DOE/ER/14057--2

DE92 040280

**NEW APPROACHES TO ESTIMATION OF
MAGNETOTELLURIC PARAMETERS**

**FINAL TECHNICAL REPORT
1 AUGUST, 1989 - 31 July, 1991**

GARY D. EGBERT

College of Oceanography
Oregon State University
Oceanography Admin Bldg 104
Corvallis, OR 97331-5503

**PREPARED FOR THE
U.S. DEPARTMENT OF ENERGY
OFFICE OF BASIC ENERGY SCIENCES
UNDER GRANT DE-FG06-89ER14057**

DISCLAIMER

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

MASTER

Received by OSTI

ce

1992

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Final Technical Report for DOE grant DE-FG06-89ER14057

New Approaches to Magnetotelluric Parameter Estimation

Summary and Overview

The primary goal of the research supported by grant DE-FG06-89ER14057 was to develop and test fully efficient robust data processing procedures for single station and remote reference magnetotelluric (MT) data. We made substantial progress on development, testing and comparison of optimal procedures for single station data. A principal finding of this phase of the research was that the simplest robust procedures can be more heavily biased by noise in the (input) magnetic fields, than standard least squares estimates. To deal with this difficulty we developed a robust processing scheme which combined the regression M-estimate with coherence pre-sorting. This hybrid approach greatly improves impedance estimates, particularly in the low signal-to-noise conditions often encountered in the "dead band" (0.1 - 10.0 hz). The methods, and the results of comparisons of various single station estimators are described in detail in this report.

For remote reference data we found that while standard procedures were often adequate, robust schemes based on generalizations of the single station methods generally performed better. However, the greatest current difficulty with wide-band remote reference MT data are the systematic biases which result from coherent EM cultural noise encountered in or near populated areas. Efforts to develop processing methods which will provide an effective remedy to this difficulty were the focus of much of our effort on remote reference processing. Our work on this difficult problem is still in the preliminary stages. In this report, we give a discussion of some theoretical developments completed under funding from this grant, and present some initial test applications of these ideas.

A secondary goal of our proposed research was to work on statistical aspects of static distortion models. On this front, we made progress on developing methods for estimating static distortion parameters, and for testing hypotheses about the underlying dimensionality of the geological section. These are described in *Egbert et al.* [1992; sec. 5, appendix], which is in press at *Journal of Geophysical Research*. Additional work completed under this grant involved

development of an impulse response prediction filter for iterative time domain cleaning of MT data. This work is described in *Egbert* [1992], which is in press in *Geophysics*. Copies of both manuscripts are attached separately.

Introduction

At crustal pressures and temperatures, rock forming minerals are good electrical insulators, with conduction in the crust dominated by mobile ions in pore fluids. Electrical conductivity thus varies greatly between sedimentary and crystalline rocks, and is very sensitive to porosity, and the presence, temperature, and salinity, of free fluids. These represent physical properties whose subsurface distribution is important in a wide range of applications. There are many circumstances in which electromagnetic (EM) methods represent the best, if not the only, way to map variations in these properties. In concert with other EM methods, MT has thus proven to be a valuable geophysical tool, both for fundamental studies of crustal structure and dynamics (e.g., *Wannamaker et al.* [1989]; *Stanley et al.* [1990]), and for solving specific problems in exploration for hydrocarbon (e.g., when prospecting below volcanic cover; (*Herisson* [1990]; *Bernard et al.* [1990]) and geothermal energy resources (e.g., *Goldstein* [1988]). Audio-frequency MT (AMT) is useful for very near surface applications and may be usefully applied to a wide range of waste management and environmental clean up problems. As our ability to collect, process and interpret MT data progresses, more refined uses of the method are likely to be developed. For instance, it is very likely that MT will prove useful for long term monitoring of temporal changes in subsurface conductivity. This could ultimately have applications to geothermal reservoir management, hazardous waste containment, and possibly, monitoring changes in ground stress for earthquake prediction.

MT is a field technique which uses naturally occurring variations in the EM fields. There are advantages and disadvantages to the use of such passive sources. On the one hand, because the sources occur naturally and more or less continuously, fieldwork is simpler and less expensive than with active source techniques. This allows for the possibility of relatively rapid, dense coverage of the earth's surface (e.g., *Morrison et al.* [1990]). This is a very important consideration, since it is becoming increasingly clear that surface EM methods require dense spatial

coverage to prevent aliasing from small scale near surface structures (e.g., *Torres-Verdin and Bostick* [1990]). Passive sources would also obviously be ideal for long term monitoring applications. On the other hand, the use of passive sources results in uncontrollable, and to some extent unpredictable, variations in signal and noise amplitudes, substantially complicating the path from raw data to interpreted structure.

The processing and interpretation of MT data can be divided into three distinct steps: (1) reduction of EM field time series ($\approx 10^6$ real numbers/site) to a frequency dependent uniform source impedance tensor (which is generally transformed into the more physically interpretable apparent resistivity (ρ) and phase (ϕ) curves, consisting of 10^2 - 10^3 real numbers/site); (2) application of imaging and/or inversion algorithms to create maps of conductivity variations; (3) interpretation in conjunction with other geological and geophysical data. In principle, the initial data reduction of step (1) can be accomplished quite simply. If the external sources are exactly uniform, then in the absence of noise, the electric ($\mathbf{e}$) and magnetic ($\mathbf{h}$) fields measured at a single MT site are simply related in the frequency domain via the linear relation

$$\mathbf{e} = \mathbf{Z}(\omega)\mathbf{h}, \quad (1)$$

where $\mathbf{Z}(\omega)$ is the frequent dependent MT impedance tensor. In the realistic case where noise and violations of the source assumptions must be considered, (1) will hold only approximately and statistical methods must be used to estimate $\mathbf{Z}$. In the simplest approach, the time series is divided into a series of short segments which are Fourier transformed to yield complex Fourier coefficients $\mathbf{e}_i, \mathbf{h}_i$. These are then assumed to be related via the linear statistical model with additive noise

$$\mathbf{e}_i = \mathbf{Z}(\omega)\mathbf{h}_i + \boldsymbol{\epsilon}_i \quad (2)$$

suggesting that $\mathbf{Z}(\omega)$ be estimated by using linear least squares (LS) to minimize the misfit to (2). Unfortunately, this simple approach often fails catastrophically, producing estimates which are heavily biased, wildly oscillatory, and generally useless (e.g., see Figures 1, 7-16). As a consequence a number of refinements to the simple single station LS approach have been proposed over the past 15 years in an effort to guarantee impedance estimates which are useful for subsequent stages in the interpretation process.

The failure of the LS impedance estimate can be traced to two fundamental inadequacies of the simple model (2). First, this model explicitly assumes that noise is restricted to the dependent, or "predicted" electric field channels, while the magnetic fields are observed without error. It was recognized long ago (e.g., *Swift* [1967], *Sims et al.* [1971]) that the violation of this assumption would result in the downward bias of estimated impedance amplitudes. To avoid these bias errors the remote reference method, in which horizontal magnetic fields recorded simultaneously at a second (remote) site are correlated with the EM fields at the local site, was proposed and developed by *Gamble et al.* [1979a, b] and *Goubau et al.* [1978, 1984]. RR data often yields substantially better (less biased and more stable) impedance estimates than the single station LS approach, and although this approach complicates field procedures, it is rapidly becoming the industry standard.

The second failure of the single station LS estimates lies in the implicit assumption of a Gaussian distribution for the errors in (2). This assumption often fails for MT data due to the non-stationarity of both signal and noise. In essence, it is quite typical to find that some portions of the signal come much closer to satisfying the idealized relationship of (1) than others. This results in a marginal error distribution in the frequency domain which is heavy tailed, or contaminated by outliers. A number of MT processing methods have been proposed which attempt to exploit this fact by adaptive screening or weighting of the frequency domain data. Early efforts in this direction exploited some sort of "coherence sorting" (using only those data segments which exhibit the highest coherence between electric and magnetic fields; *Stodt* [1983, 1986]; *Jones and Jodicke*, [1984]). More recently, the regression M-estimate (RME; c.f. *Huber* [1981]) has been adapted to yield single station and RR impedance estimation procedures which are resistant to outliers, and are robust to violations of distributional assumptions (*Egbert and Booker* [1986]; *Chave and Thomson* [1989]; *Chave et al.* [1987]; *Larsen* [1989]; *Sutarno and Vozoff* [1991]). These procedures can be rigorously justified mathematically (e.g., *Huber* [1981]; *Hampel et al.* [1986]), and are easily implemented using an adaptive iterative weighted LS procedure (so they are in some sense operationally similar to earlier, more *ad hoc* methods). A comparison of data processing methods applied to long period MT data from the EMSLAB experiment demonstrated the clear advantage of robust (and related) adaptive data weighting procedures (*Jones et al.* [1989]; Figure 1).

For the most part RMEs have been tested on long period MT and magnetovariational data. Few applications of the methods to the processing of the wide-band (.001–100 Hz) MT data relevant to imaging of near surface crustal structure have been reported to date. In this report we discuss efforts to adapt and extend the RME so that it will be useful for routine processing of wide-band MT data, both for single station and remote reference experimental configurations.

We have found that wide-band data presents problems which require novel solutions. The simplest application of the RME is frequently inadequate. The "dead band" (≈ 0.1 – 10.0 Hz) presents a particular problem, because signal-to-noise ratios (SNRs) in this frequency range are often very low, due both to low levels of ionospheric signals and high levels of cultural noise. In some circumstances the typical data is mostly noise and the unusual, or outlier, data are the rare time segments with useful signal, (which the RME may actually throw out as "outliers"). We have thus found that while single station RMEs of the impedance are typically smoother than the standard least squares estimates, they are often severely biased to very low amplitudes in the dead band. The situation is often further complicated by the occurrence of large outliers in both the input and output channels, and by coherent cultural EM noise of finite spatial scale.

We will first discuss our work on single station estimates, which is essentially complete. In brief, we have developed an estimation scheme which is a hybrid of "coherence sorting" and a generalized RME, with adaptive weighting to reduce the effect of leverage points. We present the results of extensive tests of this scheme on single station data of varying quality, and demonstrate the superiority of this hybrid method both to the standard RME, and to other data weighting schemes, such as coherence sorting by itself. Efforts to extend these methods to the case of RR data have shown some promise, but as we shall show below, much additional work is necessary in this area. This is particularly true for processing of RR data from areas of high cultural noise, which we see as the principle unsolved problem with the initial acquisition and data reduction phases of the MT method. In the final section we discuss some theoretical work on, and preliminary tests of, some possible approaches to improving impedance estimation procedures for RR (and more general multiple-station) MT data.

2. Single Station Impedance Estimates

The processing methods discussed here are based on refinements to the RME implementation described in detail in *Egbert and Booker* [1986] and *Jones et al.* [1989]. To effectively handle routine processing of wide-band MT data several enhancements to the processing code were first necessary. These included refinements to the initial time series processing, such as methods for adaptive autoregressive pre-whitening of the time series before Fourier analysis. These are discussed in a previous report [*Egbert*, 1990]. More significant refinements to the single station impedance estimates, including development and testing of a "generalized RME", which is resistant to excessive influence from a small number of high power events, and of a coherence pre-sorting scheme, which effectively improves signal to noise ratios in the dead band, are described here.

Generalized Regression M-Estimate

As discussed by *Chave and Thomson* [1989] (see also *Egbert and Booker* [1986]), the standard RME can be excessively influenced by a small number of very large amplitude data sections. This can result in the complete breakdown of estimates. In the terminology of linear statistical models, Fourier coefficients with unusually large magnetic (or more generally, reference) channel amplitudes are *leverage points*, i.e., individual data points which have an inordinate influence (leverage) on the estimate. Even if the large amplitude is real signal, this can result in failure of the robust estimates [*Huber*, 1981]. The situation can be much worse if the large amplitude sections result from contamination of the magnetic channel or reference channel by noise. For long period magnetovariational data we controlled this problem (to a great extent at least) by using a power dependent weighting scheme (*Egbert and Booker*, 1986). We could do this because we knew the spectral characteristics of our signal and noise very well, and could thus use a fixed set of empirically determined weights. For wide-band MT the situation is not so simple, since both signal and noise amplitudes are highly variable in space and time.

To deal with this problem in the adaptive fashion appropriate to routine wide-band MT processing, we have implemented and tested a scheme which automatically downweights leverage points. This is accomplished as follows:

(1) A robust estimate of the horizontal magnetic field spectral density matrix (SDM) $\hat{\mathbf{H}}$ is computed by an iterative process. To do this we start with the horizontal field SDM computed from the averaged cross products of all data:

$$\hat{\mathbf{H}}_0 = N^{-1} \sum_{i=1}^N \mathbf{h}_i \mathbf{h}_i^T \quad (3)$$

Then the effective influence of a single point on the impedance estimate (LS or RME) can be computed as

$$e_i = \mathbf{h}_i^T \hat{\mathbf{H}}_0^{-1} \mathbf{h}_i \quad (4)$$

If all horizontal field segments are of similar amplitude, $e_i \approx 1$ for each i . If e_i is much larger than 1, then the i th data segment has a relatively large influence on the estimate. In fact, it is readily seen that $\sum_{i=1}^N e_i = N$, and that e_i/N gives an estimate of the fraction of the total influence of the i th data segment on the estimate. We set a limit on e_i , $e_{\max}$ ($= 10$ in the examples discussed here), omit every point which exceeds this limit, and then recompute $\hat{\mathbf{H}}$ using the reduced data set. We then recompute the influence values e_i using the new SDM $\hat{\mathbf{H}}_1$. This process is iterated (resulting in a series of SDMs $\hat{\mathbf{H}}_n \rightarrow \hat{\mathbf{H}}$) until all e_i are less than $e_{\max}$. This effectively gives an estimate of the scale of the horizontal magnetic fields, excluding unusually high power data sections.

(2) Using the estimate of $\hat{\mathbf{H}}$ from the final iteration of step (1), we compute e_i (defined in equation (4)) for all data points. Before further processing data is multiplied by a weight of the form

$$w(e_i) = \begin{cases} 1 & \text{if } e_i < p_1 \\ (p_1/e_i)^{1/2} & \text{if } p_1 < e_i < p_2 \\ 0 & \text{if } p_2 < e_i \end{cases} \quad (5)$$

where p_1 and p_2 are parameters which depend on the sample sizes, which we took to be of the general form

$$p_1 = c_1 N^\alpha \quad p_2 = c_2 N^\alpha ,$$

with various values of c_1 , c_2 , and α tried. For the results reported below, we have used $\alpha = 0.5$, $c_1 = 2.0$, and $c_2 = 20.0$. This scheme has two effects. First, data sections of extremely high

power are treated as suspect, and omitted from the pilot LS estimate used to initialize the iterative computation of the RME. For instance if $N = 100$, data segments with power $p_2 = 200$ times larger than the typical power are omitted completely. Second, sections of moderately high power are downweighted, so that the maximum influence any single point can have on the results is bounded. With $N = 100$ this bound is $p_1 = 20$ times the influence of a typical data point. Note that the scale used to define a "typical" data point is determined by omitting all large amplitude sections. Note also that "hard rejection" of leverage points is only used for the first guess estimate of the impedance in the iterative RME scheme. These omitted leverage points (i.e., those with $w(e_i) = 0$ in (5)) can still be included in the final robust estimate (with weights $w(e_i) = (p_1/e_i)^{1/2}$) if they are not identified as outliers during the RME iterations. We will refer to the method described here as a generalized or weighted M-estimate. It is very similar to a general class of leverage resistant M-estimators described by *Huber* [1981].

In general we have found that the GRME results in only small changes in impedance estimates. Occasionally, however, this refinement to the RME improves estimates significantly. An example of this is provided by the apparent resistivity estimates from one of the EMSLAB long period MT stations (Figure 2). The horizontal magnetic fields were contaminated by a single section of very large amplitude noise (readily visible in plots of the time series, but passed to the estimation procedure by operator error). Standard LS estimates (dashed line, solid dots) are nearly worthless. The robust results (dotted line, asterisks) are much better, but are still very erratic around a period of 200 seconds. Results obtained using the method described above (solid line, triangles) are much improved. When leverage points are adaptively downweighted, the estimates are completely insensitive to the presence of the contaminated data. Note also that at shorter periods the estimate which omits leverage points is also systematically above the standard robust estimate. This reflects the bias effect of the noise in the horizontal components, which the robust single station scheme cannot completely eliminate.

In Figure 3 we compare results from applying the RME and the generalized RME to wide-band MT data. Here the improvement is considerably less spectacular, but some differences between the two estimates can be seen between 0.1 and 1.0 hz. In particular, for the GRME both phase and apparent resistivity curves are smoother at periods of several seconds, and the downward bias of the lower branch of the apparent resistivity is not as severe near 10 seconds.

However, for both the RME and the GRME estimates near 10 seconds are poor, with large error bars for phases and significant downward biases in apparent resistivity. Given these problems, the comparatively small improvements with the GRME are of little practical significance in this case. In summary, the GRME is more "fool proof", and thus better suited to rapid automatic processing than the straight RME. As such, it is a useful addition to code used for routine data processing. However, in many, if not most, cases differences in estimates obtained with the RME and GRME are small.

Coherence Pre-Sorting

The regression M-estimate of EM transfer functions described by *Egbert and Booker* [1986] and *Chave et al.* [1987] can be rigorously justified (both mathematically, in terms of various technical definitions of optimality, and in practice, as demonstrated by numerous Monte Carlo studies; *Hampel et al.*, 1986]) for a statistical model of the form

$$e_{xi} = Z_{xx}h_{xi} + Z_{yx}h_{yi} + \epsilon_i \quad (6)$$

where the random errors ϵ_i are drawn from a distribution contaminated by (possibly very large) outliers. The focus of the M-estimate approach is on *distributional robustness* for a *univariate* statistical model - one is concerned with deviations in the distribution of the noise (a one-dimensional random variable added to the output channel), from the idealized Gaussian form. The regression M-estimate is *not* necessarily robust to other violations of the model (6). The poor performance exhibited in Figure 3 by even the robust methods (standard LS estimates for this data set are in fact much worse) reflects the inadequacy of (6), even if heavy tailed distributions are allowed for.

The signal spectrum for MT data has a minimum near 0.1–10.0 *hz* (the "dead band"), while at the same time cultural noise can be very large at these frequencies. Thus typical signal-to-noise ratios are often very low in all channels (i.e., for the "input" magnetic, as well as the "output" electric, channels). This can create a situation (particular near areas of significant cultural activity) in which the typical data is mostly noise, while the unusual, or "outlier" data are the rare time segments with useful signal, which the RME downweights or throws out. The GRME can even exacerbate this problem, because the high power (and in this case best SNR) segments are also likely to be leverage points, which are further downweighted. It is thus necessary to

modify the RME to first identify, and eliminate, those time segments which are mostly noise.

To do this we have adopted a general strategy based on a combination of "coherence sorting" and the RME. In its simplest form, the coherence between electric and magnetic fields for a series of very short (64-128 points) data segments are computed in a wide-band centered on the target estimation frequency. Low coherence segments are then eliminated before use of the RME. By itself, sorting by coherence is not a new idea (e.g., Stodt, 1983; Jones and Jodicke, 1984). What is novel here is to use this as a preliminary sort, followed by robust processing of the high coherence segments. Results of applying this simple version of coherence pre-sorting to the data used for the previous figure are given in Figure 4b. When combined with the GRME this approach dramatically improves the impedance estimates. Note that by itself coherence sorting (as we have implemented it here) does not produce estimates comparable to any of the robust schemes (Figure 4a).

To understand some of the problems with the use of coherence sorting and the RME separately, it is useful to consider a simple real scalar analogue of the complex impedance tensor estimation problem - i.e., the problem of estimating the slope of a line constrained to go through the origin. Coherence sorting looks at subsets of the data points (a band of nearby frequencies from a single time segment), and omits those sets with low coherence, while retaining those with high coherence. Some schematic examples of low and high coherence subsets are given in Figure 5. For this example we assume that there is noise in both input and output components, and that signal amplitudes are highly variable, but most often very low. The line with the true slope is plotted for reference. Each set of circled points in a plot corresponds to the data from a single time segment for which the coherence is computed. The uncircled points correspond to data from other time segments. The set in (5a) has low signal strength and low coherence, and is thus eliminated. Meanwhile, the set in (5b) with high signal strength and high coherence is retained. In this sense coherence sorting tends to improve SNRs. However, as illustrated in (5c), it is possible for a set to have high coherence even when it contains one or more large outliers. Thus coherence sorting can miss large outliers which might be easily detected by other robust procedures which look at the consistency of each data point with *all* of the others.

On the other hand, noise in the input magnetic field channel causes estimated impedance amplitudes to be biased down. For the case of least squares (LS) estimates, it is easily shown that, on average, amplitudes are $S/(S+N)$ times the true amplitude, where S and N are the signal and noise powers in the input channel. Coherence sorting tends to increase SNR and hence to decrease these bias effects. The same is not necessarily true for the RME. In fact, when typical S/N ratios are low, The RME can be even more severely biased than LS estimates. To see why, we again consider a schematic scalar model of the form

$$Y = bX \quad b = 1 \quad (7)$$

where 95% of the data has low signal strength (signal variance = 1.0), 5% of the data has greater signal strength (signal variance = 81.0), and both the X and Y components have Gaussian noise (variance = 1.0) added. We generated 200 synthetic data sets of this form, and estimated the line slopes using the regression M-estimate, with Huber weights determined iteratively from the normalized residuals via

$$w(r) = \begin{cases} 1 & \text{if } |r| \leq 1.5 \\ 1.5/|r| & \text{if } |r| > 1.5 \end{cases} \quad (8)$$

Initial estimates were obtained using LS, and the final iteration was accomplished using weights of the form

$$w(r) = \begin{cases} 1 & \text{if } |r| \leq 2.5 \\ 0 & \text{if } |r| > 2.5 \end{cases} \quad (9)$$

which corresponds to a redescending influence function Ψ . [see *Egbert and Booker, 1986; Chave et al., 1987*].

An example of one synthetic data set is plotted in Figure 6a, along with the lines estimated by LS, the final iteration with Huber weights, and the final iteration with the redescending Ψ . The robust slope estimates are more severely biased downwards. In essence, the largest residuals at the first iteration are the points with the best S/N ratio. These are down-weighted, S/N is further decreased, and the bias becomes more severe. At the final iteration all of the best data is completely eliminated and the bias becomes much worse. This behavior is typical for the very low S/N model considered here. Slope estimates for the first 20 simulated data sets are plotted

as a function of iteration number in Figure 6b. The initial estimate (iteration = 0) is the LS estimate. Subsequent estimates almost always decrease with each iteration, with a particularly steep drop on the final iteration with the redescending Ψ . The average over all 200 data sets is given by the heavy line. This simple simulation demonstrates that the RME can be more severely biased than LS when typical SN ratios are low. This offers strong theoretical support for the hybrid scheme, which first uses coherence sorting to eliminate the numerically dominant low SN data segments

In Figures 7-16 we present a comparison of single station estimation methods for all sites from a small MT survey conducted by Dr. H. Waff near the Mendocino triple junction in Northern California. For this comparison we have not done any "picking and choosing" of particular sites to highlight the advantages of the methods discussed in this report. Four methods are compared - standard LS (top left), LS with coherence sorting top right), the RME (bottom left), and the GRME with coherence pre-sorting (bottom right). In virtually all cases the robust estimates in the bottom two panels are superior to the LS estimates in the top panels. In many cases the improvement is quite dramatic (e.g., Figures 8, 10, 11, 12, 14). Note that sometimes the robust estimates with and without coherence pre-sorting are nearly identical. Sometimes, however, coherence pre-sorting makes a substantial difference in the final estimate, particularly in the dead band (e.g., Figures 8, 10, 11). Note in particular Figure 8. Here the RME estimate without coherence pre-sorting is actually quite smooth, (especially compared to the non-robust estimates) but the downward bias in the dead band is very extreme. Indeed, for one branch of the apparent resistivity (open symbols in the lower left hand ρ curve) the bias is more severe than for the LS estimates, exactly as predicted by the simple theory discussed above.

Overall the data in this survey is not of the highest quality, and even with the best of estimation schemes there are still some problems evident in the dead band. With better data (which is probably more common than not these days), simpler estimation schemes will produce quite reasonable impedance estimates. However, when data is heavily contaminated by noise, the more sophisticated analysis methods discussed here can make a significant difference. We conclude that coherency pre-sorting combined with the regression M-estimate performs considerably better than either one alone, at least in the dead band. Together, the two approaches are capable of producing acceptable results from very noisy data, even when no effort is made to

remove the obviously contaminated data.

An important, and more widely applicable, lesson to be learned from these results is that the rigorously justifiable optimality, robustness etc., of any statistical procedure (such as the regression M-estimate) may be meaningless if the statistical model (i.e., Eq. (6)) is misspecified.

3. Remote Reference

Simple extensions of robust methods to remote reference

The remote reference (RR) estimate of MT impedances (Gamble *et al.* [1979a]) explicitly allows for noise in both input (magnetic) and output (electric) channels. Coherent magnetic fields recorded at a nearby remote site (we denote these as $\mathbf{r}$) are cross-correlated with the local EM fields (denoted $\mathbf{h}$ and $\mathbf{e}$) to estimate the impedance. Explicitly, the RR impedance estimate is

$$\hat{\mathbf{Z}} = \sum_{i=1}^N \mathbf{e}_i \mathbf{r}_i^\dagger \left[\sum_{i=1}^N \mathbf{h}_i \mathbf{r}_i^\dagger \right]^{-1}, \quad (10)$$

where the superscript dagger denotes the Hermitian conjugate transpose of the complex frequency domain field vectors. Provided noise at the local and remote sites is uncorrelated, it is simple to demonstrate that this estimate of the impedance tensor is asymptotically unbiased.

Chave and Thomson [1989] (hereafter CT) suggested a simple extension of the standard single station regression M-estimate to the case of RR MT data. The CT procedure is essentially a weighted version of the standard RR estimate of (10)

$$\hat{\mathbf{Z}} = \sum_{i=1}^N w_i \mathbf{e}_i \mathbf{r}_i^\dagger \left[\sum_{i=1}^N w_i \mathbf{h}_i \mathbf{r}_i^\dagger \right]^{-1} \quad (11)$$

with smaller weights w_i (determined as for the single station M-estimate) assigned to data points with large electric field residuals. The results presented in Jones *et al.* [1989] (see Figure 1) suggest that this robust RR procedure is superior both to standard RR estimates and to single station regression M-estimates, at least for long period MT data.

In Figure 17 we give examples of wide-band MT apparent resistivities and phases estimated with the CT robust remote reference method. The MT data used for this test were

collected near the site of the Loma Prieta, CA earthquake by Electromagnetic Instruments Inc., (EMI) in conjunction with T.R. Madden and R. Mackie, in October 1990. These results, which are fairly typical of the entire 10 station survey, show severe contamination by urban cultural noise sources, presumably from the nearby Santa Clara Valley, and from the San Francisco Bay area's regional DC mass transit system (BART). These data were taken with pairs of stations up to 20 km apart, a distance that was thought to be great enough for successful RR processing. However, the apparent resistivity and phase curves are extremely noisy. Furthermore, the extremely rapid variations of both ρ and ϕ at periods of 1-10 seconds are unphysical, and thus almost certainly reflect systematic biases (presumably due to contamination by coherent, but effectively non-uniform EM sources).

The difficulties in this survey arise from the high levels of cultural noise, coupled with the relatively low natural signal levels for frequencies in the dead band ($\approx 0.1-10.0$ Hz). Obtaining useful MT impedance estimates in this frequency range represents a major challenge, even in areas with only moderate cultural noise contamination. As we shall discuss in more detail below, our initial efforts on this problem, based on extending the single station methods of the previous section to the case of remote reference, have only been moderately successful. To produce usable MT interpretation parameters in difficult circumstances such as those encountered in the examples above, new and creative approaches to processing and collecting multiple station MT data will be required. We will close this section with a general discussion of some new possible approaches to this problem, and present preliminary results of initial tests of these ideas on data from the very noisy Loma Prieta survey. *Coherence sorting and other simple refinements*

The simplest refinement of the CT RR impedance estimate is to use coherence pre-sorting, just as with the single station MT data. A comparison of this approach, and simpler non-robust and robust RR estimates is given for two of the Loma Prieta sites in Figures 18 and 19. Again the best results were generally obtained using the hybrid robust/coherence pre-sort scheme. For instance, in Figure 18 the ρ and ϕ curves are smoothest for the hybrid method (particularly at the longest and shortest periods). Furthermore, the sharp unphysical peak near 10 seconds in the upper branch of the apparent resistivity is noticeably reduced by the robust estimate with coherence pre-sorting. Note also the difference between the two robust apparent resistivity curves at the longest periods (the circled portion of the curves in Figures 18 and 19), which are markedly

smoother when coherence pre-sorting is used. Similar results were obtained for other sites in the survey: the RME with coherency pre-sorting was generally better than other approaches, but only marginally so (see Figure 19). Overall, improvements obtained by these refinements do not appear to be significant, compared to the large apparent biases in all estimates between 1.0-10.0 seconds.

For single station MT data, coherence pre-sorting greatly improves impedance estimates in the dead band because it increases the effective SNR, and thus reduces bias. Because the RME can amplify biases, coherence pre-sorting is even more valuable when used in a hybrid scheme. The situation with regard to remote reference is somewhat different. Here, the bias effects of noise in the input channels are not so important. Low SNRs still can result in instability in the impedance estimates, particularly in the presence of outliers in the input magnetic field channels, but it is not clear that simple coherence sorting is the best way to deal with this problem. Moreover, the problems seen in impedance estimates from the Loma Prieta survey probably result from sources of cultural electromagnetic noise which have spatial scales which are large enough to be coherent, but too small to be treated as uniform.

Before considering other approaches to improving RR MT processing in the face of such difficulties, it is appropriate to reconsider the simple model of (6). In fact, for MT data collected in a modern RR survey a much more complicated, and richer, statistical model is appropriate. For purposes of our discussion we consider a RR configuration with two simultaneously recording five component MT sites. In fact, while it is possible to do RR MT with as few as 6 channels, current trends are toward even more channels than the 10 we consider. For example, multiple electric field dipoles could be deployed at the local site, or horizontal magnetic fields could be recorded at more than one remote site. (The later configuration has been used experimentally in recent commercial MT surveys where two 5 component sites are recorded simultaneously, together with a separate distant remote horizontal magnetic site; A. Mills, Zephyr Geophysical, personal communication.) Allowing for noise in all data channels, a reasonable statistical model for (frequency domain) RR data is *Egbert and Booker* [1989; hereafter EB]:

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{h}_{1i} \\ \mathbf{e}_{1i} \\ \mathbf{h}_{2i} \\ \mathbf{e}_{2i} \end{bmatrix} = \alpha_{1i} \begin{bmatrix} \eta_{11} \\ \chi_{11} \\ \eta_{21} \\ \chi_{21} \end{bmatrix} + \alpha_{2i} \begin{bmatrix} \eta_{12} \\ \chi_{12} \\ \eta_{22} \\ \chi_{22} \end{bmatrix} + \boldsymbol{\epsilon}_i = \mathbf{U}\boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i \quad (12)$$

Here $\mathbf{h}_{ji}$ and $\mathbf{e}_{ji}$ are the observed magnetic and electric fields for the i th time segment at the j th site and the parameters α_{ki} define the polarization of the magnetic fields for the i th time segment. The columns of $\mathbf{U}$ (i.e., η_{jk} and χ_{jk}) are the idealized magnetic and electric field vectors which would be observed at the j th site for uniform magnetic sources, linearly polarized N-S ($k=1$) and E-W ($k=2$), and of unit magnitude at the local ($j=1$) site. The impedance tensors for each site, as well as vertical field and interstation TFs are simply determined from the matrix of unknown parameters $\mathbf{U}$ (EB). In contrast to (3), this is a *multivariate statistical model* (e.g., Morrison [1967], Andersen [1984]); the random noise is represented by the (10-dimensional complex) vectors ϵ_i , whose components (in the measurement coordinate system) can be correlated. To discuss distributional aspects of the problem, we have to consider the *joint* distribution of all components together. Thus, as we consider issues of robustness, we should ask if outliers tend to occur separately in each component, or rather simultaneously, perhaps in some characteristic pattern, in multiple channels. The full multivariate problem is not equivalent to ten separate univariate problems.

The standard RR estimate (Gamble *et al.* [1979a]) and the CT robust RR estimate implicitly assume a more realistic model such as (12), which allows for noise in all data channels. However, neither of these approaches directly addresses the multivariate aspects of the estimation problem. In both cases, much information is ignored, so it would seem very likely that both the standard and CT RR estimates could be improved upon, in some cases significantly.

First, note that the CT robust RR estimate only considers the possibility of outliers in the predicted channel. However, noise at the remote site can substantially degrade the performance of the RR estimate. Indeed, the variance of the RR estimate is proportional to

$$(R_H^2 S_H^2)^{-1} \quad (13)$$

where the two terms in (13) are the squared multiple coherence between the remote and local magnetics and the total power in the local magnetic field channels, respectively (cf., Gamble *et al.* [1979]). Since outliers at the remote site can substantially reduce R_H^2 (but don't increase S_H^2), downweighting data segments for which a local-remote TF fits very poorly should improve the performance of robust RR estimates. This can be implemented using a variant on the iterative RME computation. Given estimates $\hat{\mathbf{T}}_{lr}$, $\hat{\mathbf{T}}_{rl}$ of the local-remote and remote-local magnetic field

TFs, we can calculate an estimated S/N for the i^{th} data point:

$$\hat{S}/N_i = \frac{\|\hat{\mathbf{T}}_{lr}\mathbf{h}_i\| \|\hat{\mathbf{T}}_{rl}\mathbf{r}\|}{\|\mathbf{r}-\hat{\mathbf{T}}_{lr}\mathbf{h}_i\| \|\mathbf{h}-\hat{\mathbf{T}}_{rl}\mathbf{r}\|}. \quad (14)$$

Then, for the next iteration in the RME RR impedance calculation, only data points for which $\hat{S}/N_i \geq S/N_{\min}$ are used. Note that in contrast to the standard RME iterative reweighting scheme, data points are downweighted here based on the *relative* size of residuals. Thus data with rather average noise content can be omitted if there is also little signal. With the RME (or CT RR) estimate, data is only downweighted if the *absolute* size of the residual is large. In this sense, weighting schemes based on the estimated S/N ratios of (14) can be viewed as a refinement of coherence sorting. Also note that for the CT estimate, outliers in $\mathbf{r}$ are not detected although these can also significantly degrade remote reference estimates. In contrast, the approach suggested here will tend to eliminate data segments for which the remote channels are badly contaminated with noise. An initial test of this scheme for one of the Loma Prieta RR station pairs in which the remote channels were often very noisy, is given in Figure 20. In this case data weighting based on local-remote SN ratios significantly reduced variability for periods from .1–2s. A further example of this weighting scheme, used in conjunction with initial coherence pre-sorting (here based on coherence of the local and remote magnetic channels) is given in Figure 21.

There are other simple ways in which RR impedance estimates could be improved. In RR methods proposed so far, one chooses two channels as "the" remotes (usually the horizontal magnetics) and ignores the other channels at the reference site. In fact, information from all channels could (and should) be used to maximize the effective SN ratio at the remote site (and hence maximize R_H^2). This idea becomes even more important as the total number of channels increases.

However, all of these simple extensions still fail to deal with the full statistical problem (12), with the consequence that useful data is ignored. The examples given above demonstrate that there are opportunities to improve RR MT impedance estimation procedures by using more of the data. In the subsections which follow we develop these ideas more fully and systematically, and give a general overview of some new approaches to this estimation problem, based on

multivariate and robust statistical methods. Future efforts will be directed toward bringing these ideas together in a practical estimation algorithm.

One final point: existing RR processing methods assume that noise at local and remote sites is uncorrelated. When this assumption breaks down, estimates may be biased (with little warning, unless the bias is so severe that the results are obviously wrong). In fact, as we shall discuss below, there are multivariate statistical methods which allow some indication of the presence of correlated noise. We suggest that these methods should be further developed as diagnostic tools for detecting when this is a problem.

Multivariate Errors-in-Variables Models

Initially we consider classical (Gaussian) multivariate procedures. Approaches to robustness are discussed below. Eq. (12) defines a linear multivariate model which, in various forms, has been referred to as structural relationship, factor analysis, linear functional equation, and (the name we adopt) multivariate errors-in-variables (MEV) models. These models for formal parameter estimation are closely related to more exploratory methods known as principal components (PCs) or empirical orthogonal functions (EOFs). A general review from a statistical perspective is given by *Anderson* [1984], and an extensive discussion of applications to geomagnetic array data is given by *Egbert* [1987] and EB. Simple versions of this model have been previously applied to single station MT impedance estimation by *Jupp* [1978] and *Park and Chave* [1984]. A summary of some relevant statistical aspects of estimation for the MEV model is summarized in Appendix A of EB. Here we sketch only the basic ideas, using the same notation as EB.

The parameters of interest in (12) are given by the 10×2 complex matrix $\mathbf{U}$ from which the impedance tensors and tippers for both sites can be calculated. Unbiased estimates of $\mathbf{U}$ can be obtained from the two dominant eigenvectors of $\Sigma_N^{-1/2} \mathbf{S} \Sigma_N^{-1/2}$, where $\mathbf{S} = \sum \mathbf{X}_i \mathbf{X}_i^\dagger$ is the spectral density matrix (SDM, the 10×10 matrix of all possible component cross product averages), and Σ_N is the covariance matrix of the noise, which also must be estimated. Without some restriction on the form of Σ_N it is not possible to uniquely resolve signal ($\mathbf{U}$) and noise (Σ_N) parameters (*Anderson and Rubin* [1956]; EB). For instance, for single station 5 component MT data it is possible to assume Σ_N diagonal, but not a general model allowing noise to be correlated between

channels. However, for two station RR MT data it is (in theory at least) possible to assume arbitrary local noise models, provided there is no correlation between noise at the local and remote sites (so Σ_N is block diagonal; *Egbert* [1987]). An iterative procedure for estimating both the local noise covariance and U for such a model is given in EB.

There are a number of potential advantages with this multivariate approach to RR MT impedance estimation relative to more standard approaches, but there are also possible drawbacks. On the plus side, the estimates discussed in EB are asymptotically equivalent to maximum likelihood for Gaussian errors, and are thus efficient (i.e., for Gaussian errors they effectively use all of the information in the data; *Gleser* [1981]). Furthermore, as discussed below, the estimator can be cast as a solution to a generalized least squares problem, so there are obvious approaches to robustness based on generalizations of the M-estimate. On the other hand, estimation of U and Σ_N simultaneously requires an iterative procedure which may not always converge. Furthermore, allowing the local error covariances to be completely general requires a large number of free parameters (50 real numbers). When sample sizes are small this may result in unstable estimates of Σ_N , possibly resulting in instability or biases in estimates of the MT impedances. In this case it might be better to assume a simpler model for the noise, for instance allowing only for uncorrelated noise (of variable amplitude) in each channel (requiring 10 real numbers). We consider the MEV estimation scheme outlined here as a starting point for future research. However, substantial modification of the basic method may be required before a practical scheme for routine processing of MT data is achieved. In any event, the MEV (or PC/EOF) approach has proven its value for exploratory analysis of MT array data (EB), and should find use as a diagnostic tool in routine MT surveys.

Canonical Coherence

Another classical multivariate technique which may be adapted to studies of signal and noise, and more generally to MT data processing, is *canonical correlation analysis* (e.g., *Morrison* [1967]). This is useful for exploring the correlation structure between two sets of random vectors X , Y of dimensions m and n respectively. Here we adapt the canonical correlation approach to the case of complex data appropriate to frequency domain analysis of MT data (and slightly modify the name to canonical coherence (CC) analysis). In our application the vectors

$\mathbf{X}$ and $\mathbf{Y}$ might represent the complex Fourier coefficients for EM fields observed at the local and remote sites, with $m=n=5$, or they could respectively represent all of the Fourier coefficients for the magnetic and electric field components (so $m = 6$, $n = 4$).

The canonical coherences are defined as follows: For m -vectors $\mathbf{a}$ and n -vectors $\mathbf{b}$ let

$$R^2(\mathbf{a}, \mathbf{b}) = \frac{\left[\text{Cov}[\mathbf{a}^\dagger \mathbf{X}, \mathbf{b}^\dagger \mathbf{Y}] \right]^2}{\text{Cov}[\mathbf{a}^\dagger \mathbf{X}, \mathbf{a}^\dagger \mathbf{X}] \text{Cov}[\mathbf{b}^\dagger \mathbf{Y}, \mathbf{b}^\dagger \mathbf{Y}]} \quad (15)$$

R^2 gives the squared coherence between all possible linear combinations of the components of $\mathbf{X}$ with those of $\mathbf{Y}$. Let

$$\rho_1^2 = \max_{\mathbf{a}, \mathbf{b}} R^2(\mathbf{a}, \mathbf{b})$$

and let $\mathbf{a}_1, \mathbf{b}_1$ be the arguments of R^2 for which the maximum is achieved. We refer to ρ_1^2 as the *first canonical coherence* of $\mathbf{X}$ and $\mathbf{Y}$. Next, let $\mathbf{a}_2, \mathbf{b}_2$ maximize R^2 subject to the constraints

$$\text{Cov}[\mathbf{a}_2^\dagger \mathbf{X}, \mathbf{b}_1^\dagger \mathbf{Y}] = \text{Cov}[\mathbf{a}_1^\dagger \mathbf{X}, \mathbf{b}_2^\dagger \mathbf{Y}] = 0, \quad (16)$$

and let $\rho_2^2 = R^2(\mathbf{a}_2, \mathbf{b}_2)$. ρ_2^2 is the *second canonical coherence* of $\mathbf{X}$ and $\mathbf{Y}$. This procedure can be continued up to $p = \min(m, n)$ yielding the series of canonical squared coherences

$$\rho_1^2 > \rho_2^2 > \dots > \rho_p^2 \quad (17)$$

Together with the vectors $\mathbf{a}_j, \mathbf{b}_j$ which satisfy

$$\rho_j^2 = R^2(\mathbf{a}_j, \mathbf{b}_j) \quad \text{Cov}[\mathbf{a}_j^\dagger \mathbf{X}, \mathbf{b}_k^\dagger \mathbf{Y}] = 0 \text{ if } j \neq k \quad (18)$$

these completely characterize the coherence structure between two sets of random vector valued time series.

Implementation of the method is simple: ρ_j^2 and $\mathbf{a}_j, \mathbf{b}_j$ are obtained by solving the eigenvalue problems

$$(\mathbf{S}_{\mathbf{YX}}^\dagger \mathbf{S}_{\mathbf{YY}}^{-1} \mathbf{S}_{\mathbf{YX}} - \lambda \mathbf{S}_{\mathbf{XX}}) \mathbf{a} = 0 \quad (19a)$$

$$(\mathbf{S}_{\mathbf{YX}} \mathbf{S}_{\mathbf{XX}}^{-1} \mathbf{S}_{\mathbf{YX}}^\dagger - \lambda \mathbf{S}_{\mathbf{YY}}) \mathbf{b} = 0 \quad (19b)$$

where $\mathbf{S}_{\mathbf{XX}} = \sum \mathbf{X}_i \mathbf{X}_i^\dagger$, $\mathbf{S}_{\mathbf{YY}} = \sum \mathbf{Y}_i \mathbf{Y}_i^\dagger$, and $\mathbf{S}_{\mathbf{YX}} = \sum \mathbf{Y}_i \mathbf{X}_i^\dagger$. The two systems have identical non-negative eigenvalues, which are the CCs ρ_j^2 . The corresponding eigenvectors of (19a) and (19b)

are, respectively, $\mathbf{a}_j$ and $\mathbf{b}_j$.

If $\mathbf{X}$ and $\mathbf{Y}$ are the EM fields at the local and remote sites, then the CCs tell us about the parts of the data at the two sites which have the highest inter-site coherence. For the idealized case of infinite sample sizes, it is not hard to show that if sources are uniform and noise at the local and remote sites is uncorrelated, there will be only two non-zero CCs. Furthermore, in this case, the vectors

$$\mathbf{S}_{XX}\mathbf{a}_j \quad \mathbf{S}_{YY}\mathbf{b}_j \quad j = 1, 2 \quad (20)$$

define the uniform source response of the earth at the local and remote sites respectively. With finite data contaminated only by incoherent (between sites) noise there will be two CCs near 1.0 and 3 near 0.0, and $\mathbf{S}_{XX}\mathbf{a}_j$, $\mathbf{S}_{YY}\mathbf{b}_j$, $j = 1, 2$ can be used to compute unbiased estimates of the impedance tensors at the local and remote sites respectively. If there are more than 2 large CCs, some of the "noise" is coherent between the two sites. Note that for Gaussian data formal hypothesis tests for the significance of non-zero coherence are readily available (e.g., *Morrison*, [1967]). An example application of CC analysis (and principal components analysis) to RR MT data contaminated by substantial correlated cultural noise is given in Figure 22.

It is not clear that estimates based directly on CC analysis of the impedances will be preferable to those obtained by other approaches. Nonetheless, CC analysis can be useful in several ways. First, it provides a simple and powerful diagnostic tool for identifying situations when the assumptions which underpin the RR method fail. In this case CC analysis can also be useful for identifying the nature of coherent noise, a useful first step in fixing or avoiding these problems. Second, CC identifies the two linear combinations of all 5 components at the remote site ($y_1 = \mathbf{b}_1 \mathbf{Y}$, $y_2 = \mathbf{b}_2 \mathbf{Y}$) which maximize coherence with EM fields at the local site. One should thus perhaps use y_1 and y_2 as the remote reference fields for RR processing. This would be one way to simultaneously use all of the remote channels as an optimal remote reference. This

approach could also be used to combine data from multiple remote sites.

Robustness in Multivariate Models

There are two somewhat distinct robustness problems which must be addressed for the multivariate methods discussed here. First, one can consider the problem of robust estimation of the SDM, without making specific model assumptions concerning signal or noise. This approach is needed so that exploratory analyses which use the SDM, such as canonical correlation or principal components analysis, are not unduly influenced by a small number of unusual data points. Second, given more specific assumptions about signal and noise (e.g., the MEV models discussed above) robust methods for estimating the unknown parameters are needed.

The first problem can be approached from several directions, as outlined in *Huber [1981; Chapter 8]*. One approach which is particularly appealing is analogous to the M-estimate - the covariance estimates are obtained as maximum likelihood estimates for a heavy tailed "ellipsoidal" family of multivariate densities. More specifically one assumes a family of densities of the form

$$f(\mathbf{x}; \mathbf{V}) = |\det \mathbf{V}| f(|\mathbf{V}\mathbf{x}|) \quad (21)$$

where f is a spherically symmetric probability density, and the multi-dimensional shape of the distribution is given by the "pseudo-covariance matrix" $(\mathbf{V}\mathbf{V}^T)^{-1}$. For the p-dimensional complex Gaussian distribution, $f(r) = (2\pi)^{-p} \exp[r^2/2]$. To make the procedure robust a heavy tailed f is used. Methods for choosing f are discussed in some detail in *Huber [1981]*. In this approach the robust estimate of the SDM is given by $(\mathbf{V}\mathbf{V}^T)^{-1}$, where $\mathbf{V}$ maximizes (21). The estimates have the form of a weighted cross-product matrix, where the weights are determined (iteratively) by the data.

For the second problem, the approach must depend to some extent on the noise model assumed. For purposes of our initial discussion we consider the special case where Σ_N is diagonal (with known error variances for each channel). In this case the MEV estimates minimize (*Gleser [1981]*)

$$\sum_{ij} \left[(X_{ij} - \sum_{k=1}^2 \alpha_{ik} U_{kj}) / \sigma_j \right]^2 \quad (22)$$

This can be accomplished by solving an eigenvalue problem, or by a more general alternating conditional least squares algorithm, which is developed in the context of a slightly different problem in *Egbert* [1991]. Emulating the regression M-estimate, we can replace (22) by the more general misfit penalty functional

$$\sum_i \rho_0 \left[\sum_j \rho_1 \left[(X_{ij} - \sum_{k=1}^2 \alpha_{ik} U_{kj}) / \sigma_j \right] \right] \quad (23)$$

Minimizing (23) over the parameters U_{kj} and α_{ik} leads to a series of coupled normal equations for weighted least squares:

$$\sum_i w_{ij} U_{kj} r_{ij} = 0 \quad k = 1, 2 \quad j = 1, 10 \quad (24)$$

$$\sum_j w_{ij} \alpha_{ik} r_{ij} = 0 \quad k = 1, 2 \quad i = 1, N \quad (25)$$

where r_{ij} and w_{ij} are the residuals and the weights:

$$r_{ij} = X_{ij} - \sum_{l=1}^2 \alpha_{il} U_{lj} \quad w_{ij} = \rho'_0 \left(\sum_j \rho_1(r_{ij}) \right) \rho'_1(r_{ij}) / r_{ij}. \quad (26)$$

This approach is directly analogous to the usual regression M-estimate. However, we now must compute the estimates iteratively (even for a fixed set of weights we must iterate between the two sets of weighted LS systems of (24) and (25)).

If $\rho_0(r) = r$ then $w_{ij} = \rho'_1(r_{ij})/r_{ij}$. This case, in which each component is downweighted separately, would be most appropriate if outliers occurred in each channel separately. On the other hand if $\rho_0 = r^2$ then $w_{ij} = w_i = \rho'_0(\|r_i\|^2)$. Now, weights are determined jointly for all components, which would be most appropriate if outliers occurred jointly in all channels together. In this case the estimates of U can be obtained from the eigenvectors of the weighted SDM, $\sum_i w_i X_i X_i^T$. Eq. (23) suggests a hybrid approach in which both possibilities are accommodated. Clearly other variants are possible and should be explored. For example, if noise is uncorrelated between stations we should probably downweight data at each station separately,

but we might allow for correlation of outliers between channels at each station.

Multivariate Time Domain Analysis for Outlier Detection/Clean Up

Very often noise in MT data is "spikey", sporadic, and well localized in time (e.g., Figure 23). These noise spikes may affect only a small fraction of points in the time series, but occur so frequently that it is virtually impossible to get long clean data segments to Fourier transform. If a large fraction of data segments are contaminated, each with small sections of noise, the frequency domain robust methods discussed above would be ineffective. On the other hand, methods for detecting and cleaning up outliers in the time domain could work quite well. *Larsen* [1980; 1989] and *Schultz and Larsen* [1987] have discussed applications of this idea to long period geomagnetic data. Here we consider further development of this idea, again emphasizing the multivariate nature of RR MT data, and the need to clean up noise in all channels.

Our first goal here is to construct a multivariate time domain filter which would separate signal and noise via simultaneous convolution with all channels of the time series, which we denote by $\mathbf{x}(t)$. With this approach the residuals and the predicted signal (each also 10 channels; denoted by $\mathbf{r}(t)$ and $\mathbf{p}(t)$) would take the form

$$\mathbf{p}(t) = \sum_{t'} \mathbf{T}(t') \mathbf{x}(t-t') \quad \mathbf{r}(t) = \mathbf{x}(t) - \mathbf{p}(t)$$

(Note that it may be convenient to actually do the convolution in the frequency domain, although we want to look at the residuals in the time domain.) The basic idea discussed here is very similar to the robust frequency domain methods considered above: a pilot estimate of $\mathbf{T}$ (computed from frequency domain TFs) would be used to compute predicted and residual time series. As with the regression M-estimate where outliers are pulled toward their predicted value, "cleaned" time series would be formed as a weighted combination of these two series, and the entire frequency domain TF analysis would be repeated on the cleaned data. This procedure could then be iterated, with the new TFs used to refine the estimate of the prediction filter $\mathbf{T}(t)$. The ability to separate signal and noise in the time domain will also be useful for our basic research on signal and noise characteristics.

We adapt ideas from EB and *Egbert* [1992] for the following suggested approach to estimation of $\mathbf{T}$ from a set of frequency domain TFs:

- (a) Estimate $U(\omega)$ for a discrete series of bands in the frequency domain, using the approaches outlined above (or some other variant on robust RR). This is equivalent to finding the TFs relating all components to the local horizontal magnetic channels.
- (b) Interpolate $U(\omega)$ with smoothing splines in the frequency domain to form a smooth band limited TF $\tilde{U}(\omega)$ (Egbert [1992]).
- (c) Convert the smoothed TFs into a smooth family of projection matrices, indexed by frequency, with $P(\omega)$ projecting onto the span of $\tilde{U}(\omega)$ (i.e., $P(\omega)$ projects onto smoothed estimates of the response spaces discussed by EB).
- (d) Fourier transform (the elements of) $P(\omega)$ to yield $T(t)$.

There are a number of refinements to this basic idea. If we find large residuals we would like to identify which channel or channels are contaminated and fix them. Indeed, looking at the time series in Figure 23 it is quite obvious to anyone familiar with MT data, that here the large noise spikes are in the magnetic channels. However without some special care, this identification may not be made by an automated procedure of the sort suggested here. The basic problem is that even when the true transfer function is known exactly, the projection operator $P(\omega)$ mixes the noise between channels, so that the residuals contain linear combinations of the noise from all channels. A prediction filter applied to the data in Figure 23 would thus tend to predict (smaller) spikes in all channels whenever a noise spike occurs in the magnetics. We suggest two approaches. First, the mapping between the TFs and the projection operator (step c) is not unique, but depends on the relative scaling of the data channels. Without going into details, we note that this scaling can be chosen to minimize mixing of noise from different channels in the residuals. Second, when there is noise in a few channels, but little signal or noise in the other channels (as in Figure 23), the predicted and residual components will be strongly and negatively correlated in the quiet channels, but positively correlated in the noisy channels. Time domain correlation between predicted and residual components may thus be useful for fine tuning automated procedures for filtering spikey noise.

Given estimates of the residual and predicted time series $r(t)$ and $p(t)$, there are still important issues to be addressed concerning time series cleaning. The standard approach which has been used in the past is most appropriate for white noise - one looks at and cleans each point

separately. This is far from ideal since it takes no account of the actual spectral content of the noise. As an improvement on this simple approach we suggest applying wavelet transform methods (e.g., Goupillaud *et al.*, 1984) to analyze the *residual* time series into orthonormal components which are localized both in frequency and time. Correction of outliers would then be done in the wavelet domain, with the cleaned time series constructed by inverting the cleaned wavelet coefficients.

There are a couple of points worth noting about the general strategy discussed here. The suggested method is based on smoothing the transfer functions in frequency while looking for outliers (see Egbert [1992] and Egbert *et al.* [1992]). However, the final impedance estimates would be computed in separate bands in the frequency domain, and should still be at least approximately independent. Nonetheless because information from a broad frequency range is used, the efficacy of outlier detection should be enhanced. For purposes of computing predicted field components (as one might want to do for some monitoring applications), for presenting initial pseudo-sections, or for application of imaging algorithms, the smoothed TFs (which are a by product of the scheme discussed here, but could also be computed separately) might be preferable to the point estimates of impedances. However, the point estimates (which represent quasi-independent information) are generally better suited for rigorous inversion or hypothesis testing, and in our view, should be the final output of the initial data reduction stage.

REFERENCES

- Anderson, T.W., Estimating Linear Statistical Relationships, *Ann. Statist.*, 12, 1-45, 1984.
- Anderson, T.W., and H. Rubin, Statistical inference in factor analysis, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 5, 111-150, ed., Jerzy Neyman, University of California Press, 1956.
- Chave, A.D., D.J. Thomson, and M.E. Ander, On the robust estimation of power spectra, coherences, and transfer functions, *J. Geophys. Res.*, 92, 633-648, 1987.
- Chave, A.D., and D.J. Thomson, 1989. Some comments on magnetotelluric response function estimation, *J. Geophys. Res.*, 94, 14,215-14,226.
- Egbert, G.D., *A multivariate approach to the analysis of geomagnetic array data*. Ph. D. dissertation, University of Washington, 1987.
- Egbert, G.D., On the synthesis of a large geomagnetic array from small overlapping arrays, *Geophys. Jour. Int.*, 106, 37-51, 1991.
- Egbert, G.D., Non-causality of the discrete time magnetotelluric impulse response, *Geophysics*, in press, 1992.
- Egbert, G.D., New approaches to estimation of magnetotelluric parameters, U.S. Dept. of Energy annual report, DOE/ER/14057-2, 1990.
- Egbert, G., and J.R. Booker, Robust estimation of geomagnetic transfer functions, *Geophys. J. R. Astron. Soc.*, 87, 173-194, 1986.
- Egbert, G.D. and J.R. Booker, Multivariate analysis of geomagnetic array data I: The response space, *J. Geophys. Res.*, 94, 14,227-14,248, 1989.
- Egbert, G.D. and J.R. Booker, Imaging crustal structure in Southwestern Washington with small magnetometer arrays, submitted to: *J. Geophys. Res.*.
- Egbert, G.D., J.R. Booker, and A. Schultz. Very long period magnetotellurics at Tucson observatory: Estimation of impedances, *J. Geophys. Res.*, in press, 1992.
- Egbert G.D. and D. Livelybrooks, Robust processing of wide-band magnetotelluric data, paper presented at Tenth Workshop on Geomagnetic Induction in the Earth, Internat. Assoc. Geomag. Aeron., Ensenada, Mexico, Aug. 22-29, 1990.
- Gamble, T.D., W.M. Goubau, and J. Clarke, Magnetotellurics with a remote reference,

- Geophysics*, 44, 53-68, 1979a.
- Gamble, T.D., W.M. Goubau, and J. Clarke, Error analysis for remote reference magnetotellurics, *Geophysics*, 44, 959-968, 1979b.
- Goldstein, N.E., Subregional and detailed exploration for geothermal-hydrothermal resources: *Geotherm. Sci. Tech.*, 1, 393-431.
- Goubau, W.M., T.D. Gamble, and J. Clarke, Magnetotelluric data analysis: removal of bias, *Geophysics*, 43, 1157-1166, 1978.
- Goubau, W.M., P.M. Maxton, R.H. Koch, and J. Clarke, Noise correlation lengths in remote reference magnetotellurics, *Geophysics*, 49, 433-438, 1984.
- Goupillaud, P., A. Grossman and J. Morlet, Cycle-octave and related transforms in seismic signal analysis, *Geoexploration*, 23, 85-102, 1984.
- Giri, N.C., *Multivariate Statistical Inference* Academic Press, New York, 1977.
- Gleser, L.J., Estimation in a multivariate 'errors-in-variables' regression
- Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel, *Robust Statistics; The Approach Based on Influence Functions*. John Wiley, New York.
- Harrison, C., Magnetotellurics applied to seismic "no record" area in West Java, Indonesia, paper presented at 60th Annual International Meeting, Soc. of Expl. Geophys., San Francisco, Ca., Sep. 23-27, 1990.
- Huber, P.J., *Robust Statistics*, Wiley, New York, 1981.
- Jensen, G., G.R. Jiracek, K.M. Johnson, M.M. Martinez, and J. Romo. Magnetotelluric modeling in the vicinity of the Salton Sea scientific drilling project, paper presented at 60th Annual International Meeting, Soc. of Expl. Geophys., San Francisco, Ca., Sep. 23-27, 1990.
- Jones, A.G., and H. Jodicke, Magnetotelluric transfer function estimation improvement by a coherence-based rejection technique, paper presented at 54th Annual International Meeting, Soc. of Expl. Geophys., Atlanta, Ga., Dec. 2-6, 1984.
- Jones, A. G., A.D. Chave, G.D. Egbert, D. Auld, and K. Bahr, 1989. A comparison of techniques for magnetotelluric response function estimation, *J. Geophys. Res.*, 94, 14,201-14,214.
- Larsen, J.C., 1980. Electromagnetic response functions from interrupted and noisy data, *J. Geomagn. Geoelectr.*, 32 SI, 89-103.
- Larsen, J.C., 1989. Transfer functions: smooth robust estimates by least squares and remote

- reference methods, *Geophys. J. Int.*, 99, 655-663.
- Morrison, D.F., *Multivariate Statistical Methods*, McGraw-Hill, San Francisco, 1967.
- Morrison, H.F., E.A. Nicholls, C. Torres-Verdin, J.R. Booker, and S.C. Constable. Comparison of magnetotelluric inversion techniques on a mineral prospect in Nevada, paper presented at 60th Annual International Meeting, Soc. of Expl. Geophys., San Francisco, Ca., Sep. 23-27, 1990.
- Park, J., and A.D. Chave, On the estimation of magnetotelluric response functions using the singular value decomposition, *Geophys. J. R. astr. Soc.*, 77, 683-709, 1984.
- Schultz, A., and J.C. Larsen, 1987. On the electrical conductivity of the earth's interior I: mid-mantle response function computation, *Geophys. J. Roy. astr. Soc.*, 733-761.
- Stanley, W.D., W.D. Mooney, and G.S. Fuis, Deep crustal structure of the Cascade Range and surrounding regions from seismic refraction and magnetotelluric data, *J. Geophys. Res.*, 95, 19,419-19,438, 1990.
- Sims, W.E., F.X. Bostick, and H.W. Smith, The estimation of magnetotelluric impedance tensor elements from measured data, *Geophysics*, 36, 938-942.
- Stodt, J.A., Noise analysis for conventional and remote reference magnetotellurics, Ph.D. dissertation, Univ. of Utah, 220 pp., 1983.
- Stodt, J.A., Weighted averaging and coherence sorting for least-squares magnetotelluric estimates, paper presented at Eighth Workshop on Geomagnetic Induction in the Earth and Moon, Internat. Assoc. Geomag. Aeron., Neuchatel, Switzerland, Aug. 24-31, 1986.
- Sutarno, D., and K. Vozoff, Phase-smoothed robust M-estimation of magnetotelluric impedance functions, *Geophysics* 56, 1999-2007, 1991.
- Swift, C.M., A magnetotelluric investigation of an electrical conductivity anomaly in the southwestern United States. Ph.D. dissertation, M.I.T., 211 pp, 1967.
- Torres-Verdin, C. and F.X. Bostick Jr. Principles of spatial surface electric field filtering in magnetotellurics: Electromagnetic Array Profiling (EMAP): *Geophysics*, Accepted for publication, 1990.
- Wannamaker, P.E., J.R. Booker, A.G. Jones, A.D. Chave, J.H. Filloux, H.S. Waff, and L.K. Law, Resistivity cross-section through the Juan de Fuca subduction system and its tectonic implications, *J. Geophys. Res.*, 94, 14,127-14,145, 1989b.

Warren, R.K., and L.J. Srnka. EMAP exploration in the volcanics of the Columbia River Basin, Washington, paper presented at 60th Annual International Meeting, Soc. of Expl. Geophys., San Francisco, Ca., Sep. 23-27, 1990.

TABLE 1. Short Descriptions of the Methods Used

Method	Description	Source
1	conventional spectral analysis	Bahr
2	conventional spectral analysis	Auld
3	conventional spectral analysis	Jones
4	weighted cascade decimation	Jones
5	remote reference	Chave
6	robust cascade decimation	Jones
7	robust	Egbert
8	robust remote reference	Chave

Figure 1: Comparison of apparent resistivities obtained from 8 processing methods (summarized in table 1) applied to 5 days of long period MT data from one site in the Emslab experiment (from Jones *et al.*, [1989]). Solid curves are from robust processing of 10 weeks of remote reference data from the same site. For methods 4 and 6, both upward biased and downward biased results are plotted. In all plots the lower curve (ρ_{up}) has been shifted down one decade for clarity. During this time period signal strength was very low, and conventional LS processing of single station data (methods 1-4) resulted in practically useless estimates. Much better estimates were obtained by robust methods, particularly those based on the regression M-estimate (methods 7 and 8). Note that method 7 (which uses only single station data) appears to be systematically biased downward for periods below 10^3 s. Conventional remote reference (method 5), is less biased than method 7, but much more variable at long periods.

Fig. 2

Figure 2: Apparent resistivities obtained from three estimators from long period MT data taken during the EMSLAB experiment. Dashed line and solid dots: standard least squares. Dotted line and asterisks: regression M-estimate. Solid line and triangles: generalized regression M-estimate, adapted to downweight leverage points. Severe contamination of horizontal fields in several short sections leads to the breakdown of the standard regression M-estimate. The generalized M-estimate performs well even in the presence of severe outliers in the reference magnetic field channels.

Figure 3: Apparent resistivities and phases obtained by applying RME (left) and GRME (right) estimators to data from a wide-band MT site in the Willamette valley (site EM75) collected during the EMSLAB experiment, by the University of Oregon MT system. Results are shown only for the central collection band, where errors were largest. The GRME, which adaptively downweights leverage points is somewhat smoother at the shortest periods, and less strongly biased downwards near 10 seconds period.

Figure 4: Apparent resistivity and phase estimates obtained by applying LS (left) and GRME with coherence pre-sorting (right) to the data for EM75 (as in Figure 3). The estimates obtained with the hybrid GRME/coherence pre-sort scheme are far superior to LS, and also to the straight robust estimates of Figure 3.

Figure 5: Schematic illustration of the effect of coherence sorting on the estimate of line slope when there is noise in both variables, and typical S-N ratios are low. Circled data points correspond to data from a single time segment. (a) Set of points with low signal strength has low coherence and is eliminated. (b) A good data set with high signal strength and high coherence is retained, so coherence sorting tends to improve signal-to-noise ratios. (c) However, sets with high signal strength and high coherence can still contain significant outliers.

Figure 6: A simple model to illustrate the enhanced bias seen in the robust ρ curve of fig. (4c). Here we plot results of using LS and the RME to estimate line slopes for synthetic data. There is noise in both variables, and typical signal-to-noise ratios are low. (a) Synthetic data, and lines estimated by LS, and two variants of the RME. True line slope is 1. Slopes estimated by the RMEs are more severely biased than the LS estimate. (b) Slope estimates for 20 simulated data sets plotted as a function of iteration number. The initial estimate (iteration = 0) is the LS estimate. Subsequent estimates almost always decrease with each iteration, with a particularly steep drop on the first iteration (which uses a redescending influence function). The average over 200 simulated data sets is given by the heavy line.

Figures 7-16: Apparent resistivity curves estimated by LS (left column) and a robust scheme (right column) for a sequence of 4 sites in a small 10 site survey near Cape Mendocino, CA. The robust method used here was a regression M-estimate with a coherence pre-sort (from *Egbert and Livelybrooks*, [1990]). The robust methods produce significantly smoother and less biased curves. Even when simple LS methods produce reasonable results (such as at site 8) significantly smoother estimates can often be obtained with robust methods.

Figure 7: Mendocino survey; Site 1

mag ref, non-robust, M01, cohmin = .0

Figure 8: Mendocino survey; Site 2

Figure 9: Mendocino survey; Site 3

Figure 10: Mendocino survey; Site 4

Figure 11: Mendocino survey; Site 5

Figure 12: Mendocino survey; Site 6

Figure 13: Mendocino survey; Site 7

Figure 14: Mendocino survey; Site 8

Figure 15: Mendocino survey; Site 9

Figure 16: Mendocino survey; Site 10

Figure 17: Examples of apparent resistivities and phases estimated by the robust remote reference processing scheme of *Chave and Thomson* [1989]. This MT data, which was collected near the site of the Loma Prieta, CA earthquake by EMI in conjunction with T. Madden and R. Mackie, is heavily contaminated by cultural noise from the nearby San Francisco Bay area. Although these results are somewhat better than those obtained from standard unweighted remote reference processing, they are still often very poor. In addition to the excessive variability seen at some sites, the shapes of the ρ and ϕ curves (rapid drop in ϕ to 0 degrees or less, coupled with the very sudden change of slope of ρ near 10s period) are strongly suggestive of bias resulting from contamination of uniform sources by much smaller scale cultural sources.

Figure 18: Remote reference data from the EM/MIT Loma Prieta Survey processed with (a) standard remote reference estimate, (b) remote reference estimate with coherence sorting, (c) robust remote reference estimate similar to the estimate proposed by Chave and Thomson [1989], and (d) robust estimate with coherence pre-sorting. Again the best results are obtained using the hybrid robust/coherence pre-sort scheme. Relative to more standard RR estimates this method reduces the unphysical and anomalous peak near 10 seconds in the upper branch of apparent resistivity, and gives smoother curves at both the longest and shortest periods. Note in particular the difference between long period apparent resistivity curves for robust estimates with and without coherence pre-sorting. However, while the robust estimates tend to yield smoother curves, the improvements are rather marginal, especially considering the unphysical behavior and large apparent bias of all estimates between 1.0-10.0 seconds.

Figure 19: Remote reference data from the EMI/MIT Loma Prieta Survey processed with (a) standard remote reference estimate, (b) remote reference estimate with coherence sorting, (c) robust remote reference estimate similar to the estimate proposed by *Chave and Thomson* [1989], and (d) robust estimate with coherence pre-sorting. as in Figure 17, the best results are obtained between 1.0-10.0 seconds.

Figure 20: A further example from the Loma Prieta survey (see figure 19). (a) Results from processing with a robust RR method similar to that described by CT. (b) Results from similar processing after eliminating large amplitude noise from the remote site. Variability is significantly reduced for periods from .1-2s. Weighting based only on the magnitude of residuals in the predicted local electric field channels is ineffective against outliers in the remote channels although these can also significantly degrade remote reference estimates.

Figure 21: A further example from the Loma Prieta survey (see figure 19). (a) Results from processing with a robust RR method similar to that described by CT. (b) Results from similar processing with preliminary sorting based on coherence of local and remote horizontal magnetic fields. Variability is significantly reduced in the dead band by this preliminary sorting.

Figure 23: An example of (high pass filtered) MT time series, showing contamination of horizontal magnetic fields with discrete bursts of noise, mostly in the magnetic field channels. The noise at high frequencies is much larger than the weak signal, which is hardly visible. We propose methods for detecting and cleaning up temporally well localized noise of this sort.

END

**DATE
FILMED**

10 / 27 / 92