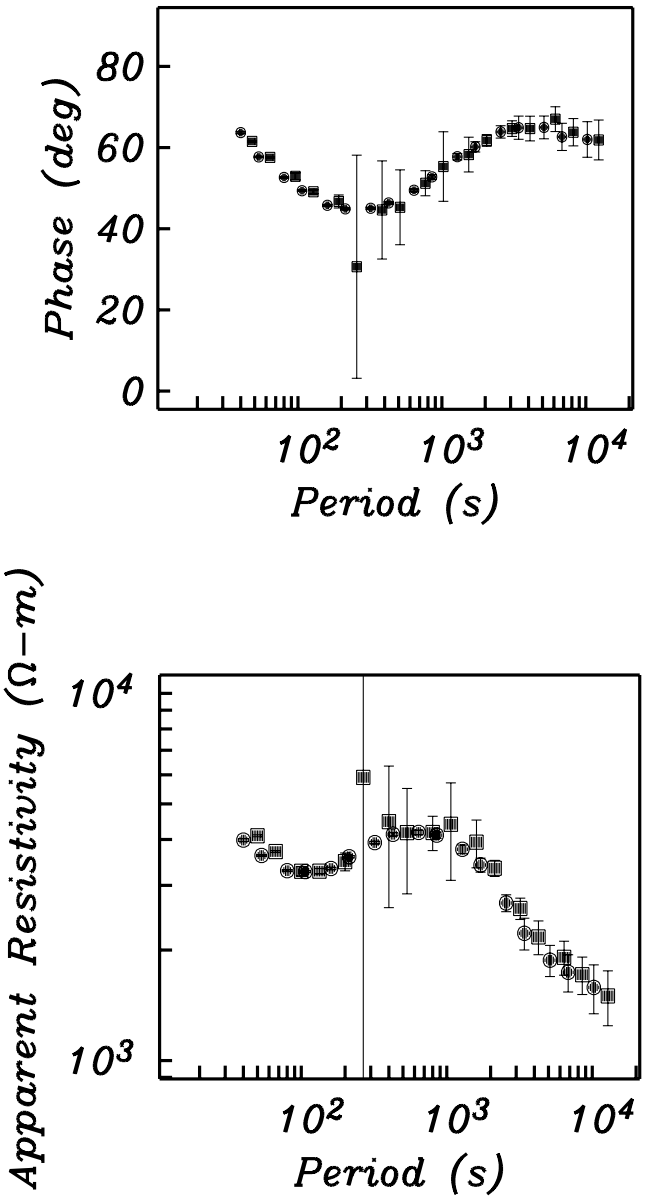

Figure 13. The Z_{yx} apparent resistivity and phase for SEA long-period MT transect site 335 with the electric and magnetic fields contaminated by additive, band-limited (200–1000 s) Cauchy-distributed noise as described in the text. In all cases, the response is computed using site 320 as a remote reference. The solid circles denote the BI estimate using the original (clean) data. This is compared with the (panel A) BI (solid squares) and (panel B) two-stage BI (solid triangles) results, both of which have been offset in period for clarity. The error bars are the double-sided 95 per cent confidence limits estimated using the jackknife.

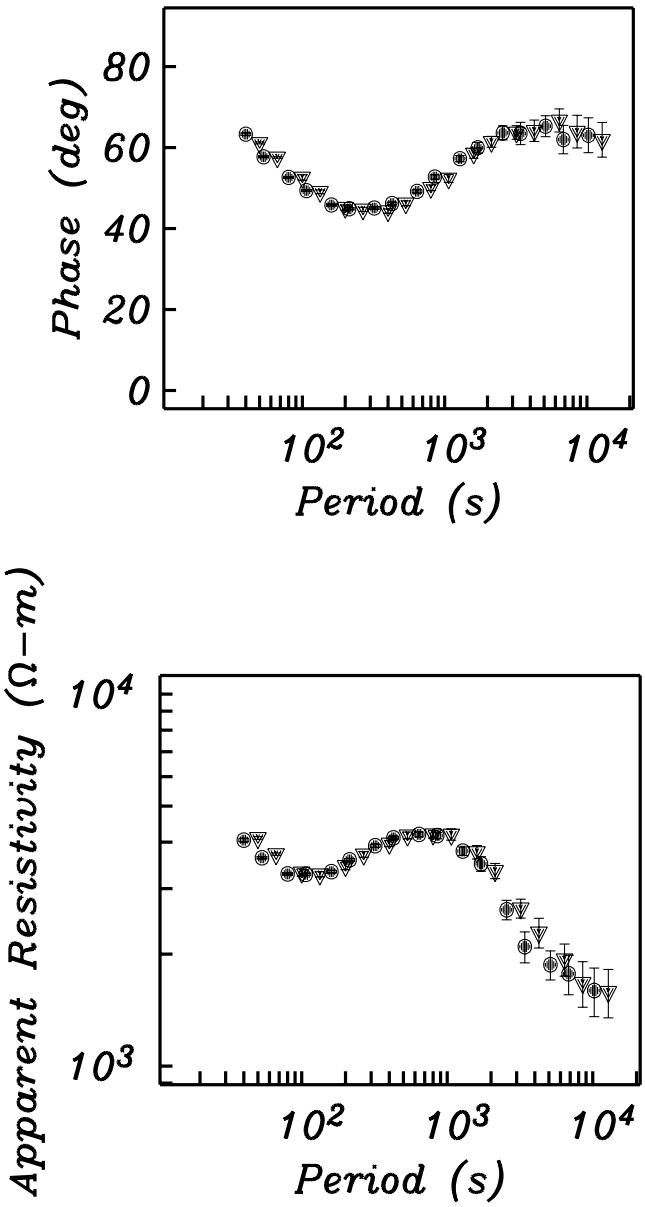

Figure 13. (*Continued.*)

of limited use because the hat matrix is effectively that from the remote reference data (see Section 5) at site 320 which are free of the noise, and the correlated noise on the local electric and magnetic field variables makes it difficult to distinguish data from noise as the noise fraction and amplitude rises. In contrast, the two-stage estimate remains stable with only modest increases in the confidence limits because the first-stage robust weights distinguish and remove the noisy sections from the local magnetic field, and then the second-stage robust weights eliminate any remaining noise in the electric field because the local magnetic field has been cleaned.

Fig. 14 compares residual q–q plots for the BI and two-stage estimators at the 853 s period, graphically illustrating the effectiveness of the latter. For both the BI and two-stage estimates, about 60 per cent of the data have been eliminated in the contaminated band by weighting; the differences in Fig. 13 reflect how completely the noise-contaminated sections have been removed. The two-stage algorithm operates effectively both with high noise amplitude and with a noise fraction approaching 50 per cent.

11 DISCUSSION

Based on the recent MT literature, it would not be controversial to suggest that all MT data processing should employ some type of robust estimator with a remote reference. The examples presented by Jones *et al.* (1989) provide graphic evidence of the value of a robust approach, and the state-of-the-art has certainly improved over the past decade to further strengthen the arguments. The authors are

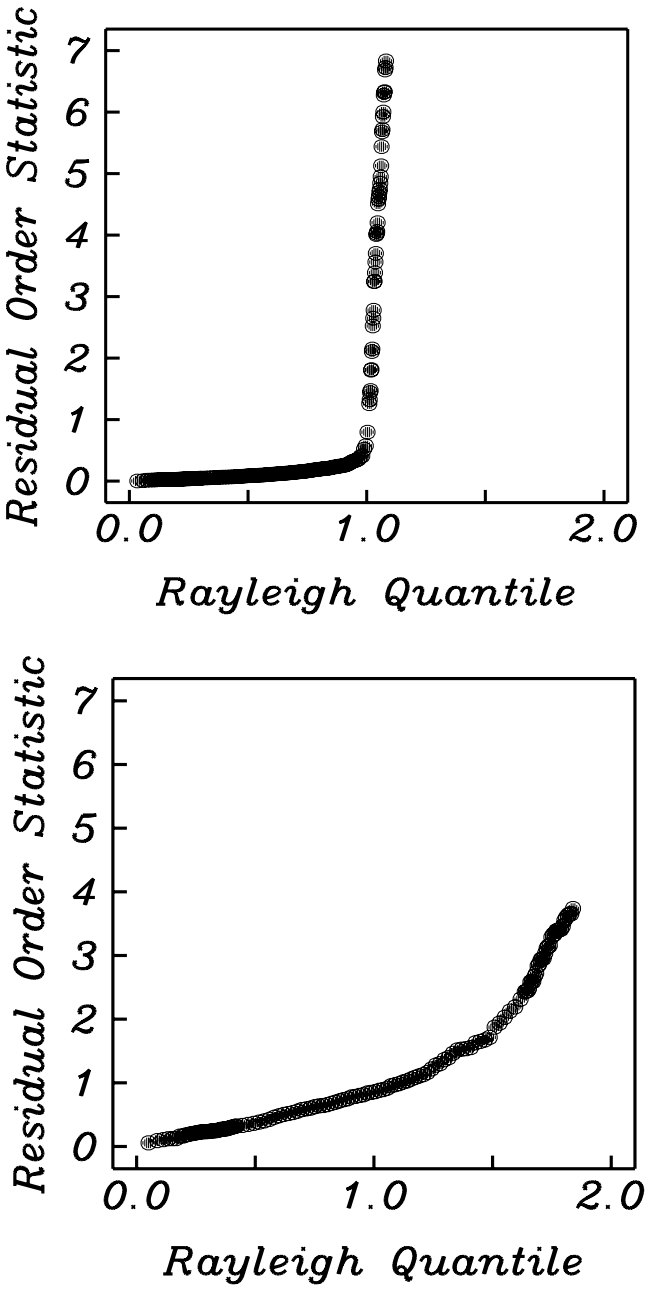

Figure 14. Quantile–quantile plots for the residuals from the BI and two-stage estimators shown in Figs 13(a) (top) and (b) (bottom) at a period of 853 s. The BI residuals are very long tailed, reflecting the inability of the estimator to discriminate between signal and correlated noise, resulting in the erratic response function seen in Fig. 13(a). The two-stage estimator very effectively removes the correlated noise in the local magnetic field during the first stage, enabling the second stage to remove it from the local electric field. This results in a clean q–q plot and a good response function estimate.

not aware of any instances where robust methods fail when applied intelligently rather than blindly. This means that individual attention must be paid to each data set, all of the time-series and their power spectra need to be inspected, and the relevant physics of the induction process must be incorporated into the analysis plan (e.g. the nature of the source fields and any noise components must be understood in at least a gross sense). These steps lead to decisions about data editing, detrending and the pre-processing stage (Section 8). A preliminary non-robust analysis is often useful because it allows the raw data coherences and residual/hat matrix statistics to be assessed (Section 9), and helps to further refine the requirement for pre-processing. It can also indicate whether a standard robust estimator will be sufficient.

Based on the results presented in this paper and the authors' experience over the past decade, it is suggested that standard use of a bounded influence remote reference estimator should supplant that of its robust counterpart. If the enumerated analysis principles are followed, the authors have not seen any examples where bounded influence estimators fail to yield results that are, at a minimum, comparable to robust estimates, and have seen numerous examples where bounded influence results are substantially better. This is especially true as the auroral oval is approached, where substorm effects can produce spectacular leverage (see Garcia *et al.* 1997, and the BC87 examples in Section 10), and in the presence of many types of cultural noise. Robust estimators which weight data sections on the basis of regression residuals frequently fail to detect leverage points, and can easily be dominated by a small fraction of high-leverage data or even a single data point. Bounded influence estimators avoid this trap in a semi-automatic fashion. The bounded influence estimator of Section 6 has a breakdown point of $\sim$0.5, and can handle data sets with an even larger fraction of high leverage values under some circumstances.

It is not difficult to rationalize the underlying log beta form for the non-BI hat matrix diagonal distribution seen in Figs 6 and 8. Ionospheric and magnetospheric processes are extremely non-linear, and more so at auroral than lower latitudes. As a result, their electromagnetic effects are the result of many multiplicative steps. This means that the statistical distribution of the magnetic variations will tend towards log normal rather than normal (Lanzerotti *et al.* 1991), and hence the resulting hat matrix diagonal will tend towards log beta rather than beta.

There are situations where correlated noise is present on both the local electric and magnetic fields where more elaborate processing algorithms may be required. The two-stage algorithm of Section 7 works well in many instances, as shown with a synthetic example in Section 10. Larsen *et al.* (1996) introduced a different two-stage procedure to deal with the correlated noise from DC electric trains that explicitly separates the MT and noise responses. In both instances, a remote magnetic reference which is free of the correlated noise source is required. In the Larsen *et al.* algorithm, a tensor relating the local and remote magnetic fields equivalent to $\mathbf{t}$ in (18) is first computed to give a correlated-noise-free $\hat{\mathbf{b}}$, so that the correlated noise in $\mathbf{b}$ is given by $\mathbf{b} - \hat{\mathbf{b}}$. Both the magnetotelluric response $\mathbf{z}$ and the correlated-noise tensor $\mathbf{z}'$ are obtained from

$$\mathbf{e} = \hat{\mathbf{b}}\mathbf{z} + (\mathbf{b} - \hat{\mathbf{b}})\mathbf{z}' \tag{35}$$

which explicitly separates $\mathbf{e}$ into magnetotelluric signal and correlated noise parts, plus an uncorrelated residual. Eq. (35) is easily solved using the methods of this paper by recasting the problem so that $\hat{\mathbf{b}}$ and $\mathbf{b} - \hat{\mathbf{b}}$ comprise a four-column $\mathbf{b}$ in (3), and $\mathbf{z}$ in (3) is replaced by a four-parameter solution containing $\mathbf{z}$ and $\mathbf{z}'$. However, uncorrelated noise in $\mathbf{b}$ is implicitly combined with the correlated noise $\mathbf{b} - \hat{\mathbf{b}}$, and hence the estimates of $\mathbf{z}$ and $\mathbf{z}'$ will always be biased to some degree. This bias will rise as the relative size of the correlated noise $\mathbf{b} - \hat{\mathbf{b}}$ to signal $\hat{\mathbf{b}}$ increases, although this is hard to quantify. The two-stage estimate is not biased because the correlated-noise tensor $\mathbf{z}'$ is never explicitly computed. A third alternative is the robust principal components method of Egbert (1997), which can detect contamination from correlated noise and yield good MT responses in some cases. It is currently unclear which

approach will yield the best results with real data; evaluating this will require direct comparisons using troublesome data sets.

Using data from a continuously operating MT array in central California, Eisel & Egbert (2001) compared the performance of parametric and jackknife estimators for the confidence limits on the MT response function. Their preferred approach was based on standard asymptotic parametric theory for a robust estimator (their eq. 16). They further argued that both the fixed-weight jackknife, as implemented by Chave & Thomson (1989) and used in this paper, and a computationally more intensive subset deletion jackknife, overestimated the errors, which led to their favouring the parametric approach. However, it appears that the residual distribution of MT data from locations near or within the auroral oval (at a minimum) are systematically longer tailed than Gaussian even after bounded influence estimators are applied; Fig. 7 is typical of such data sets. In this instance, parametric approaches will consistently underestimate the true confidence limits, which could lead to interpretation errors. Whether the residual distribution is systematically long tailed for other MT situations is unclear, although geomagnetically undisturbed data from mid-latitudes appear to more consistently yield nearly Gaussian residuals. Because of this uncertainty, as well as due to the other advantages of the jackknife enumerated by Thomson & Chave (1991), the authors continue to favour the use of the jackknife for MT estimation. In essence, this favours confidence limit conservatism at a possible and difficult to quantify slight sacrifice in reliability.

ACKNOWLEDGMENTS

This work was supported at WHOI by the Office of Basic Energy Sciences, US Department of Energy and by National Science Foundation grant EAR-0087699. This is Woods Hole Oceanographic Institution contribution 10990.

REFERENCES

Andersen, C.W., Lanzerotti, L.J. & Maclennan, C.G., 1976. Local time variation of induction vectors as indicators of internal and external current systems, *Geophys. Res. Lett.,* **3,** 495–498.

Belsley, D.A., Kuh, E. & Welsch, R.E., 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity,* John Wiley, New York.

Carroll, R.J. & Welsh, A.H., 1988. A note on asymmetry and robustness in linear regression, *Am. Stat.,* **42,** 285–287.

Chalk, J.H.H., 1950. The optimum pulse-shape for pulse communication, *Proc. IEE,* **97,** 88–92.

Chatterjee, S. & Hadi, A.S., 1988. *Sensitivity Analysis in Linear Regression,* John Wiley, New York.

Chave, A.D. & Jones, A.G., 1997. Electric and magnetic field galvanic distortion decomposition of BC87 data, *J. Geomag. Geoelectr.,* **49,** 767–789.

Chave, A.D. & Smith, J.T., 1994. On electric and magnetic galvanic distortion tensor decompositions, *J. geophys. Res.,* **99,** 4669–4682.

Chave, A.D. & Thomson, D.J., 1989. Some comments on magnetotelluric response function estimation, *J. geophys. Res.,* **94,** 14 215–14 225.

Chave, A.D. & Thomson, D.J., 1992. Robust, controlled leverage estimation of magnetotelluric response functions, *Proc. 11th Workshop on EM Induction in the Earth,* Wellington, New Zealand, paper 8.13.

Chave, A.D. & Thomson, D.J., 2003. A bounded influence regression estimator based on the statistics of the hat matrix, *J. R. Stat. Soc., Ser. C,* **52,** 307–322.

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

Coakley, C.W. & Hettmansperger, T.P., 1993. A bounded influence, high breakdown, efficient regression estimator, *J. Am. Stat. Assoc.,* **88,** 872–880.

David, H.A., 1981. Order Statistics, 2nd edn, p. 360, John Wiley, New York.

Dmitriev, V.I. & Berdichevsky, M.N., 1979. The fundamental model of magnetotelluric sounding, *Proc. IEEE,* **67,** 1033–1044.

Durbin, J., 1960. The fitting of time series models, *Rev. Int. Stat. Inst.,* **28,** 233–244.

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

Egbert, G.D., 1992. Noncausality of the discrete-time magnetotelluric impulse response, *Geophysics,* **57,** 1354–1358.

Egbert, G.D., 1997. Robust multiple station MT data processing, *Geophys. J. Int.,* **130,** 475–496.

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

Egbert, G.D. & Livelybrooks, D., 1996. Single station magnetotelluric impedance estimation: coherence weighting and the regression M-estimate, *Geophysics,* **61,** 964–970.

Egbert, G.D., Eisel, M., Boyd, O.S. & Morrison, H.F., 2000. DC trains and PC3s: source effects in mid-latitude geomagnetic transfer functions, *Geophys. Res. Lett.,* **124,** 25–28.

Eisel, M. & Egbert, G.D., 2001. On the stability of magnetotelluric transfer function estimates and the reliability of their variances, *Geophys. J. Int.,* **144,** 65–82.

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

Garcia, X. & Jones, A.G., 2002. Atmospheric sources for audio-magnetotelluric (AMT) sounding, *Geophysics,* **67,** 448–458.

Garcia, X., Chave, A.D. & Jones, A.G., 1997. Robust processing of magnetotelluric data from the auroral zone, *J. Geomag. Geoelectr.,* **49,** 1451–1468.

Geary, R.C., 1949. Determination of linear relationships between systematic parts of variables with errors of observation the variances of which are unknown, *Econometrica,* **17,** 30–58.

Ghosh, K., 1990. Robust multivariate regression analysis of complex-valued data, *PhD dissertation*, Department of Statistics, Temple University, Philadelphia, PA, USA.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. & Stahel, W.A., 1986. *Robust Statistics,* John Wiley, New York.

Handschin, E., Schweppe, F.C., Kohlas, J. & Fiechter, A., 1975. Bad data analysis for power system state analysis, *IEEE Trans. Power Appar. Syst.,* **PAS-94,** 329–337.

Harris, F.J., 1978. On the use of windows for harmonic analysis with the discrete Fourier transform, *Proc. IEEE,* **66,** 51–83.

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

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

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

Johnson, N.L., Kotz, S. & Balakrishnan, N., 1994. *Continuous Univariate Distributions,* Vol. 1, John Wiley, New York.

Jones, A.G., 1993. The BC87 data set: data and tectonic setting, *J. Geomag. Geoelectr.,* **45,** 1089–1105.

Jones, A.G. & Spratt, J., 2002. A simple method for deriving the uniform field MT responses in auroral zones, *Earth Planets Space,* **54,** 443–450.

Jones, A.G., Kurtz, R.D., Oldenburg, D.W., Boerner, D.E. & Ellis, R., 1988. Magnetotelluric observations along the LITHOPROBE southeastern Canadian Cordilleran transect, *Geophys. Res. Lett.,* **15,** 677–680.

Jones, A.G., Chave, A.D., Egbert, G.D., Auld, D. & Bahr, K., 1989. A comparison of techniques for magnetotelluric response function estimation, *J. geophys. Res.,* **94,** 14 201–14 214.

Junge, A., 1996. Characterization and correction for cultural noise, *Surv. Geophys.,* **17,** 361–391.

Koopmans, L.H., 1995. *The Spectral Analysis of Time Series,* Academic Press, San Diego.

Krasker, W.S. & Welsch, R.E., 1982. Efficient bounded-influence regression estimation, *J. Am. Stat. Assoc.,* **77,** 595–604.

Landau, H.J. & Pollak, H.O., 1961. Prolate spheroidal wave functions, Fourier analysis and uncertainty, II, *Bell Syst. Tech. J.,* **40,** 65–84.

Landau, H.J. & Pollak, H.O., 1962. Prolate spheroidal wave functions, Fourier analysis and uncertainty, III, *Bell Syst. Tech. J.,* **41,** 1295–1336.

Lanzerotti, L.J., Medford, L.V., Maclennan, C.G., Hasegawa, T., Acuna, M.H. & Dolce, S.R., 1981. Polarization characteristics of hydromagnetic waves at low geomagnetic latitudes, *J. geophys. Res.,* **86,** 5500–5506.

Lanzerotti, L.J., Gold, R.E., Thomson, D.J., Decker, R.E., Maclennan, C.G. & Krimigis, S.M., 1991. Statistical properties of shock-accelerated ions in the outer heliosphere, *Astrophys. J.,* **380,** L93–L96.

Larsen, J.C., 1989. Transfer functions: smooth robust estimates by least squares and remote reference methods, *Geophys. J. Int.,* **99,** 655–663.

Larsen, J.C., Mackie, R.L., Manzella, A., Fiordelisi, A. & Rieven, S., 1996. Robust smooth magnetotelluric transfer functions, *Geophys. J. Int.,* **124,** 801–819.

Levinson, N., 1947. The Wiener RMS (root-mean-square) error criterion in filter design and prediction, *J. Math. Phys.,* **25,** 261–278.

Mallows, C.L., 1975. On some topics in robustness, *Technical Memorandum,* Bell Telephone Laboratories, Inc., New Jersey.

Mardia, K.V., Kent, J.T. & Bibby, J.M., 1979. *Multivariate Analysis,* Academic Press, New York.

Oettinger, G., Haak, V. & Larsen, J.C., 2001. Noise reduction in magnetotelluric time-series with a new signal-noise separation method and its application to a field experiment in the Saxonian Granulite Massif, *Geophys. J. Int.,* **146,** 659–669.

Ogawa, Y. et al, 1996. Deep electrical conductivity structures of the Appalachian Orogen in the southeastern US, *Geophys. Res. Lett.,* **23,** 1597–1600.

Percival, D. & Walden, A., 1993. *Spectral Analysis for Physical Applications,* Cambridge University Press, Cambridge.

Percival, D. & Walden, A., 2000. *Wavelet Methods for Time Series Analysis,* Cambridge University Press, Cambridge.

Rao, C.R., 1973. *Linear Statistical Inference and its Applications,* John Wiley, New York.

Reiersol, O., 1941. Confluence analysis by means of lag moments and other methods of confluence analysis, *Econometrica,* **9,** 1–23.

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

Rousseeuw, P.J., 1984. Least median of squares regression, *J. Am. Stat. Assoc.,* **79,** 871–880.

Schultz, A., Kurtz, R.D., Chave, A.D. & Jones, A.G., 1993. Conductivity discontinuities in the upper mantle beneath a stable craton, *Geophys. Res. Lett.,* **20,** 2941–2944.

Shaffer, J.P., 1991. The Gauss-Markov theorem and random regressors, *Am. Stat.,* **45,** 269–273.

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

Slepian, D., 1978. Prolate spheroidal wave functions, Fourier analysis, and uncertainty-V: the discrete case, *Bell. Syst. Tech. J.,* **57,** 1371–1430.

Slepian, D. & Pollak, H.O., 1961. Prolate spheroidal wave functions, Fourier analysis and uncertainty, I, *Bell. Syst. Tech. J.,* **40,** 43–64.

Smirnov, M.Yu., 2003. Magnetotelluric data processing with a robust statistical procedure having a high breakdown point., *Geophys. J. Int.,* **152,** 1–7.

Stuart, A. & Ord, J.K., 1994. *Kendall's Advanced Theory of Statistics,* Vol. 1: *Distribution Theory*, Edward Arnold, London.

Stuart, A., Ord, J.K. & Arnold, S., 1999. *Kendall's Advanced Theory of Statistics, Vol 2A: Classical Inference and the Linear Model,* Edward Arnold, London.

Sutarno, D. & Vozoff, K., 1989. Robust M-estimation of the magnetotelluric impedance tensor., *Expl. Geophys.,* **20,** 383–398.

Sutarno, D. & Vozoff, K., 1991. Phase-smoothed robust M-estimation of magnetotelluric impedance functions, *Geophysics,* **56,** 1999–2007.

Szarka, L., 1988. Geophysical aspects of man-made electromagnetic noise in the earth-a review, *Surv. Geophys.,* **9,** 287–318.

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

Thomson, D.J., 1982. Spectrum estimation and harmonic analysis, *Proc. IEEE,* **70,** 1055–1096.

Thomson, D.J., 1990. Quadratic-inverse spectrum estimates: applications to paleoclimatology, *Phil. Trans. R. Soc. Lond.* A., **332,** 539–597.

Thomson, D.J. & Chave, A.D., 1991. Jackknifed error estimates for spectra, coherences, and transfer functions, in, *Advances in Spectrum Analysis and Array Processing,* Vol. 1, pp. 58–113, ed. Haykin, S., Prentice Hall, Englewood Cliffs, NJ.

Wald, A., 1940. The fitting of straight lines if both variables are subject to error, *Ann. Math. Stat.,* **11,** 284–300.

Walker, G., 1931. On periodicity in series of related terms, *Proc. R. Soc.,* **A131,** 518–532.

Wannamaker, P.E. *et al.*, 1996. Magnetotelluric experiment probes deep physical state of southeastern US, *EOS, Trans. Am. geophys. Un.,* **77,** 329, 332–333.

Welch, P.D., 1967. The use of the fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, *IEEE Trans. Audio Electroacoustics,* **15,** 70–73.

Wilcox, R.R., 1997. *Introduction to Robust Estimation and Hypothesis Testing,* Academic Press, New York.

Yule, G.U., 1927. On a method of investigating periodicities in disturbed series, with special reference to Wölfer's sunspot numbers, *Phil. Trans. R. Soc. Lond.* A. **226,** 267–298.