

Robust estimation of geomagnetic transfer functions

Gary D. Egbert and John R. Booker *Geophysics Program AK-50,
University of Washington, Seattle, WA 98195, USA*

Accepted 1986 March 17. Received 1986 March 12; in original form 1985 September 27

Summary. We show, through an examination of residuals, that all of the statistical assumptions usually used in estimating transfer functions for geomagnetic induction data fail at periods from 5 min to several hours at geomagnetic mid-latitudes. This failure can be traced to the finite spatial scale of many sources. In the past, workers have tried to deal with this problem by hand selecting data segments thought to be free of source effects. We propose an automatic robust analysis scheme which accounts for the systematic increase of errors with increasing power and which automatically downweights source contaminated outliers. We demonstrate that, in contrast to ordinary least squares, this automatic procedure consistently yields reliable transfer function estimates with realistic errors.

Key words: geomagnetic induction; robust transfer function estimation

1 Introduction

The first step in the interpretation of geomagnetic induction data usually involves the estimation of a frequency-dependent relationship between measured field components. In the case of geomagnetic depth sounding (GDS), a relationship between the vertical field (Z) and the horizontal field components [$\mathbf{H} = (H, D)$] of the form

$$Z(\omega) = \mathbf{T}^T(\omega)\mathbf{H}(\omega) \quad (1)$$

is sought. Here $\mathbf{T}^T(\omega)$ is the transpose of a complex two-vector called the transfer function. In the case of the magnetotelluric (MT) method a similar relationship between the horizontal electric (\mathbf{E}) and magnetic (\mathbf{H}) fields

$$\mathbf{E}(\omega) = \mathbf{Z}(\omega) \mathbf{H}(\omega) \quad (2)$$

is estimated, where $\mathbf{Z}(\omega)$ is the impedance tensor.

The use of (1) and (2) can be simply justified if the external sources of the fields are plane waves of infinite horizontal extent. It is easy to show, using the linearity of Maxwell's equations, that with such source fields (1) and (2) are exact in the absence of measurement errors. In practice (1) and (2) do not hold exactly, both because there are measurement

errors and because the plane wave source field assumption is, at best, only approximately true. It is thus necessary to estimate the transfer function or impedance tensor from less than perfect data, and the problem becomes a statistical one.

The estimation of these parameters has been discussed extensively in the geophysical literature and many variants have been proposed (e.g. Sims, Bostick & Smith 1971; Banks 1975; Jupp 1978; Beamish 1979). In general the estimation procedures are based on least squares (LS) methods; the parameter estimate is chosen so that the sum of squares of some misfit to the data is minimized. For instance, in the case of GDS the usual estimate is obtained by choosing $\hat{\mathbf{T}}(\omega)$ to minimize the error sum of squares:

$$\sum_i |Z_i(\omega) - \hat{\mathbf{T}}^T \mathbf{H}_i(\omega)|^2. \quad (3)$$

LS procedures are conceptually and computationally simple, and with the assumption of independent, identically distributed Gaussian errors, they are, in a precise sense, statistically optimal (cf. Graybill 1976, p. 173 ff). Unfortunately, at least for long-period GDS data, such assumptions about the error structure are not tenable.

The assumption that error variances are independent of signal power seems *a priori* unlikely. In part, the misfit of the data to the linear model is due to a failure of the uniform source field assumption. The magnitude of this sort of misfit seems very likely to depend on the strength of the signal, since for a given source field spatial structure, the earth response (including the part which fails to satisfy 1) will be proportional to the source magnitude. In addition, the strongest signals at mid-latitudes typically occur during the early phase of magnetic storms when the uniform source field assumption is particularly likely to be violated.

For related reasons, the Gaussian error assumption is questionable. The failure of model assumptions makes very large deviations from predicted values possible and it is thus reasonable to expect significant outliers. Such a situation is poorly modelled by a Gaussian error distribution. Finally, source related errors often occur in clusters so that the uncorrelated error assumption may fail, e.g. the Fourier transform of a data segment recorded during the initial phase of a storm may have similar source errors in a band of frequencies.

When some or all of these three assumptions fail, LS estimates can be seriously misleading; neither they, nor the associated standard error estimates are robust. As a result, consistent estimation of geomagnetic transfer functions has typically required careful preliminary screening of data to eliminate outliers. In this paper we describe a rigorously justifiable robust estimation scheme which deals with these problems in an automatic fashion. We will focus in particular on the violation of the first two assumptions, demonstrating, through an examination of residuals, that at geomagnetic mid-latitudes the error variance increases systematically with power and that substantial departures from Gaussian error structure occur.

Failure of the equal variance assumption can be corrected by using a weighted LS estimate. We briefly review this approach, along with some standard LS theory in Section 2. Non-Gaussian errors can be dealt with by using a robust alternative to LS. In Section 3 we describe such an estimate, a so called 'regression-M estimate' (Huber 1981) based on minimizing a misfit criterion which does not allow a few bad points to dominate the estimates. This can be accomplished efficiently with an iteratively reweighted LS algorithm which downweights 'bad' points. The procedure is automatic and an asymptotic theory exists which gives approximate errors for the estimates. Finally in Section 4 we compare the performance of our weighted robust estimates with the standard LS estimates.

We will see that a robust weighted LS algorithm can lead to substantial improvement in transfer function estimates. In addition we will show that, while our error estimates may still be slightly optimistic (due to the possible failure of the uncorrelated error assumption), they are much more realistic than those obtained with standard LS theory. Finally, we discuss some implications of our results for the choice of time window length for Fourier transforming geomagnetic time series. We suggest that the weighted robust estimates (and the associated error estimates) can be improved by use of shorter time segments.

While we only consider explicitly the robust estimation of the GDS transfer function at geomagnetic mid-latitudes, the ideas discussed in this paper should be useful in a much more general setting, in particular in the analysis of MT data. Note, however, that in this paper we ignore the effect of errors in the horizontal magnetic fields. It is well known that this can lead to biases in the resulting transfer function estimates. This effect is particularly important in MT analysis where several possible solutions to this problem have been suggested [remote reference, (Gamble, Goubau & Clark 1979) or the SVD estimate, (Jupp 1978; Park & Chave 1984)]. Since these approaches are again based on LS, our methods can, with suitable modifications, be extended to improve these more complicated cases.

2 Weighted least squares

In this section we review some basic facts about the statistical theory of least squares, both for the weighted and unweighted case. We present an analysis of the residuals of a least squares fit of GDS transfer functions and show that, at least for long periods, the use of weighted least squares is essential.

The model for the data which leads to the standard least squares estimate for the GDS transfer function assumes that the vertical field observations at a fixed frequency ω (obtained by Fourier transforming windowed segments of the recorded time series) can be written as a linear function of the horizontal fields plus an error. For the i th of N observations we have:

$$Z_i(\omega) = \mathbf{H}_i(\omega)^T \mathbf{T}(\omega) + e_i. \quad (4)$$

All quantities in (4) are complex, including the errors. Although it is easy to derive complex analogues of real LS procedures, for the purposes of this paper it will be simpler to recast the problem so that all quantities are real. Defining the frequency dependent matrices \mathbf{Z} , \mathbf{H} , \mathbf{e} and \mathbf{T}' by

$$\mathbf{Z} = \begin{pmatrix} \text{Re } Z_1 \\ \text{Im } Z_1 \\ \vdots \\ \text{Re } Z_N \\ \text{Im } Z_N \end{pmatrix} \quad \mathbf{H} = \begin{pmatrix} \text{Re } H_{11} & -\text{Im } H_{11} & \text{Re } H_{12} & -\text{Im } H_{12} \\ \text{Im } H_{11} & \text{Re } H_{11} & \text{Im } H_{12} & \text{Re } H_{12} \\ \vdots & \vdots & \vdots & \vdots \\ \text{Re } H_{N1} & -\text{Im } H_{N1} & \text{Re } H_{N2} & -\text{Im } H_{N2} \\ \text{Im } H_{N1} & \text{Re } H_{N1} & \text{Im } H_{N2} & \text{Re } H_{N2} \end{pmatrix}$$

$$\mathbf{e} = \begin{pmatrix} \text{Re } e_1 \\ \text{Im } e_1 \\ \vdots \\ \text{Re } e_N \\ \text{Im } e_N \end{pmatrix} \quad \mathbf{T}' = \begin{pmatrix} \text{Re } T_1 \\ \text{Im } T_1 \\ \text{Re } T_2 \\ \text{Im } T_2 \end{pmatrix} \quad (5)$$

it is easy to check that (4) (for $i = 1, \dots, N$) is equivalent to the real matrix equation

$$\mathbf{Z} = \mathbf{H}\mathbf{T}' + \mathbf{e}. \quad (6)$$

For the remainder of this paper, all quantities will be real. We thus redefine our notation slightly. We will subsequently refer to the i th element of the real vector \mathbf{Z} (\mathbf{e}) defined in (5) as Z_i (e_i) and the vectors corresponding to the i th row of \mathbf{H} as H_i ; we also drop the prime on \mathbf{T}' of (5), so that \mathbf{T} now represents a real vector of dimension 4 (rather than a complex vector of dimension 2).

The problem of estimating \mathbf{T} from the data (assuming (6) holds) is a standard problem in statistics — multiple regression (e.g. Draper & Smith 1981). The standard regression estimate of \mathbf{T} can be written, in matrix notation:

$$\hat{\mathbf{T}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Z}. \quad (7)$$

This estimate is a least squares estimate since it minimizes the sum of squared residuals

$$\sum_{i=1}^{2N} r_i^2 = \sum_{i=1}^{2N} (Z_i - \mathbf{H}_i^T \hat{\mathbf{T}})^2. \quad (8)$$

If we add more assumptions, more can be said. If the errors are uncorrelated and have equal variances, i.e. if the covariance matrix of the error vector \mathbf{e} satisfies

$$E(\mathbf{e}\mathbf{e}^T) = \Sigma = \sigma^2 \mathbf{I}$$

then $\hat{\mathbf{T}}$ is the best (i.e. minimum variance) estimate of \mathbf{T} among all estimates which are unbiased and are linear functions of the data. This is a special case of the Gauss–Markov theorem (cf. Graybill 1976, p. 219) which gives the best linear unbiased estimate:

$$\hat{\mathbf{T}} = (\mathbf{H}^T \Sigma^{-1} \mathbf{H})^{-1} \mathbf{H}^T \Sigma^{-1} \mathbf{Z} \quad (9)$$

for a general error covariance matrix Σ .

For the case of uncorrelated errors with unequal variances, $\text{Var}(e_i) = \sigma_i^2$, the estimate (9) is a weighted least squares (WLS) estimate in the sense that $\hat{\mathbf{T}}$ then minimizes a weighted sum of squared residuals

$$\sum_{i=1}^{2N} w_i r_i^2 = \sum_{i=1}^{2N} w_i (Z_i - \mathbf{H}_i^T \hat{\mathbf{T}})^2, \quad (10)$$

where the weights w_i are proportional to $1/\sigma_i^2$.

The validity of assumptions about errors can be assessed by examining the residuals more carefully (cf. Goodall 1983). In particular, by plotting r_i^2 against $\|\mathbf{H}_i\|^2$, it is easy to estimate the magnitude of the error variance as a function of signal power. In Fig. 1 we present a smoothed version of such a plot. This plot summarizes results from 12 GDS stations located in western Washington and Oregon. Preliminary data processing, including windowing and Fourier transforming the three component time series, were carried out as described in detail in Section 4. Transfer functions were then estimated by LS and average squared residual magnitudes were computed for all Fourier coefficients in a specified signal power range. These are plotted against signal power for a range of frequencies on a log-log scale.

For shorter period data (< 300 s) the error variance is independent of signal power and ordinary LS would be justified. For longer periods, however, error magnitudes tend to increase with increasing signal power. At an hour, the noise power is essentially proportional to signal power and the assumptions of standard LS are grossly violated. This strong systematic variation of error variance demands use of the WLS estimate of (9) and (10).

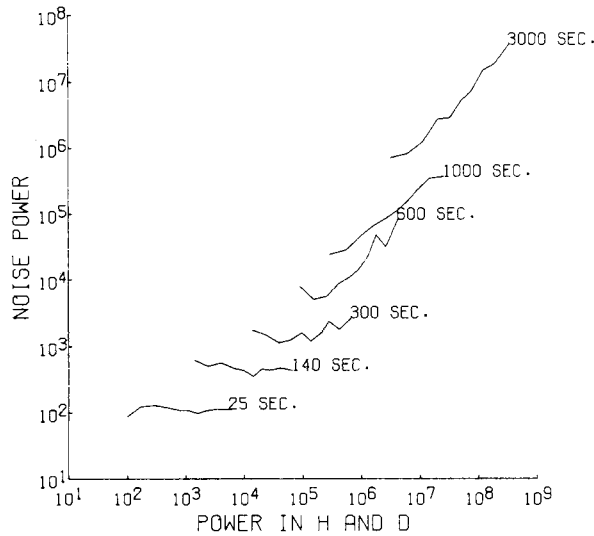


Figure 1. Noise power versus signal power ($\text{nT}^2 \text{Hz}^{-1}$). Results are averaged over 12 sites in western Washington and Oregon. Note the increasing dependence of noise power on signal power at longer periods where source effects are expected to be greatest.

The easiest way to compute a weighted estimate is to rescale the field components by dividing by the error scales setting

$$Z'_i = Z_i / \sigma_i$$

$$\mathbf{H}'_i = \mathbf{H}_i / \sigma_i$$

$$e'_i = e_i / \sigma_i$$

so that $Z'_i = \mathbf{H}'_i{}^T \mathbf{T} + e'_i$, where $\text{var}(e'_i) \equiv 1$, so that now the standard LS estimate is applicable. In order to compute the error scales, we have approximated the dependence of error variance on signal power with a linear function

$$\sigma^2(\|\mathbf{H}\|^2) = a + b\|\mathbf{H}\|^2 \quad (11)$$

where a and b depend on frequency, and have been estimated from the empirical signal-noise power curves plotted in Fig. 1.

Note that while the use of WLS can be justified by the improvement of estimation efficiency guaranteed by the Gauss-Markov theorem, there are other, perhaps more compelling justifications for its use. First, unweighted LS gives too much influence to high power events to be safe. In extreme situations the transfer function can be determined almost entirely by one or two events — events which are the most likely to be outliers. Secondly, the validity of the standard computation of estimation errors in LS is critically dependent on the equal variance assumption. It is not hard to show that with the systematic trend in error magnitudes seen in long period GDS, the error estimates obtained from unweighted LS are, on average, too small.

The results of Fig. 1 have further implications. In the estimation of geomagnetic transfer functions, it is common practice to use only data from time periods when the magnetic field is active. This practice implicitly assumes that higher signal power implies a higher signal-to-noise ratio. In fact, Fig. 1 shows that at periods of thousands of seconds at mid-latitudes, the signal-to-noise ratio is independent of signal power. For such periods, then, background

variations are at least as good for transfer function estimation as variations from disturbed times. The practice of selecting only disturbed times for analysis can be justified only for shorter periods where source Z -fields are smaller and where instrument noise is large compared to background signal levels.

3 A robust alternative to least squares

In this section we will assume that the three (Fourier transformed) component fields have been rescaled so that error variances are constant. While the Gauss–Markov theorem now guarantees that the LS estimates will be the best among all linear estimates, there is no guarantee that any linear estimate will perform reasonably. The LS estimate is, in fact, optimal among all estimates if the errors are Gaussian, but this property is exceedingly sensitive to the actual error distribution. For non-Gaussian errors, other (non-linear) estimates may have smaller variances. Worse, the LS estimate is not robust – a very small number of bad data points can lead to a catastrophic failure of the estimate.

The Gaussian error assumption can be tested by plotting the distribution of actual residuals observed against the distribution of residuals expected from a Gaussian error distribution. Such plots are called Q – Q plots or normal probability plots (cf. Goodall 1983; Wilk & Gnanadesikan 1968) and are constructed by ordering the signed, normalized (i.e. divided by the error scale) residuals and plotting the value of the i th of m ordered residuals $r_{(i)}$ against $\Phi^{-1}(i/m)$ (where Φ^{-1} is the inverse of the Gaussian distribution function). For example, with $m = 200$ the fifth residual would be expected to be $\Phi^{-1}(0.025) = -1.94$ (since 2.5 per cent of the observations from a Gaussian sample with unit variance are expected to be less than -1.94) so that this point would be plotted as $(-1.94, r_{(5)})$. If the residuals are consistent with a Gaussian error distribution, the Q – Q plot should approximate a straight line with unit slope.

In Fig. 2 we show examples of such Q – Q plots for GDS data from two stations in western Washington. In Fig. 2(a), plots for four frequencies from station SKA are given. These data included a severe storm that appeared in the magnetograms to be severely contaminated by source Z fields. The Q – Q plots show the effect of this storm on the frequency domain residuals. At short periods (60 s) the Q – Q plot is a straight line indicating that the errors are Gaussian. For longer periods, however, the Q – Q plots show that even after weighting, the error distribution becomes increasingly heavy tailed – the largest residuals are much larger than expected. We have verified that these largest residuals are mostly from the highly disturbed period. We thus see that violations of the uniform source assumption can lead to unusually large residuals, invalidating the Gaussian error assumption.

In Fig. 2(b) we show plots from station KIN. This data sequence did not include any severe storms, but was noted to be rather noisy. Here the long-period errors seem reasonably Gaussian, but the Q – Q plots show that around 300 s the error distribution is very heavy tailed. In this case the contamination is almost certainly not due to geomagnetic sources, since the problem is greatest at relatively short periods. The heavy tailed error distribution at this period is most likely due to sporadic cultural noise. Whatever the cause, the heavy tails seen in these plots demonstrate that the assumption of Gaussian errors can fail at any period.

In the past decades, substantial progress has been made in understanding the sensitivity of standard statistical procedures (such as LS estimation) to distributional assumptions and numerous robust alternatives have been proposed. Huber (1981) gives a thorough but somewhat technical treatment of the subject; Mosteller & Tukey (1977) offer a more accessible treatment; and Claerbout & Muir (1973) discuss some of the issues in a geophysical context. We discuss here a well understood, easily implemented robust alternative to the LS estimate

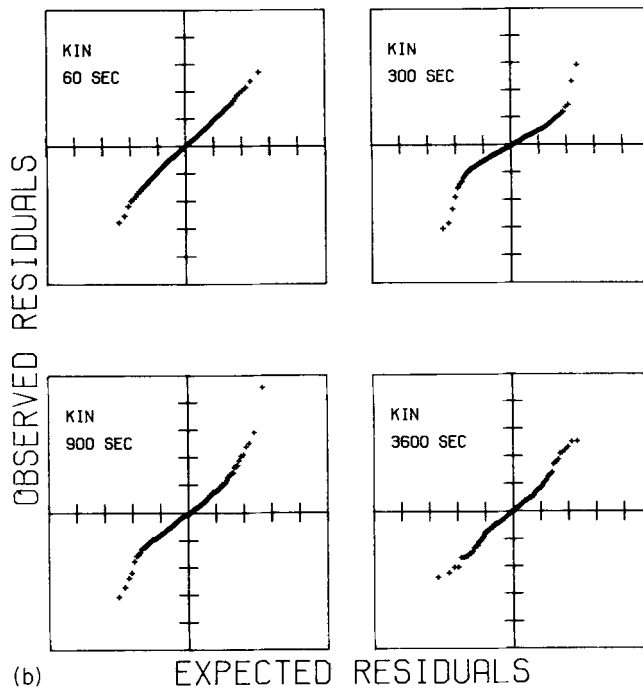
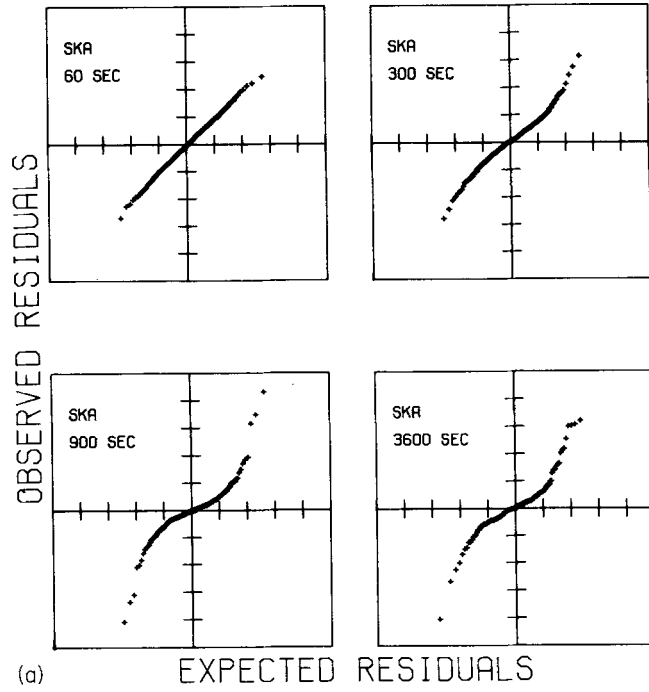


Figure 2. $Q-Q$ plots for (a) SKA – heavy tailed residual distributions for long periods due to source effects. (b) KIN – heavy tails at intermediate periods due to sporadic local noise.

of the GDS transfer function, a so-called 'regression M-estimate'. In the remainder of this section we describe this estimate briefly; its practical implementation is described in Appendix A and in the next section we give an example of its use.

The regression M-estimate is analogous to a LS estimate in that it minimizes the difference between prediction and observations, but the measure of misfit is now defined in a way that does not allow a few bad points to dominate the estimate. This general approach is familiar to geophysicists – L_1 minimization (cf. Claerbout 1976; Menke 1984) is a special case. We consider a generalization of this idea which is flexible, computationally efficient, and easily implemented. Specifically, we seek estimates of the transfer function $\hat{\mathbf{T}}$ which minimize an expression of the form

$$\sum_i \rho \left(\frac{Z_i - \mathbf{H}_i^T \hat{\mathbf{T}}}{\sigma} \right) \quad (12)$$

where $\rho(r)$ is some suitable loss function. Standard LS uses

$$\rho(r) = r^2/2$$

and L_1 minimization $\rho(r) = |r|$. For most purposes we have used a fairly standard hybrid (Huber 1981)

$$\rho(r) = \begin{cases} r^2/2 & |r| < r_0 \\ r_0 |r| - r_0^2/2 & |r| \geq r_0 \end{cases} \quad (13)$$

with $r_0 = 1.5$.

Note that (12) depends in general on the scale parameter σ used for normalizing the residuals. The loss function given by (13) corresponds to L_2 minimization for small residuals and to L_1 minimization for larger residuals. The actual transition point for the unscaled residuals is σr_0 , so the scale parameter (together with r_0) determines which residuals we consider to be large. For Gaussian errors, possibly contaminated by a small number of outliers, a transition point of $r_0 = 1.5$ times the standard deviation of the (uncontaminated) Gaussian error distribution works well. We thus use a robust estimate of the residual standard deviation for the error scale. Details are discussed in Appendix A.

Note that the minimum of (12) can be found by solving the system of equations

$$\sum_i \psi \left(\frac{Z_i - \mathbf{H}_i^T \hat{\mathbf{T}}}{\sigma} \right) \mathbf{H}_i = 0 \quad (14)$$

where $\psi(r) = \rho'(r)$. In practice the solution of (14) can be easily computed with an iterative LS algorithm which downweights outliers. Specifically, let

$$w(r) = \psi(r)/r.$$

Starting with the LS estimates of the transfer function ($\hat{\mathbf{T}}_0$) and error scale $\hat{\sigma}_0$, compute the predicted and residual Z fields:

$$\begin{aligned} \hat{Z}_{i0} &= \mathbf{H}_i^T \hat{\mathbf{T}}_0 \\ r_{i0} &= Z_i - \hat{Z}_{i0} \end{aligned} \quad (15)$$

and then the 'modified observation':

$$Z_{i1} \approx \hat{Z}_{i0} + w(r_{i0}/\hat{\sigma}_0) r_{i0}. \quad (16)$$

Note that for ρ given by (13)

$$w(r) = \begin{cases} 1 & |r| \leq r_0 \\ r_0/|r| & |r| \geq r_0 \end{cases}$$

so that the modified observations are identical to the original observations for small ($\leq r_0$ standard deviations) residuals, while for larger residuals the observations are pulled toward their predicted values. Now, using the modified observations (\mathbf{Z}_1) in place of the originals ($\mathbf{Z}_0 \equiv \mathbf{Z}$) compute new LS estimates of the transfer function

$$\hat{\mathbf{T}}_1 = (\mathbf{H}^T \mathbf{H})^{-1} (\mathbf{H}^T \mathbf{Z}_1) \quad (17)$$

together with a new error scale estimate $\hat{\sigma}_1$.

An iterative application of the above procedure is used to compute the regression M-estimate. Using the estimates $\hat{\mathbf{T}}_n$ and $\hat{\sigma}_n$ from the n th iteration (in place of $\hat{\mathbf{T}}_0$ and $\hat{\sigma}_0$) new modified observations \mathbf{Z}_{n+1} and parameter estimates $\hat{\mathbf{T}}_{n+1}$ and $\hat{\sigma}_{n+1}$ are computed exactly as above.

This procedure is iterated until convergence to the solution of (14) is achieved. Provided $\rho(r)$ satisfies the conditions

$$\rho(0) = 0 \quad \rho'(r) \geq 0 \quad 0 \leq \rho'(r) < 1$$

such convergence is guaranteed (cf. Huber 1981, p. 179 ff). This algorithm can be shown to be similar to a Gaussian-Newton scheme for solving the non-linear system (14). Note also that this algorithm is a sort of weighted least squares. At the solution $\hat{\mathbf{T}}$ to (14) we have (using $r_i = Z_i - \hat{\mathbf{T}}^T \mathbf{H}_i$)

$$0 = \sum_i \frac{\psi(r_i)}{r_i} (Z_i - \hat{\mathbf{T}}^T \mathbf{H}_i) \mathbf{H}_i = \sum_i w(r_i) (Z_i - \hat{\mathbf{T}}^T \mathbf{H}_i) \mathbf{H}_i.$$

This is identical to the equations that would arise from the weighted least squares minimization of (10), with $w_i = w(r_i)$. Here the weights are determined from the data, with points which fit poorly given smaller weights.

With sufficient regularity conditions, the estimates obtained by minimizing (12) are consistent with

$$E(\hat{\mathbf{T}}) \rightarrow \mathbf{T}$$

and they are asymptotically Gaussian with an easily approximated covariance matrix; details are given in Appendix A.

The algorithm described above guarantees convergence to a unique estimate only if $\rho(r)$ is convex ($\rho'' > 0$). This requires that ψ be a non-decreasing function. Ideally we would like to completely discard data points which are very bad. This essentially requires that $\psi(r) \approx 0$ for large r (so that the terms corresponding to large residuals in (14) are completely omitted). This in turn requires that ψ be non-monotone. For such a ψ the non-uniqueness of the solution to (14) can potentially lead to very bad results if the starting estimate is poorly determined. Hence, it is advisable (Huber 1981) to use a non-monotone ψ only for a final iteration (or two) after the algorithm has converged with a monotone ψ . To eliminate the worst points completely, we do two final iterations with

$$\psi(r) = r \exp \{-\exp [r_0(|r| - r_0)]\}$$

using $r_0 = 2.8$.

Using a non-monotone ψ for the final iterations allows this robust scheme to completely

eliminate the worst outliers. This is a significant advantage relative to L_1 minimization. Note that for the initial monotone ψ estimate, L_1 minimization could be used (it is a special case of the general method described here), but the direct computation of the L_1 estimate requires the solution of a linear programming problem. The use of the ρ given in (13) is cheaper and simpler to implement. For the large errors against which robust schemes offer protection, the two methods are similar.

4 Results

In this section we compare estimates of GDS transfer functions obtained with standard LS, WLS and the robust algorithms described above. Ten days of data, recorded at several stations in western Washington in 1984 March and April, were used for the comparison. Three components of the magnetic field were recorded using EDA fluxgate magnetometers with an 8 s sampling interval. Initial processing of the data included detection and correction of obvious isolated outliers in the time series (such as parity errors), and visual inspection of plots of the time series to check for equipment malfunction or excessive cultural noise.

Our windowing of the data used a modified cascade decimation algorithm genetically related to that used in real-time MT (Wight *et al.* 1977) which allows longer time windows for longer period data. First, 256 point data segments with a half-hour time window are chosen from the raw 8 s data. In addition, the raw data are low pass filtered and decimated by a factor of 4 to produce a time series sampled at 32 s interval. Four such decimates are collected to produce a new 256 point segment with a 2 hr time window. The filtering and decimation steps are repeated on the decimated series, producing segments sampled at 128 s which are collected to produce 256 point segments with an 8 hr time window. This process can be continued for higher levels of decimation. In practice we have obtained useful estimates at periods from 25 to at least 5000, and some times 10 000 s from the first three levels and 10 day or less of data.

At all periods, the intensity of the horizontal magnetic signal varies significantly with time. At shorter (25–250 s) periods this signal often drops below the system noise (present in all channels). Since the use of data with a low signal-to-noise ratio in the horizontal field channels can lead to biased transfer function estimates, estimation of short-period transfer functions requires screening of the data to find time periods with sufficient high frequency signal. We accomplished this automatically with a two-step procedure.

First, the undecimated and level one sets were screened in the time domain. To do this *ad hoc* indices of total signal power, based on squared second differences of the time series, were computed for each set. Power in different frequency bands was estimated by varying the spacing of points entering the second differences. Estimates from a set of logarithmically spaced frequency bands computed in this fashion were then ‘whitened’ by scaling according to typical powers found in these bands: the scaled estimates were then squared again and summed to form a scalar index of signal power for the set. Only data sets whose power index exceeded an empirically determined threshold were processed further. At the same time, using this index of signal power, data segments containing the highest power events (which included the worst storm source fields) were flagged so that they could be omitted from transfer function calculations if desired. Selected data segments were pre-whitened by first differencing (to account for the approximately $1/f$ spectrum), multiplied by a pi-prolate spheroidal data window (Thompson 1982) and Fourier transformed. Successive time windows were overlapped 30 per cent. This overlap increases estimation efficiency without significantly effecting the independence of adjacent windows (Thompson 1982).

After Fourier transformation of the time series, individual Fourier coefficients with

magnitudes below a frequency-dependent threshold were omitted. This threshold was chosen to ensure that the signal-to-noise (power) ratio was at least 10. The results presented in Fig. 1 were used to determine the appropriate power levels. This frequency domain screening affected only periods shorter than 250 s; from 250 s to the longest period looked at (10 000 s), the signal power was always well above the system noise level.

For all decimation levels, transfer functions are computed for non-overlapping bands whose width is 25 per cent of the centre frequency. Note that windowing of the data in the time domain causes correlation of Fourier coefficients in the frequency domain (due to overlap of the corresponding frequency domain window). A proper treatment of estimation errors of the band averaged estimates must allow for this. To do this we used an approximate correction which is outlined in Appendix B.

In Fig. 3 we plot the magnitude of the Parkinson vectors (i.e. the negative of the real part of the transfer function) computed by the three methods for periods from 250 to 8000 s. Results are plotted for three stations (KIN, CLA and SKA) chosen to illustrate the sort of improvements in estimates we have seen.

For station KIN (Fig. 3a) the three methods yield similar, but not identical, estimates. The robust estimate is distinctly the smoothest, while the unweighted estimate is the roughest. Plotted on an expanded scale in Fig. 4, the differences are more striking. Note that while the Parkinson vector estimated by the robust method appears to be a smoothed version of the others, the estimates are computed for the same non-overlapping bands for all methods.

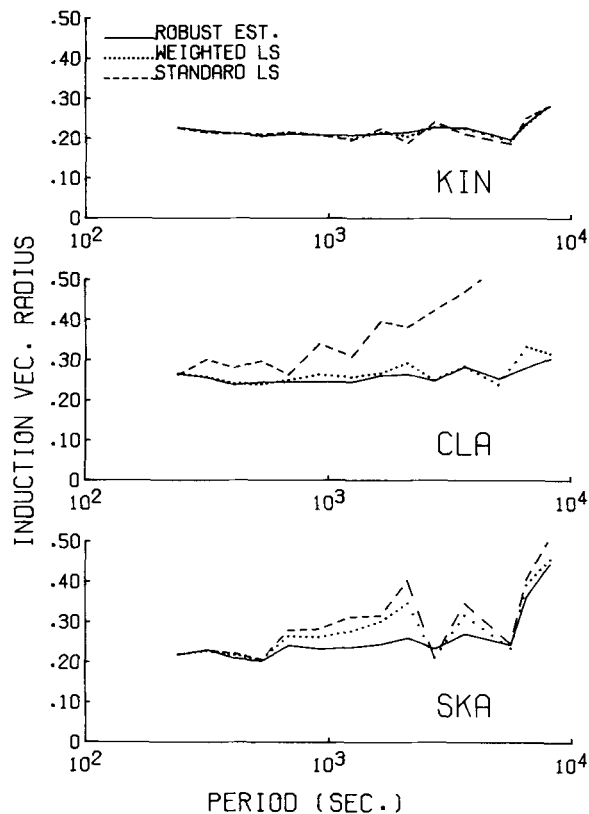


Figure 3. Magnitude of real part of induction vector, showing increased smoothness of weighted and robust estimates. (a) KIN. (b) CLA. (c) SKA.

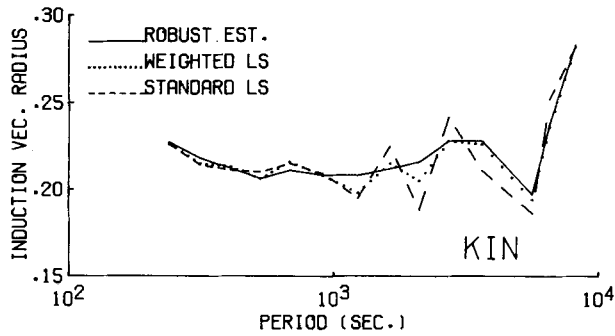


Figure 4. Magnitude of real part of induction vector for KIN on expanded scale.

We take this increase in smoothness as strong evidence of the superiority of the robust estimate. One expects that the transfer function should vary slowly with frequency (e.g. for the one-dimensional MT transfer function Wiedelt (1972) has provided rigorous bounds which severely constrain the roughness). The effect of noise can only make the estimates rougher. While smoothness does not always guarantee accuracy (systematic biases would not affect smoothness), excessive roughness always implies inaccuracy. In Fig. 3(a), where no systematic differences between curves are evident, the choice of the superior estimate is clear.

The differences between estimates at stations CLA and SKA are much greater (Fig. 3b, c). Data recorded for these stations included several large storms which severely violated the uniform source field assumption. The worst high power segments of data (identified in preliminary processing) were omitted from all transfer function estimates, but contamination of the unweighted estimates is still apparent at both stations. The weighted estimate is much smoother (and more like the robust estimate) at CLA, but is still very rough at SKA. Only the robust estimate, is smooth at both stations.

We should point out that the differences between estimates seen at SKA and CLA are not the norm. Typically, standard LS estimates work reasonably well and yield results which are similar to our weighted robust estimator. However, as the Parkinson vector estimates at CLA and SKA show, the failure of LS, when it does occur can be quite severe. Furthermore, as the results at KIN show, even when LS estimates are not obviously in error, improvement is still possible with a robust scheme.

In Fig. 5 we plot the computed estimation error for the Parkinson vectors of Fig. 3. Some of the general features of Fig. 3 are reflected in the error estimates. For instance, the poor performance (relative to the other two estimates) of the unweighted estimate at CLA shows up in the errors. At SKA, the improved performance of the robust estimate is reflected in the consistent reduction of errors beyond 1000 s.

At the same time, however, there is clear evidence that the error estimates are overly optimistic for the non-robust estimates. Most significantly, the oscillations in the non-robust estimates, as well as their deviation from the robust estimate, far exceeds the claimed two standard errors at CLA and SKA (Fig. 5). We believe that this reflects the correlated nature of errors due to severe source problems — numerous data points spread across a range of frequencies are contaminated in a similar fashion by a storm with short spatial scale source structure. While the correlation of errors may also be expected to cause underestimates in the errors for the robust algorithm, the effect is not as serious. Such source contaminated observations are precisely those likely to be thrown out by the robust estimate.

Note also that at KIN and SKA, where the weighted estimates are clearly superior, the

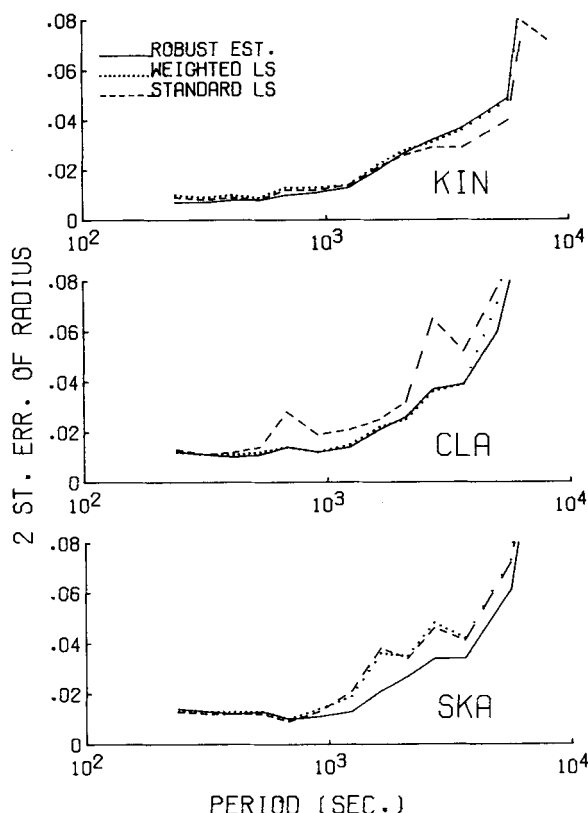


Figure 5. Two standard errors of induction vector magnitude; errors for standard LS estimates are clearly too small. (a) KIN. (b) CLA. (c) SKA.

estimation errors are almost always slightly smaller for the unweighted estimates. This reflects the downward bias in estimation error (for the unweighted estimates) caused by the systematic failure of the equal variance assumption.

The transfer function estimates presented above are based on data which were screened to delete time segments with the worst source structure. To test the robust algorithm further, we have also run station SKA without omitting these data. Parkinson vectors computed from this unscreened data by the LS, WLS, and robust algorithms are plotted in Fig. 6.

As discussed above, our method of data windowing uses a modified cascade decimation scheme. The frequency ranges for which estimates can be obtained for the two decimation levels overlap substantially, and for these frequencies a choice must be made of which estimate to use. For the results discussed above, differences between the two levels were small and the estimate with the smaller standard error was chosen. In Fig. 6 we plot estimates from both the first and second decimation levels of unscreened data, along with the robust estimate using screened data from Fig. 3(c) for reference.

The estimates computed from the two levels of decimation are comparable for the standard LS estimate, but are substantially different from the WLS and robust algorithms. In both of these cases the estimates computed from the lower decimation level (the shorter time window) are closer to the reference curve computed with the screened data. In fact, if estimates from level one are used at all possible periods, the robust algorithm produces essentially the same estimates with or without the preliminary removal of obvious outliers

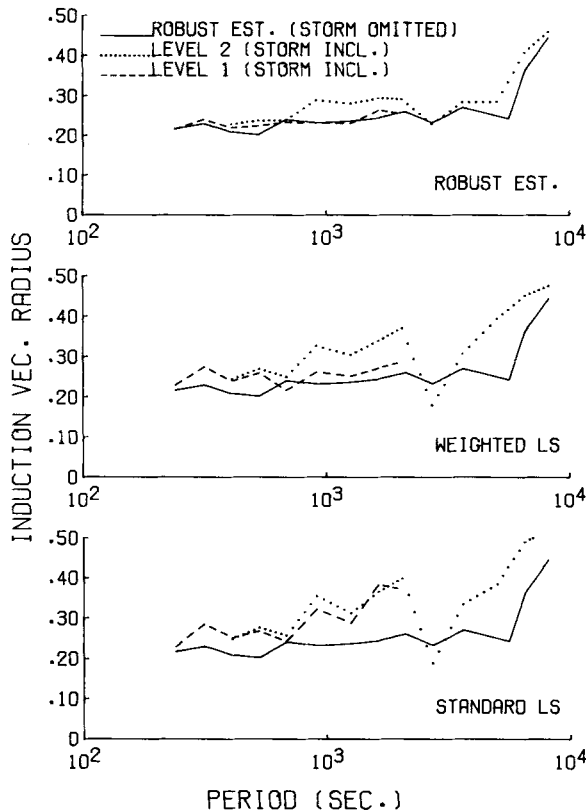


Figure 6. Magnitude of real part of induction vector at SKA with storm included. The robust and weighted estimates depend on the time window – estimates computed with the shorter window are affected less by inclusion of the storm data. (a) Robust estimate. (b) WLS. (c) LS.

due to strong storms. This result is very encouraging. With proper windowing of the time series the robust estimate is not affected by the inclusion of some very poor quality data.

The improved performance of the weighted and robust schemes with shorter time windows can be readily understood. The choice of the length of the data window involves a trade off between resolution in the time and frequency domains. Shorter time windows give sharp temporal resolution, but each Fourier coefficient (FC) represents an average over a broad frequency band; for long time windows, frequency resolution is sharp, but each FC is an average over a longer time period. With geomagnetic induction data, one is seeking to estimate a time stationary quantity which varies slowly in the frequency domain (the transfer function) from a non-stationary signal whose character may vary rapidly in time (i.e. with time-scales which are at most a few times the period of interest). This suggests that improved temporal resolution should be more important than improved frequency resolution, as we shall see.

Consider, for example, a short duration event of high power (e.g. a bay) in a much lower power background signal. If the time series is broken into short segments, the power from the event will be concentrated in a small number of FCs corresponding to the broad frequency bands for the single, short time segment containing the event. The corresponding FCs for the adjacent time segments will contain only the background signal. If a longer time

window is used, the power from the event will still be concentrated in the same frequency range which will now contain more FCs (each representing a narrower frequency band averaged over a longer time). The event's power will now, however, be averaged with the background signal. The net result is that the power from the event is spread over more of the FCs used for computation of the transfer functions when a longer time window is used. The event will thus be harder to pick out as a high powered or outlier event.

For the weighted and robust methods, then, window length can be an important consideration. Averaging high power events with the lower power background defeats the purpose of the weighting scheme. A few high power events can have their power spread over a large number of FCs so that a few short segments of data can still dominate the estimates. The same argument applies to the robust algorithm. An 'outlier event' with poor source characteristics can contaminate a number of adjacent frequencies. Since these are averaged with good data, none of these may be identifiable by themselves as outliers, and the event may slip by the robust algorithm. We thus suggest that the weighted and robust algorithms will work most effectively if the time window is kept as short as possible, consistent with the frequency resolution desired.

Note that for the standard LS estimates, the distinction between shorter and longer windows is unimportant — LS averages all the events together in essentially the same way that the longer window does. This can be seen in the similar estimates for the two levels of LS estimates in Fig. 6(c).

To test the consistency of our robust estimates, we have computed transfer functions from data collected at a single site in Oregon during two different time periods. The first (WIS064) had five and the second (WIS074) 10 days of data in 1982 August. This station was chosen because the data had both disturbed and quiet periods in each time segment and because we were confident that the sensing head had not changed orientation between segments. The two Parkinson vector estimates, together with error bars for the differences are plotted in Fig. 7. The two estimates are completely consistent. For one out of 23 frequencies, the difference between the two estimates slightly exceeds two standard errors. This is exactly what would be expected.

5 Discussion

In this paper we have considered the validity of the assumptions about errors which are inherent in the LS estimates of the GDS transfer function. Due to the nature of errors

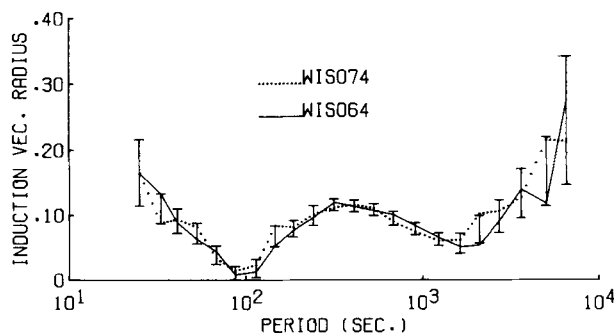


Figure 7. Comparison of induction vector estimates at WIS for two separate time periods. The vertical bars are two standard deviations of the difference between the two estimates. The two estimates are consistent within their errors.

caused by short spatial scale source fields, these assumptions often fail for long-period data, leading to poor transfer function estimates. To deal with this problem we have proposed an automatic estimation scheme which is robust to violations of the usual assumptions about errors. With several examples, we have demonstrated that this scheme, unlike standard LS, can produce reasonable transfer function estimates (with reasonable error estimates) even in the presence of poor quality data.

It could well be argued that the comparisons we have made between the standard LS and robust estimates are artificial, since any careful investigator will always screen the data to eliminate 'bad' time segments, and will then examine the resulting estimates critically. While it is certainly true that with proper care, LS estimates can work reasonably well, there are still compelling reasons to use a formal robust scheme of the sort described here.

First, outliers will not always be obvious in the time series. An informal procedure for recognizing outliers seems excessively susceptible to errors and unconscious biases. It thus seems essential to use some sort of preliminary estimate of the transfer function to facilitate recognition of bad data in a thorough and consistent manner. The method described here is a very simple, yet formally justifiable application of this idea. Secondly, with an informal procedure it is difficult to rigorously justify the estimation errors. How, for example, does one account for the effect of omitting data which does not seem to fit? This accounting is much more easily accomplished with a formal procedure. Finally, our robust procedure is automatic and is thus easier to use. With digitally recorded data and a small portable computer it is possible to get reliable estimates quickly in a field setting. This allows for any decisions about redeployment of instruments to be based on the data which have just been collected.

In this paper, we have also shown at geomagnetic mid-latitudes, that the GDS signal-to-noise ratio is independent of signal power at periods larger than 250 s. This implies that the continuum data between disturbances is at least as good for GDS work as disturbed data.

A question that we have not addressed, however, is whether our robust estimation scheme would work at geomagnetic latitudes much higher than Washington (55°). The success of our robust algorithm depends on the fact that short spatial scale sources are rare (although powerful) events. Thus they can be detected and discriminated against. We do not know whether this is also true in the auroral zone. It may turn out that at high latitudes, short spatial scales are the norm and uniform source fields are the outliers even during relatively quiet geomagnetic times. If such were the case our robust scheme would obviously fail.

Acknowledgments

We would like to acknowledge the many contributions of J. Torquil Smith to this paper – data collection, data processing and many discussions of results. We also acknowledge several useful discussions with Dr Alan Chave. This work was supported by NSF grants EAR-8307681 and EAR-8410638 and by USGS Grant 14-08-0001-20576.

References

- Banks, R. J., 1975. Complex demodulation of geomagnetic data and the estimation of transfer functions, *Geophys. J. R. astr. Soc.*, **43**, 87–101.
- Beamish, D., 1979. Source field effects on transfer functions at mid-latitudes, *Geophys. J. R. astr. Soc.*, **58**, 117–134.
- Claerbout, J. F., 1976. *Fundamentals of Geophysical Data Processing*, McGraw-Hill, New York.
- Claerbout, J. F. & Muir, F., 1973. Robust modelling with erratic data, *Geophysics*, **38**, 826–844.
- Draper, N. & Smith H. Jr., 1981. *Applied Regression Analysis*, 2nd ed, Wiley, Somerset, New Jersey.

- Gamble, T. D., Goubau, W. M. & Clarke, J., 1979. Magnetotellurics with a remote reference, *Geophysics*, **44**, 53–68.
- Goodall, C., 1983. Examining residuals, in *Understanding Robust and Exploratory Data Analysis*, eds, Hoaglin, D. C., Mosteller, F. & Tukey, J. W., Wiley, New York.
- Graybill, Franklin A., 1976. *Theory and Applications of the Linear Model*, Duxbury Press.
- Huber, Peter J., 1981. *Robust Statistics*, Wiley, New York.
- Jupp, D. L. B., 1978. Estimation of the magnetotelluric impedance functions, *Phys. Earth planet. Int.*, **17**, 75–82.
- Menke, William, 1984. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, Orlando.
- Mosteller, F. & Tukey, J. W., 1977. *Data Analysis and Regression: a Second Course in Statistics*, Addison-Wesley, Reading.
- Park, J. & Chave, A. D., 1984. On the estimation of magnetotelluric response functions using the singular value decomposition, *Geophys. J. R. astr. Soc.*, **77**, 683–709.
- Sims, W. E., Bostick, F. X. Jr & Smith, H. W., 1971. The estimation of magnetotelluric impedance tensor elements from measured data, *Geophysics*, **36**, 938–942.
- Thompson, D. J., 1982. Spectrum estimation and harmonic analysis, *Proc. IEEE*, **70**, 1055–1095.
- Wiedelt, P., 1972. The inverse problem of electromagnetic induction, *J. Geophys.*, **38**, 257–289.
- Wight, D. E., Bostick, F. X. Jr & Smith, H. W., 1977. Real Time Fourier Transformation of Magnetotelluric Data. *Final report, contract E(40-1) 5244. Energy Research and Development Administration*, Electrical Engineering Research Laboratory, University of Texas, Austin. 92 pp.
- Wilk, M. B. & Gnanadesikan, R., 1968. Probability plotting methods for the analysis of data, *Biometrika*, **55**, 1–17.

Appendix A

In this appendix we give more details about the regression M-estimate and its practical implementation.

A.1 CONCURRENT SCALE ESTIMATES

The computation of the regression M-estimate \hat{T} generally requires the concurrent estimation of the error scale σ . Assuming that the errors have a (possibly contaminated) Gaussian distribution the estimate proposed above will perform best if the scale is chosen as the standard deviation of the uncontaminated Gaussian distribution. The scale estimate should thus be an unbiased estimate of the standard deviation of a Gaussian distribution and, it should be robust. There are many possible estimates of the scale with these properties. We have used an estimate which can easily be incorporated into the iterative least squares scheme described above.

An initial scale estimate is computed, as in standard LS, from the rms residual

$$\hat{\sigma}_0 = \left[\frac{1}{(2N-4)} \sum_{i=1}^{2N} r_{i0}^2 \right]^{1/2}. \quad (\text{A1})$$

Scale estimates $\hat{\sigma}_n$ for subsequent iterations require a bit of care. On the one hand, using the actual rms of the residuals r_{in} from the computed estimate \hat{T}_n is not advisable since this estimate of σ may be severely affected by outliers. On the other hand, replacing

$$\sum_i r_{in}^2$$

by the sum of squared residuals from the modified observations

$$\sum_i (Z_{in} - \mathbf{H}_i^T \hat{T}_n)^2 = \sum_i (Z_{in} - \hat{Z}_{in})^2$$

underestimates σ because all larger residuals are decreased. A proper treatment must correct for this effect.

The rms residual obtained at the n th iteration from the modified observations are roughly

(assuming a small change in $\hat{\mathbf{T}}_n$ from the previous step so that $\hat{Z}_{in} \approx \hat{Z}_{in-1}$)

$$\begin{aligned} \frac{1}{2N-4} \sum_i (Z_{in} - \hat{Z}_{in})^2 &\approx \frac{1}{2N-4} \sum_i (Z_{in} - \hat{Z}_{in-1})^2 \\ &= \frac{1}{2N-4} \sum_i [w(r_{in-1}/\hat{\sigma}_{n-1})r_{in-1}]^2 = \frac{1}{2N-4} \sum_i (\psi(r_{in-1}/\hat{\sigma}_{n-1})^2 \hat{\sigma}_{n-1}^2 \\ &\approx E[\psi(r/\sigma)^2] \sigma^2 = \beta \sigma^2 \end{aligned} \quad (\text{A2})$$

where we have used (15) and (16). In the last line E represents the expectation operator and we have replaced sample averages by expectations and assumed $E\hat{\sigma}_{n-1}^2 = \sigma^2$ (for a Gaussian distribution). The new scale estimate is thus biased by a factor $\beta = E[\psi(r/\sigma)]^2$. Assuming a Gaussian distribution with unit variance for the loss function of (13):

$$\beta = E\psi^2 = 2 \left[\frac{1}{\sqrt{2\pi}} \int_0^{r_0} r^2 \exp(-r^2/2) dr + r_0^2 \int_{r_0^2}^{\infty} \exp(-r^2/2) dr \right]. \quad (\text{A3})$$

For $r_0 = 1.5$, $\beta = 0.7784$. A simple correction for the scale estimate is to divide by β , so that for the n th iteration the scale estimate is

$$\hat{\sigma}_n^2 = \left[\frac{1}{\beta(2N-4)} \sum_i (Z_{in} - \hat{Z}_n)^2 \right]. \quad (\text{A4})$$

This estimate is an (asymptotically) unbiased estimate of the error variance if the distribution is Gaussian, and it is robust. Note that the estimate $\hat{\sigma}$ is used only to determine the exact form of the loss function, not to compute estimation errors.

A.2 COMPUTATION OF ESTIMATION ERRORS

The asymptotic covariance matrix of the estimate $\hat{\mathbf{T}}$ is (Huber 1981)

$$\text{Cov}(\hat{\mathbf{T}}) = \frac{\sigma^2 E\psi^2}{[E\psi']^2} (\mathbf{H}^T \mathbf{H})^{-1}. \quad (\text{A5})$$

Using the natural approximations for σ^2 , $E\psi^2$, $E\psi'$ from the final (say the n th) iteration, this expression is approximated, to zero order by

$$\frac{\frac{\hat{\sigma}^2}{2N-4} \sum_i \psi(r_{in}/\hat{\sigma}_n)^2}{\left[\frac{1}{2N} \sum_i \psi'(r_{in}/\hat{\sigma}_n) \right]^2} (\mathbf{H}^T \mathbf{H})^{-1} = \frac{\frac{1}{2N-4} \sum_i r_{in}^2 (\mathbf{H}^T \mathbf{H})^{-1}}{\left[\frac{1}{2N} \sum_i \psi'(r_{in}/\hat{\sigma}_n) \right]^2} \quad (\text{A6})$$

where

$$\sum_i r_{in}^2$$

can be computed from the modified observations via

$$\sum_i r_{in}^2 = \mathbf{Z}_n^T \mathbf{Z}_n - (\mathbf{H}^T \mathbf{Z}_n)^T (\mathbf{H}^T \mathbf{H})^{-1} (\mathbf{H}^T \mathbf{Z}_n).$$

Note that the scale correction discussed above does not directly enter into the covariance estimate. For ρ given by (13), ψ' is 0 or 1 depending on residual magnitude so that the denominator in (A6)

$$\frac{1}{2n} \sum_i \psi'(r_{in}/\hat{\sigma}_n)$$

is just the fraction of 'good' data points.

Appendix B

In this appendix we derive an approximate correction to the transfer function error covariance to account for the effect of the time window on correlations between adjacent Fourier coefficients. In contrast to the rest of this paper, a complex formulation will be used. We initially consider three unwindowed time series: The horizontal field vector $\mathbf{H}(t)$, the measured vertical field vector $Z(t)$ and an error series $\epsilon(t)$ which is to represent measurement and source errors. Fourier transform these series

$$\begin{aligned}\tilde{\mathbf{H}}(\omega) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(i2\pi\omega t) \mathbf{H}(t) dt \\ \tilde{\epsilon}(\omega) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(i2\pi\omega t) \epsilon(t) dt \\ \tilde{Z}(\omega) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(i2\pi\omega t) Z(t) dt.\end{aligned}\tag{B1}$$

In the frequency domain we assume that

$$\tilde{Z}(\omega) = \mathbf{T}^T(\omega) \tilde{\mathbf{H}}(\omega) + \tilde{\epsilon}(\omega) \quad -\frac{1}{2} \leq \omega \leq \frac{1}{2}\tag{B2}$$

where $\mathbf{T}(\omega)$ is the transfer function. Note that we are assuming an idealized case here where no windowing of the data has occurred. Now we consider the effect of windowing with window $W(t)$. We assume the window is normalized so that

$$\int_{-\infty}^{\infty} [W(t)]^2 dt = 1.$$

Set

$$\tilde{W}(\omega) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp(i2\pi\omega t) W(t) dt.$$

Then at frequencies ω_i , $i = 1, \dots, I$ the Fourier coefficients of the windowed time series

will be

$$\begin{aligned}
 \mathbf{H}_i &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{W}(\omega' - \omega_i) \tilde{\mathbf{H}}(\omega') d\omega' \\
 Z_i &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{W}(\omega' - \omega_i) \tilde{Z}(\omega') d\omega' \\
 \epsilon_i &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{W}(\omega' - \omega_i) \tilde{\epsilon}(\omega') d\omega' \quad (\text{B3})
 \end{aligned}$$

and (B2) is approximately

$$Z_i = \mathbf{T}(\omega_i)^T \mathbf{H}_i + \epsilon_i \quad i = 1, \dots, I$$

this will be exactly true if $\mathbf{T}(\omega)$ is constant on the support of $\tilde{W}(\omega)$, and approximately true if $\mathbf{T}(\omega)$ changes slowly over the effective width of $\tilde{W}(\omega)$ in the frequency domain (as we will assume).

Note that in general the ϵ_i 's are not uncorrelated. We assume the error series is stationary and calculate the error covariance:

$$\begin{aligned}
 E(\epsilon_i \epsilon_j^*) &= E \left[\int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{W}(\omega' - \omega_i) \epsilon(\omega') d\omega' \int_{-\frac{1}{2}}^{\frac{1}{2}} \tilde{W}(\omega' - \omega_j)^* \epsilon(\omega')^* d\omega' \right] \\
 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} d\omega'' \tilde{W}(\omega'' - \omega_j)^* \int_{-\frac{1}{2}}^{\frac{1}{2}} d\omega' \tilde{W}(\omega' - \omega_i) E[\epsilon(\omega') \epsilon(\omega'')^*] \\
 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} d\omega'' \tilde{W}(\omega'' - \omega_j)^* \int_{-\frac{1}{2}}^{\frac{1}{2}} d\omega' \tilde{W}(\omega' - \omega_i) S_\epsilon(\omega') \delta(\omega' - \omega'') \\
 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} d\omega' \tilde{W}(\omega' - \omega_j)^* \tilde{W}(\omega' - \omega_i) S_\epsilon(\omega') \quad (\text{B4})
 \end{aligned}$$

where $S_\epsilon(\omega) = E[\epsilon(\omega) \epsilon(\omega)^*]$ is the error power spectrum and the superscript asterisk denotes complex conjugation. We further assume that the error spectrum is *locally* white so that $S_\epsilon(\omega)$ is (approximately) a constant σ^2 in the band where the integral of (B4) is concentrated. Then

$$E(\epsilon_i \epsilon_j^*) = \sigma^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} d\omega' \tilde{W}(\omega' - \omega_j)^* \tilde{W}(\omega' - \omega_i) = \sigma^2 \rho_{ij}, \quad (\text{B5})$$

where ρ_{ij} is the complex correlation between ϵ_i and ϵ_j .

Suppose now that the transfer function is estimated at

$$\frac{\omega_1 + \omega_I}{2}$$

by band averaging over the frequencies $\omega_1, \dots, \omega_I$. Then it is easily seen that for standard LS

$$\hat{T} = T + A^{-1} \sum_{i=1}^I H_i^* \epsilon_i = \text{truth} + \text{error}$$

$$\left(\text{where } A = \sum_{i=1}^n H_i^* H_i^T \right).$$

The covariance matrix of the transfer function estimate is thus

$$\begin{aligned} \Sigma &= E \left[\left(A^{-1} \sum_{i=1}^I H_i^* \epsilon_i \right) \left[\left(A^{-1} \sum_{j=1}^I H_j^* \epsilon_j \right)^* \right]^T \right] \\ &= A^{-1} \left[\sum_{ij} H_i^* H_j^T \rho_{ij} \right] A^{-1} \sigma^2. \end{aligned} \quad (B6)$$

Rather than calculate this exactly, we approximate this expression by assuming that $H(t)$ is also a realization of a stationary time series which is locally white. Then we have, using an argument analogous to that in (B5),

$$E(H_i^* H_j^T) = \rho_{ij}^* E(H_i^* H_i^T). \quad (B7)$$

If the frequencies are evenly spaced, then the correlations ρ_{ij} depend only on the difference $|i - j|$; assuming this and writing ρ_k for the correlation at (frequency) lag k ($\rho_{i-j} \equiv \rho_{ij}$), we have for (B6)

$$\Sigma = A^{-1} \left[\sum_{k=-I+1}^{I-1} \rho_k \sum_i (H_i^* H_{i+k}^T) \right] A^{-1} \sigma^2$$

replacing

$$\sum_i (H_i^* H_{i+k}^T);$$

by its expectation from (B7) and approximating the expectation $E(H_i^* H_i^T)$ by the observed average A/I we have

$$\begin{aligned} \sum_i (H_i^* H_{i+k}^T) &\approx (I - |k|) \rho_k^* E(H_i^* H_i^T) \\ &\approx \left(\frac{I - |k|}{I} \right) \rho_k^* A; \end{aligned} \quad (B8)$$

so (B6) is approximately

$$\Sigma \approx \left[\sum_{k=-I+1}^{I-1} |\rho_k|^2 \left(\frac{I - |k|}{I} \right) \right] A^{-1} \sigma^2.$$

If the window is such that $|\rho_k|^2 \approx 0$ for $|k| > 0$, this reduces to the usual expression for the covariance matrix $\Sigma = A^{-1} \sigma^2$. Assuming adjacent frequencies satisfy $\omega_{i+1} - \omega_i = 1/T$ where

T is window width, we find that for the π – prolate window

$|\rho_k|^2 \approx 0$ for $|k| \geq 2$, so (since $|\rho_k| = |\rho_{-k}|$)

$$\Sigma \approx \left(1 + \frac{2(I-1)}{I} |\rho_1|^2 \right) A^{-1} \sigma^2$$

with $|\rho_1|^2 = 0.17$ (computed numerically by Fourier transforming the pi-prolate window and computing the integral of (B5)).

Note that $2(I-1)$ is the number of pairs of adjacent frequencies in the band. If all frequencies in the band are not used (this is the case for our estimation scheme with frequency domain winowing) this should be changed slightly replacing

$$\frac{2(I-1)}{I} \quad \text{with} \quad \frac{\text{number of pairs of adjacent frequencies}}{\text{number of frequencies in band}}.$$